Submission Number: PET11-11-00266

Innovation vs. imitation and the evolution of productivity distributions

Michael David Koenig
Visiting Scholar, Department of Economics Stanford University

Jan Lorenz
Center for Social Science Methodology, Carl-von-Ossietzky University Oldenburg Germany

Fabrizio Zilibotti
Department of Economics, University of Zurich, Muhlebachstrasse 86, CH-8008 Zurich, Switzerland

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 with a tail following Zipf's law. Further, we take into account idiosyncratic shocks in firms' productivities and show that these can reduce inequality, but at a price of lowering aggregate productivity and industry performance.

Submitted: March 15, 2011.

Innovation vs. Imitation and the Evolution of Productivity Distributions

Michael D. König[a], Jan Lorenz[b], Fabrizio Zilibotti[c]

[a] *Visiting Scholar, Department of Economics, Stanford University, 579 Serra Mall, CA 94305-6072, United States*
[b] *Center for Social Science Methodology, Carl-von-Ossietzky University Oldenburg A06 3-320, Ammerländer Heerstr. 114 - 118, 26129 Oldenburg, Germany*
[c] *Department of Economics, University of Zurich, Muhlebachstrasse 86, CH-8008 Zurich, Switzerland*

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 with a tail following Zipf's law. Further, we take into account idiosyncratic shocks in firms' productivities and show that these can reduce inequality, but at a price of lowering aggregate productivity and industry performance.

Key words: innovation, growth, quality ladder, absorptive capacity, productivity differences, spillovers
JEL: O40, E10

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 techno-

Email addresses: mdkoenig@stanford.edu (Michael D. König), jalorenz@ethz.ch (Jan Lorenz), zilibott@iew.uzh.ch (Fabrizio Zilibotti)

logical knowledge [Prescott, 1998; Romer, 1993].[1] 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].[2]

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 productivity 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

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

[2] 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].

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; Kelly, 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 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 6 we analyze the conditions improving industry performance. The empirical analysis of firm productivities is given in Section 2. 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 7 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 2. 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.

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 $12,313,464$.

The resulting productivity distributions for the years 1995 to 2004 and the corresponding average productivities 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

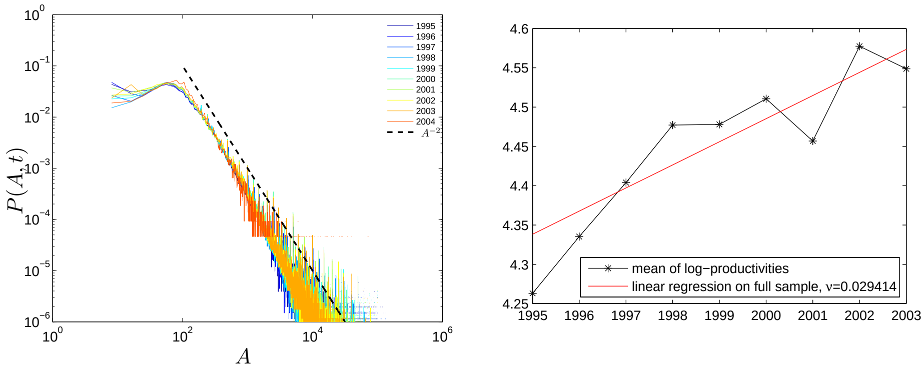

Figure 1: (Left) Productivity distribution in the years 1995 to 2004. The dashed line indicates a power-law $P(A,t) \propto A^{-2}$ with exponent minus two and support $A \geq A_{\min} = 10^2$. The figure suggest that the distributions are close to a power-law with an exponent of around two for $A \geq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

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-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].[3] 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.

We note here that Corcos et al. [2007] have estimated the productivity distributions using the same data set, while also controlling for physical and

[3]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].

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.

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

3.1. Environment

A unique final good, denoted by $Y(t)$, is produced by a representative competitive firm using labor and a set of intermediate goods $x_i(t)$, $i \in N = \{1, 2, ..., n\}$, according to the production function

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

where $x_i(t)$ is the economy's input of intermediate good i at time t, $A_i(t)$ is the technology level of sector i at time t, and $\epsilon_i\,(t)$ is a productivity shock assumed to be i.i.d. across sectors and over time (see more discussion in Section 5). We normalize the labor force to unity, $L = 1$. The final good $Y(t)$ is used for consumption, as an input to R&D and also as an input to the production of intermediate goods. The profit maximization program yields the following inverse demand function for intermediate goods,

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

where the price of the final good is set to be the numeraire.

Each intermediate good i is produced by a technology leader which can produce the best quality of the input at the unit marginal cost. The leader is subject to the potential competition of a fringe of firms that produce the same input at the constant marginal cost χ, where $1 < \chi \leq 1/\alpha$. A higher value of χ indicates less competition. Bertrand competition implies that technology leaders monopolize the market, and set the price equal to the unit cost of the fringe,

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

and sell at that price the equilibrium quantity $x_i(t) = \chi^{-\frac{1}{1-\alpha}} A_i(t)$. The profit earned by the incumbent in any intermediate sector i will then be proportional to the productivity in that sector

$$\pi_i(t) = (p_i(t) - 1)\, x_i(t) = \psi \epsilon_i(t) A_i(t), \tag{2}$$

where $\psi \geq \frac{\chi-1}{\alpha} \chi^{-\frac{1}{1-\alpha}}$ which is monotonically increasing in α and decreasing in χ. In equilibrium, output is proportional to aggregate productivity as follows

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

where aggregate productivity is $A(t) = \sum_{i=1}^{n} \epsilon_i(t) A_i(t)$.

3.2. Technological Change

The productivity of each intermediate good $i \in N$ is assumed to take on values along a quality ladder with rungs spaced proportionally by a factor $\bar{A} > 1$. 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 $A_i = \bar{A}^{a_i}$.[4]

Firm i's productivity $A_i \in \{1, \bar{A}, \bar{A}^2, \dots\}$ grows as a result of technology improvements, either undertaken in-house (innovation) 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; Kelly, 2001; Rosenberg, 1976]. At time step $t = \Delta t, 2\Delta t, 3\Delta t, \dots,\ \Delta t > 0$, a firm i is selected at random and decides either to imitate another firm or to conduct in-house R&D, depending on which of the two gives it higher expected profits.[5]

3.2.1. Innovation

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 $b = 0, 1, 2, \dots$ 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 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

[4] 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 $1 + g = \bar{A}$.

[5] 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].

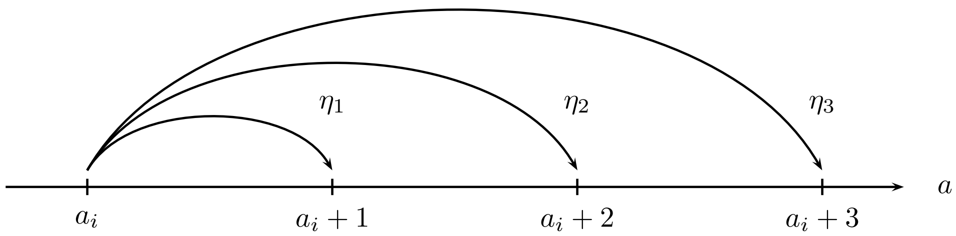

Figure 2: Illustration of the innovation process of firm i with log-productivity $\log A_i = a_i \log \bar{A} = a_i$ (setting $\log \bar{A} = 1$). With probability η_1 firm i makes one productivity improvement and advances by one log-productivity unit, with probability η_2 firm i makes two productivity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

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].

3.2.2. Imitation

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 $A_j(t)$ as long as $A_j(t) > A_i(t)$ which is equivalent to $a_j(t) > a_i(t)$. Conditional on firm i selcting a firm j with higher productivity, firm i 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. Moreover, the attempt finishes after the first failure. This reflects the fact that knowledge absorption 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 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 distri-

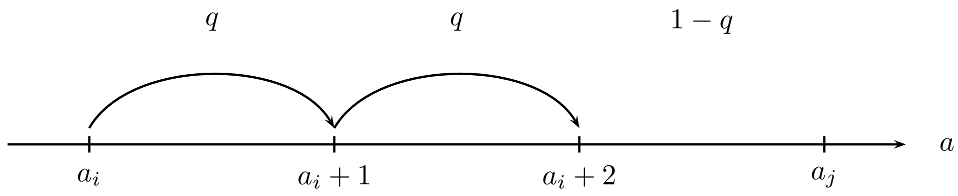

Figure 3: Illustration of the imitation of log-productivity a_j of firm j through firm i with log-productivity a_i, where the log-productivity of firm i is $\log A_i = a_i \log \bar{A} = a_i$ (setting $\log \bar{A} = 1$). Firm i successfully imitates two log-productivity units (with probability q^2)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

bution 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.

Switching to normalized log-productivity in Equation (5) we obtain

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

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

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

If $q<1$ and b is much larger than a, the following approximation holds:

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

In this case, the log-productivity firm i obtains through imitation does not depend on the log-productivity of firm j but only on 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. Relationship with the Existing Literature

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 [2002] 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 With No Innovation-vs.-Imitation Choice

In this section, we analyze the evolution of productivity distribution first in a world where all firms innovate through in-house R&D, and then in a world where all firms try to imitate more productive firms. We show that in the former case the variance of the productivity distribution increases over time. In contrast, in the latter case, the variance decreases over time converging to a mass point at the productivity level of the initially most productive firm. Neither case is consistent with the evidence discussed in Section 2. Then, in Section 5 we move to the main contribution of the paper, i.e., the characterization of productivity distribution in a world where firms choose optimally whether to innovate or to imitate existing technologies.

4.1. Notation

Consider the distribution of normalized log-productivity in the population of $N \in \mathbb{N}$ firms over time. Recall that normalized log-productivity only takes values in the set $S = \{1, 2, \ldots, a_{\max}\}$ with $a_{\max} \in \mathbb{N} \cup \{\infty\}$. This set can be finite when there exists a maximum attainable log-productivity $a_{\max}$ or equal to the integers $S = \mathbb{N}$ when $a_{\max} = \infty$. Let $P_a(t)$ indicate the fraction of firms having log-productivity $a \in S$ at time $t \in T$. 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. It holds that $P_a(t) \geq 0$ and $\sum_{a=1}^{\infty} P_a(t) = 1$.[6] In what follows we may omit for simplicity either a or t in the arguments of $P_a(t)$ whenever it is no source of confusion.

In order to analyze the evolution of $P(t)$, we borrow tools from stochastic approximation theory as they have been used in the literature on evolutionary game theory [see e.g. Benaim and Weibull, 2003; Sandholm, 2006]. Our dynamics of innovation and imitation induces a discrete time, discrete space Markov chain $\big((P^N(t))_{t \in T}\big)_{N=N_0}^{\infty}$, where the chain indexed by N takes on values in the state space (simplex) $P^N = \{P \in \mathbb{R}_+^{|S|} : PN \in \mathbb{Z}^{|S|}, \sum_{a \in S} P_a = 1\}$ indicating the fraction of firms with a certain log-productivity $a \in S$. At times $t \in T = \{0, \Delta t, 2\Delta t, \ldots\}$, with $\Delta t = 1/N$, exactly one firm in the population of N firms is selected at random and given the opportunity to introduce a technology improvement (through either innovation or imitation, as discussed below). The conditional probability $T_{ab} : P^N \to \mathbb{R}_+^{|S| \times |S|}$ that a firm with log-productivity a switches to log-productivity b at time t is given by

$$T_{ab}(P) = \mathbb{P}\left(P^N(t + \Delta t) = P + \frac{1}{N}(e_b - e_a) \middle| P^N(t) = P \right), \qquad (9)$$

[6]Note, that when $s = \mathbb{N}$ the vector $P(t)$ is unbounded to the right implying log-productivity to be not bounded from above. If we would also allow for log-productivity to decay then $P(t)$ should in principle also be unbounded to the left. However, in the following we will restrict our analysis to the semi-bounded case without decay. The case of productivity decay will be discussed in Appendix B.1.

where $e_a \in \mathbb{R}^{|S|}$ is the standard unit basis vector corresponding to log-productivity $a \in S$. The conditional transition probabilities of our Markov chain $(P^N(t))_{t\in T}$ are then given by

$$\mathbb{P}\left(P^N\left(t+\Delta t\right)=P+z \middle| P^N\left(t\right)=P\right)$$
$$= \begin{cases} P_a T_{ab}(P) & \text{if } z=\frac{1}{N}(e_b-e_a), \quad a,b\in S, \quad a\neq b, \\ 1-\sum_{b\in S}\sum_{b\neq a} P_a T_{ab}(P) & \text{if } z=0, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

In the following subsections, we derive the matrix $\mathbf{T}(P)$ with elements $T_{ab}(P)$, $a, b \in S$, under the individual firm laws of motion associated with innovation (Equation (4)) and imitation (Equation (7)), respectively.

4.2. Productivity Dynamics Under Innovation

In this section, we assume that all firms enage in R&D. More formally, this is the equilibrium outcome when firms have no absorptive capacity for imitation ($q = 0$). The random variable $\eta(t) \in S$ is restricted to natural numbers as possible realizations of the process of in-house innovations are represented by our discrete probabilistic framework.[7] Moreover, we assume that 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 and $\eta_0 = 1 - \sum_{b\geq 1}\eta_b$. Thus, the transition matrix for log-productivity due to in-house R&D corresponding to Equation (3) is

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

[7]We can however easily approach continuous imitation and innovation processes, since we can approximate the continuous case to an arbitrary precision by mapping the continuous productivities to finer grained discretizations.

The evolution of the log-productivity distribution in Equation (23) is given by

$$\frac{\partial P(t)}{\partial t} = P(t)(\mathbf{T}^{\text{in}} - \mathbf{I}).$$

Consider first particular the case where one step of innovation is achieved with probability p, thus, $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$. In matrix-vector notation the evolution of the log-productivity distribution can be written as follows

$$\frac{\partial P(t)}{\partial t} = P(t) \begin{bmatrix} -p & p & 0 & \dots & & \\ 0 & -p & p & 0 & \dots & \\ 0 & 0 & -p & p & 0 & \dots \\ & & & \ddots & \ddots & \ddots \end{bmatrix}.$$

For each level of log-productivity a this means that

$$\frac{\partial P_a(t)}{\partial t} = p(P_{a-1}(t) - P_a(t)). \tag{11}$$

Using Equation (11) we can derive the log-productivity distribution at time $t \in \mathbb{R}_+$.

The intuition conveyed by the simple process above can be generalized to any (non-decaying) process. In particular, there is a positive drift of the random innovation process. Thus, the log-productivity approaches a Gaussian shape in the limit of large times t, due to the central limit theorem. Mean and variance rise linearly with t. 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 1. *If $\mathbb{E}(\eta) > 0$ and $q = 0$ then the* log-*productivity distribution approaches a Normal distribution $\mathcal{N}(t\mu_\eta, t\sigma_\eta^2)$, 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 $\sigma_A^2 = \left(e^{t\sigma_\eta^2} - 1\right) e^{2t\mu_\eta + t\sigma_\eta^2}$.*

The key finding here is that the variance of the log-productivity distribution increases over time. It is also worth noting that for large times t, the lognormal distribution will be close to a power-law (or Pareto distribution) for a wide range of productivities, as stated in the next corollary.

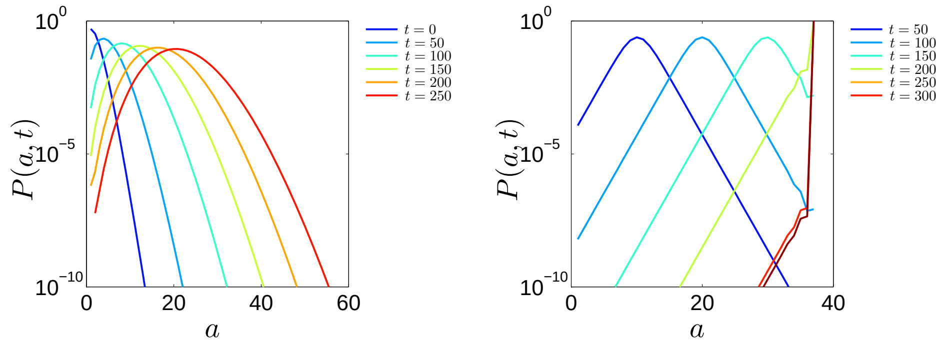

Figure 4: (Left) The log-productivity distribution $P(a, t)$ for different periods and $p = 0.1$ when firms can only increase their productivity through in-house R&D. (Right) The log-productivity distribution $P(t)$ for different periods and $q = 1$ in the case of pure imitation.

Corollary 1. *The asymptotic productivity distribution can be approximated by a power-law distribution for* $A = \mathcal{O}(e^{(\mu_\eta + 2\sigma_\eta^2)t})$.

The time evolution of the log-productivity distribution $P(t)$ 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 1.

4.3. Productivity Dynamics Under Imitation

In this section, we consider the polar opposite case when firms have not independent capacity to innovate through in-house R&D, and can only introduce technological progress through imitating other firms' technologies. More formally, this is an equilibrium outcome if $\eta_i = 0$ for $i \geq 1$. The long-run outcome is easy to guess: all firms will converge to the same productivity level, equal to the largest productivity level in the initial distribution. However, the analysis of this case is instructive, since it provides key insights for the general case in which firms face a non-trivial choice between innovation and imitation.

Consider the transition matrix for imitation. The conditional transition probability from log-productivity a to log-productivity $b > a$ is given by

$$\begin{aligned} T^{\mathrm{im}}_{ab}(P) &= q^{b-a} P_b + q^{b-a}(1-q)P_{b+1} + q^{b-a}(1-q)P_{b+2} + \dots \\ &= q^{b-a}\left(P_b + (1-q)\sum_{k=1}^{\infty} P_{b+k}\right) \\ &= q^{b-a}\left(P_b + (1-q)(1-F_b)\right), \end{aligned} \tag{12}$$

with F being the cumulative distribution of P as defined by $F_b = \sum_{c=1}^{b} P_c$. For $b < a$ imitation of b is omitted by firm a, thus $T^{\mathrm{im}}_{ab}(P) = 0$. The staying probability for $b = a$ is thus $T^{\mathrm{im}}_{aa}(P) = 1 - \sum_{b>a} T^{\mathrm{im}}_{ab}(P)$.

The transition matrix $\mathbf{T}^{\mathrm{im}}$ with elements given by Equation (12) for the imitation process in Equation (5) is interactive.[8] It depends on the current distribution of log-productivity $P_a(t)$ and it is given by

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

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

$$\mathbf{T}^{\mathrm{im}}(P) = \begin{bmatrix} F_1 & P_2 & P_3 & P_4 & \dots \\ 0 & F_2 & P_3 & P_4 & \dots \\ 0 & 0 & F_3 & P_4 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix},$$

since $S_a(P) = 1 - \sum_{b=a+1}^{\infty} P_b = F_a(t)$, where the cumulative distribution function is given by $F_a(t) = \sum_{b=1}^{a} P_b$.

The evolution of the log-productivity distribution is given by

$$\frac{\partial P(t)}{\partial t} = P(t)(\mathbf{T}^{\mathrm{im}}(P(t)) - \mathbf{I}), \tag{13}$$

[8] For an interactive Markov chain the conditional transition probabilities depend on the current distribution [Conlisk, 1976].

which can be written in vector-matrix notation as

$$\frac{\partial P(t)}{\partial t} = P(t) \begin{bmatrix} S_1(t)-1 & P_2(t) & P_3(t) & P_4(t) & \dots \\ 0 & S_2(t)-1 & P_3(t) & P_4(t) & \dots \\ 0 & 0 & S_3(t)-1 & P_4(t) & \dots \\ & & & \ddots & \ddots \end{bmatrix},$$

where $S_a = 1 - \sum_{b=a+1}^{\infty} q^{b-a} (P_b + (1-q)(1-F_b))$. From Equation (13) we can derive the ordinary differential equation (ODE) governing the evolution of the cumulative log-productivity distribution in a more compact form.

Proposition 2. *Assume firms cannot innovate in-house ($\eta_0 = 1$ and $\eta_i = 0$, for all $i \geq 1$), then the evolution of the cumulative* log-*productivity distribution $F(t)$ is given by*

$$\frac{\partial F(a,t)}{\partial t} = F(a,t)^2 - F(a,t) + (1-q)(1-F(a,t)) \sum_{b=0}^{a-1} q^b F(a-b,t). \quad (14)$$

For $q = 0$ Equation (14) is trivially satisfied, as $\partial F(a,t)/\partial t = 0$. The boundary conditions are $F(0,t) = 0$ and $F(\infty, t) = 1$. From Proposition 2 we can determine the asymptotic distribution in the limit of large times t.

Corollary 2. *If there exists a maximal initial* log-*productivity $\bar{a}$ such that $F_a(0) = 1$ for all $a \geq \bar{a}$ then the asymptotic cumulative* log-*productivity distribution of Equation (14) is given by*

$$\lim_{t\to\infty} F_a(t) = \begin{cases} 0, & \text{if } a < \bar{a}, \\ 1, & \text{if } a \geq \bar{a}. \end{cases} \quad (15)$$

Note, that Equation (15) is equivalent to $\lim_{t\to\infty} P_{\bar{a}}(t) = 1$. Thus all probability mass concentrates at $\bar{a}$ in the course of time.

When absorptive capacity limits are strong and thus climbing a rung on the quality ladder is unlikely (i.e., small q), we neglect terms of the order $\mathcal{O}(q^2)$ and derive the following corollary.

Corollary 3. *Assume firms cannot innovate in-house ($\eta_0 = 1$ and $\eta_i = 0$, for all $i \geq 1$), and 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*

$$\frac{\partial F(a,t)}{\partial t} = qP(a,t)(1 - F(a,t)). \tag{16}$$

Consequently, for $q \to 0$ we get from Equation (16) that $\partial F(a,t)/\partial t = 0$.

On the other hand, when almost no limit to absorptive capacity exists (q close to one) we neglect terms of the order $\mathcal{O}((1-q)^2)$ and we derive from Equation (14) the following corollary.

Corollary 4. *Assume firms cannot innovate in-house ($\eta_0 = 1$ and $\eta_i = 0$, for all $i \geq 1$), and 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.*[9] *Then the evolution of the cumulative* log*-productivity distribution is given by*

$$\frac{\partial F(a,t)}{\partial t} = (2q - 1)\left(F(a,t)^2 - F(a,t)\right). \tag{17}$$

In the special case of $q = 1$ we find that we recover the knowledge growth dynamics analyzed by Lucas [2008] . The result is given in the following proposition.

Proposition 3. *Assume that firms cannot innovate in-house ($\eta_0 = 1$ and $\eta_i = 0$, for all $i \geq 1$), and that there are no absorptive capacity limits for imitation ($q = 1$). Then, the cumulative* log*-productivity distribution follows the ODE*

$$\frac{\partial F(a,t)}{\partial t} = F(a,t)^2 - F(a,t). \tag{18}$$

Starting from an arbitrary initial cumulative distribution $F(a,0)$, the cumulative log*-productivity distribution at time t is given by*

$$F(a,t) = \frac{F(a,0)}{F(a,0) + e^t(1 - F(a,0))}. \tag{19}$$

Note, that the distribution $P(t)$ can always be extracted from the cumulative distribution $F(t)$ and vice versa.

The time evolution of the log-productivity distribution $P(t)$ for different periods and $q = 1$ starting from an exponential initial distribution in the

[9]For a precise notion of sufficiently smooth see the proof.

case of pure imitation can be seen in Figure 4. The figure reveals that the distribution concentrates its mass at the maximum initial productivity level as time increases.

5. Equilibrium Productivity Growth

In this section we study the case in which firms face a non-trivial choice can increase productivity through both innovation (in-house R&D) or the imitation of other firms. The decision on which of the two strategies gives it a higher expected utility from profits. The basis of expected profits is the expected value of productivity obtained through in-house R&D and through imitation, given their current log-productivity level, the realization of the productivity shock, and the current distribution of log-productivities in the population.

We first consider expected the expected productivity of firm i if it innovates. We have that

$$A_i^{\mathrm{in}}(t) := \mathbb{E}_i^{\mathrm{in}}(\bar{A}^{a_i(t)+\eta(t)}) = \bar{A}^{a_i(t)}\mathbb{P}(\eta(t)=0) + \bar{A}^{a_i(t)+1}\mathbb{P}(\eta(t)=1) + \dots$$

Assuming that $\mathbb{P}(\eta(t)=1) = p$ and $\mathbb{P}(\eta(t)=0) = 1-p$ for some $p \in (0,1)$ we obtain

$$A_i^{\mathrm{in}}(t) = \bar{A}^{a_i(t)}(1-p+\bar{A}p).$$

It follows that

$$a_i^{\mathrm{in}}(t) := \log A_i^{\mathrm{in}}(t) = a_i(t) + \log(1-p+\bar{A}p),$$

where we have assumed that $\log \bar{A} = 1$. Next, we consider imitation. Let $A_i^{\mathrm{im}}(t) := \mathbb{E}_i^{\mathrm{im}}(A_i(t))$ be the expected productivity of firm i if it imitates. We then have that[10]

$$A_i^{\mathrm{im}}(t) = e^{a_i(t)}S_{a_i(t)}(P(t)) + \sum_{b=a_i(t)+1}^{\infty} e^b q^{b-a}\left(P_b(t) + (1-q)(1-F_b(t))\right).$$

[10] For $q=1$ this simplifies to

$$A_i^{\mathrm{im}}(t) = e^{a_i(t)}F_{a_i(t)}(t) + \sum_{b=a_i(t)+1}^{\infty} e^b P_b(t) = e^{a_i(t)}\left(F_{a_i(t)}(t) + \sum_{b=a_i(t)+1}^{\infty} e^{b-a_i(t)}P_b(t)\right).$$

We then can write

$$a_i^{\mathrm{im}}(t) := \log A_i^{\mathrm{im}}(t) = a_i(t) + \log\left(S_{a_i(t)}(t) + \sum_{b=a_i(t)+1}^{\infty} e^{b-a_i(t)} q^{b-a} \left(P_b(t) + (1-q)(1-F_b(t))\right)\right).$$

The productivity shock is assumed to be strategy-specific, $\epsilon_i\left(t\right) \in \{\epsilon_i^{in}\left(t\right), \epsilon_i^{im}\left(t\right)\}$. Firms observe the realizations of the $\epsilon_i\left(t\right)$ before deciding whether to undertake in-house R&D or innovation. The expected profit of firm i when innovating in-house is given by

$$\pi_i^{\mathrm{in}}(t) = \psi A_i^{\mathrm{in}}(t)\epsilon_i^{\mathrm{in}}(t)$$

and similarly, the expected profit of firm i through imitation is

$$\pi_i^{\mathrm{im}}(t) = \psi A_i^{\mathrm{im}}(t)\epsilon_i^{\mathrm{im}}(t),$$

where $\epsilon_i^{\mathrm{in}}(t)$ and $\epsilon_i^{\mathrm{im}}(t)$ are i.i.d. non-negative random variables. Then the *ex-ante* probability that a firm's profit from innovation is larger than from imitation is given by

$$\begin{aligned}\mathbb{P}(\pi_i^{\mathrm{im}}(t) > \pi_i^{\mathrm{in}}(t)) &= \mathbb{P}(A_i^{\mathrm{im}}(t)\epsilon_i^{\mathrm{im}}(t) > A_i^{\mathrm{in}}(t)\epsilon_i^{\mathrm{in}})(t) \\ &= \mathbb{P}(\log\epsilon_i^{\mathrm{in}}(t) - \log\epsilon_i^{\mathrm{im}}(t) < a_i^{\mathrm{im}}(t) - a_i^{\mathrm{in}}(t)).\end{aligned}$$

In order to make our model more tractable we make a specific assumption on the distribution of the shocks. More precisely, we assume that $\epsilon_i^{\mathrm{in}}(t)$ and $\epsilon_i^{\mathrm{im}}(t)$ are independently type-II extreme value (Frechet) distributed with parameter $\beta \geq 0$. It then follows that $\log\epsilon_i^{\mathrm{in}}(t)$ and $\log\epsilon_i^{\mathrm{im}}(t)$ are independently type-I extreme value (Gumbel) distributed with parameter $1/\beta$. Under this assumption it follows that the probability that the firm chooses imitation rather than in-house R&D is given by [see e.g. Anderson et al., 1992][11]

$$p_\beta^{\mathrm{im}}(a_i(t),P(t)) = \mathbb{P}\left(\pi_i^{\mathrm{im}}(t) > \pi_i^{\mathrm{in}}(t)\right) = \frac{e^{\beta a_i^{\mathrm{im}}(a_i(t),P(t))}}{e^{\beta a_i^{\mathrm{im}}(a_i(t),P(t))} + e^{\beta a_i^{\mathrm{in}}(a_i(t))}} = \frac{1}{1+e^{-\beta(a_i^{\mathrm{im}}(a_i(t),P(t)) - a_i^{\mathrm{in}}(a_i(t)))}}. \tag{20}$$

With the previously derived expressions of $a^{\mathrm{in}}(a_i(t))$ and $a^{\mathrm{im}}(a_i(t),P(t))$ we can write the innovation probability as follows

$$p_\beta^{\mathrm{in}}(a_i(t),P(t)) = \frac{1}{1+\left(\frac{S_{a_i(t)}(t)+\sum_{b=a_i(t)+1}^{\infty} e^{b-a_i(t)} q^{b-a}(P_b(t)+(1-q)(1-F_b(t)))}{1-p+\bar{A}p}\right)^{\frac{\beta}{\ln(\bar{A})}}}. \tag{21}$$

Next, we define $\mathbf{D}(P)$ as the diagonal-matrix of all probabilities $p_\beta^{\mathrm{im}}(a,P)$, i.e.

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

Putting the above results together, the conditional transition matrix $\mathbf{T}(P)$ with elements $(T_{ab}(P))_{a,b\in S}$ can be written in the compact matrix form

$$\mathbf{T}(P) = (\mathbf{I} - \mathbf{D}(P))\,\mathbf{T}^{\mathrm{in}} + \mathbf{D}(P)\mathbf{T}^{\mathrm{im}}(P). \tag{22}$$

With the above definition of the transition matrix $\mathbf{T}$ we are now ready to introduce the dynamics of the log-productivity distribution $P(t)$.

Proposition 4. *In the limit of a large number N of firms, the evolution of the* log*-productivity distribution $P(t)$ for the Markov chain $(P^N(t))_{t\in T}$ is given by*

$$\frac{dP(t)}{dt} = P(t)(\mathbf{T}(P)-\mathbf{I}) = P(t)(\mathbf{D}(P)\mathbf{T}^{im}(P) + (\mathbf{I}-\mathbf{D}(P))\mathbf{T}^{in} - \mathbf{I}), \tag{23}$$

for some initial distribution $P(0) \geq 0$.

[11] If $\log\epsilon_1$ and $\log\epsilon_2$ have a cdf $F_{\log\epsilon}(x) = \exp(-e^{-\beta x})$ then their difference $\Delta\log\epsilon = \log\epsilon_2 - \log\psi_1$ has a logistic cdf $F_{\Delta\log\epsilon}(x) = 1/(1+e^{-\beta x})$.

Observe that Equation (23) can be solved for a function $P(t)$ that is continuous in $a \in \mathbb{R}_+$. The resulting solution trajectory coincides with the one for discrete values of a if we evaluate it only at the discrete points $a \in S$, because the evolution of $P(t)$ at the discrete values of a is independent of any values not coinciding with the discrete ones. Hence, in the following, we will consider $P(t)$ being continuous in both time t and log-productivity a. Accordingly, we will replace the derivative with respect to time in Equation (23) with a partial derivative.

Equation (23) can be rewritten in matrix-vector notation as

$$\begin{aligned}\frac{\partial P(t)}{\partial t} &= \left[P_1(t)p_\beta^{\mathrm{im}}(1,P), P_2(t)p_\beta^{\mathrm{im}}(2,P), \dots\right] \begin{bmatrix} S_1(t) & P_2(t) & P_3(t) & \dots \\ 0 & S_2(t) & P_3(t) & \dots \\ & & \ddots & \ddots \end{bmatrix} \\ &\quad + \left[P_1(t)(1-p_\beta^{\mathrm{im}}(1,P)), P_2(t)(1-p_\beta^{\mathrm{im}}(2,P)), \dots\right] \begin{bmatrix} 1-p & p & 0 & \dots \\ 0 & 1-p & p & \dots \\ & & \ddots & \ddots \end{bmatrix} \\ &\quad - \left[P_1(t), P_2(t), \dots\right].\end{aligned}$$

For log-productivity a the dynamics is then given by

$$\begin{aligned}\frac{\partial P_a(t)}{\partial t} &= P_a(t)\left(P_1(t)p_\beta^{\mathrm{im}}(1,P) + \dots + P_{a-1}(t)p_\beta^{\mathrm{im}}(a-1,P) + S_a(t)p_\beta^{\mathrm{im}}(a,P)\right) \\ &\quad + (1-p)P_a(t)\left(1-p_\beta^{\mathrm{im}}(a,P)\right) + pP_{a-1}(t)\left(1-p_\beta^{\mathrm{im}}(a-1,P)\right) - P_a(t) \\ &= P_a(t)\left(\sum_{b=1}^{a-1} p_\beta^{\mathrm{im}}(b,P)P_b(t) + p_\beta^{\mathrm{im}}(a,P)S_a(t)\right) + (1-p)P_a(t)(1-p_\beta^{\mathrm{im}}(a,P)) \\ &\quad + pP_{a-1}(t)(1-p_\beta^{\mathrm{im}}(a-1,P)) - P_a(t). \qquad (24)\end{aligned}$$

With Equation (24) the evolution of the productivity distribution is completely determined and can always be computed by numerical iteration.[12]

[12] Note that if $q = 1$ Equation (24) yields

$$\begin{aligned}\frac{\partial P_a(t)}{\partial t} &= P_a(t)\left(\sum_{b=1}^{a-1} p_\beta^{\mathrm{im}}(b,P)P_b(t) + p_\beta^{\mathrm{im}}(a,P)F_a(t)\right) + (1-p)P_a(t)(1-p_\beta^{\mathrm{im}}(a,P)) \\ &\quad + pP_{a-1}(t)(1-p_\beta^{\mathrm{im}}(a-1,P)) - P_a(t). \qquad (25)\end{aligned}$$

However, in order to better understand the emerging log-productivity distributions and their dependency on the model parameters, we analyze in the following sections two limit cases. In Section 5.1 we study the weak selection limit ($\beta \to 0$) where firms choose randomly between innovation and imitation, while in Section 5.2 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. In both cases we will show that the productivity distribution is a traveling wave with tails that exhibit a power law behavior.

5.1. 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 $p_\beta^{\mathrm{im}}(a) = 0.5$ for all $1 \leq a < \infty$. Inserting this into Equation (24) gives

$$\frac{\partial P_a(t)}{\partial t} = \frac{1}{2} P_a(t) \left(\sum_{b=1}^{a-1} P_b(t) + S_a(t) \right) + \frac{1-p}{2} P_a(t) + \frac{p}{2} P_{a-1}(t) - P_a(t).$$

Similar to the results obtained in Sections 4.2 and 4.3 we can derive from the above equation the dynamics of the cumulative distribution function $F(a,t)$, in the limit of q close to one

$$\frac{\partial F_a(t)}{\partial t} = \frac{2q-1}{2} (F(a,t)^2 - F(a,t)) - \frac{p}{2} \left(F(a,t) - F(a-1,t) \right). \tag{26}$$

This differential equation for $F(a,t)$ can be solved numerically subject to the boundary conditions $\lim_{a \to \infty} F(a,t) = 1$ and $\lim_{a \to 0} F(a,t) = 0$. An example of the resulting probability mass function is given in Figure 5. Our analysis reveals that the distribution obtains a stable shape moving to the right (with increasing log-productivity) over time. Such a solutions is called a *traveling wave*. More precisely, a traveling wave is a solution of the form $F_a = f(a - \nu t)$ such that for any $s \geq t$ it must hold that $F_a(t) = F_{a+\nu s}(t+s)$. This is stated in the next proposition.

Proposition 5. *Let $F_a(t)$ be a solution of Equation (26) with Heaviside initial data $F(a,0) = \Theta(a - \bar{a})$ for some $\bar{a} \geq 0$ and define $m_\epsilon(t) = \inf\{a : F_a(t) > \epsilon\}$. Then*

$$\lim_{t \to \infty} \frac{m_\epsilon(t)}{t} = \nu$$

for some constant $\nu \geq 0$, and $F_a(t)$ is a traveling wave of the form

$$F_a = f(a - \nu t)$$

for some non-decreasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$.

In the following we derive the precise shape of the traveling wave solution of Equation (26). For this purpose, it is important to observe that in the limit of $q = 1$ we can build on the model analyzed by Majumdar and Krapivsky [2001].

Proposition 6. *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. Consider Heaviside initial conditions $F(a, 0) = \Theta(a - \bar{a})$ for some $\bar{a} \geq 0$.*

(i) If we assume that the front of the traveling wave solution of Equation (26) follows an exponential distribution with exponent $\lambda \geq 0$ for a much larger than νt, i.e. $P(a, t) \propto e^{-\lambda(a-\nu t)}$, then the traveling wave velocity ν is given by

$$\nu = \frac{2q - 1 - p + pe^{\lambda}}{2\lambda}, \tag{27}$$

where the exponent λ of the front of the distribution is given by[13]

$$\lambda = 1 + W\left(\frac{2q - 1 - p}{pe}\right). \tag{28}$$

(ii) If we assume that $P(a, t) \propto e^{\rho(a-\nu t)}$, $\rho \geq 0$, for a much smaller than νt then the exponent ρ is given by

$$\rho = \frac{1}{2}\left(2q - 1 + p + 2W\left(-\frac{p}{2}e^{\frac{1-p-2q}{2}})\right)\right). \tag{29}$$

Note that the assumption of a power law tail in Proposition 6 is not very restrictive, as such tail distributions are common for a broad class of probability distributions [Alfarano and Lux, 2010]. Our numerical analysis shows that the results of Proposition 6 hold also for other initial distributions which are concentrated enough, such as an exponential distribution with an exponent that is large enough. Moreover, one can shown that the average

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

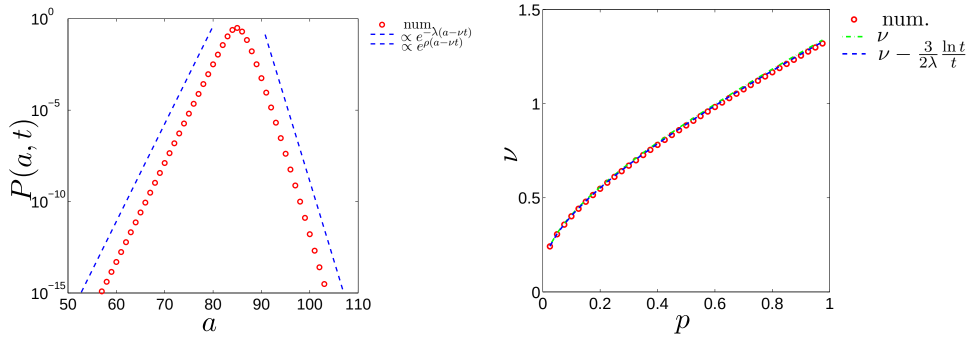

Figure 5: (Left) The log-productivity distribution $P(a,t)$ for period $t = 200$, $p = 0.119$, $q = 1$. The distribution obtained by numerical integration of Equation (38) is indicated by a circles while the theoretical predictions are shown with a dashed line. The front of the traveling wave decays as a power-law with exponent $\lambda = 2$. (Right) Traveling wave velocity ν for different values of $p \in [0,1]$ and $q = 1$ at $t = 100$.

log-productivity is given by[14]

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

By solving the continuous dynamical system corresponding to Equation (26) we can compute $F(a,t)$ for large times t. The resulting traveling wave velocity ν can be seen in Figure 5 (right) 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.[15] Further, using Equation (29) we can determine the exponent ρ for this value of p yielding $\rho = 1.231$. This can be seen in Figure 5 for time $t = 200$ and setting $q = 1$.

In the following we will derive some intuition for what happens in the case of β small but positive. As we will show, our analysis reveals that with increasing values of β the decisions of individual firms will be closer their optimal choice and this translates to the aggregate level enhancing the productivity growth rate of the economy. To simplify our analysis we set $q = 1$. We can give the following proposition.

[14]See Majumdar and Krapivsky [2001].

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

Proposition 7. *Assume that*

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

and that for β small enough ρ is given by Equation (29). Then the traveling wave velocity is given by

$$\nu = \frac{1}{\lambda}\left(\frac{1+\gamma}{2+\gamma} - \frac{\left(e^{\lambda}-1\right)\left(e^{\rho}-1\right)}{e^{\lambda+\rho}-1}\left(\sum_{b=1}^{\infty}\frac{e^{-\lambda b}}{2+\gamma\left(1+Ae^{-\lambda b}\right)} + \sum_{b=0}^{\infty}\frac{e^{-\rho b}}{2+\gamma\left(Be^{-\rho b}+Ce^{b}\right)}\right)\right.$$
$$\left. + \frac{1}{2+\gamma}\left(1+p\left(e^{\lambda}-1\right)\right)\right)$$

where λ is given by the root of $\frac{d\nu}{d\lambda}=0$,

$$A = \frac{(e-1)\left(e^{\rho}-1\right)e^{\lambda}}{\left(e^{\lambda}-e\right)\left(e^{\lambda+\rho}-1\right)},$$
$$B = \frac{\left(e^{\lambda-1}\right)(e-1)e^{\rho}}{\left(e^{\lambda+\rho}-1\right)\left(e^{1+\rho}-1\right)},$$
$$C = \frac{e\left(e^{\lambda}-1\right)\left(e^{\rho}-1\right)}{\left(e^{\lambda}-e\right)\left(e^{1+\rho}-1\right)},$$

and $\gamma = \beta/(\ln(\bar{A})(1+p(\bar{A}-1)))$.

The expression for $\frac{d\nu}{d\lambda}=0$ can be found in the proof of Proposition 7 in Appendix A. A comparison of the traveling wave velocity ν for $\beta = 0$, $\beta = 0.05$ and $\beta = 0.1$ is given in Figure 6. We find that with increasing values of β the velocity and hence the average productivity growth rate increases.

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

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, where we assume that $q = 1$.

Proposition 8. *When $p > 0, \beta > 0, q = 1$ and the initial distribution $P(0)$ has a support which is an interval of integers (possibly infinite[16]), then there*

[16] Either $\{a_1, \overset{+1}{\ldots}, a_2\}$, $\{-\infty, \overset{+1}{\ldots}, a_2\}$, or $\{a_1, \overset{+1}{\ldots}, \infty\}$ for some $a_1 < a_2$ both integers.

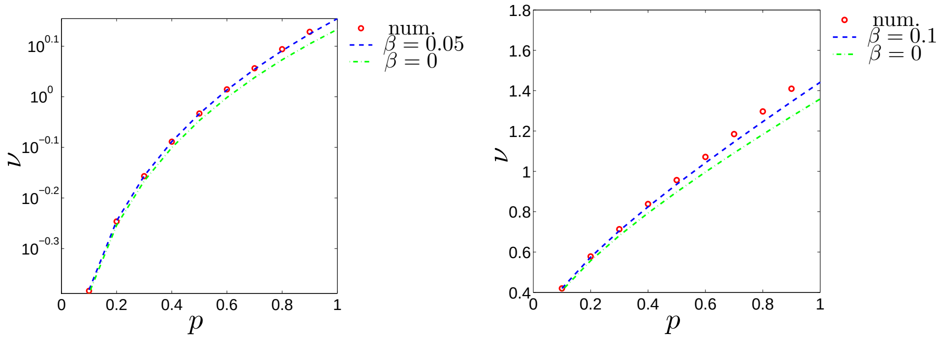

Figure 6: Traveling wave velocity $\nu(\lambda)$ as a function of $p \in (0, 1)$ assuming that $\ln(\bar{A}) = 1$. Results of numerical integration of Equation (24) are shown with circles. (Left) The solid line corresponds to $\beta = 0.05$ and (right) $\beta = 0.1$, while the dashed-dotted line corresponds to a value of $\beta = 0$. We see that the velocity for $\beta > 0$ is always higher than for $\beta = 0$.

exists a unique $a^* \in \mathbb{N}$ such that for all t it holds $p_\beta^{im}(a, P) > p_\beta^{in}(a, P)$ when $a < a^*$, and $p_\beta^{im}(a, P) < p_\beta^{in}(a, P)$ when $a > a^*$. (It might hold $p_\beta^{im}(a^*, P) = p_\beta^{in}(a^*, P)$.)

In the case of $\beta \to \infty$ it follows from Proposition 8 that there exists a threshold log-productivity a^* such that

$$\lim_{\beta \to \infty} p_\beta^{\text{im}}(a, P) = \begin{cases} 1, & \text{if } a \leq a^*, \\ 0, & \text{if } a > a^*. \end{cases} \tag{30}$$

Note that the requirement of Lipschitz continuity of $V(P)$ in Theorem 1 is violated in the limit $\beta \to \infty$ at the threshold log-productivity a^* and consequently, Equations (23) and (24), respectively, are not shown to hold. They hold, however, for any large but finite value of β. When β is large enough we can write $p_\beta^{\text{im}}(a, P) = 1 - \epsilon$ for $a \leq a^*$ and $p_\beta^{\text{im}}(a, P) = \epsilon$ for $a > a^*$ for some small $\epsilon > 0$ (which becomes arbitrarily small with increasing values of β). In this section we assume that we can neglect all terms of the order $\mathcal{O}(\epsilon)$ in the dynamics of $P(t)$ of Equation (24) as β becomes large enough (but finite). We then can state the following proposition.

Proposition 9. *Let $q = 1$ and assume that Equation (30) holds for β large enough (but finite). Then the dynamics of the cumulative* log-*productivity* *distribution is given by*

$$\frac{\partial F_a(t)}{\partial t} = \begin{cases} F_a(t)^2 - F_a(t), & \text{if } a \leq a^*, \\ F_{a^*}(t) - (1 - P_{a^*+1}(t))F_{a^*}(t) - pP_{a^*+1}(t), & \text{if } a = a^* + 1, \\ -(1 - F_a(t))F_{a^*}(t) - p(F_a(t) - F_{a-1}(t)), & \text{if } a \geq a^* + 1. \end{cases} \tag{31}$$

The above difference-differential Equation (31) for $F(a,t)$ can be solved numerically subject to the boundary conditions $\lim_{a\to\infty} F(a,t) = 1$ and $\lim_{a\to 1} F(a,t) = 0$. The log-productivity distribution for $p = 0.1$ and $q = 1$ obtained by means of numerical integration of Equation (31) can be seen in Figure 7.

Similar to the results obtained in the previous section, a numerical integration of Equation (31) reveals that the limiting log-productivity distribution is a traveling wave with power-law tails. This is further analyzed in the next proposition.

Proposition 10. *Let $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$ with $p \in [0,1]$. Consider $\beta = \infty$, $q = 1$ and Heaviside initial conditions $F(a,0) = \Theta(a - \bar{a})$ for some $\bar{a} \geq 0$. Further, assume that Equation (31) holds for β large enough (but finite) and that its solution is a traveling wave.*

(i) If we assume that the front of the traveling wave solution of Equation (31) follows an exponential distribution with exponent $\lambda \geq 0$ for all $a \geq a^* = \nu t$, i.e. $P(a,t) \propto e^{-\lambda(a-\nu t)}$, then the traveling wave velocity ν is given by

$$\nu = \frac{1}{\lambda}\left(1 + p(e^{\lambda} - 1) - \frac{p(\bar{A} - 1)(1 - e^{1-\lambda})}{e - 1}\right) \tag{32}$$

where λ is given by the root of the equation

$$e^{\lambda}(\lambda - 1) - \frac{\bar{A} - 1}{e - 1}e^{1-\lambda}(1 + \lambda) + \frac{\bar{A} + e - 2}{e - 1} = \frac{1}{p}. \tag{33}$$

(ii) If we assume that the rear of the traveling wave solution of Equation (31) follows an exponential function with exponent $\rho \geq 0$ for all $a < a^*$, i.e. $P(a,t) \propto e^{\rho(a-\nu t)}$, then the exponent ρ is given by $\rho = 1/\nu$.

Equation (33) can be solved numerically, using standard numerical root finding procedures [see e.g. Press et al., 1992, Chap. 9], to obtain the exponent

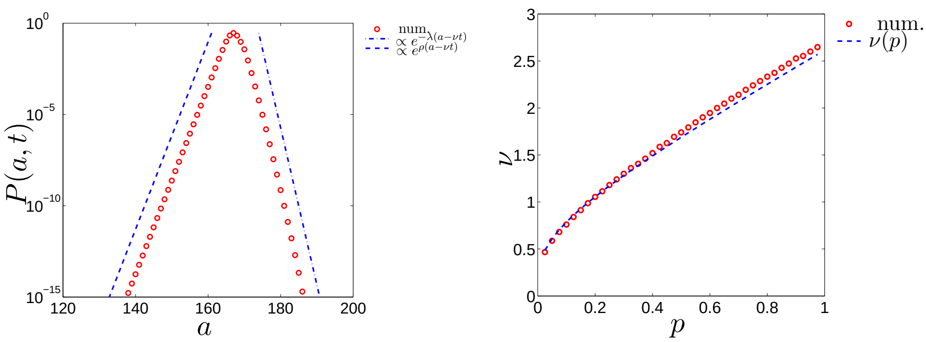

Figure 7: (Left) The log-productivity distribution $P(a,t)$ for $p = 0.125$ and $\ln \bar{A} = 1$ at $t = 400$. The distribution obtained by numerical integration of Equation (31) is indicated with circles while the theoretical predictions are shown with a dashed line. The front of the traveling wave is close to a power-law with exponent λ of 2. (Right) Traveling wave velocity ν for different values of $p \in [0,1]$ by means of numerical integration of Equation (31) and theoretical prediction indicated by the dashed line.

λ. We find that we can generate distributions with power-law tails that reproduce our findings in in Section 2. Inserting λ into Equation (32) further gives the traveling wave velocity ν. This is shown in Figure 7 (right). A comparison of the theoretical predictions for the exponents λ and ρ from Proposition 10 with the numerical log-productivity distribution $P(a,t)$ for $p = 0.1$ can be seen in Figure 7. Finally, note that our numerical analysis shows that Proposition 10 holds also for generic initial distributions which are concentrated enough, such as an exponential distribution with an exponent that is large enough.

6. Efficiency and Inequality

In this section we first turn to the analysis of industry performance and efficiency. An industry has a higher performance, measured in aggregate intermediate goods and final good production, if it has a higher average log-productivity.[17] 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

[17] 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-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.

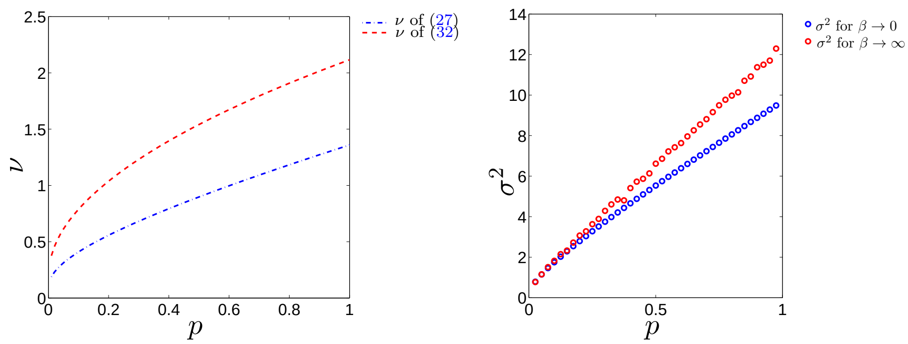

Figure 8: (Left) Traveling wave velocities ν in the strong ($\beta \to \infty$) and weak ($\beta \to 0$) selection limits. (Right) Variances σ^2 for the same values of β.

selection limits.

Proposition 11. *Assume that* $q = 1$*. Then, under the assumptions of Propositions 6 and 10, we have that for any* $p \in (0, 1)$

$$\lim_{\beta \to 0} \nu(\beta, p) < \lim_{\beta \to \infty} \nu(\beta, p).$$

Proposition 11 implies that aggregate productivity and output are higher if firms perceive expected profits from innovation and imitation with vanishing noise than with strong 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 8 (left). The figure confirms the result of Proposition 11. For all values of the innovation probability $p \in [0, 1]$, the traveling wave velocity ν is higher the better firms can evaluate the expected outcomes of their R&D strategies. Next, we can investigate the variance σ^2 in the log-productivity distribution as a measure for inequality in the economy. This is done in Figure 8 (right) with respect to the two extreme cases of $\beta \to 0$ and $\beta \to \infty$ by numerical integration of Equations (26) and (31). We see that the variance is always higher in the strong selection limit and the difference increases with increasing innovation probability p. Taking into account the previous efficiency results we hence find that a higher industry performance comes along with a higher inequality in firm's productivities. This finding supports previous works that highlight concentration as a typical characteristic of efficient industries [see e.g. Westbrock, 2010, for a recent example].

7. 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 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 reduce industry performance and efficiency while at the same time increase inequality.

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, one could extend our framework by introducing heterogeneous interactions in the form of a network in the imitation process and analyze the emerging productivity distributions, such as in Di Matteo et al. [2005]; Ehrhardt et al. [2006]; Kelly [2001]. This is beyond the scope of the present paper and we leave this avenue for future research.

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.

Alfarano, S., Lux, T., 2010. Extreme Value Theory as a Theoretical Background for Power Law Behavior, Kiel Working Papers, Kiel Institute for the World Economy,.

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.
URL http://www.nber.org/papers/w14135

Anderson, S. A., De Palma, A., Thisse, J., 1992. Discrete Choice Theory of Product Differentiation. MIT Press.

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

Barro, R. J., Sala-i Martin, X., 2004. Economic Growth. MIT Press.

Benaim, M., Weibull, J., 2003. Deterministic approximation of stochastic evolution in games. Econometrica, 873–903.

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).

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.
Howitt, P., 2000. Endogenous growth and cross-country income differences. American Economic Review 90 (4), 829–846.
Howitt, P., Mayer-Foulkes, D., 2002. R&D, Implementation and Stagnation: A Schumpeterian Theory of Convergence Clubs. SSRN eLibrary.
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.
Kelly, M., 2001. Linkages, thresholds, and development. Journal of Economic Growth 6 (1), 39–53.
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.
Kurtz, T., 1970. Solutions of ordinary differential equations as limits of pure jump Markov processes. Journal of Applied Probability 7 (1), 49–58.
Kurtz, T., 1971. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. Journal of Applied Probability 8 (2), 344–356.
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., 1980. Production sets, technological knowledge, and R & D: Fragile and overworked constructs for analysis of productivity growth? The American Economic Review 70 (2), 62–67.
Nelson, R. R., Winter, S. G., 1982. An evolutionary theory of economic change. Cambridge, Mass. : Belknap Press of Harvard University Press.
Nelson, R. R., Winter, S. G., Schuette, H. L., 1976. Technical change in an evolutionary model. The Quarterly Journal of Economics 90 (1), pp. 90–118.
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., Mar. 1997. Empirics for growth and distribution: Stratification, polarization, and convergence clubs. Journal of Economic Growth 2 (1), 27–59.
URL http://dx.doi.org/10.1023/A:1009781613339
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.
Sandholm, W., 2006. Population games and evolutionary dynamics.
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.
Stoneman, P., 2002. The economics of technological diffusion. Wiley-Blackwell.
Weitzman, M. L., 1998. Recombinant growth. The Quarterly Journal of Economics 113 (2), 331–360.
Westbrock, B. (2010). Natural concentration in industrial research collaboration. *The RAND Journal of Economics*, 41(2):351–371.

Appendix

A. Proofs of Propositions, Corollaries and Lemmas

Proof of Proposition 4. In the following we introduce the random variable ζ_P^N whose distribution describes the stochastic increments of $(P^N(t))_{t\in T}$ from the state $P \in P^N$

$$\mathbb{P}\left(\zeta_P^N = z\right) = \mathbb{P}\left(P^N(t+\Delta t) = P + z \middle| P^N(t) = P\right). \tag{34}$$

Moreover, we introduce the functions V^N, A^N and A_δ^N by

$$\begin{aligned} V^N(P) &= N\mathbb{E}(\zeta_P^N), \\ A^N(P) &= N\mathbb{E}(|\zeta_P^N|), \\ A_\delta^N(P) &= N\mathbb{E}\left(\left|\zeta_P^N I_{\{|\zeta_P^N|>\delta\}}\right|\right). \end{aligned}$$

We then can state the following theorem by Kurtz [1970, 1971]:[18]

Theorem 1. *Let $V : \mathbb{R}^{|S|} \to \mathbb{R}^{|S|}$ be a Lipschitz continuous vector field. Suppose that for some sequence $(\delta^N)_{N=N_0}^{\infty}$ with $\lim_{N\to\infty} \delta^N = 0$, it holds that*

(i) $\lim_{N\to\infty} \sup_{P\in P^N} \left|V^N(P) - V(P)\right| = 0$,

(ii) $\sup_N \sup_{P\in P^N} A^N(P) < \infty$, *and*

(iii) $\lim_{N\to\infty} \sup_{P\in P^N} A_{\delta^N}^N(P) = 0$,

and that the initial conditions $P(0)^N = P_0^N$ converge to P_0. Let $\{P(t)\}_{t\geq 0}$ be the solution of the mean-field dynamics

$$\frac{dP}{dt} = V(P) \tag{35}$$

starting from P_0. Then for each $T < \infty$ and $\epsilon > 0$, we have that

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{t\in[0,T]} \left|P(t)^N - P(t)\right| < \epsilon\right) = 1.$$

[18] See also Sandholm [2006, Chap.10.2].

In the following we prove that the conditions (i) to (iii) in Theorem 1 hold for our framework. First, observe that

$$\begin{aligned}
V^N(P) &= N\mathbb{E}(\zeta_P^N) \\
&= N \sum_{a,b\geq 1} \frac{1}{N}(e_b - e_a)\mathbb{P}\left(\zeta_P^N = \frac{1}{N}(e_b - e_a)\right) \\
&= N \sum_{a,b\geq 1} \frac{1}{N}(e_b - e_a)P_a T_{ab}(P) \\
&= \sum_{a\geq 1} e_a \left(\sum_{b\geq 1} P_b T_{ba}(P) - P_a \sum_{b\geq 1} T_{ab}(P)\right) \\
&= \sum_{a\geq 1} e_a V_a(P) = V(P)
\end{aligned}$$

which is independent of N. This implies that condition (i) in Theorem 1 is satisfied. Note also that since $T_{ab}(P)$ is continuously differentiable, $V(P)$ is a Lipschitz continuous function as required. Further, observe that since $|e_a - e_b| = \sqrt{2}$ for $a \neq b$ and 0 otherwise, $(P^N(t))_{t\in T}$ has jumps of at most $\sqrt{2}/N$. Hence, for $\delta^N = \sqrt{2}/N$

$$A^N_{\delta^N}(P) = N\mathbb{E}\left(\left|\zeta_P^N I_{\{|\zeta_P^N| > \sqrt{2}/N\}}\right|\right) = 0,$$

and condition (iii) in Theorem 1 holds. Finally, we find that

$$A^N(P) = N\mathbb{E}(|\zeta_P^N|) \leq N\frac{\sqrt{2}}{N} = \sqrt{2} < \infty,$$

and also condition (ii) in in Theorem 1 is satisfied.

Theorem 1 tells us that when the number of firms N is large, nearly all sample paths of the Markov chain $(P^N(t))_{t\in T}$ stay within a small ϵ of the solution of the mean-field dynamics of Equation (35), which can be written in the compact form $dP(t)/dt = P(t)(\mathbf{T}(P) - \mathbf{I})$. □

Proof of Proposition 1. 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 1. 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 2. In the general case of $q \in [0, 1]$ the evolution of the cumulative log-productivity distribution is given by

$$\begin{aligned}
\frac{\partial F_a(t)}{\partial t} &= P_a(1-q)(1-F_a) + P_a F_a \\
&+ P_{a-1}q(1-q)(1-F_a) + P_{a-1}(1-q)(1-F_a) + P_{a-1}F_a \\
&+ P_{a-2}q^2(1-q)(1-F_a) + P_{a-2}q(1-q)(1-F_a) + P_{a-2}(1-q)(1-F_a) + P_{a-2}F_a \\
&+ \dots \\
&- F_a.
\end{aligned}$$

This can be written as

$$\frac{\partial F_a(t)}{\partial t} = F_a(t)^2 + (1-q)(1-F_a(t)) \sum_{b=0}^{a-1} q^b F_{a-b}(t) - F_a(t).$$

□

Proof of Proposition 3. From Equation (13) we derive

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

Summation over a yields

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

The cumulative log-productivity distribution follows the recursive relation

$$\frac{\partial F(a,t)}{\partial t} = F(a,t)^2 - F(a,t). \tag{36}$$

This can be written as

$$\frac{\partial \ln F(a,t)}{\partial t} = F(a,t) - 1,$$

with the solution

$$F(a,t) = \frac{F(a,0)}{F(a,0) + e^t(1 - F(a,0))},$$

and the initial distribution $F(a,0)$. □

Proof of Corollary 2. From Equation (14) we see that for all $a \geq \bar{a}$ it must hold that $\partial F_a(t)/\partial t = 0$ and so $F_a(t) = 1$ for all $t \geq 0$. In contrast, for all $a < \bar{a}$ and $q > 0$ there exists a positive probability that a firm with log-productivity $b > a$ is imitated, leading to a decrease in $F_a(t)$. Eventually, we then have that $\lim_{t\ to\infty} F_a(t) = 0$ for all $a < \bar{a}$. □

Proof of Corollary 3. We can write Equation (14) as

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

This is

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

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

$$\frac{\partial F(a,t)}{\partial t} = qP(a,t)G(a,t),$$

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

Proof of Corollary 4. Equation (14) can be written as

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

From this we obtain

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

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

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

$\square$

In the following we derive a lemma and a corollary which will help us to show that Equation (26) admits a traveling wave solution with a stable shape. This result is given in Proposition 5.[19]

First, from Equation (26) we can derive the following lemma:

Lemma 1. *Let $F_a^{(1)}(t)$ and $F_a^{(2)}(t)$ be solutions of Equation (26) with initial data chosen such that $F_a^{(1)}(0) \geq F_a^{(2)}(0)$. Then for all $t > 0$ we have that $F_a^{(1)}(t) \geq F_a^{(2)}(t)$.*

Proof of Lemma 1. We introduce the difference

$$V_a(t) = F_a^{(2)}(t) - F_a^{(1)}(t).$$

In the following we show that if $V_a(0) \leq 0$ then $V_a(t) \leq 0$ for all $t > 0$. We can write Equation (26) as follows

$$\frac{\partial F_a(t)}{\partial t} + F_a(t) = \frac{2q-1}{2}F_a(t)^2 + \frac{3-2q-p}{2}F_a(t) + \frac{p}{2}F_{a-1}(t).$$

[19] Our results are heavily inspired by Bramson [1983], who analyzed the traveling wave solution $u(x,t) = w(x - \nu t)$ of the Kolmogorov equation $\frac{\partial u}{\partial t} = f(u) + \frac{\partial^2 u}{\partial x^2}$.

We then get for $V_a(t)$

$$
\begin{aligned}
\frac{\partial V_a(t)}{\partial t} + V_a(t) &= \frac{2q-1}{2}((F_a^{(2)}(t))^2 - (F_a^{(1)}(t))^2) + \frac{3-2q-p}{2} V_a(t) + \frac{p}{2} V_{a-1}(t) \\
&= \underbrace{\frac{2q-1}{2}}_{\geq 0} \underbrace{V_a(t)}_{\leq 0} \underbrace{(F_a^{(2)}(t) + F_a^{(1)}(t))}_{\geq 0} + \underbrace{\frac{3-2q-p}{2}}_{\geq 0} \underbrace{V_a(t)}_{\leq 0} + \underbrace{\frac{p}{2}}_{\geq 0} \underbrace{V_{a-1}(t)}_{\leq 0}.
\end{aligned}
$$

Hence, we find that if $V_a(t) \leq 0$ for all $a \geq 0$ then also $\partial V_a(t)/\partial t + V_a(t) \leq 0$.

Next, we show that if $V_a(t) \leq 0$ and $\partial V_a(t)/\partial t + V_a(t) \leq 0$ then also $V_a(t+s) \leq 0$ for all $s > 0$. For this purpose, let $\epsilon = s/n$ with $n \in \mathbb{N}$. For n being sufficiently large (and ϵ sufficiently small) we can use a first-order Taylor approximation to write

$$
\begin{aligned}
V_a(t+\epsilon) &= V_a(t) + \frac{\partial V_a(t)}{\partial t}\epsilon \\
V_a(t+2\epsilon) &= V_a(t+\epsilon) + \frac{\partial V_a(t+\epsilon)}{\partial t}\epsilon \\
&\vdots \\
V_a(t+n\epsilon) &= V_a(t+(n-1)\epsilon) + \frac{\partial V_a(t+(n-1)\epsilon)}{\partial t}\epsilon
\end{aligned}
$$

We can assume that $V_a(t) \leq 0$. If $\partial V_a(t)/\partial t \leq 0$ then we also have that $V_a(t+\epsilon) \leq 0$. Otherwise, we observe that

$$
V_a(t+\epsilon) = V_a(t) + \frac{\partial V_a(t)}{\partial t}\epsilon \leq V_a(t) + \frac{\partial V_a(t)}{\partial t} \leq 0,
$$

so that also in this case $V_a(t+\epsilon) \leq 0$. We can repeat this argument for all $\epsilon, 2\epsilon, \dots, n\epsilon = s$ and show that $V_a(t+s) \leq 0$. □

A direct consequence of Lemma 1 is the following corollary.

Corollary 5. *Let $F_a(t)$ be a solution of Equation (26) with Heaviside initial data, that is*

$$
F_a(0) = \Theta(a - \bar{a}) = \begin{cases} 0, & \text{if } a < \bar{a}, \\ 1, & \text{if } a \geq \bar{a}. \end{cases} \tag{37}
$$

Further, define $m_\epsilon(t) = \inf\{a : F_a(t) \geq \epsilon\}$ for any $\epsilon \in [0,1]$. Then we have that $F_{a+m_\epsilon(t)}(t)$ converges uniformly to some function $f_\epsilon(a)$ as $t \to \infty$.

Proof of Corollary 5. For $t_0, b \in \mathbb{R}_+$ we set for any $a \geq 0$

$$
\begin{aligned}
F_a^{(1)}(t) &= F_{a+m_\epsilon(t_0)}(t) \\
F_a^{(2)}(t) &= F_{a+m_\epsilon(t_0+b)}(t+b).
\end{aligned}
$$

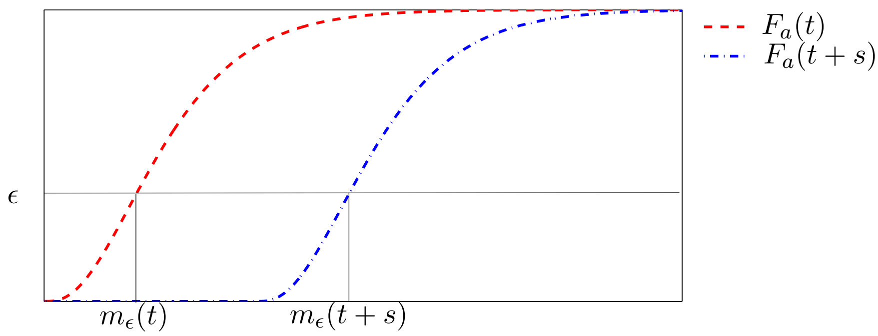

Figure 9: Illustration of distributions and at times t and $t+s$ for $s>0$.

If we start from Heaviside initial data we have that $F_a^{(1)}(0) \geq F_a^{(2)}(0)$ and Proposition 1 applies. It follows that $F_a^{(1)}(t) \geq F_a^{(2)}(t)$ for all $t>0$. We then can write

$$0 \leq F_{a+m_\epsilon(t_0+b)}(t_0+b) \leq F_{a+m_\epsilon(t_0)}(t_0) \leq 1.$$

For each value of b this is a decreasing sequence of real numbers which is bounded from below and thus its infimum is the limit. In particular, since t_0, b and ϵ were chosen arbitrarily, we obtain that $F_{a+m_0(t)}(t)$ converges to some $f(a) \geq 0$ from above as $t \to \infty$. An illustration can be seen in Figure 9. Note that since $F_a(t)$ is monotonic and $f(a)$ is continuous, convergence is uniform. $\square$

With Corollary 5 in place we are now ready to give the proof of Proposition 5.

Proof of Proposition 5. By Corollary 5 we can fix a value of $\epsilon = \frac{1}{2}$, where $m_{1/2}(t)$ is the median of $F(t)$, and have that

$$\lim_{t\to\infty} F_{a+m_{1/2}(t)}(t) = f(a),$$

for some time-independent function $f(a)$ satisfying $f(0) = 1/2$. This implies that

$$\lim_{t\to\infty} \frac{dF_{a+m_{1/2}(t)}(t)}{dt} = 0,$$

or equivalently

$$\frac{\partial F_{a+m_{1/2}(t)}(t)}{\partial t} + \frac{\partial F_{a+m_{1/2}(t)}(t)}{\partial a}\frac{dm_{1/2}(t)}{dt} = o(1).$$

Using Equation (26), the above equation can be written as follows

$$o(1) = \frac{2q-1}{2} F_{a+m_{1/2}(t)}(t)^2 + \frac{1-2q-p}{2} F_{a+m_{1/2}(t)}(t) + \frac{p}{2} F_{a+m_{1/2}(t)-1}(t) + \frac{\partial F_{a+m_{1/2}(t)}(t)}{\partial a} \frac{dm_{1/2}(t)}{dt}.$$

Integrating with respect to time over the interval $[t, t+1)$, considering a value of t large enough and integrating over $[0, a)$, we obtain

$$o(1) = \int_0^a \left(\frac{2q-1}{2} f(x)^2 + \frac{1-2q-p}{2} f(x) \frac{p}{2} f(x-1) \right) dx + f(a)(m_{1/2}(t+1) - m_{1/2}(t)).$$

The only time dependent term on the RHS from the above equation is $m_{1/2}(t+1) - m_{1/2}(t)$ while the LHS is constant so that we must have

$$\lim_{t\to\infty} (m_{1/2}(t+1) - m_{1/2}(t)) = \nu$$

for some constant $\nu \geq 0$. In particular, if $m_{1/2}(t) = \nu t$ then the above equation is trivially satisfied.

Further, we must have that $F_{m_{1/2}(t)}(t) = F_{m_{1/2}(t+s)}(t+s)$, or equivalently, $F_{\nu t}(t) = F_{\nu(t+s)}(t+s)$, and this is satisfied for $F_a(t) = f(a - \nu t)$. It follows that the solution of Equation (26) must be a traveling wave. Note that due to the stable shape of the traveling wave, the above result holds for any value of ϵ. □

Proof of Proposition 6. First, we prove part (i) of the proposition. In order to solve for the traveling wave solution of Equation (26) we observe that in terms of the complementary cumulative log-productivity it can be written as

$$\frac{\partial G(a,t)}{\partial t} = \frac{2q-1}{2}(-G(a,t)^2 + G(a,t)) - \frac{p}{2}(G(a,t) - G(a-1,t)). \quad (38)$$

Proposition 5 implies that the dynamics of the complementary cumulative log-productivity distribution $G(a,t)$ in Equation (38) admits a traveling wave solution $G(a,t) = g(x)$, $x = a - \nu t$ with velocity ν satisfying

$$\nu \frac{dg(x)}{dx} = \frac{2q-1}{2}(g(x)^2 - g(x)) + \frac{p}{2}(g(x) - g(x-1)). \quad (39)$$

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 (38).

Then we obtain from Equation (38) 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

$$\nu = \frac{2q-1-p+pe^{\lambda}}{2\lambda}. \tag{40}$$

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

$$2q-1-p+pe^{\lambda} = p\lambda e^{\lambda}. \tag{41}$$

The minimum of Equation (40) is obtained at λ solving Equation (41). This yields

$$\lambda = 1 + W\left(\frac{2q-1-p}{pe}\right), \tag{42}$$

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

Next, we show part (ii) of the proposition, where 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 (26) to obtain

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

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

$$2\rho\nu = 2q-1+p-pe^{-\rho}. \tag{43}$$

This equation can be solved numerically to obtain the exponent ρ [see e.g. Press et al., 1992, Chap. 9]. □

Proof of Proposition 7. Motivated by our analysis for the case of $\beta = 0$, we make the following assumption on the log-productivity distribution

$$P_a = N \begin{cases} e^{\rho(a-\nu t)}, & \text{if } a \leq \nu t, \\ e^{-\lambda(a-\nu t)}, & \text{if } a > \nu t, \end{cases} \tag{44}$$

with the normalization constant N given by

$$\frac{1}{N} = \frac{e^{\rho}}{e^{\rho}-1} + \frac{1}{e^{\lambda}-1}. \tag{45}$$

The average log-productivity is then given by

$$
\begin{aligned}
\mathbb{E}(a) &= \sum_{b=1}^{\infty} bP_b = N\sum_{b=1}^{\nu t} be^{\rho(b-\nu t)} + N\sum_{b=\nu t+1}^{\infty} be^{-\lambda(b-\nu t)} \\
&= N\left(\frac{e^{\lambda}+\nu t(e^{\lambda}-1)}{(e^{\lambda}-1)^2} + \frac{e^{\rho}(e^{-\rho t}-1+\nu t(e^{\rho}-1))}{(e^{\rho}-1)^2}\right) \\
&\approx \nu t,
\end{aligned}
$$

for large t. Note that both, the distribution as well as the innovation probability are translational invariant, since they only depend on the difference $a-\nu t$. We then can write Equation (21) for $a > \nu t$ as follows

$$
p_{\beta}^{\text{in}} = \frac{1}{1+\left(\frac{1+N\frac{(e-1)e^{-\lambda(a-\nu t-1)}}{(e^{\lambda}-1)(e^{\lambda}-e)}}{1+p(\bar{A}-1)}\right)^{\frac{\beta}{\ln(\bar{A})}}}.
$$

For small values of β this can be written as

$$
p_{\beta}^{\text{in}} = \frac{1}{2+\gamma\left(1+\frac{(e-1)(e^{\rho}-1)}{(e^{\lambda}-e)(e^{\lambda+\rho}-1)}e^{-\lambda(a-\nu t-1)}\right)},
$$

where we have denoted by $\gamma = \beta/(\ln(\bar{A})(1+p(\bar{A}-1)))$ and used the fact that

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

For a large enough we get $p_{\beta}^{\text{in}} \sim \frac{1}{2+\gamma}$. Similarly, for $a < \nu t$ one can show that

$$
p_{\beta}^{\text{in}} = \frac{1}{2+\gamma\left(\frac{(e^{\lambda}-1)(e-1)e^{\rho}}{(e^{\lambda+\rho}-1)(e^{\rho+1}-1)}e^{-\rho|a-\nu t|} + \frac{(e^{\lambda}-1)(e^{\rho}-1)}{(e^{\lambda-1})(e^{\rho+1}-1)}e^{|a-\nu t|}\right)}.
$$

For a much smaller than νt we get $p_{\beta}^{\text{in}} \sim 0$. Let us denote by $A = \frac{(e-1)(e^{\rho}-1)e^{\lambda}}{\left(e^{\lambda}-e\right)\left(e^{\lambda+\rho}-1\right)}$, $B = \frac{\left(-1+e^{\lambda}\right)(-1+e)e^{\rho}}{\left(-1+e^{\lambda+\rho}\right)(-1+e^{1+\rho})}$, $C = \frac{e\left(-1+e^{\lambda}\right)(-1+e^{\rho})}{\left(-e+e^{\lambda}\right)(-1+e^{1+\rho})}$. Then we can write for β small enough

$$
p_{\beta}^{\text{in}} = \begin{cases} \frac{1}{2+\gamma\left(1+Ae^{-\lambda(a-\nu t)}\right)}, & \text{if } a \leq \nu t, \\ \frac{1}{2+\gamma\left(Be^{\rho(a-\nu t)}+Ce^{-(a-\nu t)}\right)}, & \text{if } a > \nu t, \end{cases} \tag{46}
$$

With the distribution given in Equation (44) and the innovation probability from Equation (46), we obtain for a larger than νt from Equation (25)

$$\begin{aligned}\lambda\nu &= \sum_{b=1}^{\nu t}\left(1-\frac{1}{2+\gamma\left(Be^{\rho(b-\nu t)}+Ce^{-(b-\nu t)}\right)}+1-\frac{1}{2+\gamma\left(1+Ae^{-\lambda(a-\nu t)}\right)}\right)Ne^{\rho(b-\nu t)}\\ &+\sum_{b=\nu t+1}^{a-1}\left(1-\frac{1}{2+\gamma\left(1+Ae^{-\lambda(b-\nu t)}\right)}+1-\frac{1}{2+\gamma\left(1+Ae^{-\lambda(a-\nu t)}\right)}\right)Ne^{-\lambda(b-\nu t)}\\ &+\left(1-\frac{1}{2+\gamma\left(1+Ae^{-\lambda(a-\nu t)}\right)}\right)Ne^{-\lambda(a-\nu t)}\\ &+(1-p)\frac{1}{2+\gamma\left(1+Ae^{-\lambda(a-\nu t)}\right)}+pe^{\lambda}\frac{1}{2+\gamma\left(1+Ae^{-\lambda(a-1-\nu t)}\right)}-1,\end{aligned}$$

For a much larger than νt and large t the above equation can be written as

$$\begin{aligned}\lambda\nu &= \frac{1+\gamma}{2+\gamma}-\frac{\left(e^{\lambda}-1\right)\left(e^{\rho}-1\right)}{e^{\lambda+\rho}-1}\left(\sum_{b=1}^{\infty}\frac{e^{-\lambda b}}{2+\gamma\left(1+Ae^{-\lambda b}\right)}+\sum_{b=0}^{\infty}\frac{e^{-\rho b}}{2+\gamma\left(Be^{-\rho b}+Ce^{b}\right)}\right)\\ &+\frac{1}{2+\gamma}\left(1+p\left(e^{\lambda}-1\right)\right)\end{aligned}$$

Rearranging for ν and taking the derivative of ν with respect to λ yields the FOC

$$\begin{aligned}0 &= \frac{-2+p-\gamma+e^{\lambda}p(-1+\lambda)}{(2+\gamma)\lambda^2}\\ &-\sum_{b=0}^{\infty}\left(e^{-b\rho}\left(-1+e^{\rho}\right)\left(-2+\frac{1}{\left(e-e^{\lambda}\right)^2\left(-1+e^{1+\rho}\right)}e^{-b\rho}\left(-(-1+e)e^{\rho}\left(e-e^{\lambda}\right)^2\left(-1+e^{\lambda}\right)^2\gamma\right.\right.\right.\\ &+e^{b\rho}\left(-2e^{\lambda}\left(e-e^{\lambda}\right)^2\left(-1+e^{1+\rho}\right)\left(-1+e^{\lambda+\rho}+\lambda-e^{\rho}(1+\lambda)\right)\right.\\ &\left.\left.\left.\left.\left.-e^{1+b}\left(-1+e^{\lambda}\right)^2\left(-1+e^{\rho}\right)\gamma\left(e+e^{2\lambda+\rho}+e^{\lambda}(-1+\lambda)-e^{1+\lambda+\rho}(1+\lambda)\right)\right)\right)\right)\right)\\ &\times\left(\left(-1+e^{\lambda+\rho}\right)^2\left(2+\frac{\left(-1+e^{\lambda}\right)\left(\frac{e^{1+b}(-1+e^{\rho})}{-e+e^{\lambda}}+\frac{(-1+e)e^{\rho-b\rho}}{-1+e^{\lambda+\rho}}\right)\gamma}{-1+e^{1+\rho}}\right)^2\lambda^2\right)^{-1}\\ &-\sum_{b=1}^{\infty}\left(e^{-b\lambda}\left(-1+e^{\rho}\right)\left(\frac{(-1+e)e^{\lambda-b\lambda}\left(-1+e^{\rho}\right)\gamma\left(e^{\lambda}(1+e)+e^{2\lambda}(-1+\lambda)-e(1+\lambda)\right)}{\left(e-e^{\lambda}\right)^2}\right.\right.\\ &\left.\left.+(2+\gamma)\left(-1-b\lambda+e^{\lambda}\left(1+(-1+b)\lambda+e^{\rho}\left(1+\lambda+b\lambda-e^{\lambda}(1+b\lambda)\right)\right)\right)\right)\right)\\ &\times\left(\frac{(-1+e)e^{\lambda-b\lambda}\left(-1+e^{\rho}\right)\gamma\lambda}{-e+e^{\lambda}}+\left(-1+e^{\lambda+\rho}\right)(2+\gamma)\lambda\right)^{-2}.\end{aligned}$$

This equation can be solved numerically to obtain the values of λ, and from those the corresponding values of the traveling wave velocity ν. $\square$

Proof of Proposition 8. The threshold productivity satisfies $A_i^{\mathrm{im}} = A_i^{\mathrm{in}}$, or equivalently $a^* := a_i^{\mathrm{im}} = a_i^{\mathrm{in}}$. With the above values for the expected productivities from innovation and imitation this implies that

$$a^* + \log(1 - p + \bar{A}p) = a^* + \log\left(F_{a^*} + \sum_{b=a^*+1}^{\infty} e^{b-a^*} P_b\right). \tag{47}$$

This can be written as

$$1 - p + \bar{A}p = F_{a^*} + \sum_{b=1}^{\infty} e^b P_{b-a^*}, \tag{48}$$

or equivalently

$$1 - p + \bar{A}p = 1 - G_{a^*} + \sum_{b=1}^{\infty} e^b P_{b+a^*} = 1 + \sum_{b=1}^{\infty} (e^b - 1) P_{b+a^*}. \tag{49}$$

That is

$$p(\bar{A} - 1) = \sum_{b=1}^{\infty} (e^b - 1) P_{b+a^*}. \tag{50}$$

The threshold log-productivity a^* must satisfy the following condition

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

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

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

$f(a,t)$ is strictly monotonous decreasing if $f(a-1,t) - f(a,t) = (e - 1)P(a,t) > 0$. This holds for all a in the support S of $P(a,t)$ where $P(a,t) > 0$. Hence, if at time t for all $a \in S$ we have that $P(a,t) > 0$ then there exist a unique threshold log-productivity a^* satisfying the above condition.

Next, we show that if $P(b,t)$ satisfies the above condition, then it also must hold that $f(a-1, t+\Delta t) - f(a, t+\Delta t) > 0$. First, consider $a \leq a^*$.

Then for $q = 1$, $P(a,t) > 0$ and $F(a,t) > F(a-1,t)$ we get

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

On the other hand, 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. For Δt going to zero we obtain the corresponding result in continuous time. □

PROOF OF PROPOSITION 9. Under the assumption that Equation (30) holds for β large enough, we can insert Equation (30) into Equation (24) to find that the evolution of the log-productivity distribution can be written as

$$\frac{\partial P_a(t)}{\partial t} = \begin{cases} P_a(t)(F_{a-1}(t) + F_a(t)) - P_a(t), & \text{if } a \leq a^*, \\ P_a(t)F_{a^*}(t) + (1-p)P_a(t) - P_a(t), & \text{if } a = a^* + 1, \\ P_a(t)F_{a^*}(t) + (1-p)P_a(t) + pP_{a-1}(t) - P_a(t), & \text{if } a > a^* + 1. \end{cases}$$

For the dynamics of the cumulative log-productivity distribution $F(a,t) = \sum_{b=1}^{a} P(a,t)$ we then get for $a < a^*$

$$\begin{aligned}
\frac{\partial F_a(t)}{\partial t} &= \sum_{b=1}^{a} \frac{\partial P_b(t)}{\partial t} \\
&= \sum_{b=1}^{a} \left(P_b(t)(F_{b-1}(t) - F_b(t)) - P_b(t)\right) \\
&= F_a(t)^2 - F_a(t),
\end{aligned}$$

where in the last line from above we have used the results obtained in Proposition 2. Next, for $a = a^* + 1$ we get

$$\begin{aligned}
\frac{\partial F_{a^*+1}(t)}{\partial t} &= \sum_{b=1}^{a^*} \frac{dP_b(t)}{dt} + \frac{\partial P_{a^*+1}(t)}{\partial t} \\
&= F_{a^*+1}(t)^2 - F_{a^*+1}(t) + P_{a^*+1}(t)F_{a^*}(t) - pP_{a^*+1}(t) \\
&= F_{a^*}(t)^2 - F_{a^*}(t) - (F_{a^*+1}(t) - F_{a^*}(t))(p - F_{a^*}(t)) \\
&= -(1 - F_{a^*+1}(t))F_{a^*}(t) - p(F_{a^*+1}(t) - F_{a^*}(t)).
\end{aligned}$$

Similarly, for $a > a^* + 1$ we get

$$\begin{aligned}\frac{\partial F_a(t)}{\partial t} &= \sum_{b=1}^{a^*} \frac{\partial P_b(t)}{\partial t} + \frac{\partial P_{a^*+1}(t)}{\partial t} + \sum_{b=a^*+2}^{a} \frac{\partial P_b(t)}{\partial t} \\ &= F_{a^*}(t)^2 - F_{a^*}(t) + P_{a^*+1}(t)F_{a^*}(t) - pP_{a^*+1}(t) \\ &+ \sum_{b=a^*+2}^{a} \left(F_{a^*}(t)P_b(t) - p(P_b(t) - P_{b-1}(t))\right) \\ &= -(1 - F_a(t))F_{a^*}(t) - p(F_a(t) - F_{a-1}(t)).\end{aligned}$$

Putting the above results together we can write

$$\frac{\partial F_a(t)}{\partial t} = \begin{cases} F_a(t)^2 - F_a(t), & \text{if } a \leq a^*, \\ -(1 - F_a(t))F_{a^*}(t) - p(F_a(t) - F_{a-1}(t)), & \text{if } a \geq a^* + 1. \end{cases}$$

Note that for all $a \geq 1$ and $t \geq 0$ we have that $\frac{dF_a(t)}{dt} \leq 0$. □

Proof of Proposition 10. We first prove part (i) of the proposition. We assume that the log-productivity distribution for $a > a^*$ is given by $P(a,t) = Ne^{-\lambda(a-\nu t)}$ with a proportionality factor $N = P(a^*, t)$. For the complementary cumulative distribution function $G(a,t) = 1 - F(a,t) = \sum_{b=a+1}^{\infty} P(b,t)$ for $a > a^*$ this implies that

$$G(a,t) = \sum_{b=a+1}^{\infty} Ne^{-\lambda(b-\nu t)} = \frac{N}{e^\lambda - 1} e^{-\lambda(a-\nu t)}. \tag{51}$$

In terms of the complementary cumulative distribution function $G(a,t) = 1 - F(a,t)$ we then can write Equation (31) for a much larger than the threshold a^* as

$$\frac{\partial G(a,t)}{\partial t} = G(a,t)\left(1 - G(a^*,t)\right) - p\left(G(a,t) - G(a-1,t)\right)$$

Inserting Equation (51) yields

$$\lambda\nu e^{-\lambda(a-\nu t)} = e^{-\lambda(a-\nu t)}\left(1 - \frac{N}{e^\lambda - 1}\right) - p\left(e^{-\lambda(a-\nu t)} - e^{-\lambda(a-1-\nu t)}\right) \tag{52}$$

which gives

$$\lambda\nu = 1 - \frac{N}{e^\lambda - 1} - p\left(1 - e^\lambda\right). \tag{53}$$

Next, note that the threshold log-productivity a^* satisfies

$$a^* + \log\left(F(a^*,t) + \sum_{b=a^*+1}^{\infty} e^{b-a^*} P(b,t)\right) = a^* + \log\left(1 + p(\bar{A} - 1)\right). \tag{54}$$

This means that the expected log-productivity obtained through innovation equals the expected log-productivity obtained through imitation. Equation (54) can be written as

$$\sum_{b=a^*+1}^{\infty} (e^{b-a^*} - 1)P(b,t) = p(\bar{A} - 1)$$

Inserting $P(a,t) = Ne^{-\lambda(a-\nu t)}$ into the above equation and assuming that $a^* = \nu t$ yields

$$p(\bar{A} - 1) = N\sum_{b=1}^{\infty}(e^b - 1)e^{-\lambda b} = N\left(\frac{1}{e^{\lambda-1} - 1} + \frac{1}{1 - e^{\lambda}}\right),$$

so that

$$N = p(\bar{A} - 1)\left(\frac{1}{e^{\lambda-1} - 1} + \frac{1}{1 - e^{\lambda}}\right)^{-1}, \tag{55}$$

Inserting N into Equation (53) gives

$$\lambda\nu = 1 - \frac{p(\bar{A} - 1)}{e^{\lambda} - 1}\left(\frac{1}{e^{\lambda-1} - 1} + \frac{1}{1 - e^{\lambda}}\right)^{-1} - p(1 - e^{\lambda}),$$

so that we obtain

$$\nu = \frac{1}{\lambda}\left(1 + p(e^{\lambda} - 1) - \frac{p(\bar{A} - 1)(1 - e^{1-\lambda})}{e - 1}\right). \tag{56}$$

According to the selection principle we have encountered already in the proof of Proposition 6, for sufficiently steep initial conditions of $F(a,0)$ the value of λ is realized that minimizes Equation (56). The traveling wave velocity ν as a function of λ for different values of p can be seen in Figure 10 (left). The corresponding first-order condition (FOC) is given by

$$\frac{d\nu}{d\lambda} = \frac{1 - e + p(\bar{A} + e - 2) + (e-1)e^{\lambda}p(\lambda - 1) - (\bar{A} - 1)e^{1-\lambda}p(1 + \lambda)}{(-1 + e)\lambda^2} = 0$$

and Equation (33) follows. The FOC from above is equivalent to

$$\frac{e - 1}{\bar{A} + e - 2 + (e-1)e^{\lambda}(\lambda - 1) - (\bar{A} - 1)e^{1-\lambda}(1 + \lambda)} = p, \tag{57}$$

which is illustrated in Figure 10 (right).

Next, we consider part (ii) of the proposition. Observe that for values of a much smaller than $a^* = \nu t$ we can neglect the term $F(a,t)^2$ in Equation (31). We then assume that the rear of the log-productivity distribution can be described by an exponential function $P(a,t) \propto e^{\rho(a-\nu t)}$. Inserting this

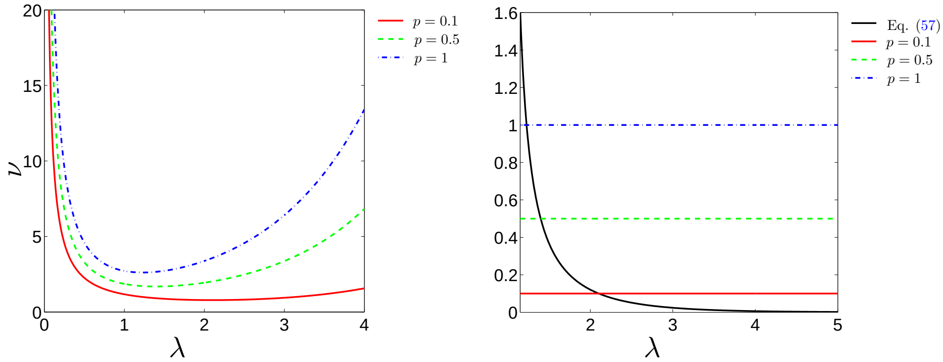

Figure 10: (Left) The traveling wave velocity ν as a function of λ for different values of $p = 0.1$, $p = 0.5$ and $p = 1$. (Right) The ν minimizing value of λ for the same values of p. The figures show that the minimizing value of λ is decreasing for increasing values of p, and consequently, the front of the traveling wave becomes steeper. Moreover, the velocity ν of the traveling wave increases with increasing values of p.

into Equation (31) for a smaller than a^* gives

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

and hence we obtain

$$\rho = \frac{1}{\nu}.$$

□

Proof of Proposition 11. The traveling wave velocity in the limit of $\beta \to 0$ follows from Equations (27) and (28) as

$$\lim_{\beta\to 0}\min_{\lambda} \nu(\lambda;\beta) = \frac{1 + p(e^{1+W\left(\frac{1-p}{pe}\right)} - 1)}{2(1 + W\left(\frac{1-p}{pe}\right))},$$

while the traveling wave velocity for $\beta \to \infty$ is given by Equation (32). We then have that

$$\begin{aligned}&\lim_{\beta\to\infty} \nu(\lambda;\beta) - \lim_{\beta\to 0}\min_{\lambda} \nu(\lambda;\beta)\\ &= \frac{1}{\lambda}\left(1 + p(e^{\lambda} - 1) - \frac{p(\bar{A}-1)(1-e^{1-\lambda})}{e-1}\right) - \frac{1 + p(e^{1+W\left(\frac{1-p}{pe}\right)} - 1)}{2(1 + W\left(\frac{1-p}{pe}\right))} > 0.\end{aligned}$$

An illustration of $\lim_{\beta\to\infty} \nu(\lambda;\beta)$ and $\lim_{\beta\to 0}\min_{\lambda} \nu(\lambda;\beta)$ can be seen in Figure 11. Since the above equation holds for all λ, it holds in particular for the value of λ minimizing $\lim_{\beta\to\infty} \nu(\lambda;\beta)$, and hence, we have that $\lim_{\beta\to\infty} \nu(\lambda;\beta) > \lim_{\beta\to 0} \nu(\lambda;\beta)$. A higher traveling wave velocity ν implies

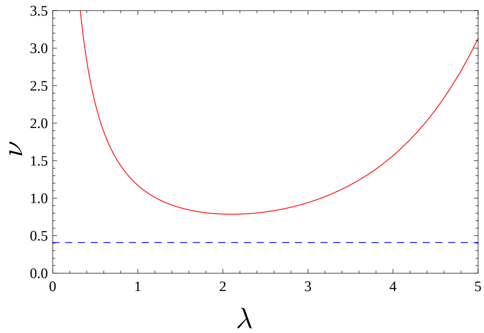

Figure 11: The traveling wave velocity $\lim_{\beta\to\infty} \nu(\lambda;\beta)$ as a function of λ shown with a solid curve and $\lim_{\beta\to 0} \min_\lambda \nu(\lambda;\beta)$ shown with a dashed line for $p = 0.1$ and $\log \bar{A} = 1$.

first-order stochastic dominance of the respective cumulative distribution functions and therefore a higher average productivity. □

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 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 in a small interval $\delta t = 1/N$ 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

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

The evolution of the log-productivity distribution in the limit of large N is then given by

$$\frac{\partial P(t)}{\partial t} = P(t)\left((1-r)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\mathrm{in}} + \mathbf{D}\mathbf{T}^{\mathrm{im}}(P(t))\right) + r\mathbf{T}^{\mathrm{dec}} - \mathbf{I}\right). \quad (58)$$

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.[20] 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 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, a randomly selected firm either decide to conduct in-house R&D or imitate other firms' technologies and, second, entry and exit takes place. Both events happen within a small time interval $[t, t+\Delta t]$. We then have to modify Equation (23) accordingly. In the case of $A_0 = 1$ we can write in the limit of large N

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

where $Q = [1 \quad 0 \quad 0 \quad \ldots]$.

[20] 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.

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

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

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

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

where similar to the previous sections we have assumed that $\Delta t = 1/N$ and taken the limit $N \to \infty$.