

Innovation vs. Imitation and the Evolution of Productivity Distributions[☆]

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

^a*SIEPR & 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 tractable dynamic model of productivity growth and technology spillovers that is consistent with the emergence of real world empirical productivity distributions. Firms can improve productivity by engaging in in-house R&D, or alternatively, by trying to imitate other firms' technologies subject to limits to their absorptive capacities. The outcome of both strategies is stochastic. The choice between in-house R&D and imitation is endogenous, and based on firms' profit maximization motive. Firms closer to the technological frontier have less imitation opportunities, and tend to choose more often in-house R&D, consistent with the empirical evidence. The equilibrium choice leads to balanced growth featuring persistent productivity differences even when starting from ex-ante identical firms. The long run productivity distribution can be described as a traveling wave with tails following Zipf's law as it can be observed in the empirical data. Idiosyncratic shocks to firms' productivities of R&D reduce inequality, but also lead to lower 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 large and persistent productivity differences not only across countries [e.g. [Durlauf, 1996](#); [Durlauf and Johnson, 1995](#); [Feyrer, 2008](#); [Quah, 1993, 1996, 1997](#)], but also across firms and plants within narrow sectors [[Baily et al., 1992](#)]. A prominent explanation for these productivity differences is that they stem from differences in technological knowledge [see, e.g. [Doms et al., 1997](#)]. On the one hand, part of these 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, their size and persistence are evidence of a slow knowledge diffusion between firms [[Eeckhout and Jovanovic, 2002](#); [Geroski, 2000](#); [Stoneman, 2002](#)].

[☆]We would like to thank Daron Acemoglu and seminar participants at the University of Zurich, the Annual Conference of the Association for Public Economic Theory in Bloomington, 2011, and the Annual Conference of the European Economic Association in Oslo, 2011, for helpful comments. Michael D. König acknowledges financial support from Swiss National Science Foundation through research grant PBEZP1-131169. Fabrizio Zilibotti acknowledges financial support from the ERC Advanced Grant IPCDP-229883.

Email addresses: mdkoenig@stanford.edu (Michael D. König), jan.lorenz@uni-oldenburg.de (Jan Lorenz), fabrizio.zilibotti@econ.uzh.ch (Fabrizio Zilibotti)

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. 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].¹ 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 propose a theory that combines process of technology development through 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. The model is shown to reproduce some empirical regularities concerning the productivity distributions of firms at both an aggregate and disaggregated levels, as well as their evolution over time. In particular, we analyze a large data set containing information about more than six million firms in the period between 1992 to 2005. We find that the productivity distributions over these firms exhibit power-law tails over all periods of time. A similar observation for a single point in time has been made in Corcos et al. [2007]; Di Matteo et al. [2005]. Moreover, we observe an increasing trend in the average productivity. The growing distribution of firms can then be describes in terms of a “traveling wave”.

The model economy is a Schumpeterian (quality ladder) growth model, in the spirit of Acemoglu et al. [2006], where differentiated intermediate goods are produced by monopolistically competitive firms. Firms producing different varieties have heterogenous productivities that change stochastically over time. The key assumptions is that there are technological spillovers across firms producing different intermediates. More specifically, a firm producing variety i can try to imitate the technology used by a firm producing variety j whenever the opportunity arises. A distinctive feature of our model is that firms make endogenous decisions of whether to undertake in-house R&D (innovation) or to imitate other firms’ technologies. The success of the imitation strategy depends on the availability of better technologies (which depends on the endogenous distribution of productivity) and their absorptive capacities. Starting from ex ante identical firms our model generates heterogeneous productivity distributions with power-law tails. Thus, evolving theoretical productivity distributions obey Zipf’s law at both tails (for small and large values) consistent with the empirical observation.³ Moreover, the theoretical distribution evolves

¹Although entry, exit and reallocation are important determinants of aggregate productivity growth, entry and exit account for only 25% of total productivity growth [Acemoglu, 2009, Chap. 18]. Therefore, a successful theory of economic growth should aim at understanding not only the process of innovation and selection across firms, but also the determinants of investments in technology adoption.

²There 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].

³Zipf’s law is also observed for distributions of several other economic variables of interest (e.g. firm size) in numerous empirical studies [e.g. De Wit, 2005; Gabaix, 1999; Saichev et al., 2009].

endogenously over time as a “traveling wave” with stable shape. The endogenous innovation-imitation choice is crucial for the result. If the population of firms consisted of a fixed proportion of innovators and imitators, the limiting distribution would feature an ever increasing variance of productivities across firms. The intuition for our main result is simple: firms that are close to the frontier do fresh innovation, driving the movement of the productivity frontier. Firms lagging behind choose to imitate and the probability of successful imitation is increasing with the distance to the frontier. This prevents that an ever growing gap emerges between more and less successful firms.

Tractability is aided by two simplifying assumptions. First, there are no sunk costs and firms can switch instantaneously across innovation-imitation strategies. Thus, although firms are fully rational and maximize the present value of their profits, their optimal choice can be expressed in terms of a repeated static maximization problem. Second, we introduce the natural assumption that idiosyncratic firm-specific shocks affect the comparative advantage of firms in performing in-house R&D relative to imitation (more precisely, the shock affects the productivity of implementing successful in-house innovations). Firms are assumed to make the period choice between in-house innovation and imitation after the shocks have been realized. This assumption allows us to provide an analytic characterization of the long-run productivity distributions for the limit case in which the variance of the shocks is sufficiently large to drive the choice between innovation and imitation, irrespective of the state of technology of the firm. In such a case, we can attain a complete characterization including a proof of existence for the traveling wave solution. In the polar opposite case of no shocks we can also achieve characterization of the long run distribution, but need to make certain functional assumptions on the shape of the distribution. A numerical analysis reveals that the log-productivity distribution is a traveling wave with exponential tails also for intermediate cases. In summary, the assumption that firms are subject to productivity shocks to R&D turns out to be inessential, except for allowing us to prove analytically the traveling wave result in particular cases.

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. In particular, [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. In his model, there is an exogenous (stochastic) productivity growth of incumbent firms that spills over to entrants that can imitate the technology of existing firms. The theoretical mechanism is different from ours, since in [Luttmer \[2007\]](#) firms face no choice between innovation and imitation. From an empirical standpoint, Luttmer’s model explains the emergence of a Zipf’s law for large firm sizes (right power tail), whereas our model captures the

power law behavior for both small and large productivities.⁴ [Acemoglu and Cao \[2010\]](#) construct, as we do, a Schumpeterian model and also obtain Zipf’s law for large firm sizes. In their model, incumbent firms engage in incremental innovations, while entry is associated with radical innovations and creative destruction (i.e., the successful entrant replaces the incumbent). Similar to [Luttmer \[2007\]](#), their model does not feature an endogenous choice of the R&D strategy. Relative to these papers, our paper shares with [Acemoglu and Cao \[2010\]](#) the implication that large productivity gains are due to R&D and productivity improvements introduced by incumbent firms, which is consistent with the empirical evidence. In addition, our model is consistent with the empirical evidence that firms closer to the technology frontier engage in more R&D investments [see [Griffith et al., 2003](#)].

[Alvarez et al. \[2008\]](#); [Lucas \[2008\]](#) and [Lucas and Moll \[2011\]](#) study models of technology diffusion using the framework of [Eaton and Kortum \[1999\]](#). Each producer draws from a random sample of firms and “copies” the technology of the firm with which it is matched whenever the latter has a better technology. These papers are related to our work, and explore dimensions that we do not consider – for instance, [Lucas and Moll \[2011\]](#) focus on the trade-off in the use of time between production and imitation and on the effects of progressive taxation. Relative to our contribution, these authors do not take into account limitations in the ability of firms to imitate external knowledge nor do they model explicitly the strategic decisions of firms whether to undertake in-house R&D or to copy other firms. Because in their model firms can only copy from existing firms (or ideas), the equilibrium dynamics would converge in the long run to a mass point corresponding to the productivity level of the most productive firm. To avoid such a degenerate long-run distribution, they assume that there exists at least one firm with unlimited productivity. This is not necessary in our model, since firms that are close to the technology frontier choose endogenously to innovate (i.e., draw from an exogenous productivity distribution) rather than to imitate. This yields an endogenous growth engine.⁵

[Ghigliano \[2011\]](#) constructs a search-based growth model where firm-level technologies are embodied in patents and new technologies are invented through building on already existing patents. Similar to our model, this model also generates Zipf’s law distributed productivity distributions, however, the underlying process does not focus on the endogenous decision of firms whether to innovate or imitate, but rather the recombination of existing technologies into novel ones.

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 affecting inequality and the growth rate of the economy. Section 5 provides

⁴[Atkeson and Burstein \[2010\]](#) also study the role of innovations by incumbents and the interactions with entrants, but focus mostly on the trade implications.

⁵[Lucas and Moll \[2011\]](#) consider an environment in which firms make full dynamic choices. Our framework has the advantage of yielding analytical solutions, while they rely on simulations. However, our simplifying assumptions do not come without costs, as the decision to innovate-imitate has in reality obvious dynamic implications that our stylized model misses. We regard the two contributions as complementary, each being better suited to different aims.

a calibration of the model’s parameters. In Section 7 we conclude. The proofs of all propositions and corollaries can be found in Appendix C. A number of possible extensions of the model is given in Appendix D.

2. Empirical Productivity Distributions

In this section, we present some empirical results about the productivity distribution across firms. We emphasize three features that are consistent with our theory. First, the distribution of high-productivity firms is well described by a power-law. Second, the distribution of low-productivity firms is also well approximated by a power-law, although this approximation is less accurate, arguably due to noisy data at low productivity levels. Third, the distribution is characterized by a constant growth rate over time, where both the right and the left power-law are fairly stable. This implies that the evolution over time of the productivity distribution can be described as a “traveling wave”. While the first property is well known [see e.g. [Corcos et al., 2007](#)], the second and the third have not been emphasized in the literature.

We compute the empirical productivity levels of firms using the Amadeus database provided by Bureau van Dijk. We extract a data set which contains a total of 5,216,989 entries from European firms in the years from 1992 to 2005.⁶ We use a balanced subsample of all firms for the years 1995 to 2003.⁷ We further include only western European countries⁸, since the predictions of our theory are about equilibrium growth properties. Eastern European economies were at the start of a transition into capitalism in the early 90’s.⁹

The productivity of each firm is estimated using a production function approach. We assume a Cobb-Douglas technology

$$Y_{it} = A_{it}C_{it}^aL_{it}^bM_{it}^c u_{it}, \quad (1)$$

where Y_{it} denotes the value added of firm i at time t , A_{it} is (total factor) productivity, C_{it} its capital in fixed assets, L_{it} its labor force in number of employees, M_{it} its cost of materials and u_{it} is an error term.¹⁰ The estimation of A_{it} follows the method introduced by [Levinsohn and Petrin \[2003\]](#).¹¹ Descriptive statistics of the estimated productivities grouped year-wise can be seen in

⁶These are firms for which data is available for the all the variables from the list: value added, operating revenue, fixed assets, number of employees, cost of materials and cost of employees. These were data points from 1,413,487 firms.

⁷We choose the time span 1995 to 2003 for the balanced panel because, these are the years with a substantial number of firms for which data exists in all years (more than 200,000). In this balanced panel are 52,837.

⁸The countries we define as western European are Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, Britain, Greece, Ireland, Iceland, Italy, Liechtenstein, Luxembourg, Monaco, Malta, Netherlands, Norway and Sweden.

⁹In total there are 49,022 firms in the balanced panel of western European firms, with data available in every year from 1995 to 2003. These are firms from Belgium (3,842), Switzerland (91), Germany (1), Spain (14,934), Finland (21), France (17,892), Italy (12,094), and Netherlands (147).

¹⁰These variables are represented by Amadeus variables value added, fixed assets, number of employees and costs of materials. The other variables are used for robustness checks only.

¹¹The parameters a, b, c in Equation (1) are assumed to be larger than zero and u_{it} is a positive error term. Taking logs delivers the regression equation

$$\log Y_{it} = \log A_{it} + a \log C_{it} + b \log L_{it} + c \log M_{it} + \epsilon_{it}. \quad (2)$$

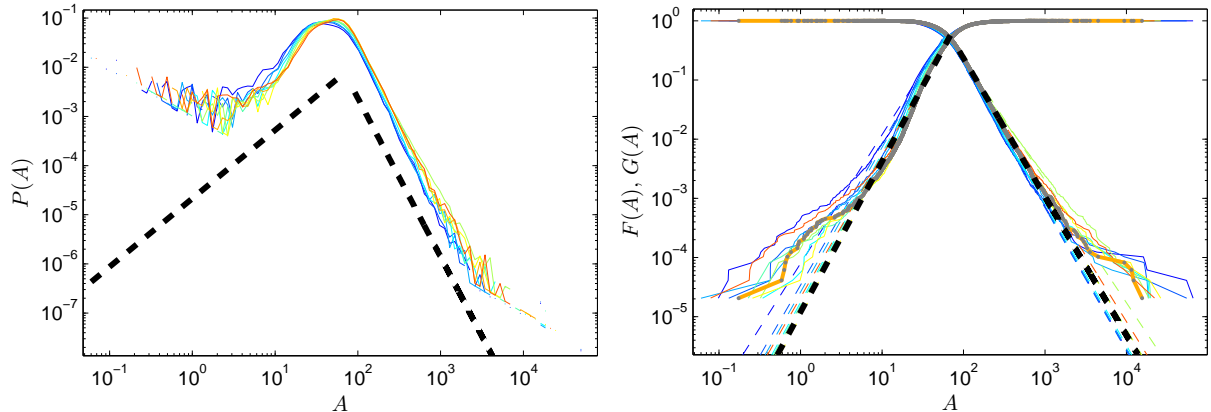


Figure 1: (Top) Probability density functions $P(A)$ (pdf), (Bottom) cumulative distribution function $F(A)$ (cdf) and complementary cumulative distribution function $G(A)$ (ccdf) of total factor productivity for the years 1995 (blue) to 2003 (red) for western European countries. Dashed lines show power-law fits. The bold black line represents the year 2002. Cf. Table 1) for the fitted exponents.

Table 2 in Appendix E.

The top panel in Figure 1 shows the empirical probability density functions $P(A)$ (pdf) over firms for each year considered.¹² The bottom panel in Figure 1 shows the corresponding cumulative distribution functions (cdf) $F(A)$, and, respectively, the complementary cumulative distribution function (ccdf) $G(A) = 1 - F(A)$ for the same time period.¹³ We observe that the left and right tails of the distributions are well approximated by power-laws $P(A) \propto e^{\rho A}$ for small A and $P(A) \propto e^{-\lambda A}$ for large A , where one notes a fatter left tail (cf. Table 1).¹⁴

The dashed lines in Figure 1 indicate fits for tail exponents for the year 2002, and Table 1 shows the estimated values for ρ and λ . From Table 1 we observe that the exponents remain relatively stable over the years of observation: The estimated right tail exponents is around $\lambda = 3.32$ and left tail exponent is around $\rho = 1.46$.

Moreover, we observe a slight rightward shift in empirical distributions over the years of observation. The descriptive statistics of Table 2 in Appendix E show a yearly increase in the average productivity. We find that average productivity grows exponentially with time at a

We use the STATA implementation `levpet` explained in Petrin et al. [2004] to estimate the Cobb-Douglas parameters of Equation (2). We gained $\hat{a} = 0.075$, $\hat{b} = 0.705$. According to Petrin et al. [2004] we estimate the productivity of firm i at time t by

$$\hat{A}_{it} = \exp\left(\log Y_{it} - \hat{a} \log C_{it} - \hat{b} \log L_{it}\right) = \frac{Y_{it}}{C_{it}^{\hat{a}} L_{it}^{\hat{b}}}. \quad (3)$$

¹²Density functions are computed on the basis of histograms of the number of firms on 100 logarithmic bins spread over the range of all data points.

¹³Both cumulative distribution functions are shown in Figure 1 as the vector of sorted total factor productivity values on the abscissa versus a vector with the frequencies (in regular steps of $1/N$) on the ordinate. Data points are less dense at the extremes.

¹⁴The fit of the left tail ($\bar{\rho}$) is performed as a linear regression on the logarithmic data on both axes for all values of the cumulative distribution function below the geometric mean of the corresponding sample. The fit of the right tail ($\bar{\lambda}$) is performed as a linear regression on the logarithmic data on both axes for all values of the complementary cumulative distribution function above the arithmetic mean of the corresponding sample. For the power law tails of the pdf (as the derivative of the cdf) it consequently holds: $\lambda = \bar{\lambda} + 1$ and $\rho = \bar{\rho} - 1$.

year	λ	$>\text{mean}(A)$	$R^2(\lambda)$	ρ	$<\text{geomean}(A)$	$R^2(\rho)$
1995	3.41	32.8%	0.99	1.17	48.7%	0.98
1996	3.42	33.3%	0.99	1.39	49.4%	0.98
1997	3.41	34.1%	1.00	1.44	49.9%	0.97
1998	3.40	33.5%	0.99	1.52	50.0%	0.98
1999	3.29	32.5%	1.00	1.53	50.9%	0.97
2000	3.07	30.1%	1.00	1.54	51.3%	0.97
2001	3.29	31.7%	0.99	1.59	50.3%	0.98
2002	3.34	32.4%	1.00	1.54	50.5%	0.98
2003	3.23	31.0%	1.00	1.39	50.7%	0.97
average	3.32			1.46		

Table 1: Fitted values according to the description in Footnote 14. The estimated power law exponents for the right and left tail of the probability density function λ and ρ . The percentage of firms on which the regression is computed is shown as well as the corresponding coefficient of determination R^2 .

rate ν .¹⁵ We further perform a KPSS test for trend stationarity of average log-productivity [Kwiatkowski et al., 1992]. Based on our estimated productivity values we cannot reject the null hypothesis of trend stationarity in our panel of average log-productivity.¹⁶ We then compute the average log-productivity ν from the data by estimating the parameters of an exponential growth function on a measures of central tendency per year $\bar{A}(t) = \exp(\nu t + \text{const.})$. Exponential growth of productivity corresponds to linear growth of log-productivity $\log \bar{A}(t) = \nu t + \text{const.}$. From our sample we estimate $\nu = 0.0227$ (see also Appendix E). This corresponds to a yearly average growth of productivity of 2.3%.

The above empirical analysis indicates that the evolution of the distribution of productivity of firms over time has a stable shape¹⁷ with an exponentially growing average and power-law tails. We refer to such a distribution as a *traveling wave*. The traveling wave is characterized by three parameters: the right tail exponent λ , the left tail exponent ρ and a growth rate ν .

In the next sections we will introduce a model which is able to generate “traveling wave” productivity distribution consistent with our empirical observations.

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 $x_i(t)$, $i \in \mathcal{N} = \{1, 2, \dots, N\}$, according to the production

¹⁵Arithmetic and geometric mean productivity per year are shown in Figure 12 in Appendix E on a semi-logarithmic plot.

¹⁶We compute a p-value of 0.0549 with a critical value of the corresponding test statistic given by 0.1434 at a 5% significance level. We also perform an Im-Pesaran-Shin unit-root panel data test which favors the hypothesis that a nonzero fraction of the log-productivity estimates for the firms represent stationary processes [see also Im et al., 2003]. Other panel data tests for stationarity typically assume that the number of years of observation is much larger than the number of firms and thus do not apply to our data set.

¹⁷“Stable shape” means that at any point in time the distribution is the same up to location and scale parameters.

function

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

where $x_i(t)$ is the economy's input of intermediate good i at time t , $A_i(t)$ is the technology level of sector i at time t , and $\epsilon_i(t)$ is a productivity shock [cf. Bloom, 2009].¹⁸ 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 $x_i(t) = \chi^{-\frac{1}{1-\alpha}} \epsilon_i A_i(t)$. The profit earned by the incumbent in any intermediate sector i will then be proportional to the productivity in that sector

$$\pi_i(t) = (p_i(t) - 1) x_i(t) = \psi \epsilon_i(t) A_i(t), \quad (4)$$

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 $i \in \mathcal{N}$ 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 $\bar{A}^1, \bar{A}^2, \bar{A}^3, \dots$. Firm i , which has achieved a_i productivity improvements then has productivity $A_i = \bar{A}^{a_i}$.¹⁹

¹⁸See Section 4.3 for a more detailed discussion.

¹⁹Consider 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 $1 + g = \bar{A}$.

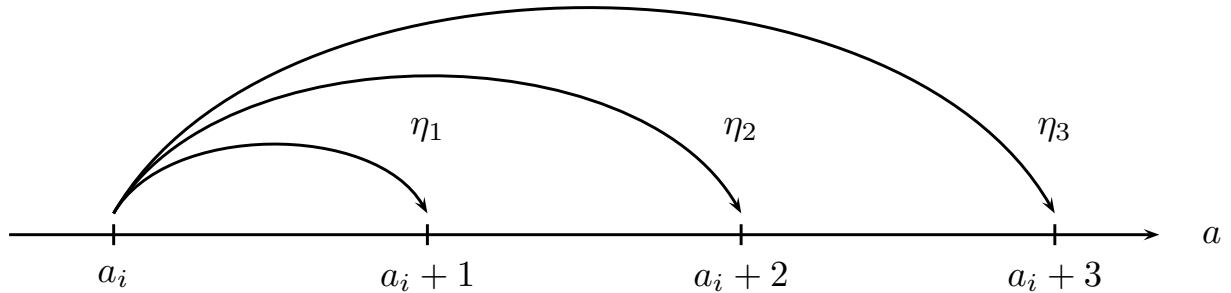


Figure 2: Illustration of the innovation process of firm i with log-productivity $\log A_i = a_i \log \bar{A} = a_i$ (setting $\log \bar{A} = 1$). With probability η_1 firm i makes one productivity improvement and advances by one log-productivity unit, with probability η_2 firm i makes two productivity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

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]. In each time interval $[t, t + \Delta t)$, $\Delta t > 0$, 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.²⁰

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

$$A_i(t + \Delta t) = \bar{A}^{a_i(t) + \eta(t)} = A_i(t) \bar{A}^{\eta(t)}. \quad (5)$$

$\eta(t) \geq 0$ is a nonnegative 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 (5) is given by

$$a_i(t + \Delta t) = a_i(t) + \eta(t). \quad (6)$$

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, with the stochastic factor being the random variable \bar{A}^η . In the limit of continuous time we obtain a geometric Brownian motion for productivity [Saichev et al., 2009, pp. 9].

²⁰We will explain in more detail the innovation and imitation process in Section 4.

3.2.2. Imitation

In the case of imitation, firm i with productivity $A_i(t)$ selects another firm $j \in \mathcal{N}$ 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 selecting 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

$$A_i(t + \Delta t) = A_i(t)\bar{A}^\kappa = \bar{A}^{a_i(t)+\kappa}, \quad (7)$$

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

$$\mathbb{P}_i(\kappa = k | a_i(t), a_j(t)) = \begin{cases} q^k(1 - q) & \text{if } 0 \leq k < a_j(t) - a_i(t), \\ q^k & \text{if } k = a_j(t) - a_i(t), \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Note, that it holds $\sum_{k=0}^{\infty} P(\kappa = k) = 1$, 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 (7) we obtain²¹

$$a_i(t + \Delta t) = a_i(t) + \kappa. \quad (9)$$

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

4. Evolution of the Productivity Distribution

In this section, we analyze the evolution of the productivity distribution. We first establish some useful notation. Then, we analyze characterize the equilibrium dynamics of the productivity distribution.

²¹If 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 $\mathbb{E}_t[a_i(t + \Delta t) | a_i(t) = a, a_j(t) = b] = \sum_{c=0}^{b-a-1} (a + c)(1 - q)^c + bq^{b-a} = a + q \frac{1 - q^{b-a}}{1 - q}$. If $q < 1$ and b is much larger than a , the following approximation holds: $\mathbb{E}_t[a_i(t + \Delta t) | a_i(t) = a, a_j(t) = b] \approx a + \frac{a}{1 - q}$. 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.

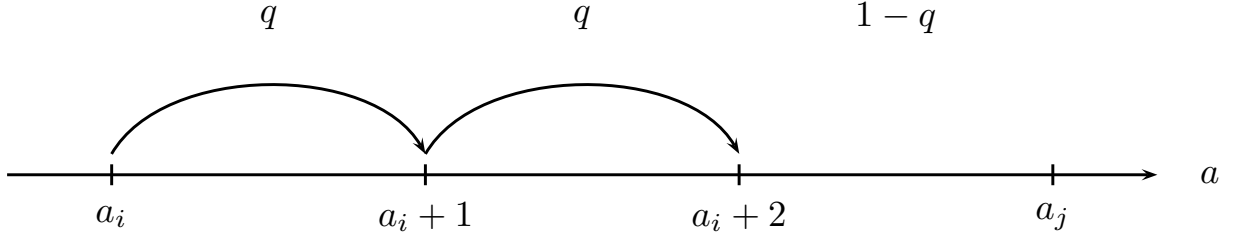


Figure 3: Illustration of the imitation of log-productivity a_j of firm j through firm i with log-productivity a_i , where the log-productivity of firm i is $\log A_i = a_i \log \bar{A} = a_i$ (setting $\log \bar{A} = 1$). Firm i successfully imitates two log-productivity units (with probability q^2)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

4.1. Characterization of the Productivity Dynamics

Consider the distribution of normalized log-productivity $a \equiv \log A$ in the population of $N \in \mathbb{N}$ firms over time. Recall that normalized log-productivity only takes values in the set $S = \{1, 2, \dots, a_{\max}\}$ 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 $P(t) = (P_1(t) \ P_2(t) \ \dots \ P_a(t) \ \dots)$ represents the distribution of log-productivity at time t . It holds that $P_a(t) \geq 0$ and $\sum_{a=1}^{\infty} P_a(t) = 1$. 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.

Our dynamics of innovation and imitation induces a discrete time, discrete space Markov chain $((P^N(t))_{t \in T})_{N=N_0}^{\infty}$, where the chain indexed by N takes on values in the state space (simplex) $P^N = \{P \in \mathbb{R}_+^{|S|} : N \cdot P \in \mathbb{Z}^{|S|}, \sum_{a \in S} P_a = 1\}$ indicating the fraction of firms with a certain log-productivity $a \in S$. At times $t \in T = \{0, \Delta t, 2\Delta t, \dots\}$, 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 in the following sections). The conditional probability $T_{ab} : P^N \rightarrow \mathbb{R}_+^{|S| \times |S|}$ that a firm with log-productivity a switches to log-productivity b at time t is given by

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

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

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

With these definitions we are able to derive the differential equation governing the evolution of the

productivity distribution by using the following proposition:

Proposition 1. *Consider the Markov chain $(P^N(t))_{t \in T}$ with a transition matrix $\mathbf{T}(P)$ which is Lipschitz continuous in P . Then in the limit of a large number of firms N , the evolution of the log-productivity distribution $P(t)$ is given by the differential equation*

$$\frac{\partial P(t)}{\partial t} = P(t)(\mathbf{T}(P) - \mathbf{I}), \quad (10)$$

for some initial distribution $P(0) : S \rightarrow [0, 1]$.

One can show that there exists a unique solution to Equation (10) which is continuous in the initial conditions $P(0)$.²²

The theory employed to derive Equation (10) does not hold for discontinuous transition matrices $\mathbf{T}(P)$ (a scenario we will encounter in Section 4.3.2), and so we need to generalize the dynamics of $P(t)$ to differential inclusions (set-valued differential equations). This is done in the following proposition:

Proposition 2. *Consider the Markov chain $(P^N(t))_{t \in T}$ with transition matrix $\mathbf{T}(P)$. Define $V(P) = P(t)(\mathbf{T}(P) - \mathbf{I})$ and let*

$$\bar{V}(P) = \bigcap_{\varepsilon > 0} cl(\text{conv}(V(\{P' \in \mathbb{R}_+^S : \|P - P'\| \leq \varepsilon\})))$$

the closed convex hull of all values of V that obtain vectors P' arbitrarily close to P . Then in the limit of a large number N of firms, the evolution of the log-productivity distribution $P(t)$ is given by the differential inclusion

$$\frac{\partial P(t)}{\partial t} \in \bar{V}(P(t)),$$

for some initial distribution $P(0) : S \rightarrow [0, 1]$.

A solution is still guaranteed to exist in this case, however, it might not be unique.²³

In the following sections, we derive the matrix $\mathbf{T}(P)$ with elements $T_{ab}(P)$, $a, b \in S$, under the individual firms' laws of motion associated with innovation in Equation (6) and imitation in Equation (9), respectively. First, in Section 4.2, we consider a world where R&D strategies are exogenous with a fixed fraction of innovators and imitators. The purpose of this section is to contrast against the general case in which these strategies are endogenous and based on profit maximizing behavior. Moreover, in the exogenous case, one can show that the log-productivity distribution of the population of the firms engaging in in-house R&D converges to a normal distribution with increasing variance over time (cf. Proposition 3). However, we do not observe such a divergence in the variance of empirically observed productivities in Section 2. In a more realistic model, it is therefore necessary to allow firms to engage in both, innovation and imitation,

²²Observe that Equation (10) can be solved for a function $P_a(t)$ 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 (10) with a partial derivative.

²³We refer the reader to Sandholm [2010, Appendix 6.A] for more discussion.

to advance their productivity levels. This is the case we are going to discuss in the subsequent Section 4.3, where the general model is introduced.

4.2. Exogenous Innovation-Imitation Strategies

We analyze in this section the evolution of the productivity distribution in a world where R&D strategies are exogenous with a fixed fraction of innovators and imitators. We consider three cases: in Section 4.2.1 all firms engage in in-house R&D, in Section 4.2.2 all firms try to imitate and in Section 4.2.3 some firms always do in-house R&D, while others always imitate.

4.2.1. Innovation Only

In this section, we assume that all firms engage in in-house R&D. More formally, this is the equilibrium outcome when firms have no absorptive capacity for imitation (corresponding to $q = 0$). As a firm with log-productivity a tries to innovate, the probability that it obtains a log-productivity $b > a$ is $T_{ab}^{\text{in}} = \eta_{b-a}$. 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 $(\eta_0 \ \eta_1 \ \dots \ \eta_m)$, with η_b representing the probability to increase the productivity by b units and $\eta_0 = 1 - \sum_{b=1}^m \eta_b$. We can then introduce the transition matrix due to in-house R&D

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

From Proposition 1 it follows that the evolution of the log-productivity distribution in Equation (10) in the limit of large N is given by

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

This is a diffusion equation with a positive drift due to stochastic productivity improvements from in-house R&D. Thus, the log-productivity approaches a Gaussian shape in the limit of time t , due to the central limit theorem. Mean and variance rise linearly with t as the following proposition states.

Proposition 3. *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 $\sigma_\eta^2 = \text{Var}(\eta)$. The productivity distribution converges to a log-normal distribution with mean $\mu_A = e^{t\mu_\eta + \frac{1}{2}t\sigma_\eta^2}$ and variance $\sigma_A^2 = (e^{t\sigma_\eta^2} - 1) e^{2t\mu_\eta + t\sigma_\eta^2}$.*

The important finding of Proposition 3 is that the variance of the log-productivity distribution increases over time.

4.2.2. Imitation Only

Next, we consider the polar opposite case when firms have no independent capacity to innovate through in-house R&D, and can only introduce technological progress through imitating other firms' technologies. This is the case 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 in the initial distribution. However, the analysis of this case is instructive, since it provides key insights for the general case with both innovation and imitation.

As a firm with log-productivity a tries to imitate other firms, the probability that it obtains a log-productivity $b > a$ is given by²⁴

$$\begin{aligned} T_{ab}^{\text{im}}(P) &= q^{b-a}P_b + q^{b-a}(1-q)P_{b+1} + q^{b-a}(1-q)P_{b+2} + \dots \\ &= q^{b-a} \left(P_b + (1-q) \sum_{k=1}^{\infty} P_{b+k} \right) \\ &= q^{b-a} (P_b + (1-q)(1 - F_b)), \end{aligned} \quad (11)$$

with F being the cumulative distribution of P as defined by $F_b = \sum_{c=1}^b P_c$. For $b < a$, the firm prefers not to imitate, thus $T_{ab}^{\text{im}}(P) = 0$. The staying probability for $b = a$ is thus $T_{aa}^{\text{im}}(P) = 1 - \sum_{b>a} T_{ab}^{\text{im}}(P)$.

Observe that the transition matrix \mathbf{T}^{im} with elements given by Equation (11) for the imitation process is interactive.²⁵ It depends on the current distribution of log-productivity $P(t)$ and it is given by

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

with $S_a(P) \equiv 1 - \sum_{b=a+1}^{\infty} T_{ab}^{\text{im}}(P) = 1 - \sum_{b=a+1}^{\infty} q^{b-a} (P_b + (1-q)(1 - F_b))$. Following Proposition 1, the evolution of the log-productivity distribution in the limit of large N is given by

$$\frac{\partial P(t)}{\partial t} = P(t)(\mathbf{T}^{\text{im}}(P(t)) - \mathbf{I}). \quad (12)$$

From Equation (12) we can derive a differential equation governing the evolution of the cumulative log-productivity distribution.

Proposition 4. *Assume firms cannot innovate in-house ($\eta_0 = 1$ and $\eta_i = 0$, for all $i \geq 1$), then in the limit of large N , the evolution of the cumulative log-productivity distribution $F(t)$ is given*

²⁴The first term in the first line of Equation (11) considers the case of a firm with log-productivity a observing a firm with log-productivity b at random and successfully climbing the $b - a$ rungs separating them, which happens with probability q^{b-a} . The second term considers the case that a firm with log-productivity a observes a firm with log-productivity $b + 1$, successfully climbs $b - a$ rungs (with probability q^{b-a}) but fails to climb the last rung (with probability $1 - q$). And so on. See also Figure 3.

²⁵For an interactive Markov chain the transition probabilities depend on the current distribution [Conlisk, 1976].

by

$$\frac{\partial F_a(t)}{\partial t} = F_a(t)^2 - F_a(t) + (1-q)(1-F_a(t)) \sum_{b=0}^{a-1} q^b F_{a-b}(t), \quad a \in S, \quad (13)$$

for some initial distribution $F(0) : S \rightarrow [0, 1]$.

The boundary conditions are $\lim_{a \rightarrow 1} F_a(t) = 0$ and $\lim_{a \rightarrow \infty} F_a(t) = 1$. Consider an initial distribution $F_a(0)$ with finite support. Then there exists a maximal initial log-productivity a_m such that $F_a(0) = 1$ for all $a \geq a_m$. From Equation (13) we see that for all $a \geq a_m$ it must hold that $\frac{\partial F_a(t)}{\partial t} = 0$ and so $F_a(t) = 1$ for all $t \geq 0$. In contrast, for all $a < a_m$ 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

$$\lim_{t \rightarrow \infty} F_a(t) = \begin{cases} 0, & \text{if } a < a_m, \\ 1, & \text{if } a \geq a_m. \end{cases} \quad (14)$$

Note, that Equation (14) is equivalent to $\lim_{t \rightarrow \infty} P_{a_m}(t) = 1$. Thus all probability mass concentrates at a_m in the course of time. Note also that in the special case of $q = 1$ we recover the knowledge growth dynamics analyzed by Lucas [2008].²⁶

4.2.3. Innovation and Imitation

Finally, we consider the evolution of the productivity distribution in a world where R&D strategies are exogenous with a fixed fraction of innovators and imitators. Since R&D strategies are exogenous, we observe that the dynamics of the innovators always follows the dynamics of the case of pure in-house R&D in Section 4.2.1. From our previous discussion, we know that for these firms we obtain a log-normal log-distribution with a variance that increases over time (see Proposition 3). Since the proportion of innovators and imitators is fixed, this implies that also the variance of the distribution of the total population of firms will diverge when we have both innovators and imitators.²⁷ However, from the empirical evidence discussed in Section 2 we do not find support for such a divergence. Hence, in order to develop a more realistic model for the evolution of the empirical productivity distribution, we need to move beyond the case of exogenous innovation-imitation strategies. This is the focus of the following section.

²⁶For $q = 1$ we can derive from Equation (13) the following differential equation $\frac{\partial F_a(t)}{\partial t} = F_a(t)^2 - F_a(t)$. This can be written as $\frac{\partial \ln F_a(t)}{\partial t} = F_a(t) - 1$, with the solution

$$F_a(t) = \frac{F_a(0)}{F_a(0) + e^t(1 - F_a(0))}, \quad a \in S,$$

and the initial distribution $F_a(0)$. We see that $\lim_{t \rightarrow \infty} F_a(t) = 0$ as long as $F_a(0) < 1$, while $F_a(t) = 1$ for all $a \in S$ with the property that $F_a(0) = 1$.

²⁷It is possible to characterize the dynamics of the cumulative log-productivity distribution in terms of a differential equation. However, this admits no closed-form solution. The analysis is deferred to Appendix A.

4.3. Endogenous Innovation-Imitation Strategies

This section contains the main contribution of our paper. We analyze the case in which firms choose whether to innovate (in-house R&D) or to imitate other firms, based on a standard value-maximization objective. In our environment this is equivalent to maximizing the expected profit flow, given by (4), in each period. We assume $\epsilon_i(t) \in \mathbb{R}_+$ to be a firm-specific shock affecting the firm's productivity in implementing in-house R&D. In particular, $\epsilon_i(t) = 1$ if a firm decides to imitate and $\epsilon_i(t) = \epsilon_i^{\text{in}}(t)$ if the firm pursues in-house R&D, where $\epsilon_i^{\text{in}}(t)$ is a stochastic variable, assumed to be independently and identically log-logistic distributed across firms, with scale parameter β [Fisk, 1961].²⁸ Thus, $\log \epsilon_i^{\text{in}}(t)$ has a logistic distribution. In a time interval $[t, t + \Delta t)$ a single firm is selected at random, makes a new draw $\epsilon_i^{\text{in}}(t)$, and is granted the opportunity to change its R&D strategy. Such events are i.i.d. across firms and over time. $\epsilon_i^{\text{in}}(t)$ is drawn from the cdf $F_{\epsilon_i^{\text{in}}(t)} : \mathbb{R}_+ \rightarrow [0, 1]$, where $F_{\epsilon_i^{\text{in}}(t)}(x) = \frac{1}{1+x^{-\beta}}$ (or, identically, $F_{\log \epsilon_i^{\text{in}}(t)}(x) = \frac{1}{1+e^{-\beta x}}$). When no event occurs, $\epsilon_i^{\text{in}}(t)$ remains constant. Note that the median of $\epsilon_i^{\text{in}}(t)$ equals 1/2, implying that, at every t , 50% of the firms have a comparative advantage in imitation and innovation, respectively. This is for simplicity and entails no loss of generality.

Firm i chooses innovation whenever, conditional on its current productivity $A_i(t)$ and the state of $\epsilon_i(t)$ (which is known at time t)

$$\epsilon_i^{\text{in}}(t) \times \mathbb{E}_i^{\text{in}} [A_i(t + \Delta t) | A_i(t)] > \mathbb{E}_i^{\text{im}} [A_i(t + \Delta t) | A_i(t)],$$

where $\mathbb{E}_i^{\text{in}}[\cdot | A_i(t)]$ and $\mathbb{E}_i^{\text{im}}[\cdot | A_i(t)]$ denote expectations conditional on innovation and imitation of firm i with current productivity level $A_i(t)$, respectively. Note that, as $\beta \rightarrow \infty$ (almost) all firms draw $\epsilon_i^{\text{in}}(t) \approx 1$, namely, productivity shocks vanish. To the opposite, as $\beta \rightarrow 0$, half of the firms are totally unable to implement innovations, and the other half are infinitely productive. Thus, this corresponds to a case in which half of the firms do in-house R&D and half imitate, irrespective of their initial productivity.²⁹

We can now analyze the expected productivity of a firm i with a given productivity $A_i(t)$ at time t if it were to imitate or innovate at time $t + \Delta t$. Let $A_i^{\text{in}}(A_i(t)) = \mathbb{E}_i^{\text{in}} [A_i(t + \Delta t) | A_i(t)]$ be the expected productivity of firm i if it innovates at time $t + \Delta t$. We have that

$$A_i^{\text{in}}(A_i(t)) = \bar{A}^{a_i(t)} \mathbb{P}(\eta(t + \Delta t) = 0) + \bar{A}^{a_i(t)+1} \mathbb{P}(\eta(t + \Delta t) = 1) + \dots$$

Similarly, let $A_i^{\text{im}}(A_i(t), P(t)) = \mathbb{E}_i^{\text{im}} [A_i(t + \Delta t) | A_i(t)]$ be the expected productivity of firm i if it imitates. Analogous to the derivation of Equation (11), one can show that the expected

²⁸The assumption that under imitation $\epsilon_i(t) = 1$ is for simplicity and entails no loss of generality, since what matters is the difference between the value of $\epsilon_i(t)$ under innovation and imitation, respectively. For instance, we can think that when a technology is copied, there is no uncertainty about the best way to implement it.

²⁹Note that even the case of $\beta \rightarrow 0$ is different from that of exogenous innovation and imitation strategies discussed in Section 4.2.3. There, each firm was permanently assigned to a given type (either innovator or imitator). In contrast, here, even in the case of $\beta \rightarrow 0$ firms change their R&D strategy over time, although the proportions of innovators and imitators remain approximately constant.

productivity of firm i if it imitates is given by

$$A_i^{\text{im}}(A_i(t), P(t)) = \bar{A}^{a_i(t)} S_{a_i(t)}(P(t)) + \sum_{b=a_i(t)+1}^{\infty} \bar{A}^b q^{b-a} (P_b(t) + (1-q)(1-F_b(t))), \quad (15)$$

with $S_a(P) = 1 - \sum_{b=a+1}^{\infty} \mathbf{T}_{ab}^{\text{im}}(P)$ from Section 4.2.2. Consequently, the expected profit of firm i when innovating in-house is given by $\pi_i^{\text{in}}(t) = \psi A_i^{\text{in}}(A_i(t)) \epsilon_i^{\text{in}}(t)$, and similarly, the expected profit of firm i through imitation is $\pi_i^{\text{im}}(t) = \psi A_i^{\text{im}}(A_i(t), P(t))$, where $\epsilon_i^{\text{in}}(t)$ are i.i.d. non-negative random variables. Let $a_i^{\text{in}}(a_i(t)) \equiv \log A_i^{\text{in}}(A_i(t))$ and $a_i^{\text{im}}(a_i(t), P(t)) \equiv \log A_i^{\text{im}}(A_i(t), P(t))$. Further, let

$$p_{\beta}^{\text{im}}(a_i(t), P(t)) \equiv \mathbb{P}(\pi_i^{\text{im}}(t) > \pi_i^{\text{in}}(t)) = \mathbb{P}(\log \epsilon_i^{\text{in}}(t) < a_i^{\text{im}}(a_i(t), P(t)) - a_i^{\text{in}}(a_i(t)))$$

denote the probability that a firm's profit from imitation is larger than from innovation. Then

$$\begin{aligned} p_{\beta}^{\text{im}}(a_i(t), P(t)) &= F_{\log \epsilon_i^{\text{in}}(t)}(a_i^{\text{im}}(a_i(t), P(t)) - a_i^{\text{in}}(a_i(t))) \\ &= \frac{1}{1 + e^{-\beta(a_i^{\text{im}}(a_i(t), P(t)) - a_i^{\text{in}}(a_i(t)))}}, \end{aligned} \quad (16)$$

where the second equality follows from the properties of the logistic function. We also define the probability of innovation as $p_{\beta}^{\text{in}}(a_i(t), P(t)) \equiv 1 - p_{\beta}^{\text{im}}(a_i(t), P(t))$.

An important results regarding the propensity of firms to conduct in-house R&D can be given for a particularly simple and intuitive case in which one step of innovation is achieved with probability p , thus, $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$. Further, assume for simplicity that firms do not have any absorptive capacity limits ($q = 1$). We then can give the following proposition:

Proposition 5. *Assume that $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$ for some $p \in (0, 1)$, $\beta > 0$ and that firms do not have any absorptive capacity limits ($q = 1$). Then for any P there exists a unique threshold log-productivity $a^* \in \mathbb{N}$ such that it holds that $p_{\beta}^{\text{im}}(a, P) > p_{\beta}^{\text{in}}(a, P)$ when $a < a^*$, and $p_{\beta}^{\text{im}}(a, P) < p_{\beta}^{\text{in}}(a, P)$ when $a > a^*$, where it might hold that $p_{\beta}^{\text{im}}(a^*, P) = p_{\beta}^{\text{in}}(a^*, P)$.*

Proposition 5 states that relatively backward firms (below the threshold a^*) are more likely to imitate, while firms with an advanced technology (above the threshold a^*) are more likely to innovate.

We now turn to the general description of the evolution of the productivity distribution. Observe that the transition matrix $\mathbf{T}(P)$ is the sum of the transition matrix for imitation $\mathbf{T}^{\text{im}}(P)$ (see Section 4.2.2) and innovation \mathbf{T}^{in} (see Section 4.2.1), each weighted with the probability of imitation $p_{\beta}^{\text{im}}(a, P)$ and innovation $p_{\beta}^{\text{in}}(a, P) = 1 - p_{\beta}^{\text{im}}(a, P)$, respectively. Further, observe that the imitation probability $p_{\beta}^{\text{im}}(a, P)$ is continuous in P for any finite β . Hence, by virtue of Proposition 1, we are able to state the dynamics of the log-productivity distribution when the population N of firms becomes large.

Proposition 6. *Assume that $\beta < \infty$ and let $\mathbf{D}(P)$ be the diagonal-matrix of all probabilities $p_{\beta}^{\text{im}}(a, P)$. Then in the limit of large N , the evolution of the log-productivity distribution $P(t)$ is*

given by

$$\frac{\partial P(t)}{\partial t} = P(t) (\mathbf{T}(P) - \mathbf{I}) = P(t) ((\mathbf{I} - \mathbf{D}(P)) \mathbf{T}^{in} + \mathbf{D}(P) \mathbf{T}^{im}(P) - \mathbf{I}), \quad (17)$$

for some initial distribution $P(0) : S \rightarrow [0, 1]$.

In the special case of $\eta_1 = p$ and $\eta_0 = 1 - p$ for some $p \in (0, 1)$ we obtain

$$\begin{aligned} \frac{\partial P_a(t)}{\partial t} = & P_a(t) \left(\sum_{b=1}^{a-1} p_\beta^{im}(b, P) P_b(t) + p_\beta^{im}(a, P) S_a(t) \right) + (1 - p) P_a(t) p_\beta^{in}(a, P) \\ & + p P_{a-1}(t) p_\beta^{in}(a - 1, P) - P_a(t), \quad a \in S. \end{aligned} \quad (18)$$

Proposition (6) provides a complete characterization of the evolution of the productivity distribution, which can be computed by direct numerical iteration.

The above discussion does not cover the case without shocks ($\beta = \infty$), i.e., $\epsilon_i(t) = \epsilon_i^{in}(t) = 1$ for all $i \in \mathcal{N}$. This case is of particular economic interest, since all firms are *ex ante* identical and the decision whether to imitate or do in-house R&D is entirely determined by the state of productivity. However, the technical analysis is somewhat more involved since the imitation probability $p_\beta^{im}(a_i(t), P(t))$ of Equation (16) has a point of discontinuity in the limit as $\beta \rightarrow \infty$ (corresponding to the case of vanishing shocks), and has $V(P) = \mathbf{T}(P) - \mathbf{I}$ where the transition matrix is given by Equation (17). In this case, the evolution of the log-productivity distribution then follows a differential inclusion (a set-valued differential equation) $\frac{\partial P(t)}{\partial t} \in \bar{V}(P)$, as stated in Proposition 2.

The model is parsimoniously parameterized by a parameter $\beta \geq 0$ governing the variance of the productivity shocks, the in-house innovation probability $p \in [0, 1]$ (restricting our analysis to the case of $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$), and the imitation probability $q \in [0, 1]$. In order to better understand the resulting productivity distributions and their dependency on the parameters of the model, we analyze in the following sections two polar opposite cases where additional analytical results can be proven. In Section 4.3.1 we study the limit of strong productivity shocks (characterized by $\beta \rightarrow 0$), while in Section 4.3.2 we consider the limit of no productivity shocks (as covered by Proposition 2). In both cases we will demonstrate that the productivity distribution follows a traveling wave with tails that exhibit a power-law decay, corresponding to a log-productivity distribution with exponential tails.³⁰ A numerical analysis (based on the numerical integration of Equation (18)) shows that the result also holds for intermediate values of the variance of shocks.

4.3.1. Equilibrium Growth with Strong Shocks

In the limit case of $\beta \rightarrow 0$, it holds imitation probability $p_\beta^{im}(a, P) = 0.5$ in Equation (16). This means that firms chooses between imitation and in-house R&D uniformly at random for all a and all P . Inserting this into Equation (18), and focusing on the case of q close to unity, yields the

³⁰Note that $P(a, t) \propto e^{-\lambda a} = e^{-\lambda \log A} = A^{-\lambda}$.

following differential equation for the dynamics of the cumulative distribution function $F(t)$

$$\frac{\partial F_a(t)}{\partial t} = \frac{2q-1}{2}(F_a(t)^2 - F_a(t)) - \frac{p}{2}(F_a(t) - F_{a-1}(t)), \quad a \in S, \quad (19)$$

subject to the boundary conditions $\lim_{a \rightarrow \infty} F_a(t) = 1$ and $\lim_{a \rightarrow 1} F_a(t) = 0$. A numerical solution for the resulting probability mass function is given in Figure 4 (left). Our analysis will reveal that the distribution obtains a stable shape moving to the right (with increasing average log-productivity) over time. Such a solution is called a *traveling wave*. More precisely, a traveling wave is a solution of the form $F_a(t) = f(a - \nu t)$ such that for any $s \geq t$ it must hold that $F_a(t) = F_{a+\nu s}(t + s)$. For specific initial conditions, we can give a more formal result, as stated in the following proposition.

Proposition 7. *Let $F_a(t)$ be a solution of Equation (19) with Heaviside initial value $F_a(0) = \Theta(a - a_m)$ for some $a_m \geq 1$ and define $m_\epsilon(t) = \inf\{a : F_a(t) > \epsilon\}$. Then*

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

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 $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$.

From a numerical integration of this differential equation we find that the limiting log-productivity distribution decays exponentially in the tails. Guided by this observation, we impose a general exponential function for the tails and derive the parameters of this function, as well as the traveling wave velocity.³¹ It turns out that in the limit of vanishing absorptive capacity limits ($q = 1$) we recover the model analyzed by Majumdar and Krapivsky [2001]. The general case can be solved in a similar way to their analysis and is stated in the following proposition.

Proposition 8. *Assume that $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$ with $p \in [0, 1]$ and strong productivity shocks with $\beta = 0$. Consider weak absorptive capacity limits such that q is close to unity and Equation (19) holds.*

- (i) *If we assume that the front of the traveling wave solution of Equation (19) follows an exponential distribution with exponent $\lambda \geq 0$, i.e. $F_a(t) \propto e^{-\lambda(a-\nu t)}$ for a much larger than νt , then the traveling wave velocity is given by*

$$\nu(p, q) = \frac{2q - 1 - p + pe^{\lambda(p, q)}}{2\lambda(p, q)}, \quad (20)$$

³¹Note that the assumption of an exponential decay is not very restrictive, as such tail distributions are common for a broad class of probability distributions. More precisely, for the log-productivity values exceeding some high threshold a^* , i.e. $a > a^*$, the distribution F_a can be written approximately as [see e.g. De Haan and Ferreira, 2006, Chap. 3]

$$1 - F_a \approx (1 - F_{a^*}) \left(1 - H_\gamma \left(\frac{a - a^*}{f(a^*)} \right) \right), \quad a > a^*,$$

for some positive function $f : \mathbb{R} \rightarrow \mathbb{R}_+$, where $H_\gamma(a) = 1 - (1 + \gamma a)^{-\frac{1}{\gamma}}$ is the generalized Pareto distribution. This is known as the Pickands–Balkema–de Haan theorem [Pickands, 1975]. For $\gamma = 0$ we obtain an exponential distribution, that is $\lim_{\gamma \rightarrow 0} H_\gamma(a) = 1 - e^{-a}$.

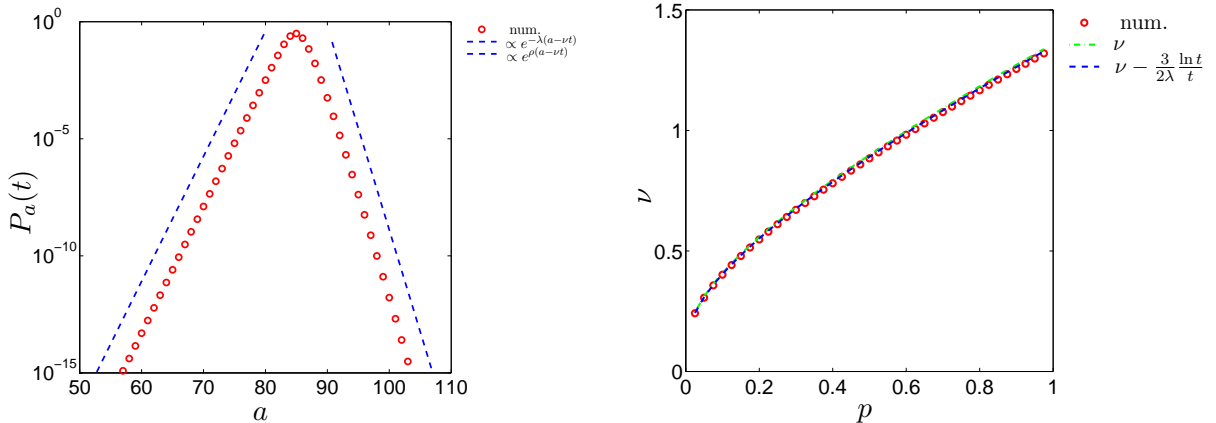


Figure 4: (Left) The log-productivity distribution $P_a(t)$ for period $t = 200$, $p = 0.119$, $q = 1$. The distribution obtained by numerical integration of Equation (35) is indicated by a circles while the theoretical predictions are shown with a dashed line. The front of the traveling wave decays 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 $t = 100$.

where the exponent $\lambda(p, q)$ of the front of the distribution is given by³²

$$\lambda(p, q) = 1 + W\left(\frac{2q - 1 - p}{pe}\right). \quad (21)$$

(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 $\rho(p, q)$ is given by

$$\rho(p, q) = \frac{1}{2} \left(2q - 1 + p + 2W\left(-\frac{p}{2} e^{\frac{1-p-2q}{2}}\right) \right). \quad (22)$$

Our numerical analysis shows that the results of Proposition 8 hold for general initial distributions which are concentrated enough, such as an exponential distribution with an exponent that is large enough. Moreover, when there are no limitations in the abilities of firms to imitate other firms' technologies ($q = 1$), one can show that transitional dynamics of the average log-productivity is given by $\mathbb{E}_t[a] = \nu(p, q)t - \frac{3}{2\lambda(p, q)} \ln t + O(1)$ [Majumdar and Krapivsky, 2001].

By solving the dynamical system corresponding to Equation (19) but continuous in state we can compute $F_a(t)$ for large times t . The resulting traveling wave velocity $\nu(p, q)$ (the productivity growth rate) for $q = 1$ can be seen in Figure 4 (right) for different values of p together with our theoretical predictions. From Equations (20) and (21) we directly find that the growth rate of the economy is increasing in the innovation probability p (see also Figure 4, right) and the absorptive capacity parameter q , i.e. $\frac{\partial \nu(p, q)}{\partial p} > 0$, $\frac{\partial \nu(p, q)}{\partial q} > 0$. Further results in the limit of strong productivity shocks (considering that β is small but positive) can be found in Appendix B. Our analysis reveals that with increasing β the growth rate ν of the economy's productivity increases.

³² $W(x)$ is the Lambert W function (or product log), which is implicitly defined by $W(x)e^{W(x)} = x$, and can be written as $W(x) = -\sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} (-x)^n$ for $|x| < \frac{1}{e}$.

4.3.2. Equilibrium Growth without Shocks

In this section we consider an economy where productivity shocks become irrelevant as $\beta \rightarrow \infty$ and for all firms $i \in \mathcal{N}$ we have that $\epsilon_i(t) = \epsilon_i^{\text{in}}(t) = 1$. In this case, all firms are ex ante identical and the decision whether to imitate or to conduct in-house R&D is entirely determined by their state of productivity. As the following lemma illustrates, there exists a critical log-productivity level a^* below which firms only imitate other firms technologies, while the firms with log-productivities above the threshold a^* conduct only in-house R&D.

Lemma 1. *Assume that firms have maximal absorptive capacity ($q = 1$). Then in the limit of vanishing productivity shocks ($\beta \rightarrow \infty$) there exists for any productivity distribution a threshold log-productivity $a^* \in \mathbb{N}$ such that*

$$\lim_{\beta \rightarrow \infty} p_{\beta}^{\text{im}}(a, P) = \begin{cases} 1, & \text{if } a \leq a^*, \\ 0, & \text{if } a > a^*. \end{cases} \quad (23)$$

The dynamics of the log-productivity distribution is given by the differential inclusion of Proposition 2. From Equation (23) we observe that the imitation probability has a point of discontinuity at a^* and is continuous for $a < a^*$ and $a > a^*$. The same holds for the function $\bar{V}(P)$ of Proposition 2, which then is given by $V(P)$ for all $a \neq a^*$, and the differential inclusion becomes a differential equation at the continuity points of $V(P)$. We then can state the following proposition.

Proposition 9. *Let $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$ with $p \in [0, 1]$. Consider vanishing productivity shocks ($\beta \rightarrow \infty$), and assume that firms have maximal absorptive capacity ($q = 1$). Then the dynamics of the cumulative log-productivity distribution is given by*

$$\frac{\partial F_a(t)}{\partial t} = \begin{cases} F_a(t)^2 - F_a(t), & \text{if } a < a^*, \\ (F_a(t) - 1)F_{a^*}(t) - p(F_a(t) - F_{a-1}(t)), & \text{if } a > a^*. \end{cases} \quad (24)$$

The above differential equation for $F_a(t)$ can be solved numerically subject to the boundary conditions $\lim_{a \rightarrow \infty} F_a(t) = 1$ and $\lim_{a \rightarrow 1} F_a(t) = 0$. The resulting log-productivity distribution for $p = 0.1$ and $q = 1$ can be seen in Figure 5 (left).

A numerical integration of Equation (24) reveals that the log-productivity distribution is a traveling wave with exponential tails. Moreover, we observe that the threshold log-productivity a^* is close to the average log-productivity νt . In the following proposition we then assume a general exponential function for the tails,³³ and further assume that the threshold log-productivity a^* is given by νt , in order to solve for the limiting log-productivity distribution.

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 vanishing productivity shocks ($\beta \rightarrow \infty$) and assume that firms do not have any absorptive capacity limits ($q = 1$).*

³³Such a behavior is typical for large productivity values. See also Footnote 31.

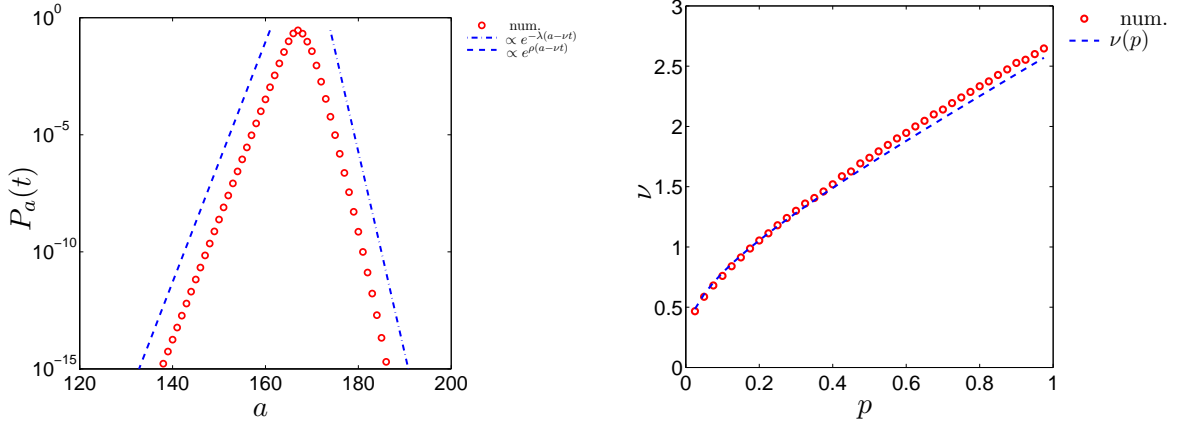


Figure 5: (Left) The log-productivity distribution $P_a(t)$ for $p = 0.125$ and $\log \bar{A} = 1$ at $t = 400$. The distribution obtained by numerical integration of Equation (24) is indicated with circles while the theoretical predictions are shown with a dashed line. The front of the traveling wave 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 (24) and theoretical prediction indicated by the dashed line.

- (i) If we assume that the front of the traveling wave solution of Equation (24) follows an exponential distribution with exponent $\lambda \geq 0$ for all $a \geq a^*$, i.e. $P_a(t) \propto e^{-\lambda(a-\nu t)}$, and we assume that the threshold log-productivity a^* is given by the average νt , then the traveling wave velocity is given by

$$\lim_{\beta \rightarrow \infty} \nu^\beta(p) = \frac{1}{\lambda} \left(1 + p(e^\lambda - 1) - \frac{p(\bar{A} - 1)(1 - e^{1-\lambda})}{e - 1} \right) \quad (25)$$

where λ is the root of the equation

$$e^\lambda(\lambda - 1) - \frac{\bar{A} - 1}{e - 1} e^{1-\lambda}(1 + \lambda) + \frac{\bar{A} + e - 2}{e - 1} = \frac{1}{p}. \quad (26)$$

- (ii) If we assume that the left tail of the traveling wave solution of Equation (24) follows an exponential function with exponent $\rho \geq 0$, i.e. $P_a(t) \propto e^{\rho(a-\nu t)}$, for all a much smaller than the threshold a^* then the exponent ρ is given by $\rho = 1/\nu$.

Equation (26) can be solved numerically, using standard numerical root finding procedures [see e.g. Press et al., 1992, Chap. 9], to obtain the exponent λ . Inserting λ into Equation (25) further gives the traveling wave velocity ν . This is shown in Figure 5 (right) together with the numerical values for $p = 0.125$ and $\log \bar{A} = 1$ at $t = 400$. The figure shows that the traveling wave velocity ν is increasing with the innovation success probability p . This means that as firms' in-house R&D success probability increases, also the growth rate of the economies productivity increases when absorptive capacity of firms is maximal ($q = 1$).³⁴

³⁴Numerical integration for $\beta \rightarrow \infty$ and q intermediate between zero and one suggest the hypothesis that there is a transition in the (p, q) -space between travelling-wave behavior and log-normal-with-rising-variance behavior. For $q > 5p$ the traveling-wave behavior is visually clear. For $q \leq p$ the log-normal-with-rising-variance behavior is visually clear.

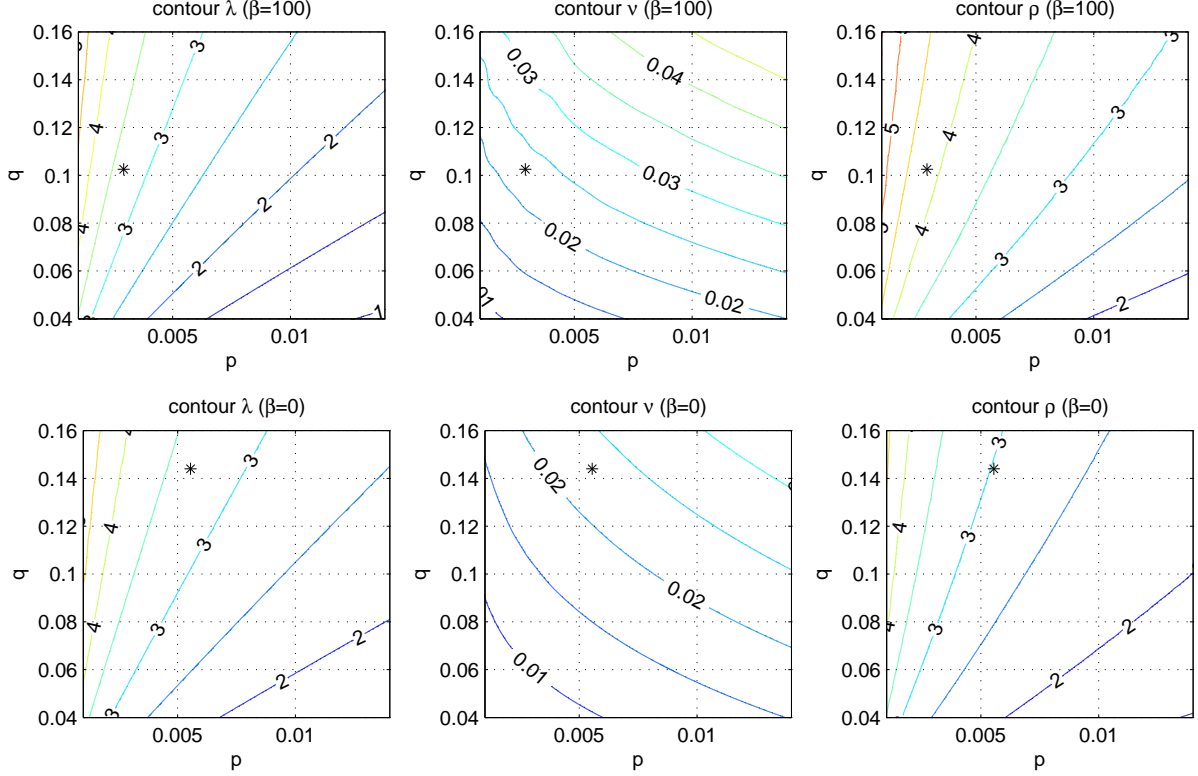


Figure 6: Exploration of impact of innovation probability p , imitation probability q , on the dependent power-law parameters λ , and ρ , and on the productivity growth rate ν . In the top row the parameter $\beta = 100$ is relatively high corresponding to the case where productivity shocks are almost absent. The bottom row corresponds to $\beta = 0$ with maximally large shocks. The contour plots are based on numerical integrations of the ODE in Eq. (17) as explained in Section 5. The black asterisk marks the calibrated (p, q) -points which approximates λ and ν closest (but not ρ).

5. Calibration of the Model's Parameters

The goal of this section is to calibrate the model's parameters given by the innovation success probability p and the imitation success probability q , such that the empirically observed right tail exponent λ and the growth rate of the traveling wave ν can be reproduced.

Ideally, also the parameter β should be included in the calibration procedure. But it turned out that its impact is not as strong, and we put less emphasis on it. Further on, ideally, also the left tail exponent ρ shall fit. But we also regard the fit of ρ as second order priority for two reasons. First, the empirical data is less reliable in the lower region. Second, from a practical point of view, it is much more interesting to understand the growth of productivity and the spread of the distribution to the few most innovative firms than how the distribution expands to the least innovative firms.

Our theoretical results on the computation of λ, ν and ρ cover only parts of the (β, p, q) -parameter space. Further on, the interdependence we know is quite complex and non-linear. Thus, a simple regression estimation procedure is ruled out.

We developed a hands on method to estimate λ, ρ and ν for computed trajectories with parameters p, q and β , based on some heuristics which we derived from sorrow observations. The method works as follows: Start with initial distribution $P_0 = (1 \ 0 \dots)$ on a long enough

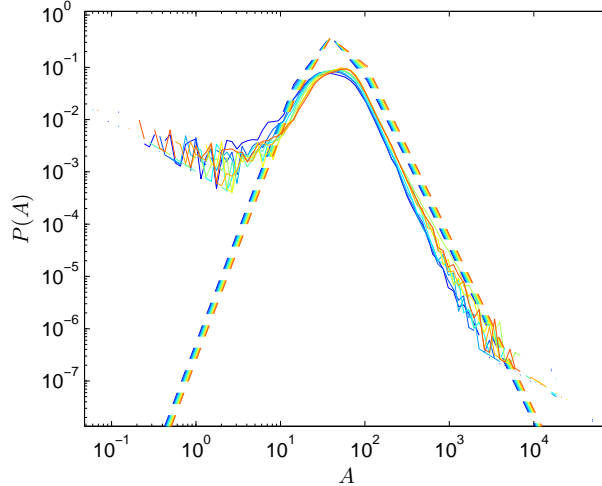


Figure 7: ODE trajectories for calibrated parameters ($p = 0.002925, q = 0.1025$) compared to the empirical pdf. The parameter $\beta = 100$ is set close to dynamics without shocks. The parameters p and q are calibrated to fit $\lambda = 3.32$ (from Table 1) and $\nu = 0.0227$ (from Table 3) closest. This leads to a left power-law tail exponent $\rho = 2.46$. The empirically estimated value is $\rho = 1.46$ (see Table 1), but this might be an artifact of a distortion in the very low productivity region, as the slope fits the distribution in the intermediate low level well. The solution of the ODE of log-productivity is computed starting with $P = (1 \ 0 \ \dots)$ until its peak fits to the peak of the empirical distribution from the year 2003. The solution is shifted upwards to enable better comparison.

vector (we used length 30). All distributions mentioned here are handled as pdf's. Decide on an appropriate T_{\max} and compute the distributions numerically (with `Matlab`'s function `ode45`) along the trajectory at time steps $t = 0, 1, 2, \dots, T_{\max}$. Heuristics for the choice of T_{\max} where experimentally quantified such that the peak of the distribution at T_{\max} lies well in the center of the support of P_0 .

Then compute the arithmetic mean of productivity and the geometric mean of productivity for the distribution in each time step t . The arithmetic means build the lower bounds for the support of the distribution where λ is fitted by linear regression on the logarithm of productivity and the logarithm of the distribution function. The geometric means build the upper bounds for the support of the distribution where ρ is fitted by linear regression on the logarithm of productivity and the logarithm of the distribution function.³⁵ Based on this we are able to fit λ and ρ for each time step t . We compute an estimate for ν for each time step t by looking at the differences in average log-productivity for time step t and $t - 1$.

We observed that for large enough T_{\max} the fitted values stabilize, but some regular fluctuations remained due to the discreteness of the support of the distribution. To minimize the effect we averaged several values of λ and ρ along an interval of values of t of a certain length until T_{\max} .³⁶

Based on this calibration method we computed values of λ, ρ and ν for the theoretical distributions of the ODE as a function of p, q and β on the grid $p = 0.001, ^{+0.0002}, 0.014,$

³⁵Support for fitting was further restricted to the region where the distribution function was larger than a certain accuracy to avoid distortion from border effect which appear when floating point precision achieves its limits.

³⁶We found reasonable heuristics for assigning such a "wavelength" that the slight fluctuations could be averaged out well.

$q = 0.04, +0.002, 0.16$ and $\beta = 0, 100$. After computation of the field we improved accuracy of the grid (using Matlab's function `interp2`). We improved the accuracy of p to steps of length 0.000025 and the accuracy of q to steps of length 0.00025. Within this grid we computed the values of p and q which minimized the quadratic difference of empirical and theoretical λ plus the quadratic difference of the empirical and theoretical ν .

In Figure 6 and 7 we report the calibrated values of $p = 0.002925, q = 0.1025$, (for the case of $\beta = 100$ where shocks are almost absent) and show in the latter figure the computed theoretical trajectories of the distribution of productivity together with the time evolution of empirical distribution. The figure reveals that the model can well reproduce the observed pattern.

6. Growth, Inequality and Policy Implications

Our model is parsimoniously parameterized by a parameter $\beta \geq 0$ governing the variance of the productivity shocks, the in-house innovation probability $p \in [0, 1]$ and the parameter $q \in [0, 1]$ measuring the absorptive capacity of the firms in the economy. In this section we study the effects of each of the three parameters (β, p, q) on (i) the speed of growth and (ii) the inequality implied by the productivity distribution. This will allow us to analyze the effects of R&D policies that impact the innovation success probability p and the imitation success probability q . Examples for the first are R&D subsidy programs that foster the development of in-house innovations while policies that weaken the intellectual property protection regime (and hence make it easier to imitate others' technologies) are examples for the latter.

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.³⁷ Equivalently, this corresponds to a higher average log-productivity per unit of time, as measured by the growth rate ν . From our analysis in the previous sections, we can derive the following result for the growth rate (and thus for efficiency) comparing the two extreme cases of vanishing and strong productivity shocks.

Proposition 11. *Assume that there are no absorptive capacity limits ($q = 1$). Then, under the assumptions of Propositions 8 and 10, we have that $\lim_{\beta \rightarrow 0} \nu^\beta(p) < \lim_{\beta \rightarrow \infty} \nu^\beta(p)$ for any $p \in (0, 1)$.*

Proposition 11 shows that the growth rate (and hence aggregate productivity and output) are higher when shocks are small. In this proposition we have assumed that firms do not face any absorptive capacity limits (by setting $q = 1$). By means of a numerical integration of Equation (17) we can also study general values of q . We do this for two possible cases: (a) we keep the value of the absorptive capacity parameter q at its calibrated value of 0.079 and analyze the impact of changes in the innovation success probability p , or (b) we set p to its calibrated value of 0.0054 and study the effects of a change in q (see also Section 5 for the estimation of these parameters).

³⁷We will consider the average productivity measured by the geometric mean $\mu = \sqrt[n]{A_1 A_2 \cdots A_n} = (\prod_{i=1}^n A_i)^{1/n}$, which is related to the arithmetic average of the log-productivity values via $\frac{1}{n} \sum_{i=1}^n \log A_i = \log \mu$. However, our results also hold for the arithmetic average of the productivity values.

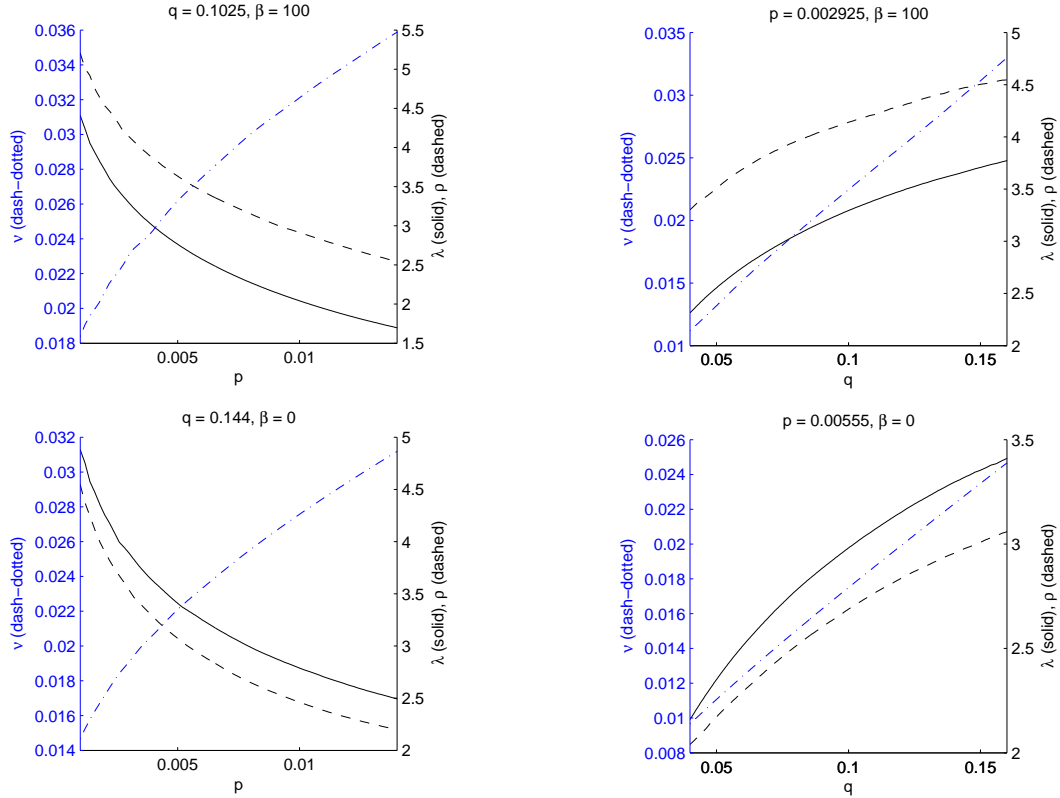


Figure 8: Plots of λ, ρ and ν for p (resp. q) when q (resp. p) is fixed to the value from the calibrated parameters illustrated in Figure 6 (at the black asterisk). In the top row the parameter $\beta = 100$ is relatively high corresponding to the case where productivity shocks are almost absent. The bottom row corresponds to $\beta = 0$ with maximally large shocks.

For both cases we provide a numerical analysis when productivity shocks are small ($\beta = 100$) and when productivity shocks are being dominant ($\beta = 0$). The results are shown in Figure 8.

In case (a) in Figure 8 (left panels) we find that for both values of β an increase in the innovation success probability p increases ν and hence accelerates growth. Thus, an R&D subsidy program which increases firms' in-house R&D success probability p leads to a higher growth rate of the economy. A similar analysis, but with varying values of the absorptive capacity (i.e. the imitation success probability q) in case (b) is shown in Figure 8 (right panels). The figure reveals that an increase in the absorptive capacity q always increases the growth rate ν . Thus, an implication of our model is that policies which positively affect the absorptive capacity q , for example by weakening the intellectual patent protection of incumbent technologies in an industry, can have a positive effect on the growth rate ν of the economy.

A complete numerical analysis of the growth rate ν for general values of q in the case of productivity shocks being small ($\beta = 100$) and in the case of productivity shocks being dominant ($\beta = 0$) is shown in Figure 6 (middle panels). The figure confirms the result of Proposition 11. For all values of the innovation probability p , the productivity growth rate ν is higher the smaller the productivity shocks are (comparing the cases of $\beta = 100$ and $\beta = 0$). The same holds for an increase in the absorptive capacity parameter q . Moreover, an increase in p or q leads to a higher growth rate ν .

Further, we can investigate the degree of inequality in the economy. As our measure of inequality we take the exponent λ of the right power-law tail of the distribution. A smaller value of λ corresponds to a more dispersed distribution with a higher degree of inequality. For both cases (a) and (b) we provide a numerical analysis for small productivity shocks ($\beta = 100$) and dominant productivity shocks ($\beta = 0$) in Figure 8 (left panel). In case (a) we see that the exponent λ is always higher in the limit of strong productivity shocks and the difference increases with increasing innovation success probability p . However, in case (b) the reverse relationship holds: an increase in the absorptive capacity q yields a higher value of λ and thus reduces inequality.

We can derive the following policy implications from the preceding analysis. We find that both types of policies, those that enhance the in-house innovation success probability p as well as those that facilitate the imitation and diffusion of existing technologies (increasing the value of q) increase the growth rate ν of the economy. However, while the first leads to an increase in inequality (smaller values of λ), the latter has the opposite effect of decreasing inequality (higher values of λ). It must be noted, however, that an economy in which technologies can easily be imitated (high q) but there is no in-house R&D ($p \rightarrow 0$) does not generate growth. Thus, a balanced approach is required, fostering both, the capacities of firms to generate innovations in-house and an environment in which these innovations can diffuse throughout the economy.

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 firms productivities are exposed to exogenous shocks [cf. Bloom, 2009]. We show that these shocks 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 D. First, in Appendix D.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 D.2 we depart from the assumption of a fixed population of firms and instead allow for firm entry and exit. Third, in Appendix D.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., 2009. Introduction to modern economic growth. Princeton.
- 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., Cao, D., 2010. Innovation by entrants and incumbents. Tech. rep., NBER.
- Acemoglu, D., Gancia, G., Zilibotti, F., 2010. Competing Engines of Growth: Innovation and Standardization. *Journal of Economic Theory*. forthcoming.
- Acemoglu, D., Zilibotti, F., 2001. Productivity Differences. *Quarterly Journal of Economics* 116 (2), 563–606.
- Aghion, P., Bloom, N., Blundell, R., Griffith, R., Howitt, P., 2005. Competition and Innovation: An Inverted-U Relationship. *Quarterly Journal of Economics* 120 (2), 701–728.
- Alvarez, F. E., Buera, F. J., Robert E. Lucas, J., June 2008. Models of idea flows, NBER Working Paper No. 14135.
- Atkeson, A., Burstein, A., 2010. Innovation, firm dynamics, and international trade. *Journal of political economy* 118 (3), 433–484.
- Baily, M., Hulten, C., Campbell, D., Bresnahan, T., Caves, R., 1992. Productivity dynamics in manufacturing plants. *Brookings Papers on Economic Activity. Microeconomics* 1992, 187–267.
- Barro, R., Sala-i Martin, X., 1997. Technological diffusion, convergence, and growth. *Journal of Economic Growth* 2 (1), 1–26.
- Benaim, M., Weibull, J., 2003. Deterministic approximation of stochastic evolution in games. *Econometrica*, 873–903.
- Bloom, N., 2009. The impact of uncertainty shocks. *Econometrica* 77 (3), 623–685.
- Bramson, M., 1983. Convergence of solutions of the Kolmogorov equation to traveling waves. American Mathematical Society.
- 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., 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 Haan, L., Ferreira, A., 2006. Extreme value theory: an introduction. Springer.
- 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.
- Doms, M., Dunne, T., Troske, K., 1997. Workers, wages, and technology. *The Quarterly Journal of Economics* 112 (1), 253.
- 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., 1999. International technology diffusion: Theory and measurement. *International Economic Review* 40 (3), 537–570.
- 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.
- Fisk, P., 1961. The graduation of income distributions. *Econometrica*, 171–185.
- 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* 102 (52), 18801.
- Gabaix, X., 1999. Zipf’s Law For Cities: An Explanation. *Quarterly Journal of Economics* 114 (3), 739–767.
- Gast, N., Gaujal, B., 2010. Mean field limit of non-smooth systems and differential inclusions. *ACM SIGMETRICS Performance Evaluation Review* 38 (2), 30–32.
- Geroski, P. A., 2000. Models of technology diffusion. *Research Policy* 29 (4-5), 603–625.
- Ghiglino, C., 2011. Random walk to innovation: why productivity follows a power law. *Journal of Economic Theory*.
- Gibrat, R., 1931. *Les inégalités économiques*. Librairie du Recueil Sirey, Paris.
- Griffith, R., Redding, S., Van Reenen, J., 2003. R&D and absorptive capacity: Theory and empirical evidence. *Scandinavian Journal of Economics* 105 (1), 99–118.
- 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.
- Im, K., Pesaran, M., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. *Journal of econometrics* 115 (1), 53–74.
- 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.
- Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics* 54 (1-3), 159–178.
- Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. *Review of economic studies* 70 (2), 317–341.
- Lucas, R., Moll, B., 2011. Knowledge growth and the allocation of time, university of Chicago Working Paper.
- 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).
- Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71 (6), 1695–1725.
- 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.
- Pareto, V., 1896. *Cours d’Economie Politique*. F. Rouge, Lausanne 250.
- Petrin, A., Poi, B., Levinsohn, J., 2004. Production function estimation in stata using inputs to control for unobservables. *Stata Journal* 4 (2), 113–123.
- Pickands, J., 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, 119–131.
- Powell, W., Grodal, S., April 2006. *Oxford Handbook of Innovation*. Oxford University Press, USA, Ch. Networks of Innovators.
- Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. *Numerical recipes in C*. Cambridge University Press.
- Quah, D., 1993. Galton’s fallacy and tests of the convergence hypothesis. *The Scandinavian Journal of Economics*, 427–443.
- Quah, D., 1996. Twin peaks: growth and convergence in models of distribution dynamics. *The Economic Journal*, 1045–1055.
- Quah, D. T., 1997. Empirics for growth and distribution: Stratification, polarization, and convergence clubs. *Journal of Economic Growth* 2 (1), 27–59.

- Rosenberg, N., 1976. Perspectives on technology. Cambridge University Press.
- Roth, G., Sandholm, W. H., 2010. Stochastic approximations with constant step size and differential inclusions, university of Neuchatel and University of Wisconsin.
- Saichev, A., Malevergne, Y., Sornette, D., 2009. Theory of Zipf's law and beyond. Springer Verlag.
- Sandholm, W., 2003. Evolution and equilibrium under inexact information. *Games and Economic Behavior* 44 (2), 343–378.
- Sandholm, W., 2010. Population games and evolutionary dynamics. MIT Press.
- Simon, H. A., 1955. On a class of skew distribution functions. *Biometrika* 42 (3/4), 425–440.
- 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. Growth with Exogenous R&D Strategies

Denote by $P_a^{(1)}(t)$ the fraction of innovators (with a total of N_1 innovators) with log-productivity a at time t and similarly denote by $P_a^{(2)}(t)$ the fraction of imitators (with a total of N_2 imitators) with log-productivity a at time t . The total fraction of firms with log-productivity a at time t can then be written as

$$P_a(t) = \frac{N_1 P_a^{(1)}(t) + N_2 P_a^{(2)}(t)}{N_1 + N_2} = n_1 P_a^{(1)}(t) + n_2 P_a^{(2)}(t),$$

where we have introduced the population shares of innovators $n_1 = N_1/N$ and imitators $n_2 = N_2/N$ with $N = N_1 + N_2$. The evolution of the log-productivity distribution $P^{(1)}(t)$ of innovating firms is independent of the imitating firms and, by virtue of Proposition 1, it is given by (see also Section 4.2.1)

$$\frac{\partial P^{(1)}(t)}{\partial t} = P^{(1)}(t)(\mathbf{T}^{\text{in}} - \mathbf{I}).$$

In contrast, the evolution of the log-productivity distribution $P_a^{(2)}(t)$ of imitating firms is given by

$$\frac{\partial P_a^{(2)}(t)}{\partial t} = P_a(t) \sum_{b=1}^a P_b^{(2)}(t) - P_a^{(2)}(t) \left(1 - \sum_{b=1}^{a-1} P_b(t) \right). \quad (27)$$

The first term in the above equation takes into account the fraction of imitating firms with log-productivities smaller or equal to a that imitate a firm with log-productivity a . The second term considers the imitating firms with log-productivity a that imitate a firm with log-productivity larger than a . This is equivalent to the residual firms that fail to imitate a firm with log-productivity larger than a .

For simplicity, assume that one step of innovation is achieved with probability p , thus, $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$. Summation over a and rearranging terms, one can then derive from Equation (27) the dynamics of the cumulative log-productivity distribution $F_a(t)$, which is given by

$$\frac{\partial F_a(t)}{\partial t} = F_a(t)^2 - F_a(t) - n_1 F_a^{(1)}(t) F_a(t) + n_1 F_a^{(1)}(t) - n_1 p P_a^{(1)}(t). \quad (28)$$

Given the solution for $P_a^{(1)}(t)$ (and $F_a^{(1)}(t)$, respectively) and a fixed value of a , Equation (28) is a Riccati first-order, linear differential equation with non-constant, nonlinear coefficients, for which no closed form solution exists.³⁸

B. The Limit of Strong Productivity Shocks

In the following we derive some intuition for what happens in the case of β small but positive, that is, when the productivity shocks are strong but do not completely dominate the R&D decision of the firms. To simplify our analysis assume that firms face no absorptive capacity limits ($q = 1$). We then can give the following proposition.

³⁸For a fixed log-productivity a , denote by $y(t) = F_a(t)$. Then one can write from Equation (28) the following differential equation $\frac{dy(t)}{dt} + ay(t)^2 + b(t)y(t) = c(t)$, where $a = -1$, $b(t) = 1 + n_1 F_a^{(1)}(t)$ and $c(t) = n_1 (F_a^{(1)}(t) - p P_a^{(1)}(t))$.

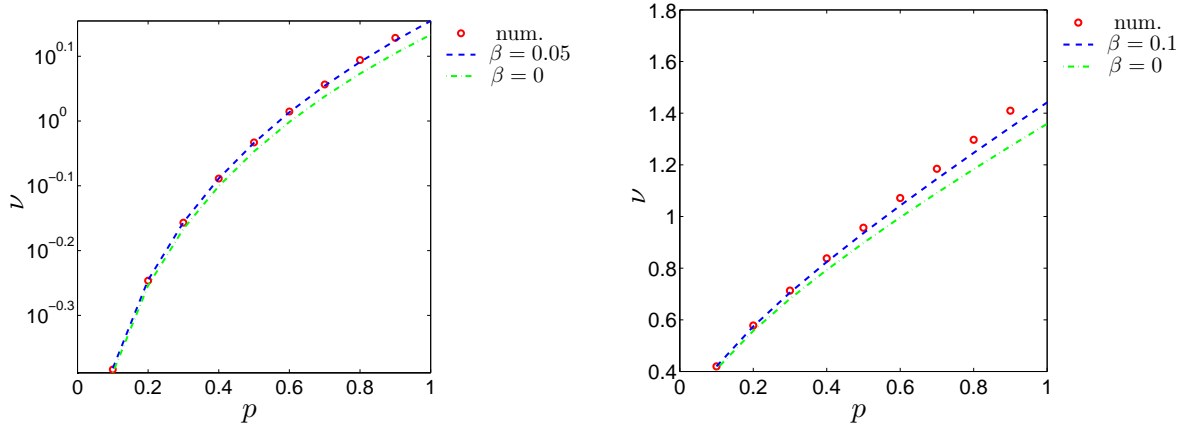


Figure 9: Traveling wave velocity $\nu^\beta(\lambda)$ as a function of $p \in (0, 1)$ assuming that $\log \bar{A} = 1$. Results of numerical integration of Equation (18) are shown with circles. (Left) The dashed line corresponds to $\beta = 0.05$ and (right) $\beta = 0.1$, while the dashed-dotted line corresponds to a value of $\beta = 0$. We see that the velocity for $\beta > 0$ is always higher than for $\beta = 0$.

Proposition 12. Let $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$ with $p \in [0, 1]$. Assume that firms do not have any absorptive capacity limits ($q = 1$). Further, assume that for β small enough the solution of Equation (19) admits a traveling wave of the form

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

and that for β small enough ρ is given by Equation (22). Then the traveling wave velocity is given by

$$\nu^\beta(p) = \frac{1}{\lambda} \left(\frac{1 + \gamma(\beta, p)}{2 + \gamma(\beta, p)} - \frac{(e^\lambda - 1)(e^\rho - 1)}{e^{\lambda+\rho} - 1} \left(\sum_{b=1}^{\infty} \frac{e^{-\lambda b}}{2 + \gamma(\beta, p)(1 + A(\lambda, \rho)e^{-\lambda b})} + \sum_{b=0}^{\infty} \frac{e^{-\rho b}}{2 + \gamma(\beta, p)(B(\lambda, \rho)e^{-\rho b} + C(\lambda, \rho)e^b)} \right) + \frac{1}{2 + \gamma(\beta, p)} (1 + p(e^\lambda - 1)) \right)$$

and λ is given by the root of $\frac{d}{d\lambda} \nu^\beta(p) = 0$, where we have denoted by

$$\begin{aligned} A(\lambda, \rho) &= \frac{(e - 1)(e^\rho - 1)e^\lambda}{(e^\lambda - e)(e^{\lambda+\rho} - 1)}, & B(\lambda, \rho) &= \frac{(e^{\lambda-1})(e - 1)e^\rho}{(e^{\lambda+\rho} - 1)(e^{1+\rho} - 1)}, \\ C(\lambda, \rho) &= \frac{e(e^\lambda - 1)(e^\rho - 1)}{(e^\lambda - e)(e^{1+\rho} - 1)}, & \gamma(\beta, p) &= \frac{\beta}{\log(\bar{A})(1 + p(\bar{A} - 1))}. \end{aligned}$$

The precise expression for $\frac{d}{d\lambda} \nu^\beta(p) = 0$ can be found in the proof of Proposition 12 in Appendix C. A comparison of the traveling wave velocity $\nu^\beta(p)$ for $\beta = 0$, $\beta = 0.05$ and $\beta = 0.1$ is given in Figure 9. We find that with increasing values of β the velocity and hence the average productivity growth rate increase, and that this effect is stronger, the larger is the innovation success probability p .

C. Proofs of Propositions, Corollaries and Lemmas

PROOF OF PROPOSITION 1. 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 $P \in P^N$

$$\mathbb{P}(\zeta_P^N = z) = \mathbb{P}(P^N(t + \Delta t) = P + z \mid P^N(t) = P). \quad (29)$$

Moreover, following the notation in Sandholm [2010, Chap.10.2] we introduce the functions V^N , A^N and A_δ^N by

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

When $\beta < \infty$, we then can state the following theorem [Benaim and Weibull, 2003; Kurtz, 1970; Sandholm, 2003]:

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

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

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

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

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

$$\frac{dP}{dt} = V(P) \quad (30)$$

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

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

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 as long as $\beta < \infty$, $V(P)$ 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 $\delta^N = \sqrt{2}/N$

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

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 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 (30), which can be written in the compact form $\frac{P(t)}{dt} = P(t)(\mathbf{T}(P) - \mathbf{I})$, for any finite $\beta < \infty$. \square

PROOF OF PROPOSITION 2. We consider the case of $\beta = \infty$. In this case the imitation probability $p_{\beta}^{\text{im}}(a, P)$ of Equation (16) has a point of discontinuity, and so does $V(P) = \mathbf{T}(P) - \mathbf{I}$. Let $\|P\|$ denote the L^2 norm in $\mathbb{R}_+^{|\mathcal{S}|}$. Define

$$\bar{V}(P) = \bigcap_{\epsilon > 0} \text{cl} \left(\text{conv} \left(V \left(\{P' \in \mathbb{R}_+^{\mathcal{S}} : \|P - P'\| \leq \epsilon\} \right) \right) \right) \quad (31)$$

as the closed convex hull of all values of V that obtain vectors P' arbitrarily close to P . We then can state the following theorem [Gast and Gaujal, 2010]:^{39,40}

Theorem 2. *Let $\bar{V}(P)$ be upper semi-continuous and assume that there exists an $c > 0$ such that $\|\bar{V}(P)\| \leq c$. Then for all $T > 0$*

$$\inf_{P \in D_T(P(0))} \sup_{0 \leq t \leq T} \|P^N(t) - P(t)\| \xrightarrow{P} 0,$$

where $P(t)$ is a solution of the differential inclusion

$$\frac{dP}{dt} \in \bar{V}(P) \quad (32)$$

with initial conditions $P(0)$ for any $t \in [0, T]$, and $D_T(P(0))$ denotes the set of all solutions of Equation (32) starting from $P(0)$.

For any P where $V(P)$ is continuous, also $\bar{V}(P) = \{V(P)\}$, while if $V(P)$ is discontinuous, $\bar{V}(P)$ is the set-valued function defined in Equation (31). Since $V(P)$ is bounded, $\bar{V}(P)$ is bounded and upper semi-continuous. Hence, the requirements of Theorem 2 are satisfied and Equation (32) describes the dynamics of the log-productivity distribution in the limit of N being large for any $t \in [0, T]$ when $\beta = \infty$. \square

PROOF OF PROPOSITION 3. Observe that in the case of pure innovation the log-productivity $a_i(t) = \log A_i(t)$ of firm i grows according to Equation (5), from which we get $a_i(t) = a_i(0) + \sum_{j=1}^t \eta(t_j)$, where $t_j \geq 0$ denotes the time at which the j -th innovation arrives. Assuming that

³⁹See also Roth and Sandholm [2010].

⁴⁰The set $\bar{V}(P)$ is upper semi-continuous if for any $P \in \mathbb{R}^{|\mathcal{S}|}$ and any open set O containing $\bar{V}(P)$, there exists a neighborhood N of P such that $\bar{V}(N) \subset O$.

the random variables $\eta(t)$ 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, $\sum_{j=1}^t \eta(t_j)$ converges to a normal distribution $\mathcal{N}(\mu_\eta t, \sigma_\eta^2 t)$. Consequently, $A_i(t)$ converges to a log-normal distribution with mean $\mu_A = e^{t\mu_\eta + \frac{1}{2}t\sigma_\eta^2}$ and variance $\sigma_A^2 = \left(e^{t\sigma_\eta^2} - 1\right) e^{2t\mu_\eta + t\sigma_\eta^2}$. \square

PROOF OF PROPOSITION 4. Inserting Equation (11) into the differential Equation (12), and summation over a yields the evolution of the cumulative log-productivity distribution $F(t)$ in the general case of $q \in [0, 1]$ as 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

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

and the proposition follows. \square

PROOF OF PROPOSITION 5. When $\beta > 0$ we see from the definition of the imitation probability in Equation (16) that $p_\beta^{\text{im}}(a, P) > p_\beta^{\text{in}}(a, P)$ is equivalent to $a_{\text{im}}(a, P) > a_{\text{in}}(a)$. Equality holds at the threshold log-productivity $a^* \equiv a^{\text{im}}(a^*, P) = a^{\text{in}}(a^*)$. This can be written as

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

Rearranging terms yields

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

or equivalently

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

That is

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

The condition $p_\beta^{\text{im}}(a, P) > p_\beta^{\text{in}}(a)$ for all $a > a^*$ can then be written as follows

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

The validity of this inequality, as well as 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.

Consider a small time interval $\Delta t > 0$. 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

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

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 $\Delta t \rightarrow 0$ we then obtain the corresponding result in continuous time. \square

PROOF OF PROPOSITION 6. The transition matrix $\mathbf{T}(P)$ is the sum of the transition matrix for imitation $\mathbf{T}^{\text{im}}(P)$ and innovation \mathbf{T}^{in} , each weighted with the probability of imitation $p_{\beta}^{\text{im}}(a, P)$ and innovation $p_{\beta}^{\text{in}}(a, P) = 1 - p_{\beta}^{\text{im}}(a, P)$, that is

$$\mathbf{T}(P) = (\mathbf{I} - \mathbf{D}(P)) \mathbf{T}^{\text{in}} + \mathbf{D}(P) \mathbf{T}^{\text{im}}(P), \quad (33)$$

where $\mathbf{D}(P)$ is the diagonal-matrix of all imitation probabilities $p_{\beta}^{\text{im}}(a, P)$. For any finite value of β , the imitation probability $p_{\beta}^{\text{im}}(a, P)$ is continuous in P , and so is $\mathbf{T}(P)$. Hence, we can apply Proposition 1, and derive the differential equation for the evolution of the log-productivity distribution stated in Equation (17). \square

In the following we derive a lemma and a corollary which will help us to show that Equation (19) admits a traveling wave solution with a stable shape.⁴¹

First, from Equation (19) we can derive the following lemma:

Lemma 2. *Let $F_a^{(1)}(t)$ and $F_a^{(2)}(t)$ be solutions of Equation (19) 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 2. 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 (19) as follows

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

⁴¹Our results follow 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

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

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 $V_a(t+s) \leq 0$. \square

A direct consequence of Lemma 2 is the following corollary.

Corollary 1. *Let $F_a(t)$ be a solution of Equation (19) with Heaviside initial data, that is*

$$F_a(0) = \Theta(a - a_m) = \begin{cases} 0, & \text{if } a < a_m, \\ 1, & \text{if } a \geq a_m. \end{cases} \quad (34)$$

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 to some function $f_\epsilon(a)$ as $t \rightarrow \infty$.

PROOF OF COROLLARY 1. For $t_0, b \in \mathbb{R}_+$ we set for any $a \geq 0$

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

If we start from Heaviside initial data we have that $F_a^{(1)}(0) \geq F_a^{(2)}(0)$ and Proposition 2 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 $t \rightarrow \infty$. An illustration can be seen in Figure 10. \square

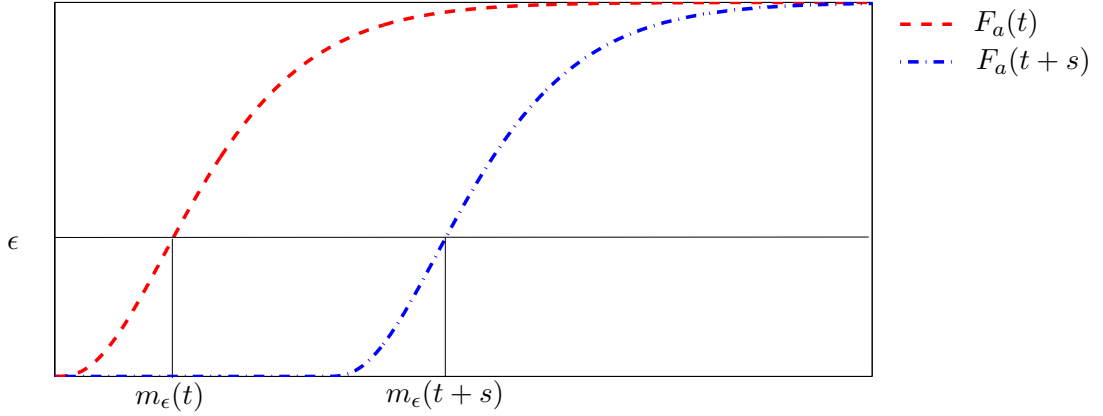


Figure 10: Illustration of distributions and at times t and $t + s$ for $s > 0$.

We are now in place to give a proof of Proposition 7.

PROOF OF PROPOSITION 7. By Corollary 1 we can fix a value of $\epsilon = \frac{1}{2}$, where $m_{1/2}(t)$ is the median of $F_a(t)$, and have that

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

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

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

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 (19), the above equation can be written as follows

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

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

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

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

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

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 (19) 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 . \square

PROOF OF PROPOSITION 8. We first give a proof of part (i) of the proposition. In order to solve for the traveling wave solution of Equation (19) we observe that in terms of the complementary cumulative log-productivity it can be written as

$$\frac{\partial G_a(t)}{\partial t} = \frac{2q-1}{2}(-G_a(t)^2 + G_a(t)) - \frac{p}{2}(G_a(t) - G_{a-1}(t)). \quad (35)$$

For appropriate initial conditions Proposition 7 implies that the dynamics of the complementary cumulative log-productivity distribution $G_a(t)$ in Equation (35) admits a traveling wave solution $G_a(t) = g(x)$, $x = a - \nu t$ with velocity ν satisfying

$$\nu \frac{dg(x)}{dx} = \frac{2q-1}{2}(g(x)^2 - g(x)) + \frac{p}{2}(g(x) - g(x-1)). \quad (36)$$

Numerical studies confirm that this is also true for general initial conditions which are concentrated enough.

We then 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 (35). Then we obtain from Equation (35) 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

$$\nu = \frac{2q-1-p+pe^\lambda}{2\lambda}. \quad (37)$$

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

$$2q-1-p+pe^\lambda = p\lambda e^\lambda. \quad (38)$$

The minimum of Equation (37) is obtained at λ solving Equation (38). This yields

$$\lambda = 1 + W\left(\frac{2q-1-p}{pe}\right), \quad (39)$$

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 left tail of the traveling wave. For a much smaller than νt we can neglect the term $F(a, t)^2$ in Equation (19) 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

$$2\rho\nu = 2q-1+p-pe^{-\rho}. \quad (40)$$

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

PROOF OF COROLLARY 1. From Proposition 5 we know that there exists a unique threshold log-productivity a^* such that $a_{\text{im}}(a, P) > a_{\text{im}}(a)$ for all $a < a^*$ and $a_{\text{im}}(a, P) < a_{\text{im}}(a)$ or all $a > a^*$. From the properties of the logistic function in the definition of the imitation probability $p_{\beta}^{\text{im}}(a, P)$ in Equation (16) it follows that $p_{\beta}^{\text{im}}(a, P) \rightarrow 1$ as $\beta \rightarrow \infty$ for all $a < a^*$, while $p_{\beta}^{\text{im}}(a, P) \rightarrow 0$ as $\beta \rightarrow \infty$ for all $a > a^*$. \square

PROOF OF PROPOSITION 9. Under the assumption that Equation (23) holds for β large enough, we can insert Equation (23) into Equation (18) 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 $F_a(t) = \sum_{b=1}^a P_b(t)$ we then get for $a < a^*$

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

where in the last line from above we have used the results obtained in Proposition 4. 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

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

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 $\frac{dF_a(t)}{dt} \leq 0$. \square

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

$$G_a(t) = \sum_{b=a+1}^{\infty} Ne^{-\lambda(b-\nu t)} = \frac{N}{e^\lambda - 1} e^{-\lambda(a-\nu t)}. \quad (41)$$

In terms of the complementary cumulative distribution function $G_a(t) = 1 - F(a, t)$ we then can write Equation (24) for a much larger than the threshold a^* as

$$\frac{\partial G_a(t)}{\partial t} = G_a(t) (1 - G_{a^*}(t)) - p(G_a(t) - G_{a-1}(t))$$

Inserting Equation (41) yields

$$\lambda \nu e^{-\lambda(a-\nu t)} = e^{-\lambda(a-\nu t)} \left(1 - \frac{N}{e^\lambda - 1} \right) - p \left(e^{-\lambda(a-\nu t)} - e^{-\lambda(a-1-\nu t)} \right) \quad (42)$$

which gives

$$\lambda \nu = 1 - \frac{N}{e^\lambda - 1} - p(1 - e^\lambda). \quad (43)$$

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

$$a^* + \log \left(F_{a^*}(t) + \sum_{b=a^*+1}^{\infty} e^{b-a^*} P_b(t) \right) = a^* + \log(1 + p(\bar{A} - 1)). \quad (44)$$

This means that the expected log-productivity obtained through innovation equals the expected log-productivity obtained through imitation. This Equation 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

$$N = p(\bar{A} - 1) \left(\frac{1}{e^{\lambda-1} - 1} + \frac{1}{1 - e^\lambda} \right)^{-1}, \quad (45)$$

Inserting N into Equation (43) 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

$$\nu = \frac{1}{\lambda} \left(1 + p(e^\lambda - 1) - \frac{p(\bar{A} - 1)(1 - e^{1-\lambda})}{e - 1} \right). \quad (46)$$

According to the selection principle we have encountered already in the proof of Proposition 8,

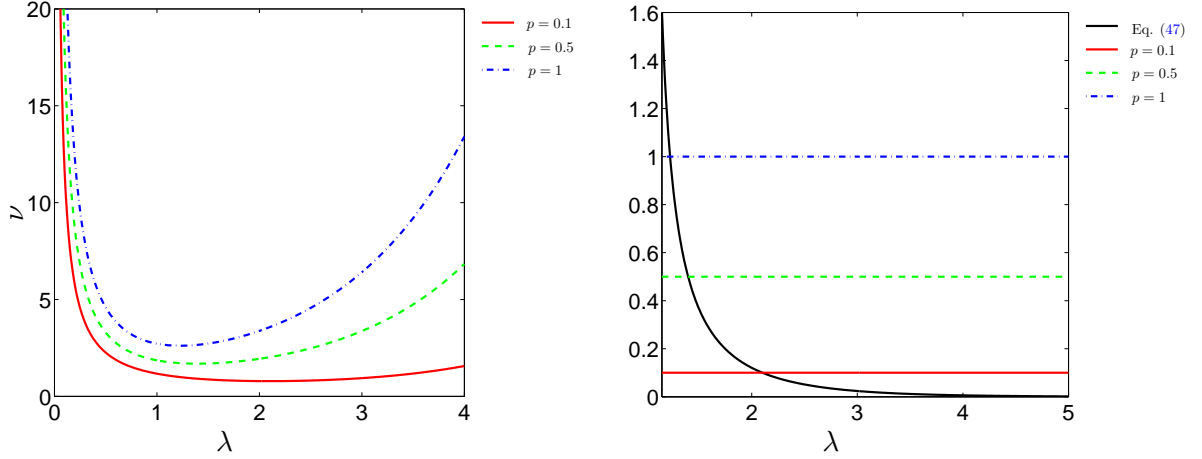


Figure 11: (Left) The traveling wave velocity ν as a function of λ for different values of $p = 0.1$, $p = 0.5$ and $p = 1$. (Right) The ν minimizing value of λ for the same values of p . The figures show that the minimizing value of λ is decreasing for increasing values of p , and consequently, the front of the traveling wave becomes steeper. Moreover, the velocity ν of the traveling wave increases with increasing values of p .

for sufficiently steep initial conditions of $F_a(0)$ the value of λ is realized that minimizes Equation (46). The traveling wave velocity ν as a function of λ for different values of p can be seen in Figure 11 (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 (26) follows. The FOC from above is equivalent to

$$\frac{e - 1}{\bar{A} + e - 2 + (e - 1)e^\lambda(\lambda - 1) - (\bar{A} - 1)e^{1-\lambda}(1 + \lambda)} = p, \quad (47)$$

which is illustrated in Figure 11 (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 (24). We then assume that the left tail of the log-productivity distribution can be described by an exponential function $P_a(t) \propto e^{\rho(a-\nu t)}$. Inserting this into Equation (24) for a smaller than a^* gives

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

and hence we obtain $\rho = 1/\nu$. □

PROOF OF PROPOSITION 11. The traveling wave velocity in the limit of $\beta \rightarrow 0$ follows from Equations (20) and (21) as

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

while the traveling wave velocity for $\beta \rightarrow \infty$ is given by Equation (25). We then have that

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

Since the above equation holds for all λ , it holds in particular for the value of λ minimizing $\lim_{\beta \rightarrow \infty} \nu^\beta(\lambda)$, and hence, we have that $\lim_{\beta \rightarrow \infty} \nu^\beta(\lambda) > \lim_{\beta \rightarrow 0} \nu^\beta(\lambda)$. A higher traveling wave velocity ν implies first-order stochastic dominance of the respective cumulative distribution functions and therefore a higher average productivity. \square

PROOF OF PROPOSITION 12. Motivated by our analysis for the case of $\beta = 0$, we make the following assumption on the log-productivity distribution

$$P_a = N \begin{cases} e^{\rho(a-\nu t)}, & \text{if } a \leq \nu t, \\ e^{-\lambda(a-\nu t)}, & \text{if } a > \nu t, \end{cases} \quad (48)$$

with the normalization constant N given by

$$\frac{1}{N} = \frac{e^\rho}{e^\rho - 1} + \frac{1}{e^\lambda - 1}. \quad (49)$$

The average log-productivity is then given by

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

for large t . In the case in which $\mathbb{P}(\eta(t) = 1) = p$ and $\mathbb{P}(\eta(t) = 0) = 1 - p$ for $p \in (0, 1)$ we obtain:

$$a^{\text{in}}(a(t)) = a(t) + \log(1 - p + \bar{A}p). \quad (50)$$

Plugging in the expressions for $a^{\text{in}}(a_i(t))$ and $a^{\text{im}}(a(t), P(t))$ given by Equations (50) and (15), respectively, and rearranging terms, yields

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

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 for $a > \nu t$ as follows

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

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 $\gamma = \beta / (\log \bar{A}(1 + p(\bar{A} - 1)))$ 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 $A = \frac{(e-1)(e^{\rho}-1)e^{\lambda}}{(e^{\lambda}-e)(e^{\lambda+\rho}-1)}$, $B = \frac{(-1+e^{\lambda})(-1+e)^{\rho}}{(-1+e^{\lambda+\rho})(-1+e^{1+\rho})}$, $C = \frac{e(-1+e^{\lambda})(-1+e^{\rho})}{(-e+e^{\lambda})(-1+e^{1+\rho})}$. Then we can write for β small enough

$$p_{\beta}^{\text{in}} = \begin{cases} \frac{1}{2+\gamma(1+Ae^{-\lambda(a-\nu t)})}, & \text{if } a \leq \nu t, \\ \frac{1}{2+\gamma(Be^{\rho(a-\nu t)}+Ce^{-(a-\nu t)})}, & \text{if } a > \nu t. \end{cases} \quad (51)$$

With the distribution given in Equation (48) and the innovation probability from Equation (51), we obtain for a larger than νt from Equation (18)

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

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

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

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

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

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

D. Model Extensions

In this appendix we sketch three possible extensions of our model. First, in Appendix D.1 we allow for productivity shocks that can also lead to a decline in the productivity of a firm [cf. Melitz, 2003]. Next, in Appendix D.2 we provide a basic mechanism for firm entry and exit. Finally, Appendix D.3 introduces an absorptive capacity limit with an upper cutoff which bounds the relative productivity a firm can imitate from above.

D.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 $\mathbb{P}(\delta = 1) = \delta_1$, $\mathbb{P}(\delta = 2) = \delta_2, \dots$, we

can introduce the matrix

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

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

$$\frac{\partial P(t)}{\partial t} = P(t) \left((1-r) ((\mathbf{I} - \mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t))) + r\mathbf{T}^{\text{dec}} - \mathbf{I} \right). \quad (52)$$

D.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.⁴² 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,⁴³ keeping a balanced in- and outflow of firms. This means that a monopolist in sector i at time t is replaced with a new firm in that sector that starts with productivity $A_0(t)$.

We assume that in each period, first, a randomly selected firm either decides 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 (10) accordingly. In the case of $A_0 = 1$ we can write in the limit of large N

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

where $Q = (1 \ 0 \ 0 \ \dots)$.

D.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 .

$$A_i(t + \Delta t) = \begin{cases} A_j(t) & \text{if } A_j/A_i \in]1, 1 + D], \\ A_i(t) & \text{otherwise.} \end{cases} \quad (53)$$

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 (53) governing the imitation process reads as

$$a_i(t + \Delta t) = \begin{cases} a_j(t) & \text{if } a_j - a_i \in]0, d], \\ a_i(t) & \text{otherwise.} \end{cases} \quad (54)$$

⁴²In 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.

⁴³Similarly, Melitz [2003] assumes that firms can be hit with a bad productivity shock at random and then are forced to leave the market.

year	N	$\min(A)$	geomean(A)	mean(A)	max(A)	std(A)
1995	48,667	0.098	59.20	78.48	65,527	414.88
1996	48,720	0.132	59.90	77.44	54,698	296.26
1997	48,762	0.060	61.47	77.78	18,526	143.72
1998	48,761	0.343	64.04	80.89	18,983	160.77
1999	48,739	0.228	66.72	85.47	19,706	166.80
2000	48,755	0.321	69.13	90.95	18,880	180.00
2001	48,684	0.257	69.48	87.66	26,382	193.43
2002	48,529	0.174	69.76	86.98	15,553	142.55
2003	48,353	0.079	70.41	89.40	22,089	167.85

Table 2: Descriptive statistics for total factor productivity as estimated by Equations (2) and (3). The number of firms in the balanced panel of western European countries is 49,022. The different numbers of N come due to the fact that productivity can not be computed when one of the input variables is nonpositive. This happens occasionally for the variables output (as added value, 1,913 times nonpositive), capital (as fixed assets, 285 times zero), and material cost (1,030 times nonpositive) for reasons which are possibly not directly related to productivity.

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}^{im} . Following Equation (53) 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}^{im} depends only on the current distribution of log-productivity $P(t)$ and simplifies to

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

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

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

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

E. Additional Empirical Results

Descriptive statistics for the estimated productivity levels for our sample are shown in Table 2. From the table we see that both the geometric as well as the arithmetic mean of the productivity increase with time. These are also shown in Figure 12. We then estimate the traveling wave velocity (growth rate) ν by estimating the parameters of an exponential growth function on a measures of a central tendency per year $\bar{A}(t) = \exp(\nu t + \text{const})$. Exponential growth of productivity corresponds to linear growth of log-productivity $\log \bar{A}(t) = \nu t + \text{const}$. The results of five fits are shown in Table 3. The fits on all data points show that a non-zero growth is highly significant (due to narrow 95% confidence intervals). The fits on arithmetic and geometric mean show an appropriate goodness of the fits. Such a goodness can naturally not be expected by fits on all data points.

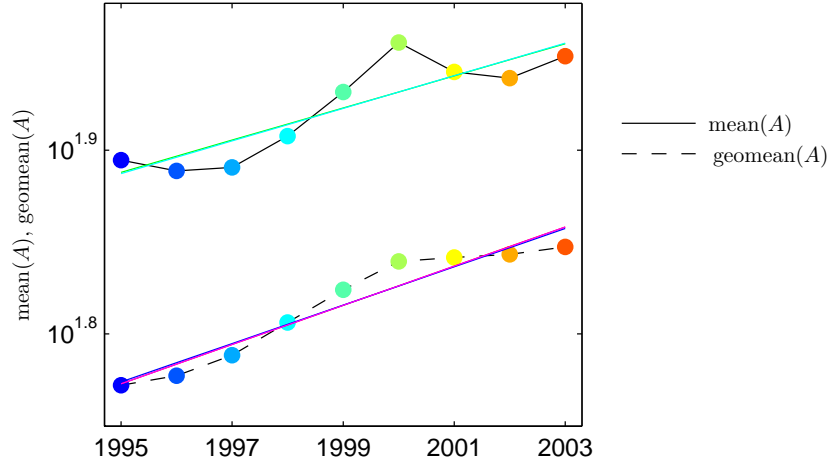


Figure 12: The arithmetic mean of productivity (points connected with a solid line) and geometric mean of productivity (points connected with a dashed line) per year. The ordinate has a logarithmic scale. Thus, the fitted lines correspond to exponential functions. The upper line corresponds to regression R5 and the lower line to regression R3 in Table 3.

	regression Eq.	$\hat{\nu}$ (95%-conf. int.)	N	R^2
R1	$\log A_{it} = \nu t + c$	0.0245 (0.0238,0.0253)	437,970	0.0097
R2	$\text{geomean}_i(A_{it}) = \exp(\nu t + c)$	0.0240 (0.0180,0.0300)	9	0.9285
R3	$\log \text{geomean}_i(A_{it}) = \nu t + c$	0.0245 (0.0187,0.0303)	9	0.9341
R4	$\text{mean}_i(A_{it}) = \exp(\nu t + c)$	0.0201 (0.0103,0.0300)	9	0.7711
R5	$\log(\text{mean}_i(A_{it})) = \nu t + c$	0.0204 (0.0108,0.0301)	9	0.7821
	average	0.0227		

Table 3: Five fits for the exponential growth of productivity. The parameter ν is the speed of the traveling wave of log-productivity. Fits R1, R2, and R3 fit a function which approximates the geometric mean of productivity, R4 and R5 approximate the arithmetic mean of productivity. Fitted parameters, their 95%-confidence intervals and their coefficient of determination R^2 are computed with the `matlab` function `fit` and the default configuration of the library models `poly1` (=linear) for R1, R3 and R5 and `exp1` for R2 and R4.