Theoretical Economics 11 (2016), 1053–1102 1555-7561/20161053

Innovation vs. imitation and the evolution of productivity distributions

Michael D. König
Department of Economics, University of Zurich

Jan Lorenz
Department of Psychology and Methods, Jacobs University Bremen

Fabrizio Zilibotti
Department of Economics, University of Zurich

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 research and developmenmt (R&D) or, alternatively, by trying to imitate other firms' technologies, subject to the limits of their absorptive capacities. The outcome of both strategies is stochastic. The choice between in-house R&D and imitation is endogenous, and is based on firms' profit maximization motive. Firms closer to the technological frontier face fewer imitation opportunities, and choose in-house R&D, while firms farther from the frontier try to imitate more productive technologies. The resulting balanced-growth equilibrium features 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 a Pareto distribution as can be observed in the empirical data.

Keywords. Imitation, innovation, growth, quality ladder, absorptive capacity, productivity differences, spillovers.

JEL classification. E10, O40.

Michael D. König: michael.koenig@econ.uzh.ch
Jan Lorenz: post@janlo.de
Fabrizio Zilibotti: fabrizio.zilibotti@econ.uzh.ch

We would like to thank a co-editor, three anonymous referees, Chad Jones, Pete Klenow, Daron Acemoglu, and seminar participants at Stanford University, 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 the Swiss National Science Foundation through research Grants PBEZP1–131169 and 100018_140266, and thanks SIEPR and the Department of Economics at Stanford University for their hospitality during 2010–2012. Fabrizio Zilibotti acknowledges financial support from the Swiss National Science Foundation through research Grant 100018_140266. This article is based on material from Michael König's thesis at the University of Zurich, under Fabrizio Zilibotti's supervision (cf. König 2010).

Copyright © 2016 Michael D. König, Jan Lorenz, and Fabrizio Zilibotti. Licensed under the Creative Commons Attribution-NonCommercial License 3.0. Available at http://econtheory.org.
DOI: 10.3982/TE1437

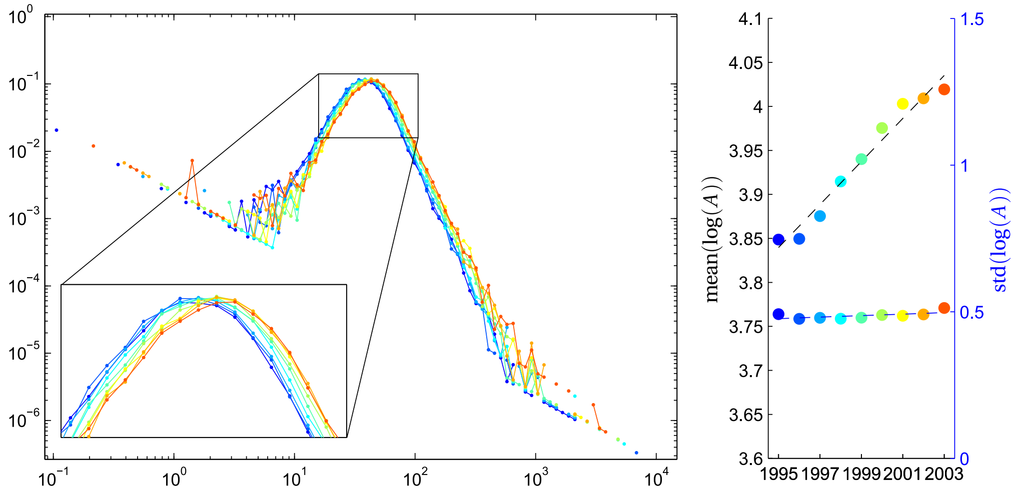

Figure 1. The left-hand panel shows the TFP distribution of French firms over the years from 1995 to 2003 (vertical axes in logarithmic scale). The right-hand panel shows the mean and standard deviation of the log TFP, with fitted regression lines.

1. Introduction

There are large and persistent productivity differences not only across countries (e.g., Feyrer 2008, Quah 1997, Durlauf 1996), but also across firms and plants within countries (Baily et al. 1992). Such differences largely reflect the use of different technologies and managerial practices (see, e.g., Bloom and Van Reenen 2011, Doms et al. 1997). Consider, for instance, the distribution of total factor productivity (TFP) from a balanced panel of 17,404 French firms in the periods between 1995 to 2003.[1] Figure 1 shows how the empirical distribution evolves over time. (Note that many of the figures are shown in color in the online version.) Three main features emerge. First, the distribution of high-productivity firms is well described by a power law.[2] Second, the distribution of low-productivity firms also is approximated by a power law, although this approximation is less accurate, arguably due to noisy data at low productivity levels. Third, the distribution is well approximated by a distribution that shifts in an affine way at a constant rate over time. We call a distribution with the latter characteristics a *traveling wave*.[3] While entry, exit, and reallocation are important determinants of firm dynamics, they altogether account for only 25% of total productivity growth (Acemoglu 2009, Chapter 18). Therefore, a theory of firm-level productivity dynamics must explain the determinants of the accumulation of technical knowledge among incumbent firms. To further the understanding of these factors, in this paper we propose a theory, related to Acemoglu et al.

[1]The data are from the Amadeus data base provided by Bureau van Dijk. The firm-level TFPs are estimated following the method introduced by Levinsohn and Petrin (2003). A detailed description of the estimation method and additional details about the data can be found in Appendix B.2 in the Technical Appendix.

[2]Pareto distributions are also observed for distributions of several other economic variables of interest (e.g., firm size) in numerous empirical studies (e.g., Gabaix 1999, Saichev et al. 2010, de Wit 2005).

[3]In Section 5, we provide a formal definition (Definition 1) of a traveling wave.

(2006), where firms can upgrade productivity over time through two alternative strategies: either by carrying out "in-house research and development" (R&D) or by imitating technologies used by other firms. The choice is driven by a standard profit-maximizing motive. The focus of the theory on the innovation-vs.-imitation margin is motivated by two observations. On the one hand, an important source of differences in technological know-how is the large variation across firms in R&D investments and in their success (Coad 2009, Cohen and Klepper 1992, 1996). On the other hand, many firms do not invest at all in R&D; their productivity increases through the adoption of technology already in use from other firms. Thus, technical knowledge diffuses over time, albeit only slowly (Griliches 1957, Eeckhout and Jovanovic 2002, Geroski 2000, Stoneman 2002, Comin and Mestieri 2014). Our theory can reproduce, both qualitatively and quantitatively, the empirical regularities outlined above.

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 heterogeneous productivities that increase over time driven by firms' endeavors to improve their technologies. For simplicity, we abstract from resource costs of R&D or imitation—the two strategies for increasing productivity. Since a firm cannot pursue both R&D and imitation at the same time, the opportunity cost of imitating is the return from R&D and vice versa. R&D activity is modeled as a draw from an exogenous distribution of productivity upgrades. Imitation is modeled as a "matching process" whereby each imitating firm is randomly matched with another firm, and can then succeed or fail in imitating the other firm's technology. The optimal choice between the two strategies hinges on the firm's position in the overall productivity distribution. Firms far from the technology frontier are more likely to be matched with higher productivity firms and optimally choose imitation. In contrast, firms close to the technology frontier are less likely to find better firms from which they can learn and, therefore, are more prone to choosing in-house R&D.[4] Our model yields a steady-state productivity distribution with trending productivity resembling the empirical distribution of Figure 1. More formally, the theoretical distribution is a traveling wave with an exponentially growing average and power-law tails. We obtain an analytical representation of the equilibrium law of motion of the distribution in terms of a system of ordinary differential equations (ODEs), and even a complete analytical characterization of the steady-state distribution (traveling wave) consistent with the equilibrium law of motion. This characterization is the main contribution of our paper.

We contrast the results with alternative environments. We show that, on the one hand, a traveling wave would not emerge in an economy where some firms always do R&D and others always imitate. In such an economy, the variance of the productivity distribution would grow over time, counterfactually. The reason is that the subpopulation of innovating firms would be excluded from any spillover from the growth at the

[4]More formally, in our model there exists a relative productivity threshold below which firms always imitate and above which they always innovate. This prediction of 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).

frontier, causing an ever-growing lower tail. On the other hand, the traveling wave would emerge in an economy in which each firm is assigned randomly to an innovation strategy in every period. Thus, what matters for our result is not that firms choose optimally between R&D and imitation, but that there is some "mixing" so that in every period firms lagging behind resort to imitation with some probability. More generally, the crux of the result is that all firms end up benefiting, sooner or later, from the spillovers accruing from the frontier productivity growth. Such spillovers ensure that a firm whose productivity is relatively low can grow more quickly as the frontier moves farther away. The case of profit-maximizing firms choosing between innovation and imitation is an economically interesting example of this mechanism: any repeatedly unsuccessful firm pursuing R&D can avoid falling too far behind by switching to imitation.

As an important extension, we study an economy in which firms have a limited capacity to absorb knowledge through imitation (Cohen and Levinthal 1989, Kogut and Zander 1992, Nelson and Phelps 1966). Namely, when a firm is matched with a more productive one, it can absorb only a (stochastic) share of the knowledge possessed by the other firm. The assumption of a limited absorptive capacity has no major bearing on the qualitative characterization of the equilibrium. However, this realistic feature turns out to improve significantly the quantitative fit of the theory, e.g., relative to the empirical distribution of Figure 1. Intuitively, in the model with an unlimited absorptive capacity, laggard firms benefit strongly from the spillovers arising from progress at the frontier. Thus, if one calibrates the model so as to fit the productivity spread observed in the data, the model (which is very parsimonious in the number of parameters) overpredicts productivity growth. In contrast, the model with a limited absorptive capacity slows down convergence within the distribution, yielding a much better fit with the empirical distribution. Another insight (hinging on numerical analysis) is that when the absorptive capacity is sufficiently small relative to the return to innovation, one obtains an ever-growing variance rather than a traveling wave.

The explicit formulation of firms' R&D behavior and the endogenous choice between innovation and imitation distinguishes our model from most of the previous literature. Klette and Kortum (2004) model the R&D decisions of multiproduct firms, but do not discuss imitation. Luttmer (2007) focuses on entry, exit, and selection in a world where incumbent firms are subject to exogenous productivity shocks, and entrant firms can imitate incumbents. His model, like ours, generates a traveling wave. There are two main differences relative to our paper. First, we focus on the endogenous decision of innovation vs. imitation by incumbent firms. Second, from a technical standpoint, Luttmer (2007) proposes an environment with continuous firm sizes, while here we analyze a Schumpeterian quality-ladder model with discrete productivity steps. Nevertheless, despite the differences, in both cases a traveling wave solutions can be obtained. Moreover, Acemoglu and Cao (2015) construct, as we do, a Schumpeterian model. They obtain Zipf's law for large firm sizes, while we focus on productivity. 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). As in Luttmer (2007), their model does not feature an endogenous choice of the R&D strategy. Ghiglino (2012) constructs a search-based growth model that generates

Pareto-distributed productivity levels focusing on the recombination of existing technologies into novel ones. In Perla and Tonetti (2014) firms can choose either to produce, or to search for existing technologies to imitate.[5] Differently from our model, their paper features no in-house R&D. Other papers focusing on innovation and imitation include Eeckhout and Jovanovic (2002) and Atkeson and Burstein (2010). None of these focuses on the innovation-vs.-imitation trade-off.

Alvarez et al. (2008), Lucas (2009), and Lucas and Moll (2014) 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 (2014) 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 neither model explicitly the strategic decisions of firms whether to undertake in-house R&D or to copy other firms, nor do they take into account limitations in the ability of firms to imitate external knowledge. 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 an unbounded distribution of knowledge. This is not necessary in our model, since here firms that are close to the technology frontier choose endogenously to innovate (i.e., draw from an exogenous productivity distribution) rather than to adopt technologies from a pool of existing ideas.

Our paper is also related to two recent contributions that were written simultaneously and independently of our paper. Benhabib et al. (2014) study a simplified deterministic framework where agents make an optimal portfolio choice between investments in innovation and adoption. Luttmer (2012) extends the model of selection and growth of Luttmer (2007) to an environment in which also incumbent firms can perform imitation. He obtains, as we do, convergence to a stable (balanced growth) productivity distribution. However, both the environment and the technique of analysis are different. In particular, in his model productivity growth is governed by a Brownian motion while we consider a standard Schumpeterian quality-ladder model. In this respect, our paper also relates to earlier Schumpeterian growth literature where firms make a choice between innovation and imitation, including Cheng and Dinopoulos (1996), Segerstrom (1991), Jovanovic and Rob (1990), and Acemoglu et al. (2006). These papers, however, do not study the endogenous evolution of the productivity distribution of firms.

The paper is organized as follows. The static model environment is introduced in Section 2. Section 3 discusses the law of motion of the productivity distribution. Section 4 studies the evolution of the distribution in an economy where the innovation strategy (in-house R&D vs. imitation) is a deterministic fixed effect of each firm. Section 5 yields the main result, characterizing the productivity distribution in a model where firms choose optimally whether to perform in-house R&D or to imitate. Sections 6 and 7 consider two extensions, and Section 8 concludes. The proofs of all propositions

[5] For a recent extension of this model, see Benhabib et al. (2016).

and lemmas, together with some additional results referred to in the text, are provided in Appendix A. Additional technical material, including extensions and details of the calibration, are provided in Technical Appendix B, available in a supplementary file on the journal website, http://econtheory.org/supp/1437/supplement.pdf.

2. The model

In the following sections we provide a microfoundation of our model based on a monopolistically competitive environment with a competitive fringe in each sector (see Section 2.1), and we introduce the basic processes of innovation and imitation (see Section 2.2) leading to productivity improvements.

2.1 *Environment*

The model economy is a version of Acemoglu et al. (2006) comprising a competitive final good sector and a continuum of unit measure of monopolistic sectors producing differentiated intermediated goods. The final good, denoted by $Y(t)$, is produced by a representative firm using labor and a set of intermediate goods $x_i(t)$, $i \in \mathcal{N} = \{1, 2, \dots, N\}$. Its technology is represented by the production function

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

where t denotes time, x_i is the intermediate good i, and A_i is the technology level of industry i. We normalize the labor force to unity, $L = 1$. The final good can be used for consumption, as an input to R&D, and also as an input to the production of intermediate goods. Its price is set to be the numeraire. The profit maximization program yields the inverse demand function for intermediate goods:

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

Each intermediate good i is produced by a technology leader who has access to the best technology. By this best-practice technology the marginal cost of producing any intermediate input equals one unit of the final good. The leader is subject to the potential competition of a fringe of firms that can produce the same input albeit at a higher constant marginal cost, χ, where $1 < \chi \leq 1/\alpha$. Note that a higher value of χ indicates less competition. Bertrand competition implies that each technology leader monopolizes its market, sets the price equal to the unit cost of the fringe, $p_i(t) = \chi$, and sells the quantity $x_i(t) = \chi^{-1/(1-\alpha)} A_i(t)$. Namely, the equilibrium entails a limit price strategy and an inactive fringe as in Acemoglu et al. (2006). The profit earned by the incumbent in any intermediate sector i is then proportional to productivity,

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

where we denote $\psi \equiv ((\chi - 1)/\alpha)\chi^{-1/(1-\alpha)}$. In equilibrium, gross output is proportional to aggregate productivity,

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

where aggregate productivity is $A(t) = \sum_{i=1}^{N} A_i(t)$. Similarly, net aggregate output, defined as final output minus the cost of intermediate production, is given by $Y^{\text{net}}(t) = Y^{\text{tot}}(t) - \sum_{i=1}^{N} x_i(t) = \zeta A(t)$, where $\zeta \equiv (\chi - \alpha)(1/\alpha)\chi^{-1/(1-\alpha)}$.[6]

Throughout the rest of the paper, when referring to firm i we always mean the most efficient producer in sector i. Moreover, our population of firms comprises only the set of technology leaders in each sector. These choices are not a source of confusion since fringe firms are inactive in equilibrium.

2.2 *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 $\bar{A}^1$, $\bar{A}^2$, $\bar{A}^3, \ldots$. Firm i, which has achieved a_i productivity improvements, then has productivity $A_i = \bar{A}^{a_i}$.

Firm i's productivity $A_i \in \{1, \bar{A}, \bar{A}^2, \ldots\}$ grows as a result of technology improvements, either undertaken in-house (innovation) or due to the imitation and absorption of other firms' technologies. The technology comes from firms in other sectors that were successful in innovating in their area of activity (Kelly 2001, Rosenberg 1976, Fai and Von Tunzelmann 2001). We consider a discrete time model where in each time period from t to $t + \Delta t$, $\Delta t > 0$, a firm i is selected at random and decides either to imitate another firm or to conduct in-house R&D, depending on which option yields higher expected profits.[7]

2.2.1 *Innovation* If firm i conducts in-house R&D at time t, then it makes $\vartheta(t)$ productivity improvements and its productivity changes as

$$A_i(t + \Delta t) = \bar{A}^{a_i(t)+\vartheta(t)} = A_i(t)\bar{A}^{\vartheta(t)}, \tag{2}$$

where $\vartheta(t) \geq 0$ is a nonnegative integer-valued random variable with a certain distribution. Let us denote $\eta_b \equiv \mathbb{P}(\vartheta(t) = b)$ for $b = 0, 1, 2, \ldots$ to quantify the distribution, satisfying $\sum_{b=0}^{\infty} \eta_b = 1$. From the productivity growth dynamics above we can go to an equivalent system by changing to the log productivity $a_i(t) = \log A_i(t)/\log \bar{A}$. We can simplify the notation if we take $\bar{A}$ as the base of the logarithm, so that $\log \bar{A} = 1$. This

[6]Given the proportionate relationship between productivity and output, all results we derive on productivity also hold for firm size (as measured, e.g., by value added). However, since we are mainly interested in the process of technological change and productivity growth, we focus on productivity instead of firm size dynamics.

[7]We explain the innovation and imitation process in more detail in Section 3 below.

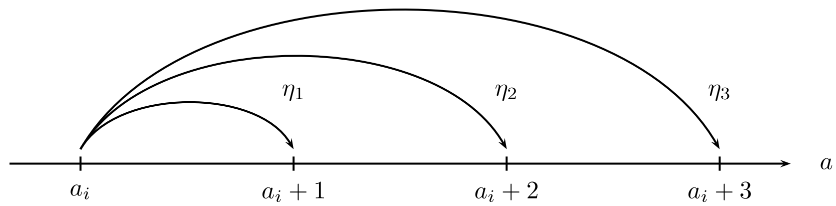

FIGURE 2. Illustration of the innovation process of firm i with log productivity $\log A_i = a_i \log \bar{A} = a_i$ (setting $\log \bar{A} = 1$). With probability η_1 firm i makes one productivity improvement and advances by one log-productivity unit, with probability η_2 firm i makes two productivity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

allows us to write log productivity as $a_i(t) = \log A_i(t)$. Then taking logs of the in-house update map in (2) gives

$$a_i(t + \Delta t) = a_i(t) + \vartheta(t). \tag{3}$$

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 $\bar{A}^{\vartheta}$. In the limit of continuous time we obtain a geometric Brownian motion for productivity (see, e.g., Saichev et al. 2010, p. 9).

In our analysis below, we restrict attention to the case in which innovation is an incremental step-by-step process , i.e., $\eta_0 = 1 - p$, $\eta_1 = p$, $\eta_b = 0$ for $b = 2, 3, \ldots$. This is for simplicity. All results can be extended to the case in which $\eta_b > 0$ for all $b \leq B < \infty$.

2.2.2 *Imitation* In the case of imitation, firm i with productivity $A_i(t)$ selects another firm j at random from the population of firms, $\mathcal{N}$, 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 that separates it from $a_j(t)$. We assume that each firm climbs each rung with a success probability $q \in [0, 1]$. 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 (Weitzman 1998, Kogut and Zander 1992).

Taking the above-mentioned process of imitation more formally, firm i's productivity changes according to

$$A_i(t + \Delta t) = A_i(t)\bar{A}^{\kappa} = \bar{A}^{a_i(t)+\kappa}, \tag{4}$$

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

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

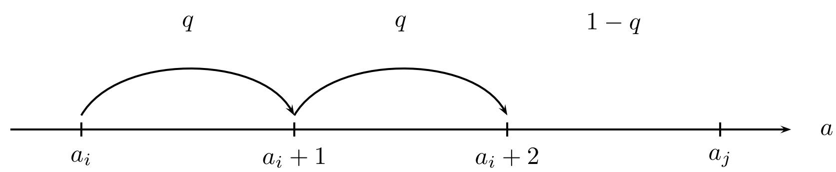

Figure 3. Illustration of the imitation of log productivity a_j of firm j through firm i with log productivity a_i, where the log productivity of firm i is $\log A_i = a_i \log \bar{A} = a_i$ (setting $\log \bar{A} = 1$). Firm i successfully imitates two log-productivity units (with probability q^2)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

Note that $\sum_{k=0}^{\infty} \mathbb{P}(\kappa = k) = 1$, as necessary for a proper probability measure. Moreover, for $q = 0$ we have that $A_i(t + \Delta t) = A_i(t)$, for $q = 1$ we have $A_i(t + \Delta t) = A_j(t)$ 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 firms' absorptive capacities. The higher is q, the better firms are able to climb rungs on the quality ladder.

Switching to log productivity and setting $\log \bar{A} = 1$ in (4) we obtain[8]

$$a_i(t + \Delta t) = a_i(t) + \kappa. \tag{5}$$

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

3. Evolution of the productivity distribution

In this section, we analyze the evolution of the productivity distribution. We first establish some useful notation. We then proceed by characterizing the equilibrium dynamics of the productivity distribution.

3.1 *Characterization of the productivity dynamics*

Consider the distribution of log productivity $a_i(t) = \log A_i(t)$ in the population of $N \in \mathbb{N}$ firms over time, where N is assumed to be a large number. Let $\mathcal{S}$ denote the set of log-productivity values, that is, $\mathcal{S} = \{\log \bar{A}, 2 \log \bar{A}, \ldots\}$. Assuming that $\log \bar{A} = 1$ this is simply the set of positive integers, $\mathbb{N}$. Further, let $P_a(t)$ indicate the fraction of firms having log productivity $a \in \mathcal{S}$ at time $t \in \mathbb{T}$. Thus, the row vector $P(t) = (P_1(t), P_2(t), \ldots, P_a(t), \ldots)$ represents the distribution of log productivity at time t. Notice that the vector is infinite to the right. It holds that $P_a(t) \geq 0$ and $\sum_{a=1}^{\infty} P_a(t) = 1$. In

[8]If firm i with log productivity $a_i(t)$ attempts to imitate firm j with log productivity $a_j(t) > a_i(t)$, then the expected log productivity i obtains is given by $\mathbb{E}_t[a_i(t + \Delta t)|a_i(t) = a, a_j(t) = b] = \sum_{c=0}^{b-a-1} (a + c)(1 - q)q^c + bq^{b-a} = a + q(1 - q^{b-a})/(1 - q)$. If $q < 1$ and b is much larger than a, the approximation $\mathbb{E}_t[a_i(t + \Delta t)|a_i(t) = a, a_j(t) = b] \approx a + q/(1 - q)$ holds. 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 that could be imitated.

what follows we may omit for simplicity either a or t in the arguments of $P_a(t)$ whenever it causes no confusion.

Our dynamics of innovation and imitation induces a discrete time, discrete space family of Markov chains $((P^N(t))_{t\in\mathbb{T}})_{N=N_0}^{\infty}$, where each member $(P^N(t))_{t\in\mathbb{T}}$ indexed by $N \geq N_0$ ($N_0 \in \mathbb{N}$ being some arbitrary lower bound on the number of firms) is a Markov chain that takes on values in the state space $P^N = \{P \in \mathbb{R}_+^{|\mathcal{S}|} : N \cdot P \in \mathbb{N}^{|\mathcal{S}|}, \sum_{a\in\mathcal{S}} P_a = 1\}$, i.e., the state space of frequency vectors for a specified N indicating the fraction of firms with a certain log productivity $a \in \mathcal{S}$. At times $t \in \mathbb{T} = \{0, \Delta t, 2\Delta t, \ldots\}$, with $\Delta t = 1/N$, exactly one firm in the population of N firms is selected at random and given the opportunity to introduce a technology improvement (through either innovation or imitation, as discussed in the following sections). The probability $T_{ab} : P^N \to \mathbb{R}_+^{|\mathcal{S}|\times|\mathcal{S}|}$ that a firm that is selected with log productivity a switches to log productivity b at time t is given by

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

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

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

With these definitions we are able to derive the differential equation governing the evolution of the productivity distribution by using the following proposition.[9]

Proposition 1. *Consider the Markov chain $(P^N(t))_{t\in\mathbb{T}}$ with transition matrix $\mathbf{T}(P)$. Define $V(P) \equiv P(t)(\mathbf{T}(P) - \mathbf{I})$ and let*

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

be 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

$$\frac{\partial P(t)}{\partial t} \in \bar{V}(P(t)) \tag{6}$$

for some initial distribution $P(0) : \mathcal{S} \to [0, 1]$. Moreover, if $\mathbf{T}(P)$ is Lipschitz continuous in P, then the evolution of the log*-productivity distribution $P(t)$ is given by the differential equation*

$$\frac{\partial P(t)}{\partial t} = V(P(t)) = P(t)\big(\mathbf{T}(P(t)) - \mathbf{I}\big). \tag{7}$$

[9] This proposition is an application of deterministic approximation theorems for discrete time Markov chains (cf. Kurtz 1970, Sandholm 2010). We refer in particular to Chapter 10 of Sandholm (2010) for a more detailed discussion of these approximation techniques.

Note that Proposition 1 covers the general case of the transition matrix $\mathbf{T}(P)$ not being Lipschitz continuous. Then the evolution of the log-productivity distribution follows a differential inclusion (i.e., a set-valued differential equation) as in (6).[10] In the case of a Lipschitz continuous $\mathbf{T}(P)$, we can simply write the evolution of the productivity distribution as a differential equation, which is stated in (7). Moreover, at all points of continuity of $\mathbf{T}(P)$ the differential inclusion is actually a differential equation.

In the following sections, we derive the matrix $\mathbf{T}(P)$ with elements $T_{ab}(P)$, $a, b \in \mathcal{S}$, under the individual firms' laws of motion associated with innovation in (3) and imitation in (5), respectively.

In Section 4 we look at the case where the decision to innovate vs. imitate is exogenous *and fixed*; this will be in contrast to the case in which a given firm will either imitate or innovate at different times, as will naturally occur when the choice is endogenous. 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 2). However, we do not observe such a divergence in the variance of empirically observed productivities as illustrated in Figure 1. In a more realistic model, it is therefore necessary to allow firms to engage in both innovation and imitation so as to advance their productivity levels. This is the case we are going to discuss in the subsequent Section 5, where the general model is introduced.

4. Exogenous innovation–imitation strategies

In this section, we introduce some key notation and provide an analysis of 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 4.1 all firms engage in in-house R&D, in Section 4.2 all firms try to imitate, and in Section 4.3 some firms always do in-house R&D, while others always imitate. We are not interested per se in these environments. However, they provide a useful contrast with (and intuition for) the results of Section 5, where firms choose optimally between in-house R&D and imitation, and where we present the main contribution of the paper. The reader who is more interested in the productivity dynamics with endogenous innovation choice might however skip these sections and start directly with Section 5.

4.1 *Innovation only*

Assume that all firms do in-house R&D or, equivalently, that firms have no absorptive capacity for imitation ($q = 0$). Innovation is assumed to yield a stochastic return and to have an incremental step-by-step nature. Namely, a firm engaging in R&D either moves one step upward in the productivity ladder or experiences no productivity change.[11] The

[10] See also Aubin and Cellina (1984).

[11] The assumption of step-by-step innovation is for simplicity. In the working paper version (König et al. 2014), we consider a more general formulation where firms doing R&D face a positive probability of making $0, 1, 2, \ldots, m$ steps forward, where $m < \infty$.

probability of success is given by $p > 0$, assumed to be independent of the firm's initial productivity. More formally, we can write the transition matrix due to in-house R&D as

$$\mathbf{T}^{\text{in}} = \begin{pmatrix} 1-p & p & 0 & \dots & 0 & \dots \\ 0 & 1-p & p & 0 & \dots & 0 \\ 0 & 0 & 1-p & p & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

From Proposition 1 it follows that, as $N \to \infty$, the evolution of the log-productivity distribution in (7) follows the ODE $\partial P(t)/\partial t = P(t)(\mathbf{T}^{\text{in}} - \mathbf{I})$. This is a diffusion equation with a positive drift. The central limit theorem implies then that the log productivity approaches a Gaussian shape as t grows. Both the mean and the variance rise linearly with t, as stated more formally by the following proposition.

Proposition 2. *Assume $q = 0$ and $p > 0$. Then, for large N, the* log-*productivity distribution approaches a normal distribution $\mathcal{N}(tp, tp(1-p))$ for large t. The productivity distribution converges to a* log-*normal distribution with mean $\mu_A = e^{tp(1+(1/2)(1-p))}$ and variance $\sigma_A^2 = (e^{tp(1-p)} - 1)e^{2tp+tp(1-p)}$.*

4.2 *Imitation only*

Next we consider the polar opposite case in which firms have no capacity to innovate through in-house R&D, and can progress only by imitating other firms' technologies. More formally, we assume $q > 0$ and $p = 0$. 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. In spite of this counterfactual implication, this is an instructive warm-up case, as it provides key insights for our main result.

The probability that a firm with log productivity a attains through imitation a log productivity $b > a$ is given by

$$\begin{aligned} T_{ab}^{\text{im}}(P) &= q^{b-a}P_b + q^{b-a}(1-q)P_{b+1} + q^{b-a}(1-q)P_{b+2} + \cdots \\ &= q^{b-a}\left(P_b + (1-q)\sum_{k=1}^{\infty} P_{b+k}\right) \\ &= q^{b-a}(P_b + (1-q)(1-F_b)), \end{aligned} \tag{8}$$

where F is the cumulative distribution of P, $F_b = \sum_{c=1}^{b} P_c$. The first term in the sum corresponds to a firm with log productivity a being matched with a firm with log productivity $b > a$ and successfully climbing up all the $b - a$ rungs. This happens with probability q^{b-a}; the second term describes the case in which the firm is matched with a firm with log productivity $b + 1$, but climbs only $b - a$ rungs, failing to climb the last rung; and so on. See also Figure 3. If $b < a$, the firm has nothing to imitate, thus $T_{ab}^{\text{im}}(P) = 0$. The probability for the firm not to make any improvement is, therefore, $T_{aa}^{\text{im}}(P) = 1 - \sum_{b>a} T_{ab}^{\text{im}}(P)$.

The transition matrix $\mathbf{T}^{\text{im}}$ with elements given by (8) is "interactive" and is given by[12]

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

where $S_a(P) \equiv 1 - \sum_{b=a+1}^{\infty} \mathbf{T}_{ab}^{\text{im}}(P) = 1 - \sum_{b=a+1}^{\infty} q^{b-a}(P_b + (1-q)(1-F_b))$. In the case of $q = 1$, which will be the benchmark of our analysis below, this simplifies to $S_a(P) = F_a$. In accordance with Proposition 1, for large N, the evolution of the log-productivity distribution is given by

$$\frac{\partial P(t)}{\partial t} = P(t)\big(\mathbf{T}^{\text{im}}(P(t)) - \mathbf{I}\big). \tag{9}$$

From (9) we can derive a system of differential equations governing the evolution of the cumulative log-productivity distribution.

Proposition 3. *Assume $q > 0$ and $p = 0$. Then, for large N, the evolution of the cumulative* log-*productivity distribution $F(t)$ is given by*

$$\frac{\partial F_a(t)}{\partial t} = F_a(t)^2 - F_a(t) + (1-q)(1-F_a(t))\sum_{b=0}^{a-1} q^b F_{a-b}(t), \quad a \in \mathcal{S}, \tag{10}$$

for some initial distribution $F(0) : \mathcal{S} \to [0, 1]$ with finite support. Then there exists a maximal initial log *productivity a_{m} such that $F_a(0) = 1$ for all $a \geq a_{\text{m}}$, and as $t \to \infty$, the distribution converges to*

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

i.e., $\lim_{t\to\infty} P_{a_{\text{m}}}(t) = 1$

In the special case of $q = 1$, we recover the knowledge growth dynamics analyzed by Lucas (2009).

4.3 *Innovation and imitation*

Consider next the evolution of the productivity distribution in a world where innovation strategies are exogenous, i.e., N_1 firms do in-house R&D while $N_2 = N - N_1$ firms imitate, where $N_1 \in \{0, 1, \dots, N\}$. In this case, the dynamics of the productivity frontier is governed by the firms engaged in in-house R&D. The resulting evolution of the productivity distribution is as analyzed in Section 4.1.[13] There we show that the productivity distribution of firms doing R&D converges to a log-normal distribution with an

[12] A Markov chain is interactive if the transition probabilities depend on the current distribution (Conlisk 1976).

[13] A more formal analysis of the case in which there are both innovators and imitators is provided in Appendix A.1.

ever increasing variance (see Proposition 2). Since the proportion of innovators and imitators is fixed, this implies that also the variance of the distribution of the total population of firms must diverge.[14] Since the empirical evidence discussed in the introduction (cf. Figure 1) suggests that there is no such increase in the variance of the distribution, a model with an exogenous proportion of innovators and imitators yields counterfactual predictions.

5. Endogenous choice of the innovation strategy

This section contains the main result of the paper. We assume that firms choose whether to innovate through 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 in every period. In turn, (1) shows that the profit is linearly increasing in the technology level. Thus, profit-maximizing firms endeavor simply to maximize the expected level of technology every period.[15] The intuitive reason for this equivalence is that there are no sunk costs: The opportunity cost of innovation is the return from imitation, and vice versa, and firms can switch back and forth between innovation and imitation with no adjustment cost. Hence, forward-looking firms simply choose the strategy (either in-house R&D or innovation) so as to maximize the expected number of improvements along the quality ladder.

Let $\mathbb{E}_i^{\text{in}}[A_i(t+\Delta t)|A_i(t)]$ and $\mathbb{E}_i^{\text{im}}[A_i(t+\Delta t)|A_i(t), P(t)]$ denote the expected productivity for a firm whose current productivity is $A_i(t)$, conditional on choosing in-house R&D and imitation, respectively. Recall that expected profits are proportional to expected productivities (see (1) in Section 2.1). Thus, the profit-maximizing firm i chooses in-house R&D whenever

$$\mathbb{E}_i^{\text{in}}[A_i(t+\Delta t)|A_i(t)] > \mathbb{E}_i^{\text{im}}[A_i(t+\Delta t)|A_i(t), P(t)], \tag{11}$$

where the expected productivity from innovation is given by

$$\mathbb{E}_i^{\text{in}}[A_i(t+\Delta t)|A_i(t)] = A_i(t)((1-p)+p\bar{A}),$$

while the expected productivity from imitation is

$$\begin{aligned}&\mathbb{E}_i^{\text{im}}[A_i(t+\Delta t)|A_i(t), P(t)]\\ &\quad = A_i(t)\left(S_{a_i(t)}P(t) + \sum_{b=a_i(t)+1}^{\infty} \bar{A}^{b-a_i(t)} q^{b-a_i(t)}\big(P_b(t)+(1-q)(1-F_b(t))\big)\right),\end{aligned}$$

[14]In particular, there is divergence in the subpopulation of firms carrying out R&D, as these do not benefit from the spillover associated with the progress in the frontier technology. It is possible to characterize the dynamics of the cumulative log-productivity distribution in terms of a differential equation, although this admits no closed-form solution. The analysis is deferred to Appendix A.1.

[15]For a formal proof, see Proposition 8 in Appendix A.2, showing that the firm's value function is increasing in its technology level.

and $S_a(P) = 1 - \sum_{b=a+1}^{\infty} \mathbf{T}_{ab}^{\text{im}}(P)$, as defined in Section 4.2. The decision rule in (11) can alternatively be captured by the indicator function

$$\chi^{\text{im}}(a_i(t), P(t)) = \begin{cases} 1 & \text{if } a_i^{\text{im}}(a_i(t), P(t)) \geq a_i^{\text{in}}(a_i(t)) \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where $a_i^{\text{in}}(a_i(t)) \equiv \log \mathbb{E}_i^{\text{in}}[A_i(t+\Delta t)|A_i(t)]$ and $a_i^{\text{im}}(a_i(t), P(t)) \equiv \log A_i^{\text{im}}(A_i(t), P(t))$. In words, $\chi^{\text{im}}(a_i(t), P(t)) \in \{0, 1\}$ is the indicator variable being 1 if firm i pursues imitation and being 0 if the firm pursues in-house R&D. Similarly, we define $\chi^{\text{in}}(a_i(t), P(t)) \equiv 1 - \chi^{\text{im}}(a_i(t), P(t))$.

To achieve a complete analytical characterization, in the rest of this section we restrict our attention to economies in which firms have no absorptive capacity limits, $q = 1$. We shall return to the more general case in Section 6.

Proposition 4. *Assume that $q = 1$. Then for any P there exists a unique threshold* log *productivity $a^*(P) \in \mathcal{S}$ such that (i) $\chi^{\text{im}}(a, P) = 1$ (and $\chi^{\text{in}}(a, P) = 0$) for $a \leq a^*(P)$ and (ii) $\chi^{\text{im}}(a, P) = 0$ (and $\chi^{\text{in}}(a, P) = 1$) for $a > a^*(P)$.*

Proposition 4 establishes that the decision about the innovation strategy has a threshold property: relatively backward firms (i.e., those weakly below the threshold $a^*(P)$) optimally choose to imitate, while more advanced firms (i.e., those above the threshold $a^*(P)$) choose to innovate.

We now turn to the equilibrium dynamics. The transition matrix $\mathbf{T}(P)$ is the sum of the transition matrices for innovation and imitation given in Sections 4.1 and 4.2, respectively, each weighted by the respective indicator function from (12). The equilibrium dynamics of the log-productivity distribution can be represented by the differential inclusion in (6) in Proposition 1. However, it is not possible to express the equilibrium dynamics in terms of the ODE (7). The reason is that whenever $a_i^{\text{in}}(a^*(P)) = a_i^{\text{im}}(a^*(P), P)$, i.e., firms at the productivity level a^* are indifferent between in-house R&D and imitation, the indicator function $\chi^{\text{im}}(a^*(P), P)$ is discontinuous in P. This violates the standard continuity condition under which we can represent the dynamics as an ODE. Since proving our main result using the theory of differential inclusions would be more involved, we roundabout this technical complication by replacing the discontinuous indicator function by a continuous approximation. This allows us to the express the equilibrium dynamics in terms of an ODE (see (14) below). More formally, we define the continuous logistic function

$$\chi_{\beta}^{\text{im}}(a_i(t), P(t)) = \frac{1}{1 + e^{-\beta(a_i^{\text{im}}(a_i(t), P(t)) - a_i^{\text{in}}(a_i(t)))}}, \tag{13}$$

with the property that $\lim_{\beta \to \infty} \chi_{\beta}^{\text{im}}(a_i(t), P(t)) = \chi^{\text{im}}(a_i(t), P(t))$. For large β, we then have that $\chi_{\beta}^{\text{im}}(a_i(t), P(t)) \approx \chi^{\text{im}}(a_i(t), P(t))$. In the working paper version (König et al. 2014), we propose an explicit microfoundation for such a formulation, whereby firms are subject to stochastic shocks affecting their productivity in performing in-house R&D, and these shocks then create a time-varying comparative advantage for different firms.

Replacing χ^{im} by χ^{im}_{β} and assuming a large population of firms ($N \to \infty$) allows us to write the evolution of the log-productivity distribution as[16]

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

for some initial distribution $P(0) : \mathcal{S} \to [0, 1]$, where $\mathbf{D}(P)$ denotes the diagonal matrix with diagonal elements given by $\chi^{\text{im}}_{\beta}(a, P)$ for all $a \in \mathcal{S}$. Making explicit the individual equation for each relative frequency, P_a, yields

$$\begin{aligned} \frac{\partial P_a(t)}{\partial t} = {} & P_a(t)\left(\sum_{b=1}^{a-1} \chi^{\text{im}}_{\beta}(b, P)P_b(t) + \chi^{\text{im}}_{\beta}(a, P)S_a(P)\right) + (1-p)P_a(t)(1 - \chi^{\text{im}}(a, P)) \\ & + pP_{a-1}(t)(1 - \chi^{\text{im}}_{\beta}(a-1, P)) - P_a(t), \quad a \in \mathcal{S}. \end{aligned} \tag{15}$$

The system of ODEs in (15), expressed in terms of P_a, can be turned into a system of ODEs in terms of the complementary cumulative productivity distribution, $G_a(t) = 1 - F_a(t)$, as indicated in the following proposition.

Proposition 5. *Assume a large population of firms with unlimited absorptive capacity limits ($q = 1$). Let the decision rule $\chi^{\text{im}}(a_i(t), P(t))$ be approximated by the continuous (logistic) function $\chi^{\text{im}}_{\beta}(a_i(t), P(t))$ given by (13). Then, in the limit of $\beta \to \infty$, for all $a \in \mathcal{S}$, the dynamics of the cumulative* log-*productivity distribution is*

$$\frac{\partial G_a(t)}{\partial t} = \begin{cases} G_a(t) - G_a(t)^2 & \text{if } a \leq a^*(P) \\ (1 - G_{\lfloor a^*(t) \rfloor}(t))G_a(t) - p(G_a(t) - G_{a-1}(t)) & \text{if } a > a^*(P). \end{cases} \tag{16}$$

The system of ODEs (16) can be solved numerically subject to the boundary conditions $\lim_{a\to\infty} G_a(t) = 0$ and $\lim_{a\to 1} G_a(t) = 1$. More interestingly, it is possible to characterize analytically a *steady-state* distribution consistent with (16).[17] Contrary to the case in which firms are assigned exogenously to in-house R&D and innovation, and consistent with the empirical evidence, this distribution has a constant variance. Moreover, contrary to the case of pure imitation this productivity distribution grows over time at a constant rate. Next we provide a formal definition of a *traveling wave*:

Definition 1. The log-productivity distribution $G_a(t)$ is a traveling wave if it is of the form $G_a(t) = g(a - \nu t)$ for some nonincreasing function $g : \mathbb{R} \to [0, 1]$, where $\nu \geq 0$ is the traveling wave velocity.

Note that Definition 1 implies that a traveling wave has the property that $G_a(t) = G_{a+\nu s}(t + s)$ for any $s \geq 0$. The following proposition shows that a traveling wave with two exponential tails is a solution for the log-productivity distribution satisfying (16).

[16] Note that this representation is legitimate for a $\beta < \infty$, although later we will focus on the limit in which $\beta \to \infty$, which is the economically interesting case. See also Section 7 for further discussion.

[17] By a steady-state distribution, we mean a distribution whose shape is preserved over time, up to changes in its mean. See the more formal definition of a *traveling wave* in Definition 1.

Proposition 6. *A function $g : \mathbb{R} \to [0,1]$ and a traveling wave velocity $\nu \geq 0$ exist such that a traveling wave $G_a(t) = g(a - \nu t)$ is a steady-state solution of (16), with a threshold given by $a^*(t) = a_0^* + \nu t$, for a constant a_0^* determined by the initial condition, $a_0^* = a^*(0)$. The shape of the traveling wave for $a \leq a^*(t)$ is*

$$G_a(t) = \frac{1}{1 + \left(\frac{1}{g_0} - 1\right) e^{(a - a_0^* - \nu t)/\nu}}, \tag{17}$$

with $g_0 = g(0)$. For $a > a^*(t)$ there exists a $p^* > 0$ such that for all $0 < p < p^*$ the two inequalities

$$\sum_{k=-\infty}^{\infty} \underline{c}_k e^{-\underline{\lambda}_k (a - \nu t)} \leq G_a(t) \leq \sum_{k=-\infty}^{\infty} \overline{c}_k e^{-\overline{\lambda}_k (a - \nu t)} \tag{18}$$

hold with appropriate constants $\underline{c}_k$, $\overline{c}_k$, and exponents $\underline{\lambda}_k$, $\overline{\lambda}_k$ having strictly positive real parts. Consequently, the following asymptotic results hold for the associated probability mass function $P_a(t) = G_{a-1}(t) - G_a(t)$:[18]

$$P_a(t) = \begin{cases} e^{\frac{a - \nu t}{\nu}} + o(1) & \text{if } a \ll a^*(t) \\ O(e^{-\overline{\lambda}_0 (a - \nu t)}) & \text{if } a \gg a^*(t). \end{cases} \tag{19}$$

The first part of the proposition establishes that if the log-productivity distribution follows the equilibrium law of motion dictated by (15) (or, identically, by (16)), then in the stationary state, the distribution reproduces itself over time, up to a trend in $a^*(t)$ whose growth is pinned down by ν.[19] The distribution is a traveling wave with velocity ν, i.e., a distribution whose second and higher moments remain constant over time.

Observe that the second part of Proposition 6 requires that the in-house R&D success probability p is bounded from above. While this assumption is necessary for the proof of this part of the proposition, in all the numerical simulations shown in the following sections we did not find a departure of the exponential decay of the right tail of the distribution.

For $a \leq a^*(t)$ in (17) we can provide an exact characterization of the solution of (16), while above the threshold in (18) we can only provide a lower and an upper bound to the exact solution. This is because the second part of (16) (for $a > a^*(t)$) is more complicated to analyze. To see this, note that the mass of firms with log productivity a below the threshold $a^*(t)$ can only change through imitation of firms with higher log productivities, where the mass of such firms is given by $G_a(t)$. In contrast, the change in the mass of firms above the threshold has two different components: First, it can change due to productivity gains from innovation, which are determined by the innovation success probability p. Second, there is an influx of imitating firms that become innovating firms in the next period, and in every period the mass of these imitating firms is given

[18] We define $g(x) = O(f(x))$ if and only if $|g(x)/f(x)|$ is bounded from above by a constant (which in our case is 1) as $x \to \infty$. Moreover, $g(x) = o(f(x))$ if and only if $g(x)/f(x) \to 0$ as $x \to \infty$, and $g(x) \sim f(x)$ if and only if $g(x)/f(x) \to 1$ as $x \to \infty$. The latter can also be written as $g(x) = f(x) + o(f(x))$.

[19] Note that $a^*(t)$ in Proposition 6 can be related to the cutoff $a^*(P)$ by using the floor function, where $a^*(P(t)) = \lfloor a_0^* + \nu t \rfloor = \lfloor a^*(t) \rfloor$, that is, the largest integer no greater than the threshold $a^*(t)$.

by $F_{\lfloor a^*(t)\rfloor}(t) = 1 - G_{\lfloor a^*(t)\rfloor}(t)$. This is why $G_{\lfloor a^*(t)\rfloor}(t)$ appears only in the second part of (16), and because of these two components and the term proportional to $G_{\lfloor a^*(t)\rfloor}(t)$ this part of (16) is more difficult to analyze.

The bounds in (18) for values of the log productivity above the threshold $a^*(t)$ exploit recent results in the mathematics literature for the analysis of so-called delay differential equations (DDE) (cf. Bellman and Cooke 1963, Driver 1977, Smith 2010), showing that the solutions to such DDE can be written as a linear combination of exponential functions (cf. Asl and Ulsoy 2003, Yi and Ulsoy 2006). More precisely, one can show that due to the appearance of the term $G_{\lfloor a^*(t)\rfloor}(t)$ in the second part of (16) we need to solve a linear DDE with nonconstant coefficients. We can, however, establish upper and lower bounds to the solution to this equation that are themselves solutions to linear DDEs with constant coefficients. Asl and Ulsoy (2003) have shown that the latter can be expressed as sums of exponential functions with well defined exponents. For log productivities far above the threshold only the dominating exponential terms in these sums remain, and so they provide exponential upper and lower bounds for the tail of the distribution. The details (including a more explicit characterization of the constants $\underline{c}_k$, $\overline{c}_k$, and exponents $\underline{\lambda}_k$, $\overline{\lambda}_k$) can be found in the proof of Proposition 6 in Appendix A.3.

The productivity distribution characterized by (17) and (18) features both a right-hand and a left-hand power-law tail, similar to what we observe in the data (see Figure 1).[20] More precisely, the lower tail of the distribution follows immediately from the logistic expression in (17); the upper tail of the distribution corresponds to the approximation of the sum $\sum_{k=-\infty}^{\infty} \underline{c}_k e^{-\underline{\lambda}_k(a-\nu t)}$ in the lower bound of (18) where only the term for $k = 0$ is retained, whereas all other terms of the sequence become negligible when a is far above the threshold $a^*(t)$, and the upper and lower bounds in (18) get arbitrarily close to each other. A numerical analysis of the solution shows that only a few terms in the sum are sufficient to obtain a good approximation of the whole stationary distribution.[21,22] Moreover, even considering only the dominant exponent (i.e., $\underline{\lambda}_0$) in the lower bound in (18) yields a fairly accurate approximation. In this case, the solution becomes very simple: $\underline{\lambda}_0$ turns out to be the unique root of the transcendental equation[23]

$$(e-1)e^{\underline{\lambda}_0}(\underline{\lambda}_0 - 1) - (\bar{A}-1)e^{1-\underline{\lambda}_0}(1+\underline{\lambda}_0) + \bar{A} + e - 2 - \frac{e-1}{p} = 0, \tag{20}$$

while the traveling wave velocity ν is given by

$$\nu = \frac{1}{\underline{\lambda}_0}\left(1 + p(e^{\underline{\lambda}_0} - 1) - \frac{p(\bar{A}-1)(1-e^{1-\underline{\lambda}_0})}{e-1}\right). \tag{21}$$

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

[21] Figure 10 in Appendix A.3 compares the solution obtained from a direct numerical integration of (16) with that obtained from the analytical solution of (17) and (18), after truncating the sequence of exponents $\underline{\lambda}_k$ to $k \in \{0, -1\}$. The numerical solution is very well approximated over the entire support; additional terms would not alter the distribution in any visible way.

[22] The Lambert function has always at most two real roots, corresponding in our notation to $k = 0$ (the "dominant root") and $k = -1$. See, e.g., Asl and Ulsoy (2003), Corless et al. (1996).

[23] The details of this derivation can be found in Remark 3 in Appendix A.3.

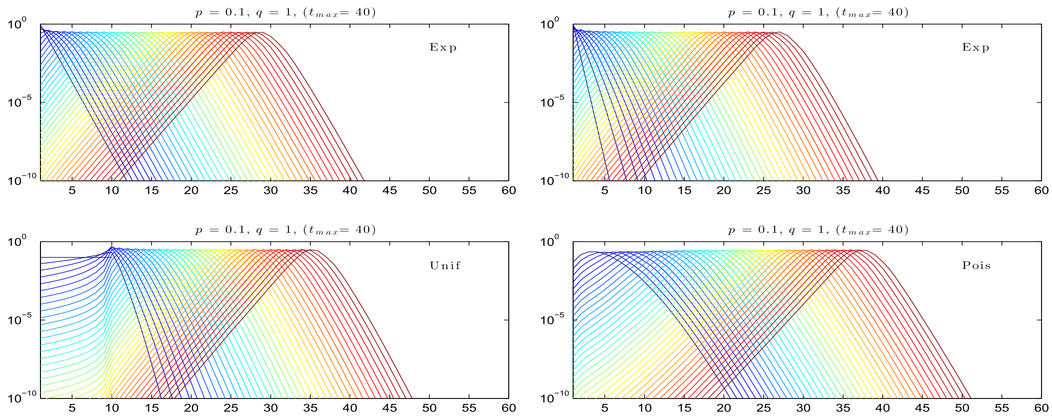

Figure 4. Examples of numerical solutions of the system of ODEs in (14) with different initial conditions. In all cases, we set $p = 0.1$, $q = 1$, and $\log \bar{A} = 1$, yielding $\lambda = 2.1$ (cf. (20)). The top left panel shows the traveling wave with an exponential decay with $\lambda = 2.1$. The other three panels show the transition from different initial conditions. The top right panel shows an initial distribution decaying exponentially with an exponent $\lambda = 5$. The tail of the distribution increases until it reaches the stationary value of $\lambda = 2.1$, as in the top left panel. The bottom left panel shows the transition from a uniform initial distribution on the interval $[0, 1]$. The bottom right panel shows the transition from a Poisson initial distribution with parameter 3. In all cases, the distribution converges to the traveling wave in the top left panel.

Proposition 6 yields an existence result: a traveling wave with an associated particular probability mass function is a steady-state solution for the log-productivity distribution.[24] For other initial distributions different from the steady-state distribution there will be transitional dynamics. We are unable to establish formal conditions that guarantee that the distribution converges to the traveling wave in Proposition 6. However, we have obtained convergence in numerically computed solutions of the system of ODEs in (16) with a variety of initial distributions. Figure 4 shows three such cases. The top left panel shows, for reference, a simulation in which the initial condition is consistent with the steady-state distribution—no transitional dynamics.[25]

The top right panel considers an initial exponential distribution with a steeper tail than in (17) and (18). As the figure shows, the tail of the distribution increases during the transition. The bottom left panel shows the case of a uniform initial distribution. Finally, the bottom right panel shows a simulation starting from a Poisson distribution. In all cases, the distributions converge to the stationary distribution shown in the top left panel.[26]

[24]However, we are unable to make any claim about the uniqueness of the steady-state distribution. Luttmer (2012) proves uniqueness in a related setup. However, the model is different, and it is not clear whether similar techniques can be extended to our framework.

[25]Note that the imitation–innovation threshold lies to the right of the maximum of the distribution. In the region around the maximum, firms imitate and the distribution is characterized by the logistic expression $(1 + (1/g_0 - 1)e^{(a-a_0^*-\nu t)/\nu})^{-1}$. In the region where firms innovate, the log productivity is well approximated by an exponentially decaying function.

[26]The code can be obtained upon request from the authors.

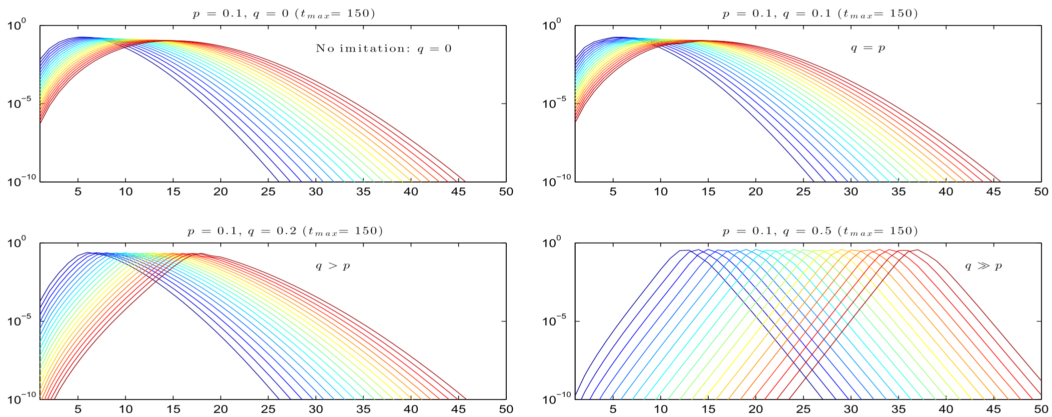

Figure 5. Examples of numerical solutions of the system of ODEs in (14) with different values of q. In all cases we set $\log \bar{A} = 1$ and $p = 0.1$. The top left panel shows an economy where growth is driven by innovation only ($q = 0$). The top right panel shows the case in which $p = q = 0.1$. The bottom panels show, respectively, the case of $q = 0.2$ and $q = 0.5$.

6. Limited absorptive capacity

In this section we consider the more general model in which firms have a limited ability to absorb other firms' technologies. We are motivated by the observation that the steady-state distribution characterized in Proposition 6 fits the data well in a qualitative but not in a quantitative sense. Intuitively, if one calibrates the key parameter of the model, p, to fit the tails of the empirical distribution in Figure 1 (and, in particular, to fit the variance of the distribution), the model overpredicts the growth rate. The intuitive reason is that the convergence rate of imitating firms is too high. Then, so as to fit the spread of the distribution, one must increase the rate of success of innovation, inducing fast growth. Alternatively, if one targets the growth rate by setting a lower value of p, the model yields too low a variance.[27]

To address this quantitative failure, we extend the model to allow for $q \leq 1$. The analysis of the case in which $q < 1$ can only be done with the aid of numerical methods (i.e., by numerical integration of (15)). Figure 5 shows numerically computed solutions of the system of ODEs in (15) for a probability of success of innovation $p = 0.1$. The figure shows four cases corresponding to different values of q.[28] As shown more formally in the analysis of Section 4, the solution in the case without imitation, $q = 0$, features a log-normal shape (i.e., a parabola in the semi-log plot) with a growing variance over time (see the top left panel). The same qualitative property extends to the case of $q = p$, i.e., when a step of imitation is as likely as a step of innovation ($q = p$). However, for q sufficiently large the distribution converges to a traveling wave with stable exponential

[27]Recall from our discussion in Section 4.2 that in the extreme case of $p \to 0$ the distribution shrinks to a degenerate distribution with mass 1 localized at the highest initial productivity value.

[28]All computations started with the initial distribution $P(0) = (1, 0, \ldots, 0)$ and levels of log productivity ranging over $a = 1, \ldots, 50$. Twenty time steps are shown ($t = 55, \overset{+5}{\ldots}, 150$; in colors from blue to red in the online version of Figure 5).

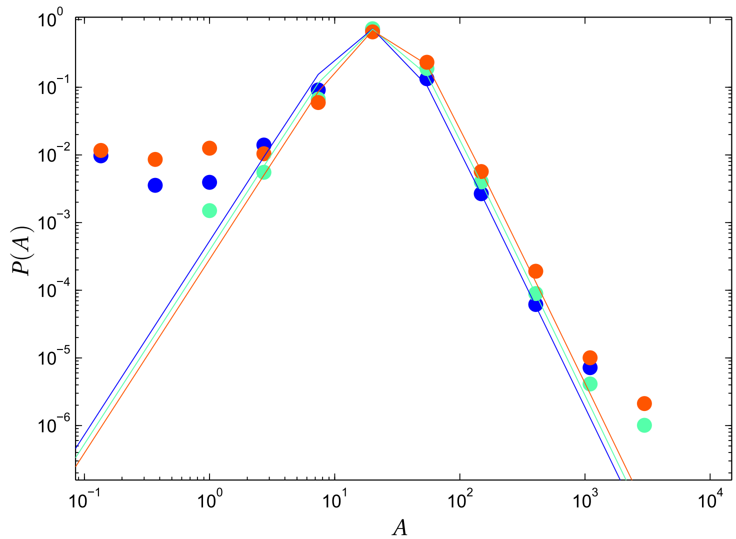

Figure 6. Comparison of the empirical distributions of Figure 1 with the calibrated model ($p = 0.0049$ and $q = 0.106$) for the years 1995, 1999, and 2003. The empirical productivity values have been binned to produce the histogram shown in the figure, using 11 bins across all observed productivity values.

tails. This is clearly visible in the bottom right panel, where the exponential tails are straight lines in the logarithmic scale of the plot.[29] Hence, our analysis suggests that a value of q considerably larger than p is necessary to match the data in Figure 1.

Next, we calibrate the parameters of our model to match the empirical productivity distribution. The details of our calibration procedure are in Appendix B.3. The best match is obtained by setting $p = 0.0049$ and $q = 0.106$. Figure 6 displays a comparison of the empirical distributions with the calibrated model for the years 1995, 1999, and 2003. The comparison between the simulated and the empirical distributions shows that the model can reproduce the observed pattern well.

7. Noisy choice of innovation and imitation

In this section, we generalize the results of Section 5 to the case in which the noise in the firm's choice of innovation strategy is noninfinitesimal (cf. (13)). The main goal of this extension is to provide a robust intuition for the driving force behind the emergence of a traveling wave. We show, in particular, that the optimal choice of innovation and imitation, is not essential. Rather, the traveling wave emerges whenever the model features a stochastic switching of firms between innovation and imitation strategies.[30]

We assume that the probability that a firm with log productivity $a_i(t)$ pursues imitation is given by (13). The decision rule in (13) can be motivated by assuming that

[29] Additional numerical analysis suggests that such traveling waves with exponential tails also emerge for lower innovation probabilities whenever $q \geq 5p$.

[30] We would like to thank the co-editor for pointing this out.

firms' profits from in-house R&D are exposed to stochastic shocks (see the accompanying working paper, König et al. (2014), for further details), while the limiting case in which $\beta \to \infty$ is analyzed in Section 5. Allowing for nonnegligible noise has no major qualitative implications. Since the innovation strategy is chosen less and less efficiently as we decrease β, the model predicts a lower productivity growth rate. While the general case can only be analyzed numerically, analytical results can be obtained for the polar case in which we let $\beta \to 0$. This yields $\chi_\beta^{\mathrm{im}}(a, P) \to 0.5$, namely, every firm chooses randomly between imitation and in-house R&D, irrespective of a and P.[31]

Setting $\chi_\beta^{\mathrm{im}}(b, P) = 0.5$ in (15) and summing over a yields the equilibrium dynamics governed by the system of ODEs

$$\frac{\partial F_a(t)}{\partial t} = \frac{1}{2}(F_a(t)^2 - F_a(t)) - \frac{p}{2}(F_a(t) - F_{a-1}(t)) \tag{22}$$

for all $a \in \mathcal{S}$. The next proposition establishes that there exists a traveling wave solution to (22).

Proposition 7. *Let $F_a(t)$ be a solution of (22) with a Heaviside initial distribution $F_a(0) = \Theta(a - a_{\mathrm{m}})$ for some $a_{\mathrm{m}} \geq 1$ and define $m_\epsilon(t) = \inf\{a : F_a(t) > \epsilon\}$. Then $\lim_{t\to\infty} \frac{m_\epsilon(t)}{t} = \nu$, for some constant $\nu \geq 0$, and $F_a(t)$ is a traveling wave of the form $F_a(t) = f(a - \nu t)$ for some nondecreasing function $f : \mathbb{R}_+ \to [0, 1]$.*

In addition, one can show that the limiting log-productivity distribution decays exponentially in the tails, similar to what we have found in Proposition 6.[32] Figure 7 illustrates examples of numerically computed solutions of the system of ODEs in (22) for $p = 0.1$, $q = 1$, and $\log \bar{A} = 1$, showing the transition from the same initial conditions as in Figure 5 to a traveling wave with stable shape. We observe that the distribution moves more slowly to the right than in Figure 5 due to the suboptimal random mixing between in-house R&D and imitation.

While we do not view a model in which firms choose their innovation strategy randomly as particularly appealing, its analysis yields interesting insights about the formal properties of the model. In particular, the existence of a traveling wave contrasts sharply with the result of the model in Section 4 where a fixed number of firms imitate and the rest do in-house R&D. In that model, the variance of productivity grows over time, whereas in the model of this section the variance does not blow up—despite the fact that in both cases the proportion of innovators and imitators is assumed to be constant. The key difference is that in the case of deterministic innovation strategies the variance increases over time within the population of in-house innovators that are permanently barred from the spillovers. In this section's model, in contrast, even firms failing repeatedly to innovate through in-house R&D are assigned, sooner or later, to imitation. When this happens, they can benefit from the productivity spillovers generated by successful firms. The fact that laggard innovators switch with positive probability into imitation, prevents the emergence of an ever-growing tail of the distribution.

[31]This model is similar to the one analyzed in Majumdar and Krapivsky (2001).

[32]The proof is available upon request.

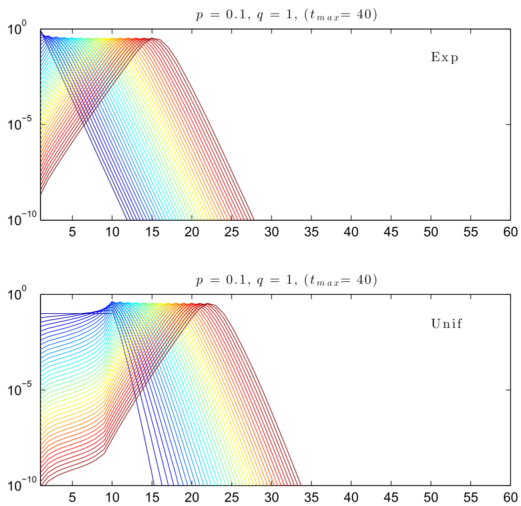

FIGURE 7. Examples of numerical solutions of the system of ODEs in (22) with $\beta = 0$ (random choice of imitation vs. in-house R&D), given different initial conditions. In all cases, we set $p = 0.1$, $q = 1$, and $\log \bar{A} = 1$. The initial conditions are the same as in Figure 5. In all cases, we observe the transition to a traveling wave with a stable shape.

In conclusion, it is not per se the optimal choice of innovation vs. imitation that yields a stable distribution. What matters is productivity spillovers coupled with the assumption that all firms can benefit from them with a positive probability. The profit-maximizing behavior of firms is a particular case of this model featuring an efficient sorting of firms into the two strategies.

8. Conclusion

In this paper we construct a model of endogenous technological change, productivity growth, and technology spillovers that 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 other firms' existing technologies. The emerging productivity distributions can be described as traveling waves with a constant shape and power-law tails, matching the empirically observed distributions.

The current model can be extended in a number of directions. We sketch three extensions in Appendix B.1. First, we outline a model of productivity growth and technology adoption that includes the possibility that a firm's productivity may also be reduced due to exogenous events such as the expiration of a patent. Second, we allow for entry and exit. Third, we consider an alternative model of capacity constraints on the ability of firms to adopt and imitate external knowledge, whereby below a relative productivity threshold firms become unable to imitate. In this case, the model can generate "convergence clubs" such as those documented in empirical studies of cross-country income differences (e.g., Feyrer 2008, Quah 1997, Durlauf and Johnson 1995).

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 Kelly (2001), Di Matteo et al. (2005), Ehrhardt et al. (2006), König (2014). We leave this avenue for future research.

Appendix A: Additional results

A.1 *Analysis of Section 4.3: Exogenous innovation strategies*

In Section 4.3 we consider a model in which the innovation strategy (either in-house R&D or imitation) is a fixed characteristic of firms. We state that in this case the productivity distribution has an ever-increasing variance. In this appendix we provide the details of the analysis. In particular, in (24) below we provide a differential equation that completely characterizes the dynamics of the log-productivity distribution.

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 4.1)

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

Thus, the variance of the distribution increases over time.

For completeness, we also characterize the evolution of the log-productivity distribution $P_a^{(2)}(t)$ of imitating firms. This is given by (see also Section 4.2)

$$\frac{\partial P_a^{(2)}(t)}{\partial t} = P_a(t) \sum_{b=1}^{a} P_b^{(2)}(t) - P_a^{(2)}(t)\left(1 - \sum_{b=1}^{a-1} P_b(t)\right). \tag{23}$$

The first term in the above equation takes into account the fraction of imitating firms with log productivities smaller than 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.

Summing over a and rearranging terms, one can then derive from (23) the dynamics of the cumulative log-productivity distribution $F_a(t)$, which is given by

$$\frac{\partial F_a(t)}{\partial t} = F_a(t)^2 - F_a(t) - n_1 F_a^{(1)}(t) F_a(t) + n_1 F_a^{(1)}(t) - n_1 p P_a^{(1)}(t). \tag{24}$$

Given the solution for $P_a^{(1)}(t)$ (and $F_a^{(1)}(t)$, respectively) and a fixed value of a, (24) is a Riccati first-order, linear differential equation with nonconstant, nonlinear coefficients, for which no closed-form solution exists.[33]

A.2 *Analysis of Section 5: The dynamic problem of the firm*

In the text we state that when a firm maximizes its expected productivity increase, it also maximizes its present value. Thus, the static optimization studied in the text is equivalent to a dynamic value-maximization problem. We consider for simplicity time increments of $\Delta t = 1$. The dynamic problem of the firm is then given by

$$V_0(A_i(0), P(0)) = \max_{(s_i(t)\in\{\text{im},\text{in}\})_{t=0}^{T-1}} \mathbb{E}\left[\sum_{t=0}^{T-1} \delta^t \pi_i^{s_i(t)}(t) \middle| A_i(0), P(0)\right],$$

where $\pi_i^{s_i(t)}(t) = \psi A_i(t)\bar{A}^{\vartheta_{s_i(t)}}$ is the per period profit of firm i choosing the R&D strategy $s_i(t) \in \{\text{im}, \text{in}\}$, $\vartheta_{s_i(t)}$ are the random increments along the quality ladder under strategy $s_i(t)$, and δ is a discount factor. The corresponding Bellman equation is given by

$$\begin{aligned} V_t(A_i(t), P(t)) = \max_{s_i\in\{\text{im},\text{in}\}} \big\{ & \psi A_i(t)\mathbb{E}[\bar{A}^{\vartheta_{s_i}} | A_i(t), P(t)] \\ & + \delta\mathbb{E}\big[V_{t+1}(A_i(t)\bar{A}^{\vartheta_{s_i}}, P(t+1)) | A_i(t), P(t)\big]\big\}. \end{aligned}$$

This can be written as

$$\begin{aligned} V_t(A_i(t), P(t)) = \max\bigg\{ & \int dF^{\text{in}}(\vartheta)\big(\psi A_i(t)\bar{A}^{\vartheta} + \delta V_{t+1}(A_i(t)\bar{A}^{\vartheta}, P(t+1))\big), \\ & \int dF^{\text{im}}(\vartheta | A_i(t), P(t))\big(\psi A_i(t)\bar{A}^{\vartheta} + \delta V_{t+1}(A_i(t)\bar{A}^{\vartheta}, P(t+1))\big)\bigg\}. \end{aligned} \tag{25}$$

Similar to Theorem 1 in Lippman and McCall (1976), we can state the following lemma.

Lemma 1. *The value function $V_t(A_i(t), P(t))$ of (25) is increasing in the productivity of firm i, $A_i(t)$, for all $i = 1, \dots, n$ and $t \geq 0$.*

With the above lemma we are now able to state the following proposition.

Proposition 8. *Consider the value function of (25). Then for each period t it is optimal for firm i to choose the strategy $s_i(t) \in \{\text{im}, \text{in}\}$ that gives it the highest expected productivity in that period.*

[33] For a fixed log productivity a, denote $y(t) = F_a(t)$. Then one can write from (24) the differential equation $dy(t)/dt + ay(t)^2 + b(t)y(t) = c(t)$, where $a = -1$, $b(t) = 1 + n_1F_a^{(1)}(t)$, and $c(t) = n_1(F_a^{(1)}(t) - pP_a^{(1)}(t))$.

A.3 *Proofs of propositions and lemmas*

In this section, we provide a formal proof of the propositions and lemmas in the text. It is convenient to introduce the random variable ζ_P^N whose distribution describes the stochastic increments of $(P^N(t))_{t\in\mathbb{T}}$ from the state $P \in P^N$:

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

Moreover, following the notation in Sandholm (2010, Chapter 10.2), we introduce the functions

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

We then can state the following lemma.

Lemma 2. *Consider some sequence* $(\delta^N)_{N=N_0}^\infty$ *with* $\lim_{N\to\infty} \delta^N = 0$. *Then we have that*

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

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

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

Proof. In the following text, we prove conditions (i)–(iii). 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) is satisfied. Further, observe that since $|e_a - e_b| = \sqrt{2}$ for $a \neq b$ and 0 otherwise, $(P^N(t))_{t\in\mathbb{T}}$ has jumps of at most $\sqrt{2}/N$. Hence, for $\delta^N = \sqrt{2}/N$,

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

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

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

and condition (ii) is also satisfied. □

We now can give the proof of Proposition 1.

Proof of Proposition 1. Note that the indicator function for imitation, $\chi^{\text{im}}(a, P)$, of (12) has a point of discontinuity at the threshold log productivity a^*, and so does $V(P) = \mathbf{T}(P) - \mathbf{I}$. Let $\|P\|$ denote the L^2 norm in $\mathbb{R}_+^{|S|}$. Define

$$\bar{V}(P) = \bigcap_{\epsilon>0} \text{cl}\Big(\text{conv}\Big(V\big(\{P' \in \mathbb{R}_+^S : \|P - P'\| \leq \epsilon\}\big)\Big)\Big) \tag{26}$$

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).[34,35]

Theorem 1. *Let $\bar{V}(P)$ be upper semicontinuous and assume that there exists a $c > 0$ 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

$$\frac{\partial P}{\partial t} \in \bar{V}(P), \tag{27}$$

with initial conditions $P(0)$ for any $t \in [0, T]$, $T \in \mathbb{R}_+$, and $D_T(P(0))$ denotes the set of all solutions of (27) 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 (26). By Lemma 2, $V(P)$ is bounded, and so we have that $\bar{V}(P)$ is bounded and upper semicontinuous. Hence, the requirements of Theorem 1 are satisfied and (27) describes the dynamics of the log-productivity distribution in the limit of N being large for any $t \in [0, T]$. □

Proof of Proposition 2. Observe that in the case of pure innovation the log productivity $a_i(t) = \log A_i(t)$ of firm i grows according to (2), from which we get $a_i(t) = a_i(0) + \sum_{j=1}^{t} \vartheta(t_j)$, where $t_j \geq 0$ denotes the time at which the jth innovation arrives. Assuming that the random variables $\vartheta(t)$ are independent and identically distributed with finite mean $\mu_\vartheta < \infty$ and variance $\sigma_\vartheta^2 < \infty$, then by virtue of the central limit theorem, $\sum_{j=1}^{t} \vartheta(t_j)$ converges to a normal distribution $\mathcal{N}(\mu_\vartheta t, \sigma_\vartheta^2 t)$. Consequently, $A_i(t)$ converges to a log-normal distribution with mean $\mu_A = e^{t\mu_\vartheta + (1/2)t\sigma_\vartheta^2}$ and variance $\sigma_A^2 = (e^{t\sigma_\vartheta^2} - 1)e^{2t\mu_\vartheta + t\sigma_\vartheta^2}$. Setting $\eta_0 = 1 - p$, $\eta_1 = p$, and $\eta_b = 0$ for $b = 2, 3, \ldots$, and noting that $\mu_\vartheta = p$ and $\sigma_\vartheta^2 = p(1 - p)$ yields the desired proposition. □

[34]See also Roth and Sandholm (2013).

[35]The set $\bar{V}(P)$ is upper semicontinuous if for any $P \in \mathbb{R}^{|S|}$ and any open set O containing $\bar{V}(P)$, there exists a neighborhood N of P such that $\bar{V}(N) \in O$.

Proof of Proposition 3. Inserting (8) into the differential equation (9) and summing 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

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

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 first part of the proposition follows.

Next, 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 $a \geq a_{\mathrm{m}}$. From (10) we see that for all $a \geq a_{\mathrm{m}}$ it must hold that $\partial F_a(t)/\partial t = 0$ and so $F_a(t) = 1$ for all $t \geq 0$. In contrast, for all $a < a_{\mathrm{m}}$ 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

$$\lim_{t\to\infty} F_a(t) = \begin{cases} 0 & \text{if } a < a_{\mathrm{m}} \\ 1 & \text{if } a \geq a_{\mathrm{m}}. \end{cases}$$

This concludes the proof of the proposition. □

Proof of Proposition 4. We see from the definition of the imitation indicator function in (12) that $\chi^{\mathrm{im}}(a, P(t)) = 1$ is equivalent to $a_{\mathrm{im}}(a, P) > a_{\mathrm{in}}(a)$. This can be written as

$$\underbrace{a + \log(1 - p + \bar{A}p)}_{\text{innovation}} \leq \underbrace{a + \log\left(F_a(t) + \sum_{b=a+1}^{\infty} e^{b-a} P_b(t)\right)}_{\text{imitation}}.$$

Rearranging terms yields

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

or, equivalently,

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

That is,

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

The existence of a threshold a^* such that $\chi^{\text{im}}(a, P(t)) = 1$ for all $a \leq a^*$ and $\chi^{\text{im}}(a, P(t)) = 0$ for all $a > a^*$ can then be written as

$$\sum_{b=a+1}^{\infty}(e^{b-a}-1)P_b(t) \begin{cases} \geq p(\bar{A}-1) & \text{if } a \leq a^* \\ < p(\bar{A}-1) & \text{if } a > a^*. \end{cases} \tag{28}$$

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

$$f(a,t) \equiv \sum_{b=a+1}^{\infty}(e^{b-a}-1)P_b(t). \tag{29}$$

The function $f(a,t)$ is strictly monotonically decreasing if $f(a-1,t) - f(a,t) = (e-1)P_a(t) > 0$. This holds for all a in the support $\mathcal{S}$ of $P_a(t)$, where $P_a(t) > 0$. Hence, if at time t for all $a \in \mathcal{S}$ we have that $P_a(t) > 0$, then there exists 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

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

Alternatively, 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 monotonically decreasing. For $\Delta t \to 0$ we then obtain the corresponding result in continuous time. □

Remark 1. Assume that we can extend $P_a(t)$ to real-valued a, which is identical to $P_a(t)$ at the discrete $a \in \mathcal{S}$, but allows $P_a(t)$ to be evaluated at $a \in \mathbb{R}$, using the same functional form of $P_a(t)$ also for real values of a. Then at all points of continuity of $f(a,t) \equiv \sum_{b=a+1}^{\infty}(e^{b-a}-1)P_b(t)$ we can identify a threshold log productivity $a^*(t) \in \mathbb{R}$ satisfying

$$f(a^*(t),t) = \sum_{b=a^*(t)+1}^{\infty}(e^{b-a^*(t)}-1)P_b(t) = p(\bar{A}-1), \tag{30}$$

that is, evaluated at $a = a^*(t)$, the inequality in (28) becomes an equality (see also Figure 8). At the points of discontinuity of $f(a,t) = \sum_{b=a+1}^{\infty}(e^{b-a}-1)P_b(t)$, the threshold

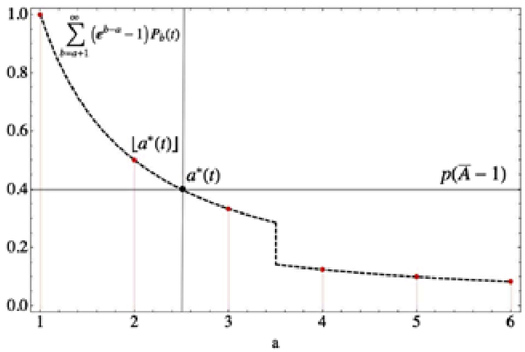

Figure 8. An illustration of the monotonically decreasing function $f(a,t) \equiv \sum_{b=a^*(t)+1}^{\infty}(e^{b-a^*(t)}-1)P_b(t)$ of (29) in the proof of Proposition 4, where its continuous extension is shown with a dashed line while the function values at the discrete values $a \in \mathcal{S}$ are indicated with vertical lines.

condition becomes

$$a^*(t) = \max\left\{a \in \mathbb{R}_{\geq 1} : \sum_{b=a+1}^{\infty}(e^{b-a}-1)P_b(t) \geq p(\bar{A}-1)\right\}. \tag{31}$$

Because $f(a,t)$ is monotonically decreasing, and the original function and its extension on continuous a evaluated at the discrete values of a are always identical, it must hold that the largest discrete value of a such that $f(a,t) \geq p(\bar{A}-1)$ from (28) must be equivalent to $\lfloor a^*(t) \rfloor$, where $a^*(t)$ is obtained from (30) for all continuity points of $f(a,t)$ and from (31) for all discontinuity points of $f(a,t)$. This observation will be useful for the proof of Proposition 6.

Proof of Proposition 5. From (15) we find that in the limit of $\beta \to \infty$ the evolution of the log-productivity distribution can be written as

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

where we have omitted the dependency on P in $a^*(P)$ to simplify the notation. For the dynamics of the cumulative log-productivity distribution $F_a(t) = \sum_{b=1}^{a} P_a(t)$ we then get, for $a \leq a^*$,

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

where in the last line we have used the results obtained in Proposition 3. 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

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

Putting the above results together we can write

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

Note that for all $a \ge 1$ and $t \ge 0$ we have that $\partial F_a(t)/\partial t \le 0$. Finally, note that from the above equation it follows that the dynamics of the complementary cumulative distribution function (cdf), $G_a(t) = 1 - F_a(t)$, is given by

$$\frac{\partial G_a(t)}{\partial t} = \begin{cases} -(G_a(t)^2 - G_a(t)) & \text{if } a \le a^*\\ (1 - G_{a^*}(t))G_a(t) - p(G_a(t) - G_{a-1}(t)) & \text{if } a > a^*. \end{cases} \qquad \square$$

Before proceeding with the proof of Proposition 6, the following lemma will be useful.[36]

Lemma 3. *Consider the delay differential equations $g'(x) = G(x, g(x), g(x-1))$ and $f'(x) = F(x, f(x), f(x-1))$ for $x > -1$, with identical preshape functions $g(x) = f(x) = \phi(x)$ for $x \in [-1, 0]$ and F being a continuous function satisfying a Lipschitz condition with respect to f. If $G \le F$, then $g(x) \le f(x)$. Analogously, if $G \ge F$, then $g(x) \ge f(x)$.*

Proof. We proceed by the "method of steps" (Smith 2010, Section 3). For $x \in [0, 1)$, both $g(x)$ and $f(x)$ must satisfy the ODEs

$$g'(x) = G(x, g(x), \phi(x-1)) \tag{32}$$

[36]A similar result can be found in Theorem 3.6 in Smith (2010).

and

$$f'(x) = F(x, f(x), \phi(x-1)). \tag{33}$$

By the "comparison lemma" (see Theorem 3.2 in Waltman 2004 or Lemma 3.4 in Khalil 2002) for ordinary differential equations (ODEs) it follows from the fact that $G \leq F$ and that, by assumption, F is a continuous function satisfying a Lipschitz condition with respect to f, that on the interval $[0, 1)$ we must have that $f(x) \geq g(x)$. We may repeat the above argument to extend the inequality still further to the right. Indeed, for $1 \leq x < 2$, $g(x)$ must satisfy the ODE

$$g'(x) = G(x, g(x), g(x-1)),$$

where $g(x-1)$ in the interval $[1, 2)$ is the predetermined solution of the ODE (32), and $f(x)$ must satisfy the ODE

$$f'(x) = F(x, g(x), g(x-1)),$$

where $f(x-1)$ in the interval $[1, 2)$ is the predetermined solution of the ODE (33). Similarly, by the comparison lemma for ODEs we then must have that $f(x) \geq g(x)$ for $x \in [1, 2)$. We then can repeat this argument to establish the inequality $f(x) \geq g(x)$ for all $x > -1$. A similar reasoning can be applied to the case of $F \leq G$ showing that $f(x) \leq g(x)$ for all $x > -1$. □

We are now able to prove Proposition 6.

Proof of Proposition 6. In the following discussion, we show that the stationary log-productivity distribution $F_a(t)$ is a traveling wave, $f(a - a^*(t))$ with $a^*(t) = a_0^* - \nu t$, consistent with Definition 1.[37] Note that this is equivalent to assuming that the complementary distribution, $G_a(t) = 1 - F_a(t)$, has a traveling wave form $g(a - a^*(t)) = g(a - a_0^* - \nu t) = 1 - f(a - a_0^* - \nu t)$. We then proceed by showing that there exists a $p^* > 0$ such that for $p < p^*$, the distribution has asymptotic exponential tails. Note that as the function $f(\cdot)$ takes real-valued arguments, it can be thought of as an underlying continuous distribution such that at each date t and for each $a \in \mathcal{S}$, the fraction of firms with probability less than or equal to a at date t, denoted $F_a(t)$, is equal to $f(a - a_0^* - \nu t)$ for some constant ν.[38]

We first check that a traveling wave satisfies the threshold condition of Proposition 4. By definition, for the threshold log productivity $a^*(t)$ (possibly real-valued) it must hold that the expected productivity gains from innovation are equal to the expected productivity gains from imitation at all continuity points of the distribution, which is equivalent to (see (30) in Remark 1)

$$F_{a^*(t)}(t) + \sum_{b=a^*(t)+1}^{\infty} e^{b-a^*(t)} P_b(t) = 1 + p(\bar{A} - 1). \tag{34}$$

[37] The constant a_0^* in the argument of $f(a - a_0^* - \nu t)$ does not change its dependency on $a - \nu t$ characterizing a traveling wave.

[38] We would like to thank the co-editor for pointing this out.

We now show that if the cdf $F_a(t)$ has a traveling wave form $f(a - a^*(t))$ and the threshold log productivity $a^*(t)$ grows linearly with t, i.e., $a^*(t) = a_0^* + \nu t$, for an appropriate traveling wave velocity ν, then the threshold condition in (34) is always satisfied. Time invariance of the left-hand side of (34) requires that[39]

$$F_{a^*(t+1)}(t+1) + \sum_{b=a^*(t+1)+1}^{\infty} e^{b-a^*(t+1)} P_b(t+1) = F_{a^*(t)}(t) + \sum_{b=a^*(t)+1}^{\infty} e^{b-a^*(t)} P_b(t).$$

With our guess for the traveling wave we have that $F_{a^*(t)}(t) = f(0) = F_{a^*(t+1)}(t+1)$. Hence, what remains to be shown is that

$$e^{-a^*(t+1)} \sum_{b=a^*(t+1)+1}^{\infty} e^b P_b(t+1) = e^{-a^*(t)} \sum_{b=a^*(t)+1}^{\infty} e^b P_b(t).$$

We then have that

$$\begin{aligned}
e^{-a^*(t+1)} & \sum_{b=a^*(t+1)+1}^{\infty} e^b P_b(t+1) \\
&= e^{-a^*(t)-\nu} \sum_{b=a^*(t)+\nu+1}^{\infty} e^b \big(F_b(t+1) - F_{b-1}(t+1)\big) \\
&= e^{-a^*(t)-\nu} \sum_{b=a^*(t)+\nu+1}^{\infty} e^b \big(f(b - a^*(t) - \nu) - f(b - 1 - a^*(t) - \nu)\big) \\
&= e^{-a^*(t)-\nu} \sum_{b=a^*(t)+1}^{\infty} e^{b+\nu} \big(f(b - a^*(t)) - f(b - a^*(t) - 1)\big) \\
&= e^{-a^*(t)-\nu} \sum_{b=a^*(t)+1}^{\infty} e^{b+\nu} \big(F_b(t) - F_{b-1}(t)\big) \\
&= e^{-a^*(t)} \sum_{b=a^*(t)+1}^{\infty} e^b P_b(t),
\end{aligned}$$

and the equality follows. Hence, we have shown that a threshold $a^*(t)$ grows linearly with t as $a^*(t) = a_0^* + \nu t$ and the assumption of a traveling wave is consistent with the threshold condition.

In the following discussion, we show that there exists a solution of the traveling wave form $g(a - a^*(t))$ to (16) and (34) (or equivalently, (28)) with $a^*(t) = a^*(0) + \nu t$ by analyzing the solution of (16) for both cases of the log productivity a above and below the threshold $a^*(t)$. We then proceed by showing that the stationary distribution has exponential tails.

Case 1: $a \le a^*(t)$. We assume that the log-productivity distribution for values of a below the threshold $a^*(t)$ has a traveling wave form. Inserting $g(a - a^*(t)) = G_a(t)$ into

[39] Without loss of generality (w.l.o.g.) we consider a time increment $\Delta t = 1$.

(16), where $a^*(t) = \nu t + a_0^*$, and denoting $x = a - a^*(t) = a - a_0^* - \nu t$ then gives for $x \leq 0$ (corresponding to $a \leq a^*(t)$) that

$$-\nu g'(x) = g(x) - g(x)^2$$

or, equivalently, the logistic differential equation

$$g'(x) = -\frac{1}{\nu}(g(x) - g(x)^2).$$

The standard solution of this logistic differential equation is given by

$$g(x) = \frac{1}{1 + \left(\frac{1}{g_0} - 1\right)e^{x/\nu}}, \tag{35}$$

with the boundary condition $g_0 = g(0)$. Thus, we have that $\lim_{x\to-\infty} g(x) = 1$. In particular, for $x \to -\infty$ we have that $g(x) \sim e^{-x/\nu}$ and the solution decays exponentially. Now (35) establishes (17) as

$$G_a(t) = g(a - a^*(t)) = \frac{1}{1 + \left(\frac{1}{g_0} - 1\right)e^{(a-a_0^*-\nu t)/\nu}}.$$

We then have that $P_a(t) = G_{a-1}(t) - G_a(t) \sim e^{(a-\nu t)/\nu}$, which is equivalent to writing $P_a(t) = e^{(a-\nu t)/\nu} + o(1)$ for a much smaller than $a^*(t) = \nu t + a_0^*$, and we have shown the first part of (19).

Case 2: $a > a^*(t)$. In the following discussion we focus on the case of $a > a^*(t)$ and assume that the threshold $a^*(t)$ grows linearly with t, that is, $a^*(t) = a_0^* + \nu t$. Moreover, we assume that $G_a(t) = g(a - a^*(t))$. Substituting $x \equiv a - a^*(t) = a - a_0^* - \nu t$ in (16) for $a > a^*(t)$ and noting that[40]

$$\begin{aligned} G_{\lfloor a^*(t)\rfloor}(t) &= g(\lfloor a^*(t)\rfloor - a^*(t)) \\ &= g(\lfloor a_0^* + \nu t\rfloor - (a_0^* + \nu t)) \\ &= g(\lfloor a - x\rfloor - (a - x)) \\ &= g(x + \lfloor -x\rfloor) \end{aligned}$$

for any integer a, by introducing g_0 from above as a constant we then get

$$\begin{aligned} -\nu g'(x) &= \big(1 - g(x + \lfloor -x\rfloor)\big)g(x) - p(g(x) - g(x-1)) \\ &= (1 - g_0)g(x) - p(g(x) - g(x-1)) - \big(g(x + \lfloor -x\rfloor) - g_0\big)g(x) \\ &= (1 - g_0)g(x) - p(g(x) - g(x-1)) - \varepsilon(x)g(x) \end{aligned} \tag{36}$$

for $x > 0$, $x + \lfloor -x\rfloor \in [-1, 0]$, where we have used the fact that $\partial G_a(t)/\partial t = -\nu g'(x)$ and we have denoted

$$\varepsilon(x) \equiv g(x + \lfloor -x\rfloor) - g_0. \tag{37}$$

[40]We would like to thank an anonymous referee for pointing this out.

Next note that due to the monotonicity of $g(x)$ we have that $\varepsilon(x) \geq 0$. Further note that the DDE (36) depends on values of the function $g(x)$ in the interval $x \in [-1, 0]$, which is given by (35) and is thus predetermined for computing the solution of (36). Rearranging terms, we can write (36) in the form

$$g'(x) + \frac{1 - g_0 - p}{\nu} g(x) + \frac{p}{\nu} g(x-1) = \frac{\varepsilon(x)}{\nu} g(x). \tag{38}$$

Denoting $a \equiv (1 - g_0 - p)/\nu$ and $b \equiv p/\nu$, the solution of (38) can be written as the solution of the integral equation (cf. Bellman and Cooke 1963, Eq. (9.3.2), p. 267)

$$g(x) = g_0 h(x) - b \int_{-1}^{0} h(x - y - 1)\phi(y)\, dy + \frac{1}{\nu} \int_0^x h(x - y)\varepsilon(y) g(y)\, dy \tag{39}$$

for $x > 0$ and $h(x)$ being the solution to the homogeneous part of the DDE (38), i.e., where the right-hand side is set to zero,[41] and $\phi(x)$ is the predetermined solution for $g(x)$ in the interval $x \in [-1, 0]$ from (35). For any $x > 0$, $g(x)$ in the left-hand side of (39) is determined by $g(y)$ for values of $y < x$. So recursively, (39) completely specifies g at any point x as a function of g evaluated at points y smaller than x. This shows existence of the solution. A more detailed discussion can be found in Section 9 in Bellman and Cooke (1963) and the method of steps introduced in Section 3 in Smith (2010), where the existence of solutions to DDEs is proven in a recursive manner. The existence of such a solution to the DDE (38) thus justifies our assumption of a traveling wave.

Hence, we have shown that there exists a solution to (16) and (34) (or, equivalently, (28)) with $a^*(t) = a^*(0) + \nu t$ for some constant ν, where we set $G_a(t)$ equal to $g(a - a^*(t)) = g(a - a_0^* - \nu t)$ for any $a \in \mathcal{S}$. This justifies our assumption of a traveling wave.[42]

In what follows we derive upper and lower bounds for the solution of (36), and from these bounds we analyze its asymptotic behavior in the limit of large x. In particular we will show that there exists a $p^* > 0$ such that for all $p < p^*$, the tail of $g(x)$ can be bounded from above and from below by exponentially decaying functions.[43]

Let us denote

$$\varepsilon \equiv \sup_{x \geq 0} \varepsilon(x) = \sup_{x \geq 0} \bigl\{ g(x + \lfloor -x \rfloor) - g_0 \bigr\} = \sup_{y \in [-1, 0]} \{ g(y) - g_0 \} = g(-1) - g_0, \tag{40}$$

and define $\overline{g}(x)$ as the solution to the delay differential equation

$$\overline{g}'(x) + \frac{1 - g_0 - p}{\nu} \overline{g}(x) + \frac{p}{\nu} \overline{g}(x-1) = \frac{\varepsilon}{\nu} \overline{g}(x). \tag{41}$$

[41] This solution is analyzed in (42) below.

[42] Observe that while $\varepsilon(x)$ is only piecewise continuous, $g(x)$ in (39) is continuous in x as the last term in (39) is an integral over a piecewise continuous function, which is continuous (cf., e.g., Shilov 1996, paragraph 9.39). As also the logistic function in (35) is continuous, we obtain that $g(x)$ is continuous for all x. Consequently, $P_b(t) = G_{b-1}(t) - G_b(t)$ is continuous, and $f(a, t) = \sum_{b=a+1}^{\infty} (e^{b-a} - 1) P_b(t)$ in Remark 1 is continuous in a, as it is the composition of continuous functions.

[43] Observe that this rules out, for example, any polynomially decaying functions.

By virtue of Lemma 3, the solution $\overline{g}(x)$ of (41) then is an upper bound to the solution $g(x)$ of (38).[44] Next, Lemma 3 implies that $\underline{g}(x) \leq g(x)$, where $\underline{g}(x)$ solves the delay differential equation

$$\underline{g}'(x) + \frac{1-g_0-p}{\nu}\underline{g}(x) + \frac{p}{\nu}\underline{g}(x-1) = 0. \tag{42}$$

Note that both (41) and (42) are instances of a first-order linear homogeneous delay differential equation (DDE) with constant coefficients (cf. Bellman and Cooke 1963, Driver 1977, Smith 2010). In the following text we first solve (41), while (42) can be solved in an analogous way.

Recall that the DDE (41) depends on values of the function $g(x)$ in the interval $x \in [-1, 0]$, which is given by (35) and thus is predetermined. Asl and Ulsoy (2003) call this the *preshape function*, which we have denoted by $\phi(x)$. Inserting the definition of ϵ from (40) into (41) we then have to solve the DDE

$$\begin{aligned} &\overline{g}'(x) + \frac{1-p-g(-1)}{\nu}\overline{g}(x) + \frac{p}{\nu}\overline{g}(x-1) = 0, \quad x \in (0, \infty) \\ &\overline{g}(x) = \phi(x) = \frac{1}{1+(\frac{1}{g_0}-1)e^{\frac{x}{\nu}}}, \quad x \in [-1, 0]. \end{aligned} \tag{43}$$

Asl and Ulsoy (2003) have shown that such a DDE admits a solution of the form[45]

$$\overline{g}(x) = \sum_{k=-\infty}^{\infty} \overline{c}_k e^{-\overline{\lambda}_k x}, \tag{44}$$

with appropriate constants $\overline{c}_k$. That is, the solution to the DDE in (43) is a linear combination of exponential functions. We have that $\overline{g}'(x) = -\sum_{k=-\infty}^{\infty} \overline{c}_k \lambda_k e^{-\overline{\lambda}_k x}$, and inserting into the DDE (43) yields

$$\nu \sum_{k=-\infty}^{\infty} \overline{c}_k \overline{\lambda}_k e^{-\overline{\lambda}_k x} = (1-g(-1)-p) \sum_{k=-\infty}^{\infty} \overline{c}_k e^{-\overline{\lambda}_k x} + p \sum_{k=-\infty}^{\infty} \overline{c}_k e^{-\overline{\lambda}_k (x-1)}.$$

This can be written as

$$\sum_{k=-\infty}^{\infty} \overline{c}_k e^{-\overline{\lambda}_k x}(\overline{\lambda}_k \nu - (1-g(-1)-p) - p e^{\overline{\lambda}_k}) = 0. \tag{45}$$

The coefficients $\overline{\lambda}_k$ in (45) are the roots of the characteristic equation (cf. Asl and Ulsoy 2003)

$$\overline{\lambda}_k \nu = 1 - g(-1) - p(1-e^{\overline{\lambda}_k}). \tag{46}$$

[44] In particular, we can write $\overline{G}(x, \overline{g}(x), \overline{g}(x-1)) \equiv \overline{g}'(x) = -((1-g_0-p)/\nu)\overline{g}(x) - (p/\nu)\overline{g}(x-1) + (\varepsilon/\nu)\overline{g}(x)$ and $G(x, g(x), g(x-1)) \equiv g'(x) = -((1-g_0-p)/\nu)g(x) - (p/\nu)g(x-1) + (\varepsilon(x)/\nu)g(x)$. Because $\varepsilon \geq \varepsilon(x)$ we must have that $\overline{G} \geq G$. Moreover, we have that $\overline{G}$ is continuous and linear in g, and hence Lipschitz in g. It follows that Lemma 3 applies.

[45] See in particular (3) and (15) in Asl and Ulsoy (2003).

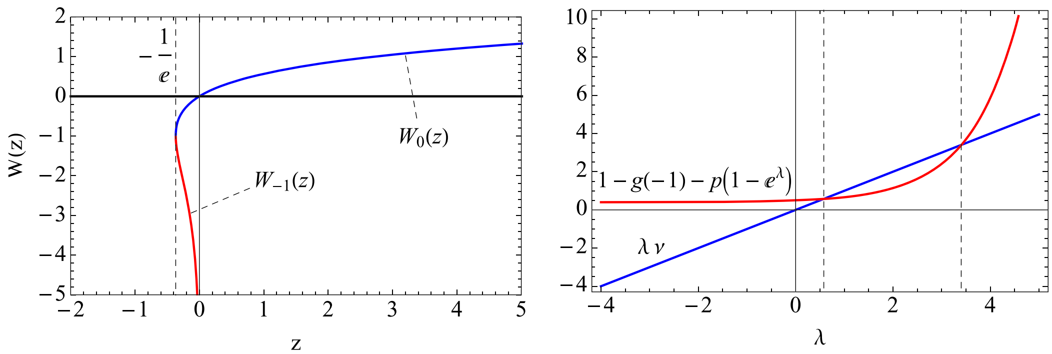

Figure 9. Left panel: The two real Lambert W functions. Right panel: The two real roots $\overline{\lambda}_0$ and $\overline{\lambda}_1$ solving (46) for $\nu = 1$, $g(-1) = 0.5$, and $p = 0.1$ indicated with dashed lines, and given by $\overline{\lambda}_0 = 0.5783$ and $\overline{\lambda}_{-1} = 3.4018$.

The roots of (46) can be written in closed form as

$$\overline{\lambda}_k = \frac{g(-1)+p-1}{\nu} + W_k\left(-\frac{p}{\nu}e^{-(g(-1)+p-1)/\nu}\right), \tag{47}$$

where $W_k(z)$ is the kth branch of the Lambert W function satisfying $W_k(z)e^{W_k(z)} = z$ for $k = 0, \pm 1, \pm 2, \ldots$ (cf. Corless et al. 1996). Note that there can be at most two real roots $W_0(z)$ and $W_{-1}(z)$. An illustration is given in the left panel of Figure 9.[46] The real parts of the higher order roots are dominated by those of $W_0(z)$ and $W_{-1}(z)$ (Asl and Ulsoy 2003). As $(p/\nu)e^{-(g(-1)+p-1)/\nu} \geq 0$, there exist either two real roots or we have the case that both coincide, namely when the argument $-(p/\nu)e^{-(g(-1)+p-1)/\nu}$ of the Lambert function in (47) equals $-\frac{1}{e}$ and when $\overline{\lambda}_0 = \overline{\lambda}_{-1} = (g(-1)+p-1-\nu)/\nu$. The further the two roots are separated from each other, the closer is the argument $-(p/\nu)e^{-(g(-1)+p-1)/\nu}$ of the Lambert function to zero (see Figure 9, left panel), which is the case for example when the innovation success probability p is small. Moreover, the existence of real roots requires that $(p/\nu)e^{-(g(-1)+p-1)/\nu} \leq 1/e$ or, equivalently, $p/\nu \leq e^{-(1-g(-1))/\nu}e^{p/\nu-1}$. An illustration for the two real roots $\overline{\lambda}_0$ and $\overline{\lambda}_{-1}$ solving (46) is shown in the right panel of Figure 9.

We next show that all the roots of the characteristic equation (46) have positive real parts. Corollary 4.10 in Smith (2010, p. 56)[47] shows that a sufficient condition for all roots x of the equation $x - b - ce^x = 0$ to have a positive real part is $b > 0$ and $|b| > |c|$. We can write (46) as $\overline{\lambda}_k - (1-p-g(-1))/\nu - (p/\nu)e^{\overline{\lambda}_k} = 0$, so that the corresponding coefficients are $b = (1-p-g(-1))/\nu$ and $c = p/\nu$. The sufficient condition then becomes $1 - p - g(-1) > p$ or, equivalently, $\frac{1}{2}(1-g(-1)) > p$. First, assume that $g_0 < 1$. Because $g(-1)$ is determined by the logistic function in (35), which is strictly smaller than 1 if $g_0 < 1$, we

[46] Note further that $|W_0(z)| \leq |W_{-1}(z)|$ while the bounds $\ln z - \ln\ln z \leq W_0(z) \leq \ln z - \frac{1}{2}\ln\ln z$ for every $z \geq e$ and $1 < -W_{-1}(z) \leq -1/z$ for every $z \in (-1/e, 0)$ hold.

[47] In particular, part (i) of Corollary 4.10 in Smith (2010) considers the equation $y + b + ce^{-ry} = 0$, with b, c being real coefficients and $r > 0$. Then, if $b > 0$ and $|b| > |c|$, all the roots have negative real parts for all $r \geq 0$. Substituting $x = -y$ and setting $r = 1$ gives $x - b - ce^y = 0$, which is the equation that we consider. Finally, note that if all the roots y have negative real parts, then all the roots $x = -y$ must have positive real parts.

have that $\frac{1}{2}(1-g(-1))>0$. Let $p^*>0$ be the smallest possible value of $\frac{1}{2}(1-g(-1))$. We then can always find a (real-valued) p between p^* and 0 such that the inequality holds for all p less than p^*. Next assume that $g_0=1$. From the logistic function in (35) we know that $g_0=1$ implies that also $g(-1)=1$. Moreover, from (40) we can conclude that $\varepsilon=0$. In this case the solutions to the upper and lower bounds in (41) and (42) coincide, and must be equivalent to the solution to the original equation (36), which is uniformly bounded by 1 as it is a complementary cumulative distribution function. Thus there cannot be any positive real parts in the characteristic roots. This shows that all the roots $\overline{\lambda}_k$ of the characteristic equation (46) have positive real parts for p small enough.[48]

The coefficients $\overline{c}_k$ in (44) follow from the preshape function, which can be written as (see (77) in Asl and Ulsoy 2003)[49]

$$\phi(x)\equiv\frac{1}{1+\left(\frac{1}{g_0}-1\right)e^{x/\nu}}=\sum_{k=-\infty}^{\infty}\overline{c}_k e^{-\overline{\lambda}_k x},\quad x\in[-1,0]. \tag{48}$$

So as to compute the Lambert coefficients, $\overline{c}_k$, consider a $2K+1$ discretization

$$\left\{-1,-\frac{2K-1}{2K},-\frac{2K-2}{2K},\dots,-\frac{2}{2K},-\frac{1}{2K},0\right\}$$

of the interval $[-1,0]$. Taking into account only $2K+1$ Lambert coefficients in (48), such that

$$\phi(x)\approx\sum_{k=-K}^{K}\overline{c}_k e^{-\overline{\lambda}_k x},\quad x\in[-1,0],$$

we get

$$\underbrace{\begin{bmatrix}\phi(0)\\ \phi(-\frac{1}{2K})\\ \phi(-\frac{2}{2K})\\ \vdots\\ \phi(-\frac{2K-2}{2K})\\ \phi(-\frac{2K-1}{2K})\\ \phi(-1)\end{bmatrix}}_{\boldsymbol{\phi}}\approx\underbrace{\begin{bmatrix}e^{-\overline{\lambda}_{-K}\cdot 0} & \dots & e^{-\overline{\lambda}_{K}\cdot 0}\\ e^{-\overline{\lambda}_{-K}\cdot(-1/(2K))} & \dots & e^{-\overline{\lambda}_{K}\cdot(-1/(2K))}\\ e^{-\overline{\lambda}_{-K}\cdot(-2/(2K))} & \dots & e^{-\overline{\lambda}_{K}\cdot(-2/(2K))}\\ \vdots & & \\ e^{-\overline{\lambda}_{-K}\cdot(-(2K-2)/(2K))} & \dots & e^{-\overline{\lambda}_{K}\cdot(-(2K-2)/(2K))}\\ e^{-\overline{\lambda}_{-K}\cdot(-(2K-1)/(2K))} & \dots & e^{-\overline{\lambda}_{K}\cdot(-(2K-1)/(2K))}\\ e^{-\overline{\lambda}_{-K}\cdot(-1)} & \dots & e^{-\overline{\lambda}_{K}\cdot(-1)}\end{bmatrix}}_{\boldsymbol{\Omega}_K}\times\underbrace{\begin{bmatrix}\overline{c}_{-K}\\ \overline{c}_{-K+1}\\ \overline{c}_{-K+2}\\ \vdots\\ \overline{c}_{K-2}\\ \overline{c}_{K-1}\\ \overline{c}_{K}\end{bmatrix}}_{\overline{\mathbf{c}}}.$$

We then have that $\overline{\mathbf{c}}\approx\boldsymbol{\Omega}_K^{-1}\boldsymbol{\phi}$, which becomes exact in the limit of $K\to\infty$, and the Lambert coefficients $\overline{c}_k$ are given by

$$\overline{c}_k=\lim_{K\to\infty}(\boldsymbol{\Omega}_K^{-1}\boldsymbol{\phi})_k.$$

[48]In our numerical simulations we find that this condition actually holds for any value of p that we have considered.

[49]Any continuous function $\phi(x)$ can be represented as an infinite series using the Lambert coefficients, $\overline{c}_k$, and the Lambert modes, $e^{-\overline{\lambda}_k x}$ (Asl and Ulsoy 2003).

Note that for large x the dominant term in (44) is the one with the smallest exponent, so that asymptotically it holds that[50]

$$\overline{g}(x) \sim e^{-\overline{\lambda}_0 x}, \quad x \to \infty,$$

where $\overline{\lambda}_0$ is the smallest root of the characteristic (46).

Similarly, the lower bound from the solution of the DDE (41) is given by

$$\underline{g}(x) = \sum_{k=-\infty}^{\infty} \underline{c}_k e^{-\underline{\lambda}_k x}, \tag{49}$$

with appropriate constants $\underline{c}_k$, where the exponents $\underline{\lambda}_k$ solve the characteristic equation

$$\underline{\lambda}_k = \frac{g_0 + p - 1}{\nu} + W_k\left(-\frac{p}{\nu} e^{-(g_0+p-1)/\nu}\right). \tag{50}$$

Hence, we have that $\underline{g}(x) \leq g(x) \leq \overline{g}(x)$, and we have shown (18). Observe further that $\underline{g}(x) - \overline{g}(x) \sim e^{-\underline{\lambda}_0 x} - e^{-\overline{\lambda}_0 x} \sim e^{-\overline{\lambda}_0 x} \to 0$ for large x when $\underline{\lambda}_0 > \overline{\lambda}_0$.[51] Moreover, we have that $g(x) = O(e^{-\overline{\lambda}_0 x})$ for large x,[52] so that we can write $G_a(t) = O(e^{-\overline{\lambda}_0(a-\nu t)})$ as a becomes much larger than $a^*(t)$. As $P_a(t) = G_{a-1}(t) - G_a(t)$, the same asymptotic behavior holds for $P_a(t)$. This proves the second part of (19). □

Remark 2. In our numerical simulations we find that the perturbation $\varepsilon(x)$ in (37) is typically small and can be neglected to obtain a fairly good approximation. A comparison of the numerical solution of (16) with the analytical predictions from (35) below the threshold and the solution of the DDE (42) above the threshold, together with the solution of (49) with exponents from (50) for $K = 3$ Lambert modes and the exponent $\underline{\lambda}_0$ obtained from (20), are shown in Figure 10. The figure shows fairly good agreement between the theoretical predictions and a direct numerical integration of (16).

Remark 3. In the following discussion we show how (20) and (21) are computed. Motivated by Remark 2, we assume that the perturbation $\varepsilon(x)$ can be neglected, so that the solution to the original DDE (38) is sufficiently well approximated by the solution to the DDE (42). Observe that the exponents $\underline{\lambda}_k$ in (50) depend on the endogenous variables ν and g_0, and so (50) cannot be used to compute $\underline{\lambda}_k$ directly. In the following text we avoid this problem by assuming that the solution to the DDE (42), given in (49), is dominated by the smallest exponent $\underline{\lambda}_0$ (corresponding to the term with the smallest decay as x increases), and then we proceed by computing this exponent.

First note that from the threshold condition in (34) we obtain

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

[50] Recall that we have shown above that all λ_k have positive real parts.

[51] Because $\underline{g}(x) \leq \overline{g}(x)$ we must have that $e^{-\underline{\lambda}_0 x} \leq e^{-\overline{\lambda}_0 x}$ for large x, implying that $\underline{\lambda}_0 > \overline{\lambda}_0$.

[52] This is because $\lim_{x\to\infty} g(x)/\overline{g}(x) \leq 1$. See also the definition in footnote 18.

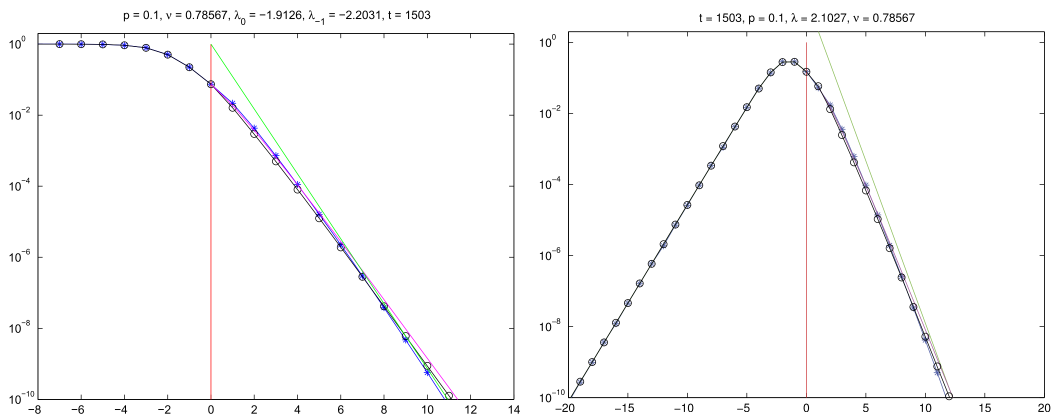

FIGURE 10. The stable shape of the complementary cumulative distribution function $G_a(t)$ for $p = 0.1$ (left panel) and the corresponding probability mass function $P_a(t)$ (right panel). The traveling wave has been detrended such that $\lfloor a^*(t) \rfloor$ coincides with the origin. (Recall that the artwork is in color in the online version.) The red vertical line indicates the threshold $\lfloor a^*(t) \rfloor$. The blue stars indicate the numerical solution of (16). The black line for values below the threshold is computed with the analytical solution from (35), where $G_{\lfloor a^*(t) \rfloor}$ is taken from the numerical solution of (16). The black line for values above the threshold is obtained from a numerical integration of the DDE (42) (using Matlab's `dde23` solver) with the preshape function from (35). The magenta line indicates the solution for values above the threshold obtained from (49) with exponents from (50) and $K = 3$ Lambert modes. The green line indicates the exponent $\underline{\lambda}_0$ obtained from (20).

Using (49) we have that $G_a(t) = \sum_{k=-\infty}^{\infty} \underline{c}_k e^{-\underline{\lambda}_k (a-\nu t)}$, and we can write

$$
\begin{aligned}
P_a(t) &= G_{a-1}(t) - G_a(t) \\
&= \sum_{k=-\infty}^{\infty} \underline{c}_k (e^{\underline{\lambda}_k} - 1) e^{-\underline{\lambda}_k (a-\nu t)} \\
&= \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{-\underline{\lambda}_k (a-\nu t)},
\end{aligned}
$$

where we have denoted $\tilde{c}_k \equiv \underline{c}_k (e^{\underline{\lambda}_k} - 1)$. It then follows that

$$
\begin{aligned}
p(\bar{A} - 1) &= \sum_{b=1}^{\infty} (e^b - 1) \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{-\underline{\lambda}_k (b + a^*(t) - \nu t)} \\
&= \sum_{b=1}^{\infty} (e^b - 1) \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{-\underline{\lambda}_k (b + a_0^*)} \qquad (51) \\
&= \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{-\underline{\lambda}_k a_0^*} \sum_{b=1}^{\infty} (e^b - 1) e^{-\underline{\lambda}_k b}
\end{aligned}
$$

$$= \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{-\underline{\lambda}_k a_0^*} \left(\frac{1}{e^{\underline{\lambda}_k - 1} - 1} + \frac{1}{1 - e^{\bar{\lambda}_k}} \right).$$

As discussed in the proof of Proposition 6, the more the principal root $\underline{\lambda}_0$ is separated from the other roots, the closer is the argument $-(p/\nu)e^{-(g_0+p-1)/\nu}$ of the Lambert W function in (50) to zero. Then only the principal Lambert mode dominates in (51), and we can write

$$p(\bar{A} - 1) = \tilde{c}_0 e^{-\underline{\lambda}_0 a_0^*} \left(\frac{1}{e^{\underline{\lambda}_0 - 1} - 1} + \frac{1}{1 - e^{\underline{\lambda}_0}} \right) + o(1). \tag{52}$$

The complementary cumulative distribution function (ccdf) evaluated at the threshold can be written as

$$g_0 = G_{a^*(t)}(t) = \sum_{k=0}^{\infty} c_k e^{-\underline{\lambda}_k a_0^*} = \sum_{k=-\infty}^{\infty} \frac{\tilde{c}_k}{e^{\underline{\lambda}_k} - 1} e^{-\underline{\lambda}_k a_0^*}.$$

Similarly, when the principal Lambert mode dominates in the above equation we obtain

$$g_0 = \frac{\tilde{c}_0}{e^{\underline{\lambda}_0} - 1} e^{-\underline{\lambda}_0 a_0^*} + o(1),$$

so that (46) can be written as

$$\underline{\lambda}_0 \nu = 1 - \frac{\tilde{c}_0}{e^{\underline{\lambda}_0} - 1} e^{-\underline{\lambda}_0 a_0^*} - p(1 - e^{\underline{\lambda}_0}) + o(1). \tag{53}$$

Inserting $\tilde{c}_0 e^{-\underline{\lambda}_0 a_0^*}$ from (52) into (53) (and dropping terms of $o(1)$) then gives

$$\underline{\lambda}_0 \nu = 1 - \frac{p(\bar{A} - 1)}{e^{\underline{\lambda}_0} - 1} \left(\frac{1}{e^{\underline{\lambda}_0 - 1} - 1} + \frac{1}{1 - e^{\underline{\lambda}_0}} \right)^{-1} - p(1 - e^{\underline{\lambda}_0}).$$

Hence, simplifying this expression we obtain the traveling wave velocity ν as a function of the principal exponent $\underline{\lambda}_0$ given by

$$\nu = \frac{1}{\underline{\lambda}_0} \left(1 + p(e^{\underline{\lambda}_0} - 1) - \frac{p(\bar{A} - 1)(1 - e^{1 - \underline{\lambda}_0})}{e - 1} \right). \tag{54}$$

The traveling wave velocity ν as a function of $\underline{\lambda}_0$ for different values of p can be seen in Figure 11 (left panel). Of particular interest will be the smallest admissible value of ν.[53] Note that the right-hand side of (54) is a convex function of $\underline{\lambda}_0$, which is characterized by a unique global minimum (see also the left panel in Figure 11). The corresponding value of $\underline{\lambda}_0$ minimizing ν can be found from the corresponding first-order condition (FOC) given by

$$\frac{d\nu}{d\underline{\lambda}_0} = \frac{1 - e + p(\bar{A} + e - 2) + (e - 1)e^{\underline{\lambda}_0} p(\underline{\lambda}_0 - 1) - (\bar{A} - 1)e^{1 - \bar{\lambda}_0} p(1 + \underline{\lambda}_0)}{(e - 1)\underline{\lambda}_0^2} = 0.$$

[53] A generic selection principle applies, where an extremal value for ν is realized from sufficiently steep initial conditions (Van Saarloos 2003, Bramson 1983).

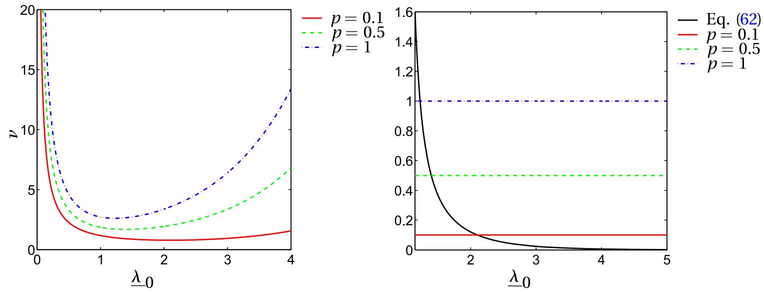

FIGURE 11. Left panel: The traveling wave velocity ν as a function of $\underline{\lambda}_0$ for different values of $p = 0.1$, $p = 0.5$, and $p = 1$. Right panel: The ν minimizing value of $\underline{\lambda}_0$ for the same values of p. The figures show that the minimizing value of $\underline{\lambda}_0$ 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.

The FOC from above is equivalent to

$$\frac{e-1}{\bar{A}+e-2+(e-1)e^{\underline{\lambda}_0}(\underline{\lambda}_0-1)-(\bar{A}-1)e^{1-\underline{\lambda}_0}(1+\underline{\lambda}_0)} = p,$$

which is illustrated in Figure 11 (right panel). This equation can be further simplified to

$$e^{\underline{\lambda}_0}(\underline{\lambda}_0-1)-\frac{\bar{A}-1}{e-1}e^{1-\underline{\lambda}_0}(1+\underline{\lambda}_0)+\frac{\bar{A}+e-2}{e-1} = \frac{1}{p}. \tag{55}$$

A comparison of the exponentially decaying solution with $\underline{\lambda}_0$ obtained from (55) and the numerical solution of (16) is shown in Figure 10.

PROOF OF LEMMA 1. Let T be the terminal period. Then we have that

$$\begin{aligned}
&V_{T-1}(A_i(T-1), P(T-1)) \\
&\quad = \max\left\{\int dF^{\text{in}}(\vartheta)\psi A_i(T-1)\bar{A}^{\vartheta}, \int dF^{\text{im}}(\vartheta|A_i(T-1), P(T-1))\psi A_i(T-1)\bar{A}^{\vartheta}\right\} \\
&\quad = \psi A_i(T-1)\max\left\{\int dF^{\text{in}}(\vartheta)\bar{A}^{\vartheta}, \int dF^{\text{im}}(\vartheta|A_i(T-1), P(T-1))\bar{A}^{\vartheta}\right\}.
\end{aligned}$$

Observe that the expected productivity gain from imitation, $\int dF^{\text{im}}(\vartheta|A_i(T-1), P(T-1))\bar{A}^{\vartheta}$, is increasing in $A_i(T-1)$, and the expected productivity gain from innovation, $\int dF^{\text{in}}(\vartheta)\bar{A}^{\vartheta}$, is nondecreasing in $A_i(T-1)$. Hence, $V_{T-1}(A_i(T-1), P(T-1))$ is increasing in $A_i(T-1)$. Next, as the induction hypothesis, assume that $V_{T-t}(A_i(T-t), P(T-t))$ is increasing in $A_i(T-t)$. Then we have that

$$\begin{aligned}
&V_{T-t-1}(A_i(T-t-1), P(T-t-1)) \\
&\quad = \max\left\{\int dF^{\text{in}}(\vartheta)\psi A_i(T-t-1)\bar{A}^{\vartheta} + \delta V_{T-t}(A_i(T-t-1)\bar{A}^{\vartheta}, P(T-t)),\right.
\end{aligned}$$

$$\int dF^{\mathrm{im}}(\vartheta|A_i(T-t-1), P(T-t-1))\psi A_i(T-t-1)\bar{A}^{\vartheta}$$
$$+ \delta V_{T-t}(A_i(T-t-1)\bar{A}^{\vartheta}, P(T-t))\Big\}.$$

As both, $V_{T-t}(\cdot,\cdot)$ and the per period profit $\int dF^{s_i(t)}(\vartheta|\cdot)\psi A_i(\cdot)\bar{A}^{\vartheta}$ are increasing in the productivity $A_i(\cdot)$ (for both strategies, innovation, $s_i(t) = \mathrm{in}$, and imitation, $s_i(t) = \mathrm{im}$), it follows that also $V_{T-t-1}(\cdot,\cdot)$ is increasing in the productivity. This proves the induction step. □

Proof of Proposition 8. Assume for a contradiction that the value function is increasing in the productivity $A_i(t)$, but that the optimal strategy is not to maximize the expected productivity gain in that period. Then not only the current expected per period profit is smaller, but, because of the monotonicity of the value function, also the expected value function in the next period is lower. However, this contradicts the assumption that the strategy is optimal, and thus cannot be the solution to the Bellman equation (25). □

In the following text, we derive a lemma and a corollary that will help us to show that (22) admits a traveling wave solution with a stable shape.[54] First, from (22) we can derive the following lemma.

Lemma 4. *Let $F_a^{(1)}(t)$ and $F_a^{(2)}(t)$ be solutions of (22) 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. We introduce the difference

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

In the following discussion, we show that if $V_a(0) \leq 0$, then $V_a(t) \leq 0$ for all $t > 0$. We can write (22) as

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

We then get for $V_a(t)$,

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

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

[54]Our results follow Bramson (1983), who analyzed the traveling wave solution $u(x,t) = w(x-\nu t)$ of the Kolmogorov equation $\partial u/\partial t = f(u) + \partial^2 u/\partial x^2$.

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 sufficiently large (and ϵ sufficiently small) we can use a first-order Taylor approximation to write

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

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

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

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

A direct consequence of Lemma 4 is the following corollary.

Corollary 1. *Let $F_a(t)$ be a solution of (22) with Heaviside initial data, that is,*

$$F_a(0) = \Theta(a - a_{\mathrm{m}}) = \begin{cases} 0 & \text{if } a < a_{\mathrm{m}} \\ 1 & \text{if } a \geq a_{\mathrm{m}}. \end{cases}$$

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 $t \to \infty$.

Proof. 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 Lemma 4 applies.[55] It follows that $F_a^{(1)}(t) \geq F_a^{(2)}(t)$ for all $t > 0$. We then can write

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

[55] To see this, note that $m_\epsilon(t)$ is increasing in t as $F_a(t)$ is decreasing in t because the right-hand side of (22) is always less than zero. Consequently, it holds that $F_a^{(1)}(0) = F_{a-m_\epsilon(t_0)}(0) = \Theta(a - (a_{\mathrm{m}} + m_\epsilon(t_0))) \geq \Theta(a - (a_{\mathrm{m}} + m_\epsilon(t_0+b))) = F_{a-m_\epsilon(t_0+b)}(0)$. The latter is, by definition of the Heaviside function, an upper bound for any probability distribution function with support restricted to the interval $[a_{\mathrm{m}} + m_\epsilon(t_0+b), \infty)$. This applies in particular to $F_a^{(2)}(0) = F_{a-m_\epsilon(t_0+b)}(b)$ for any $b \in \mathbb{R}_+$, so that we have that $F_a^{(1)}(0) \geq F_a^{(2)}(0)$. Hence, we can make use of Lemma 4.

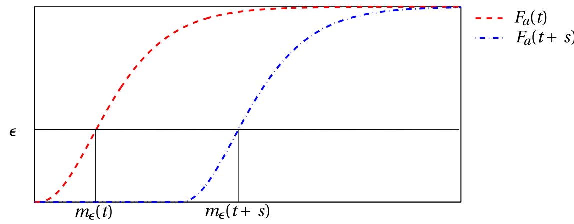

Figure 12. Illustration of distributions $F_a(t)$ and $F_a(t+s)$ at times t and $t+s$ for $s>0$.

For each value of b this is a decreasing sequence of real numbers that is bounded from below and thus its infimum is the limit. In particular, since t_0, b and ϵ were chosen arbitrarily, we obtain that $F_{a-m_0(t)}(t)$ converges to some $f(a)\geq 0$ from above as $t\to\infty$. An illustration can be seen in Figure 12. □

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

Proof of Proposition 7. Let $\epsilon>0$. Then by Corollary 1 it holds that

$$\lim_{t\to\infty} F_{a-m_\epsilon(t)}(t)=f_\epsilon(a).$$

Because of convergence it holds for the total derivative that

$$\lim_{t\to\infty}\frac{dF_{a-m_\epsilon(t)}(t)}{dt}=0$$

or, equivalently,

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

Using (22), the above equation can be written as

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

Integrating over $[0,\alpha)$, we obtain

$$o(1)=\int_0^\alpha\left(\frac{2q-1}{2}F_{a-m_\epsilon(t)}(t)^2+\frac{1-2q-p}{2}F_{a-m_\epsilon(t)}(t)+\frac{p}{2}F_{a-m_\epsilon(t)-1}(t)\right)da+(F_{\alpha-m_\epsilon(t)}(t)-F_{0-m_\epsilon(t)}(t))\frac{dm_\epsilon(t)}{dt}.$$

Looking at the limit over time ($\lim_{t\to\infty}$ on both sides), we obtain

$$o(1)=\int_0^\alpha\left(\frac{2q-1}{2}f_\epsilon(a)^2+\frac{1-2q-p}{2}f_\epsilon(a)+\frac{p}{2}f_\epsilon(a-1)\right)da+(f_\epsilon(a)-f_\epsilon(0))\lim_{t\to\infty}\frac{dm_\epsilon(t)}{dt}.$$

As everything except $\lim_{t\to\infty} dm_\epsilon(t)/dt$ does not depend on t, we can conclude that there is a constant ν such that $\lim_{t\to\infty} dm_\epsilon(t)/dt = \nu$.

Further, we must have that $F_{m_\epsilon(t)}(t) = F_{m_\epsilon(t+s)}(t+s)$ 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 (22) 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. □

References

Acemoglu, Daron (2009), *Introduction to Modern Economic Growth*. Princeton University Press, Princeton. [1054]

Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti (2006), "Distance to frontier, selection, and economic growth." *Journal of the European Economic Association*, 4, 37–74. [1054, 1055, 1057, 1058]

Acemoglu, Daron and Dan Vu Cao (2015), "Innovation by entrants and incumbents." *Journal of Economic Theory*, 157, 255–294. [1056]

Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas (2008), "Models of idea flows." Working Paper, NBER w14135. [1057]

Asl, Farshid Maghami and A. Galip Ulsoy (2003), "Analysis of a system of linear delay differential equations." *Journal of Dynamic Systems, Measurement, and Control*, 125, 215–223. [1070, 1088, 1089, 1090]

Atkeson, Andrew and Ariel Tomás Burstein (2010), "Innovation, firm dynamics, and international trade." *Journal of Political Economy*, 118, 433–484. [1057]

Aubin, Jean Pierre and Arrigo Cellina (1984), *Differential Inclusions: Set-Valued Maps and Viability Theory*. Springer-Verlag, Berlin. [1063]

Baily, Martin Neil, Charles Hulten, David Campbell, Timothy Bresnahan, and Richard E. Caves (1992), "Productivity dynamics in manufacturing plants." *Brookings Papers on Economic Activity. Microeconomics*, 1992, 187–267. [1054]

Bellman, Richard Ernest and Kenneth L. Cooke (1963), *Differential-Difference Equations*. RAND Corporation, Santa Monica, California. [1070, 1087, 1088]

Benhabib, Jess, Jesse Perla, and Christopher Tonetti (2014), "Catch-up and fall-back through innovation and imitation." *Journal of Economic Growth*, 19, 1–35. [1057]

Benhabib, Jess, Jesse Perla, and Christopher Tonetti (2016), "The growth dynamics of innovation, diffusion, and the technology frontier." Report, New York University. [1057]

Bloom, Nicholas and John Van Reenen (2011), "Human resource management and productivity." In *Handbook of Labor Economics*, Vol. 4, Part B (David Card and Orley Ashenfelter, eds.), 1697–1767, Elsevier, San Diego, California. [1054]

Bramson, Maury (1983), *Convergence of Solutions of the Kolmogorov Equation to Traveling Waves*. Memoirs of the American Mathematical Society. American Mathematical Society, Providence, Rhode Island. [1093, 1095]

Cheng, Leonard K. and Elias Dinopoulos (1996), "A multisectoral general equilibrium model of Schumpeterian growth and fluctuations." *Journal of Economic Dynamics and Control*, 20, 905–923. [1057]

Coad, Alex (2009), *The Growth of Firms: A Survey of Theories and Empirical Evidence*. Edward Elgar, Cheltenham, UK. [1055]

Cohen, Wesley M. and Steven Klepper (1992), "The anatomy of industry R&D intensity distributions." *The American Economic Review*, 82, 773–799. [1055]

Cohen, Wesley M. and Steven Klepper (1996), "A reprise of size and R&D." *The Economic Journal*, 106, 925–951. [1055]

Cohen, Wesley M. and Daniel A. Levinthal (1989), "Innovation and learning: The two faces of R&D." *Economic Journal*, 99, 569–596. [1056]

Comin, Diego A. and Marti Mestieri (2014), "Technology diffusion: Measurement, causes and consequences." In *Handbook of Economic Growth*, Vol. 2 (Philippe Aghion and Steven Durlauf, eds.), 565–622, Elsevier, Amsterdam. [1055]

Conlisk, John (1976), "Interactive Markov chains." *Journal of Mathematical Sociology*, 4, 157–185. [1065]

Corless, Robert M., Gaston H. Gonnet, David E. G. Hare, David J. Jeffrey, and Donald E. Knuth (1996), "On the Lambert W function." *Advances in Computational Mathematics*, 5, 329–359. [1070, 1089]

de Wit, Gerrit (2005), "Firm size distributions: An overview of steady-state distributions resulting from firm dynamics models." *International Journal of Industrial Organization*, 23, 423–450. [1054]

Di Matteo, Tiziana, Tomaso Aste, and Mauro Gallegati (2005), "Innovation flow through social networks: Productivity distribution in France and Italy." *The European Physical Journal B*, 47, 459–466. [1076]

Doms, Mark, Timothy Dunne, and Kenneth R. Troske (1997), "Workers, wages, and technology." *The Quarterly Journal of Economics*, 112, 253–290. [1054]

Driver, Rodney David (1977), *Ordinary and Delay Differential Equations*, volume 20 of Applied Mathematical Sciences. Springer-Verlag, New York. [1070, 1088]

Durlauf, Steven N. (1996), "On the convergence and divergence of growth rates." *The Economic Journal*, 106, 1016–1018. [1054]

Durlauf, Steven N. and Paul A. Johnson (1995), "Multiple regimes and cross-country growth behaviour." *Journal of Applied Econometrics*, 10, 365–384. [1075]

Eaton, Jonathan and Samuel Kortum (1999), "International technology diffusion: Theory and measurement." *International Economic Review*, 40, 537–570. [1057]

Eeckhout, Jan and Boyan Jovanovic (2002), "Knowledge spillovers and inequality." *American Economic Review*, 92, 1290–1307. [1055, 1057]

Ehrhardt, George, Matteo Marsili, and Fernando Vega-Redondo (2006), "Diffusion and growth in an evolving network." *International Journal of Game Theory*, 34, 383–397. [1076]

Fai, Felicia and Nicholas Von Tunzelmann (2001), "Industry-specific competencies and converging technological systems: Evidence from patents." *Structural Change and Economic Dynamics*, 12, 141–170. [1059]

Feyrer, James D. (2008), "Convergence by parts." *The BE Journal of Macroeconomics*, 8, 1–35. [1054, 1075]

Gabaix, Xavier (1999), "Zipf's law for cities: An explanation." *Quarterly Journal of Economics*, 114, 739–767. [1054]

Gast, Nicolas and Bruno Gaujal (2010), "Mean field limit of non-smooth systems and differential inclusions." *ACM SIGMETRICS Performance Evaluation Review*, 38, 30–32. [1079]

Geroski, Paul A. (2000), "Models of technology diffusion." *Research Policy*, 29, 603–625. [1055]

Ghiglino, Christian (2012), "Random walk to innovation: Why productivity follows a power law." *Journal of Economic Theory*, 147, 713–737. [1056]

Griffith, Rachel, Stephen Redding, and John Van Reenen (2003), "R&D and absorptive capacity: Theory and empirical evidence." *Scandinavian Journal of Economics*, 105, 99–118. [1055]

Griliches, Zvi (1957), "Hybrid corn: An exploration in the economics of technological change." *Econometrica*, 25, 501–522. [1055]

Jovanovic, Boyan and Rafael Rob (1990), "Long waves and short waves: Growth through intensive and extensive search." *Econometrica*, 58, 1391–1409. [1057]

Kelly, Morgan (2001), "Linkages, thresholds, and development." *Journal of Economic Growth*, 6, 39–53. [1059, 1076]

Khalil, Hassan K. (2002), *Nonlinear Systems*, third edition. Prentice Hall, Upper Saddle River, New Jersey. [1084]

Klette, Tor Jakob and Samuel Kortum (2004), "Innovating firms and aggregate innovation." *Journal of Political Economy*, 112, 986–1018. [1056]

Kogut, Bruce and Udo Zander (1992), "Knowledge of the firm, combinative capabilities, and the replication of technology." *Organization Science*, 3, 383–397. [1056, 1060]

König, Michael D. (2010), *Limited Absorptive Capacities and the Emergence of Convergence Clubs*. Master's thesis, University of Zurich. [1053]

König, Michael D. (2014), "Technology cycles in dynamic R&D networks." Working Paper, University of Zurich, Department of Economics 178. [1076]

König, Michael D., Jan Lorenz, and Fabrizio Zilibotti (2014), "Innovation vs. imitation and the evolution of productivity distributions." Unpublished, Stanford University, Jacobs University Bremen, University of Zurich. [1063, 1067, 1074]

Kurtz, Thomas G. (1970), "Solutions of ordinary differential equations as limits of pure jump Markov processes." *Journal of Applied Probability*, 7, 49–58. [1062]

Levinsohn, James and Amil Petrin (2003), "Estimating production functions using inputs to control for unobservables." *Review of Economic Studies*, 70, 317–341. [1054]

Lippman, Steven A. and John J. McCall (1976), "Job search in a dynamic economy." *Journal of Economic Theory*, 12, 365–390. [1077]

Lucas, Robert E., Jr. (2009), "Ideas and growth." *Economica*, 76, 1–19. [1057, 1065]

Lucas, Robert E., Jr. and Benjamin Moll (2014), "Knowledge growth and the allocation of time." *Journal of Political Economy*, 122, 1–51. [1057]

Luttmer, Erzo G. J. (2007), "Selection, growth, and the size distribution of firms." *The Quarterly Journal of Economics*, 122, 1103–1144. [1056, 1057]

Luttmer, Erzo G. J. (2012), "Eventually, noise and imitation implies balanced growth." Working Paper, Federal Reserve Bank of Minneapolis No. 699. [1057, 1071]

Majumdar, Satya N. and P. L. Krapivsky (2001), "Dynamics of efficiency: A simple model." *Physical Review E*, 63, 4. [1074]

Nelson, Richard R. and Edmund S. Phelps (1966), "Investment in humans, technological change, and economic growth." *American Economic Review*, 56, 69–75. [1056]

Perla, Jesse and Christopher Tonetti (2014), "Equilibrium imitation and growth." *Journal of Political Economy*, 122, 52–76. [1057]

Quah, Danny T. (1997), "Empirics for growth and distribution: Stratification, polarization, and convergence clubs." *Journal of Economic Growth*, 2, 27–59. [1054, 1075]

Rosenberg, Nathan (1976), *Perspectives on Technology*. Cambridge University Press, New York. [1059]

Roth, Grégory and William H. Sandholm (2013), "Stochastic approximations with constant step size and differential inclusions." *SIAM Journal on Control and Optimization*, 51, 525–555. [1079]

Saichev, Alexander I., Yannick Malevergne, and Didier Sornette (2010), *Theory of 'Z'ipf's Law and Beyond*. Springer-Verlag, Berlin. [1054, 1060]

Sandholm, William H. (2010), *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge, Massachusetts. [1062, 1078]

Segerstrom, Paul S. (1991), "Innovation, imitation, and economic growth." *Journal of Political Economy*, 99, 807–827. [1057]

Shilov, Georgi E. (1996), *Elementary Real and Complex Analysis* (Richard A. Silverman, ed.). Dover Publications, New York. [1087]

Smith, Hal (2010), *An Introduction to Delay Differential Equations With Applications to the Life Sciences*, volume 57 of Texts in Applied Mathematics. Springer, New York. [1070, 1083, 1087, 1088, 1089]

Stoneman, Paul (2002), *The Economics of Technological Diffusion*. Blackwell Publishers, Oxford. [1055]

Van Saarloos, Wim (2003), "Front propagation into unstable states." *Physics Reports*, 386, 29–222. [1093]

Waltman, Paul (2004), *A Second Course in Elementary Differential Equations*. Dover Books on Mathematics Series. Dover Publications, Mineola, New York. [1084]

Weitzman, Martin L. (1998), "Recombinant growth." *The Quarterly Journal of Economics*, 113, 331–360. [1060]

Yi, Sun and A. Galip Ulsoy (2006), "Solution of a system of linear delay differential equations using the matrix Lambert function." In *2006 American Control Conference*, 6. IEEE, Minneapolis, Minnesota. [1070]

Co-editor Gadi Barlevy handled this manuscript.

Submitted 2013-1-23. Final version accepted 2015-10-21. Available online 2015-10-21.