Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier*

Jess Benhabib — NYU and NBER

Jesse Perla — UBC

Christopher Tonetti — Stanford GSB and NBER

January 6, 2021

Abstract

We study how endogenous innovation and technology diffusion interact to determine the shape of the productivity distribution and generate aggregate growth. We model firms that choose to innovate, adopt technology, or produce with their existing technology. Costly adoption creates a spread between the best and worst technologies concurrently used to produce similar goods. The balance of adoption and innovation determines the shape of the distribution; innovation stretches the distribution, while adoption compresses it. On the balanced growth path, the aggregate growth rate equals the maximum growth rate of innovators. While innovation drives long-run growth, changes in the adoption environment can influence growth by affecting innovation incentives, either directly, through licensing of excludable technologies, or indirectly, via the option value of adoption.

Keywords: Endogenous Growth, Technology Diffusion, Adoption, Imitation, Innovation, Technology Frontier, Productivity Distribution

JEL Codes: O14, O30, O31, O33, O40

* We would especially like to thank Hugo Hopenhayn for suggesting the analysis of licensing, in addition to many other very helpful comments. We would also like to thank Philippe Aghion, Ufuk Akcigit, Paco Buera, Sebastian Di Tella, Pablo Fajgelbaum, Mildred Hager, Bart Hobijn, Chad Jones, Boyan Jovanovic, Pete Klenow, Bob Lucas, Erzo Luttmer, Kiminori Matsuyama, Ben Moll, Ezra Oberfield, Richard Rogerson, Tom Sargent, Nancy Stokey, Mike Waugh, and various seminar participants for useful comments and suggestions. Aref Bolandnazar and Brad Hackinen provided excellent research assistance. Jesse Perla gratefully acknowledges support from the University of British Columbia Hampton Grant.

1 Introduction

This paper studies how the interaction between adoption and innovation determines the shape of the productivity distribution, the expansion of the technology frontier, and the aggregate economic growth rate. Empirical estimates of productivity distributions tend to have a large range, with many low-productivity firms and few high-productivity firms within even very narrowly defined industries and products (Syverson (2011)). The economy is filled with firms that produce similar goods using different technologies, and different firms invest in improving their technologies in different ways. Some firms are innovative, bettering themselves while simultaneously pushing out the frontier by creating technologies that are new to the world. There are, however, many firms that instead purposefully choose to adopt already invented technologies.[1] The main contribution of this paper is to develop a model of aggregate growth that delivers a productivity distribution with an endogenous expanding frontier, range, and tail index that result from optimal firm adoption and innovation behavior.

In the model, innovation pushes out the frontier, creating the technologies that will eventually be adopted, and stretches the distribution. Adoption helps to compress the distribution by keeping the laggards from falling too far behind. Beyond affecting the shape of the distribution, both adoption and innovation affect the aggregate growth rate. Long-run growth is driven by innovation, but that does not necessarily mean that adoption cannot affect long-run growth. Rather, it means that adoption affects growth by affecting the incentives to innovate. Changes in the adoption environment can affect innovation incentives either because innovators may one day become adopters or because adopters may directly pay to license technologies from innovators.

Model Overview and Main Results. We first build a simple model of exogenous innovation and growth to isolate how innovation and adoption jointly determine the shape of the productivity distribution. We then extend the model by having firms choose an amount to invest in innovation. In the extended model, the long-run aggregate growth rate is determined by the innovation activity of high-productivity firms. Our analysis focuses on firms' decisions to adopt and innovate and how, together, these actions generate aggregate growth and shape the productivity distribution. Thus, the costs and benefits of adoption and innovation are at the core of the model.

Firms are heterogeneous in productivity, and a firm's technology is synonymous with its productivity. Adoption is modeled as paying a cost to instantaneously receive a draw of a new technology; this is a model of adoption because the new technology is drawn from the existing distribution of technologies currently in use for production. To represent innovation,

[1] See surveys, e.g., the Community Innovation Survey, in which firms report improving their products by using technologies that are either new to market or just new to the firm.

we model firms as being in either a creative or a stagnant innovation state; when creative, costly investment in innovation generates geometric growth in productivity. The more a firm invests in innovation, the faster it grows. A firm's innovation state evolves according to a two-state Markov process.[2] At each point in time, any firm has the ability to invest in innovation or adoption, and firms optimally choose whether and how to improve their productivity. Since adoption is a function of the distribution of available technologies, the productivity distribution is the aggregate state variable that moves over time, and this movement is driven by firms' adoption and innovation activity.

In equilibrium, low-productivity firms invest in adopting technologies; stagnant firms fall back relative to creative firms; medium-productivity creative firms invest small amounts to grow a bit through innovation; and higher-productivity creative firms invest a lot in innovation and push out the productivity frontier. König et al. (2020) use firm micro-data on R&D expenditure and productivity growth to provide evidence that adoption is more productive for low-productivity firms and that higher-productivity firms invest more in innovation, consistent with the behavior of firms in our model.

Easy adoption, in the sense of low cost or high likelihood of adopting a very productive technology, tends to compress the productivity distribution, as the low-productivity firms are not left too far behind. A low cost of innovation tends to spread the distribution, as the high-productivity firms can more easily escape from the pack. Thus, the shape of the distribution, which typically looks like a truncated Pareto with finite support, is determined by the relative efficiency of adoption and innovation. The stochastic innovation state ensures that some firms that have bad luck and are uncreative for a stretch of time fall back relative to adopting and innovating firms, generating a non-degenerate stationary distribution with adoption existing in the long run.

Adoption and innovation are not two completely independent paths, with some firms perpetual adopters and some perpetual innovators. Rather, the ability of all firms to invest in both activities generates general equilibrium interactions between activities. A key spillover between adoption and innovation can be seen in the option value of adoption. For high-productivity firms which are far from being low-productivity adopters, the value of having the option to adopt is small. The lower a firm's productivity, the closer it is to being an adopter and, thus, the higher the option value of adoption. The higher the option value of adoption, the lower the incentive to spend on innovating to grow away from entering the adoption region and exercising that option. Thus, the value of adoption, which is determined by the

[2]Modeling stochastic innovation with finite idiosyncratic growth rates is the key technical feature that delivers many of the desired model properties in a tractable framework. For example, we want the productivity distribution to have finite support so we can study how the commonly used abstraction of infinite support affects key properties of balanced growth path equilibria. Finite support with a maximum that is growing represents better technologies being invented over time.

cost of acquiring a new technology and the probability of adopting a good technology, affects incentives to innovate and, therefore, the aggregate growth rate. Through this channel, the better is adoption, the more tempting it is to free ride on other firms pushing out the frontier by investing less in innovation.

We conclude the paper by exploring how the excludability of technology affects the interplay of adoption and innovation. We model adopters as having to pay a fee to the firm whose technology they adopt. Hence there is an additional direct link between adoption behavior and innovation incentives that affects the shape of the distribution and the aggregate growth rate. Through this licensing channel, easier adoption leads to more licensing, which increases investments in innovation and aggregate growth.

1.1 Recent Literature

Our paper is closely related to the idea diffusion literature, including Luttmer (2007), Alvarez et al. (2008), Lucas (2009), Alvarez et al. (2013), Perla and Tonetti (2014), and Lucas and Moll (2014). Buera and Lucas (2018) provides a survey of this literature.

In an early paper on technology diffusion Nelson and Phelps (1966) develop a model that specifies a differential equation that determines technology diffusion as a function of the distance between the leader and the follower. Lucas (2009) is a key paper in the idea diffusion literature that allows ideas to diffuse not just from a leader to followers, but potentially from and to all agents in the economy. In this idea-diffusion model, the probability of receiving a particular idea depends on the frequency of that idea in the whole population. Perla and Tonetti (2014) and Lucas and Moll (2014) advance the literature by modeling agents who make a choice to invest in technology diffusion. Thus, the amount of diffusion is no longer exogenous (either as an exogenous function of the distance to the frontier or as an exogenous arrival rate of draws from a source distribution), which allows for the study of incentives, externalities, and welfare-improving policies. In this paper, we build on these models of idea diffusion, but allow agents to grow not just through diffusion, but also through innovation.

We contribute to the literature that studies both innovation and technology diffusion. Buera and Oberfield (2019) is a related semi-endogenous growth model of the international diffusion of technology and its connection to trade in goods. The authors combine the processes of idea diffusion and innovation, in the spirit of Jovanovic and Rob (1989). They model productivity upgrading according to a single process that mixes components of innovation and adoption. In contrast, our paper models these as distinct actions potentially undertaken by different firms. Furthermore, their focus is not on the endogenous determination of the shape of the distribution, since it is given exogenously by the distribution from which innovation increments are drawn.[3] The industry evolution model of Jovanovic and MacDonald (1994) features firms that grow from innovation and imitation, but a firm cannot purposely target its investment to innovation or imitation. Benhabib et al. (2014) explicitly model a choice between investing in innovation and adoption, but in a Nelson-Phelps style model in which agents' decisions do not depend on the entire productivity distribution.

Perhaps the most closely related paper to ours is König et al. (2016) (KLZ). There, as in our paper, firms make an optimal choice between two stochastic processes: one that is related to the existing productivity distribution (imitation) and one that is not (innovation). As in our paper, higher-productivity firms choose to focus more on innovation than imitation.[4] Both models produce sustained growth through the interaction of innovation and imitation. A key difference is our analysis of a finite relative technology frontier (i.e., analysis in a model in which the ratio of the maximum to minimum productivity is finite in the long-run). This allows us to study the intensive margin of innovation of frontier firms and how the option value of adoption induces a free-riding incentive and slows growth. We also study excludable technology by modeling licensing, which introduces another mechanism through which adoption affects long-run growth.[5] The importance of the tension between endogenous innovation and imitation is emphasized in König et al. (2020), which uses that perspective to analyze the recent transformation of the Chinese economy.

Acemoglu et al. (2006), Chu et al. (2014), and Stokey (2014, 2017) also explore the relationship between innovation and diffusion from different perspectives.[6] We share a similarity with those papers, as there is an advantage to backwardness in the sense of option value from the ability to adopt. The crucial element that enables the interesting trade-off between innovation and technology diffusion in our model is that the incumbents internalize some of the value from the evolving distribution of technologies, thus distorting their innovation

[3] Eeckhout and Jovanovic (2002) also model technological spillovers that are a function of the distribution of firm productivity. Acemoglu et al. (2007) also model spillovers across firms in innovation, captured by the number of firms that have attempted to implement a technology before.

[4] There are a few other differences to consider when comparing to König et al. (2016). KLZ assume no cost of either innovation or imitation and use a limit to firms' absorptive capacity to induce a tradeoff between innovation and adoption, whereas in our paper there is a cost of innovation and a cost of imitation. Furthermore, as the arrival rate of imitation is not immediate in KLZ, they have an asymptotically power-law left tail. Our model has a sweeping barrier, which results from the limit of a rapid imitation rate. Finally, the monopolistic competition and differentiated goods KLZ model provides an economic foundation for a profit function that increases in productivity, which we specify exogenously.

[5] The closest paper to ours in terms of modeling licensing is Hopenhayn and Shi (2020), which provides a search-theoretic framework for analyzing bargaining over technology transfers in an environment with congestion externalities and creative destruction. See also Jovanovic and Wang (2020), who study the impact of technology diffusion on the incentive to innovate over the dynamic path of an industry's evolution.

[6] An alternative line of literature studies the diffusion of technology from incumbents to entrants, as in Luttmer (2007), Acemoglu and Cao (2015), Sampson (2015), and Lashkari (2016).

choices. That is, incumbent firms not adopting today realize they may adopt in the future, and they derive positive value from this option to adopt.

Staley (2011) and Luttmer (2012a) explore the interaction of innovation and technology adoption in models in which both processes are exogenous.[7] Their approach connects to the well-understood KPP-Fisher equation, which allows for a clear formal analysis and sharp characterization of model properties. In our paper we explore how adoption and innovation choices respond to economic incentives and generate endogenous growth. Luttmer (2020) diagnoses the issue of multiple equilibria and hysteresis in idea diffusion models and shows how the introduction of a finite maximum growth rate for agents that learn from technology diffusion can generate uniqueness.

One question that arises in diffusion models is where do new ideas come from, and how does the discovery of those new ideas determine the aggregate growth rate.[8] To study this question, we model long-run growth that occurs through the interaction of innovation and diffusion. In our model with endogenous innovation, the interaction between innovation and adoption is especially interesting when the distribution has a finite frontier. The rate at which frontier firms invest in discovering new ideas determines the long-run aggregate growth rate. If there were no innovation, there would be no long-run growth. When the frontier is finite, the efficiency of adoption affects long-run aggregate growth in part because frontier firms realize they may become adopters in the future.

2 Baseline Model with Exogenous Stochastic Innovation

We first analyze an exogenous growth model to simplify the introduction of the environment and to isolate the economic forces that determine the shape of the stationary normalized productivity distribution. The only choice that a firm makes in this exogenous-innovation version of the model is whether to adopt a new technology or to continue producing with its existing technology. In Section 4, we develop the full version of the model in which a firm chooses its innovation rate. Endogenous innovation allows us to study how adoption and innovation activities interact to determine not only the shape of the productivity distribution, but also the aggregate growth rate.

[7] Luttmer (2012a, 2015a,b, 2020) provide careful analysis of the role of hysteresis, including the important interaction of the stochastic innovation process with initial conditions.

[8] Using an initial distribution with infinite support, such as in Perla and Tonetti (2014) and the baseline model of Lucas and Moll (2014), may provide a good approximation for the contribution of adoption to growth in the medium run, even absent a theory of innovation that pushes out a frontier. See Figure 2 of Perla and Tonetti (2014) and Figure 9 of Buera and Lucas (2018).

2.1 The Baseline Model

Time is continuous, starting at $t = 0$, and the horizon is infinite. A continuum of firms produce a homogeneous product and are heterogeneous over their productivity, Z, and innovation ability, $i \in \{\ell, h\}$. For simplicity and clarity of exposition, firm output equals firm profits equals firm productivity.[9] The measure of firms with productivity less than Z in innovation state i at time t is denoted by $\Phi_i(t, Z)$. The maximum productivity of any firm, $\bar{Z}(t) \equiv \sup \{\text{support}\{\Phi_\ell(t,\cdot)\} \cup \text{support}\{\Phi_h(t,\cdot)\}\}$ is interpreted as the technology frontier. There is a unit measure of firms, so that $\Phi_\ell(t, \bar{Z}(t)) + \Phi_h(t, \bar{Z}(t)) = 1$. The minimum of the support of the distribution is denoted by $M_i(t)$, so $\Phi_i(t, M_i(t)) = 0$, and $M_i(0) > 0$. This minimum of support is a key variable that is determined endogenously. Define the distribution unconditional on type as $\Phi(t, Z) \equiv \Phi_\ell(t, Z) + \Phi_h(t, Z)$.

A firm with productivity Z can choose to continue producing with its existing technology, in which case it would grow stochastically according to the exogenous innovation process, or it can choose to adopt a new technology.

Stochastic Process for Innovation. In the high innovation ability state (h), a firm is innovating and its productivity is growing at an exogenous rate $\gamma \geq 0$ (for now, but a choice made by firms in Section 4). In the low innovation ability state (ℓ), it has zero productivity growth from innovation (without loss of generality). Sometimes firms have good ideas or projects that generate growth and sometimes firms are just producing using their existing technology. The jump intensity from low to high is $\lambda_\ell > 0$ and from high to low is $\lambda_h > 0$. Since the Markov chain has no absorbing states and there is a strictly positive flow between the states for all Z, the support of the distribution conditional on ℓ or h is the same. With support $\{\Phi(t,\cdot)\} \equiv [M(t), \bar{Z}(t)]$, define the growth rates of the lower and upper bounds as $g(t) \equiv M'(t)/M(t)$ and $g_{\bar{Z}}(t) \equiv \bar{Z}'(t)/\bar{Z}(t)$.

We model innovation according to this continuous-time two-state Markov process because it allows for the existence of balanced growth path equilibria with finite-support productivity distributions in which adoption persists in the long run and growth is driven by the innovation choices of frontier firms. Persistence of the innovation state is modeled primarily for technical reasons related to continuous time rather than being of direct economic interest. Thus, we will perform our numerical exercises calibrating the model with high transition rates. Loosely speaking, shocks to the innovation state are like IID growth rate shocks that avoid continuous-time measurability issues with IID stochastic processes.

[9] See Perla et al. (2021) for guidance on how to enrich the model with monopolistically competitive firms that hire labor at an equilibrium wage to produce differentiated products.

Adoption and Technology Diffusion. A firm has the option to adopt a new technology by paying a cost. Adoption means changing production to use a technology that some other firm is using. We model this adoption process as undirected search across firms, as in Perla and Tonetti (2014) and Lucas and Moll (2014).

For simplicity, we model the adopting firm as drawing a new productivity, Z, from the unconditional distribution, $\Phi(t, Z)$, and starting in the low-innovation state, ℓ. That is, adopters cannot innovate immediately after adoption.[10]

The cost of adoption grows as the economy grows, and it is parameterized by $\zeta > 0$. The scale of the economy at time t can be summarized by the minimum of the support of the productivity distribution $M(t)$. Thus, we model the cost of adoption as $\zeta M(t)$.

Leapfrogging to the Frontier. Finally, firms can leapfrog to the frontier of the productivity distribution with arrival rate $\eta > 0$. The possibility that innovators can jump to the frontier represents some chance that the innovation process generates a big insight, instead of steady incremental progress. Adopters jumping to the frontier captures the possibility that sometimes adopters get lucky and their search for a new technology finds the best one available.

For tractability, we model such a jump as temporarily disruptive to innovation, such that all leapfrogging firms become ℓ-types and must wait for the Markov transition to h before they become innovators again.[11]

Firm Value Functions. Firms discount at rate $r > 0$. Let $V_i(t, Z)$ be the continuation value function for type i—i.e., the value at time t of being an i-type firm and producing with productivity Z.[12]

[10] Model variations on conditional vs. unconditional draws and maintaining or switching innovation types have few quantitative or qualitative consequences, especially given the high switching rates between states used in our numerical examples. The model setup here keeps the formulas cleaner since adopters only show up as a source in the Kolmogorov forward equation for the ℓ-type distribution, which saves on notation.

[11] The assumption that all leapfroggers switch to the ℓ state is purely for analytical convenience and can be changed without introducing qualitative differences. Similarly, the jumps to the frontier could occur exclusively for adopting firms instead of operating firms with no qualitative differences.

[12] To ease notation, we define the differential operator $\boldsymbol{\partial}$ such that $\boldsymbol{\partial}_z \equiv \frac{\partial}{\partial z}$ and $\boldsymbol{\partial}_{zz} \equiv \frac{\partial^2}{\partial z^2}$. When a function is univariate, derivatives are denoted with a prime, e.g., $M'(z) \equiv \boldsymbol{\partial}_z M(z) \equiv \frac{\mathrm{d}M(z)}{\mathrm{d}z}$.

$$rV_\ell(t,Z) = \underbrace{Z}_{\text{Production}} + \underbrace{\lambda_\ell\left(V_h(t,Z) - V_\ell(t,Z)\right)}_{\text{Switch to } h} + \underbrace{\eta(V_\ell(\bar{Z}) - V_\ell(Z))}_{\text{Jump to Frontier}} + \underbrace{\boldsymbol{\partial}_t V_\ell(t,Z)}_{\text{Capital Gains}} \tag{1}$$

$$rV_h(t,Z) = \underbrace{Z}_{\text{Production}} + \underbrace{\gamma Z \boldsymbol{\partial}_Z V_h(t,Z)}_{\text{Exogenous Innovation}} + \underbrace{\lambda_h\left(V_\ell(t,Z) - V_h(t,Z)\right)}_{\text{Switch to } \ell} + \underbrace{\eta(V_\ell(\bar{Z}) - V_h(Z))}_{\text{Jump to Frontier}}$$
$$+ \underbrace{\boldsymbol{\partial}_t V_h(t,Z)}_{\text{Capital Gains}}. \tag{2}$$

A firm's continuation value derives from instantaneous production plus capital gains as well as productivity growth if in the high innovation ability state, and it accounts for the intensity of jumps between innovation abilities i and jumps to the frontier.

The value of adoption is the continuation value of an ℓ-type firm having a new productivity drawn from $\Phi(t,Z)$ less the cost of adoption:

$$\text{Net Value of Adoption} = \underbrace{\int_{M(t)}^{\bar{Z}(t)} V_\ell(t,Z)\ \mathrm{d}\Phi(t,Z)}_{\text{Gross Adoption Value}} - \underbrace{\zeta M(t)}_{\text{Adoption Cost}}. \tag{3}$$

The Optimal Adoption Policy and the Minimum of the Support of the Productivity Distribution. The optimal firm policy is given by a threshold rule $M_a(t)$ such that all firms with productivity $Z \leq M_a(t)$ will choose to adopt and firms with $Z > M_a(t)$ will not adopt. Since the value of continuing is increasing in Z, and the net value of adopting is independent of Z, the firm's optimal adoption policy takes the form of a reservation productivity rule. While the adoption threshold could conceivably depend on the innovation type i, see Appendix A.5 for a proof showing that ℓ- and h-type firms choose the same threshold, $M_a(t)$, since the net value of adoption is independent of the current innovation type.

As draws are instantaneous, for any $t > 0$ this endogenous $M_a(t)$ becomes the evolving minimum of the $\Phi_i(t,Z)$ distributions, $M(t)$, and in a slight abuse of notation, we will refer to both the minimum of the support and the firm adoption policy as $M(t)$ going forward.[13]

In principle, there may be adopters of either innovation type with productivity in the common adoption region $Z \leq M(t)$. Define $S_i(t) \geq 0$ as the flow of i-type firms entering the adoption region at time t and denote the total flow of adopting firms as $S(t) \equiv S_\ell(t) + S_h(t)$.

The Firm Problem. A firm's decision problem can be described as choosing an optimal stopping time of when to adopt. Equivalently, it can be described as a free boundary problem,

[13]To see why the minimum of the support is the endogenous threshold, consider instantaneous adoption as the limit of a Poisson arrival rate of draw opportunities approaching infinity. In any positive time interval, firms wishing to adopt would gain an acceptable draw with probability approaching 1, so that $Z > M(t)$ almost surely. A heuristic derivation of this limit is given in Appendix A.6.

choosing the productivity level at which to adopt. Necessary conditions for the free boundary problem include the continuation value functions and, at the endogenously chosen adoption boundary $M(t)$, value matching conditions,

$$\underbrace{V_i(t, M(t))}_{\text{Continuation Value at Threshold}} = \underbrace{\int_{M(t)}^{\bar{Z}(t)} V_\ell(t, Z)\ \mathrm{d}\Phi(t, Z) - \zeta M(t)}_{\text{Net Adoption Value}}, \tag{4}$$

and smooth-pasting conditions,

$$\boldsymbol{\partial}_Z V_\ell(t, M(t)) = 0 \quad \text{if } M'(t) > 0 \tag{5}$$

$$\boldsymbol{\partial}_Z V_h(t, M(t)) = 0 \quad \text{if } M'(t) - \gamma M(t) > 0. \tag{6}$$

Value matching states that at the optimal adoption reservation productivity, the value of producing with the reservation productivity is equal to the value of adopting a new productivity. Smooth pasting is a technical requirement that can be interpreted as an intertemporal no-arbitrage condition—a necessary condition only if firms at the boundary are moving backwards relative to the boundary over time.

The Technology Frontier. Given the adoption and innovation processes, if $\bar{Z}(0) < \infty$, then $\bar{Z}(t)$ will remain finite for all t, as it evolves from the innovation of firms in the interval infinitesimally close to $\bar{Z}(t)$ and the growth rate of innovating firms is finite. Indeed, the frontier grows at rate γ since there is always some firm arbitrarily close to the threshold that grows at rate γ.

Lemma 1 (Growth Rate of the Finite Frontier)**.** *If $\bar{Z}(0) < \infty$, then $\bar{Z}(t) < \infty\ \forall\ t < \infty$ and $g_{\bar{Z}} \equiv \bar{Z}'(t)/\bar{Z}(t) = \gamma$.*

If $\left[\Phi_h(t, \bar{Z}(t)) - \Phi_h(t, \bar{Z}(t) - \epsilon)\right] > 0\ \ \forall\ \ \epsilon > 0$, then there are always firms arbitrarily close to the frontier. If some of them are type h, then they will push the frontier out at rate γ. Given that there are a continuum of firms and that the arrival rate of changes in type i is a memoryless Poisson process, for all finite t there will always be some h-type firms that are arbitrarily close to the frontier and have never jumped to the low state, so the growth rate of the frontier is always γ.

Law of Motion of the Productivity Distribution. The Kolmogorov forward equation (KFE) describes the evolution of the productivity distribution for productivities above the minimum of the support. The KFEs in the CDFs for ℓ- and h-type firms are

$$\boldsymbol{\partial}_t \Phi_\ell(t,Z) = \underbrace{-\lambda_\ell \Phi_\ell(t,Z) + \lambda_h \Phi_h(t,Z)}_{\text{Net Flow from Type Change Jumps}} \underbrace{- \eta \Phi_\ell(t,Z) + \eta \mathbb{H}\left(Z - \bar{Z}(t)\right)}_{\text{Leapfroggers}} \tag{7}$$
$$+ \underbrace{(S_\ell(t) + S_h(t))}_{\text{Flow of Adopters}} \underbrace{\Phi(t,Z)}_{\text{Draw} \leq Z} - \underbrace{S_\ell(t)}_{\ell\text{-Adopters}}$$

$$\boldsymbol{\partial}_t \Phi_h(t,Z) = \underbrace{-\gamma Z \boldsymbol{\partial}_Z \Phi_h(t,Z)}_{\text{Innovation}} - \underbrace{\lambda_h \Phi_h(t,Z) + \lambda_\ell \Phi_\ell(t,Z)}_{\text{Net Flow from Type Change Jumps}} \underbrace{- \eta \Phi_h(t,Z)}_{\text{Leapfroggers}} - \underbrace{S_h(t)}_{h\text{-Adopters}}, \tag{8}$$

where $\mathbb{H}(\cdot)$ is the Heaviside step function. For each type i, the KFEs keep track of inflows and outflows of firms with a productivity level at or below Z. An i-type firm with productivity less than or equal to Z stops being in the i-distribution at or below Z if it keeps its type and increases its productivity above Z or if it changes its type.

A firm can keep its type and increase its Z in three ways: adoption, innovation, or leapfrogging. An adopting firm has probability $\Phi(t,Z)$ of becoming type ℓ and drawing a productivity less than or equal to Z and the number of firms adopting is $(S_\ell(t)+S_h(t))$. Hence $(S_\ell(t) + S_h(t))\Phi(t,Z)$ is added to the ℓ-distribution. Additionally, the flow of adopters of type i, $S_i(t)$, is subtracted from the corresponding distribution. The $S_i(t)$ term is subtracted from the CDFs for all Z because adoption occurs at the minimum of the support. Intuitively, the adoption reservation productivity acts as an absorbing barrier sweeping through the distribution from below. As it moves forward, it collects adopters at the minimum of the support, removes them from the distribution, and inserts them back into the distribution according to Φ. Recognizing that the jumps in type at intensity λ_i are of measure 0 when calculating the number of firms that cross the boundary in any infinitesimal time period, the flow of adopters comes from the flux across the moving boundary $M(t)$.

The KFE for the h-types has a term that subtracts the firms that grow above Z through innovation: there are $\boldsymbol{\partial}_Z \Phi_h(t,Z)$ amount of h-type firms at productivity Z, and because innovation is geometric, they grow above Z at rate γZ.

Firms of productivity Z switching from type i to type i' leave the i-distribution and enter the i'-distribution at rate λ_i. For example, the amount of ℓ-type firms with productivity less than or equal to Z is $\Phi_\ell(t,Z)$, and they leave the ℓ-distribution at rate λ_ℓ and enter the h-distribution at the same rate.

Firms jump to the frontier at rate η, so they are subtracted from the CDF. For analytical tractability, all leapfrogging firms become type ℓ, so they are added to the ℓ-distribution at $\bar{Z}(t)$.

The firms are owned by a representative consumer who values the undifferentiated good with log utility and a discount rate $\rho > 0$. If the growth rate on the balanced growth path is g, then the interest rate faced by firms is $r = \rho + g$.

2.2 Normalization, Stationarity, and Balanced Growth Paths

In this paper, we study economies in equilibrium on balanced growth paths (BGPs), in which the distribution is stationary when properly rescaled and aggregate output grows at a constant rate. The economy is characterized by a system of equations defining the firm problem, the laws of motion of the productivity distributions, and consistency conditions that link firm behavior and the evolution of the distributions. To compute BGP equilibria, it is convenient to transform this system to a set of stationary equations. While we could normalize by any variable that grows at the same rate as the aggregate economy, it is convenient to normalize variables relative to the endogenous boundary $M(t)$. Define normalized productivity, normalized distributions, and normalized value functions as

$$z \equiv \log(Z/M(t)), \tag{9}$$

$$F_i(t,z) = F_i(t, \log(Z/M(t))) \equiv \Phi_i(t,Z), \tag{10}$$

$$v_i(t,z) = v_i(t, \log(Z/M(t))) \equiv \frac{V_i(t,Z)}{M(t)}. \tag{11}$$

The adoption threshold is normalized to $\log(M(t)/M(t)) = 0$, and the relative technology frontier is $\bar{z}(t) \equiv \log(\bar{Z}(t)/M(t))$. See Figure 1 for an illustration of the normalized and unnormalized distributions. Define the normalized unconditional distribution as $F(z) \equiv F_\ell(z) + F_h(z)$, which is a valid CDF (i.e., $F(0) = 0$ and $F(\bar{z}(t)) = 1$).

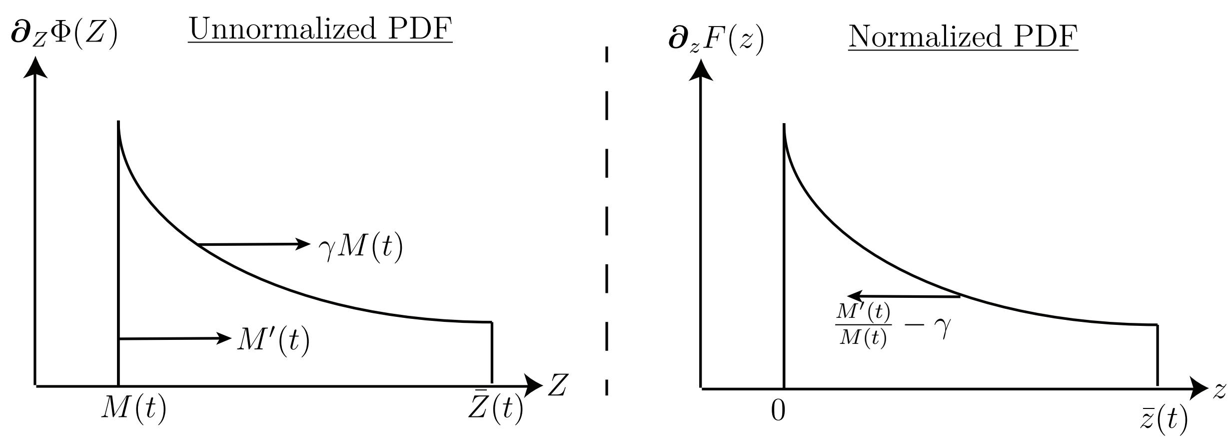

Figure 1: Normalized and Unnormalized Distributions

With the above normalizations, it is possible that the value functions, productivity distributions, and growth rates are stationary—i.e., independent of time.

Summary of Stationary KFEs and Firm Problem. A full derivation of the normalized system is given in Appendix A.1. Here, we summarize the resulting equations that characterize the laws of motion for the normalized productivity distributions and the normalized firm problem. The equations that determine the stationary productivity distributions are:

$$0 = gF_\ell'(z) + \lambda_h F_h(z) - \lambda_\ell F_\ell(z) - \eta F_\ell(z) + \eta \mathbb{H}\left(z - \bar{z}\right) + SF(z) - S_\ell \tag{12}$$

$$0 = \left(g - \gamma\right) F_h'(z) + \lambda_\ell F_\ell(z) - \lambda_h F_h(z) - \eta F_h(z) - S_h \tag{13}$$

$$0 = F_\ell(0) = F_h(0) \tag{14}$$

$$1 = F_\ell(\bar{z}) + F_h(\bar{z}) \tag{15}$$

$$S_\ell = \begin{cases} gF_\ell'(0) & \text{if } g > 0 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

$$S_h = \begin{cases} \left(g - \gamma\right) F_h'(0) & \text{if } g > \gamma \\ 0 & \text{otherwise} \end{cases}. \tag{17}$$

The necessary conditions of the normalized firm problem are:

$$\rho v_\ell(z) = e^z - g v_\ell'(z) + \lambda_\ell(v_h(z) - v_\ell(z))) + \eta(v_\ell(\bar{z}) - v_\ell(z)) \tag{18}$$

$$\rho v_h(z) = e^z - (g - \gamma) v_h'(z) + \lambda_h(v_\ell(z) - v_h(z)) + \eta(v_\ell(\bar{z}) - v_h(z)) \tag{19}$$

$$v(0) = \frac{1}{\rho} = \int_0^{\bar{z}} v_\ell(z) \,\mathrm{d}F(z) - \zeta \tag{20}$$

$$v_\ell'(0) = 0 \text{ if } g > 0 \tag{21}$$

$$v_h'(0) = 0 \text{ if } g > \gamma. \tag{22}$$

Given that the two types of firms choose the same adoption threshold, we drop the type index for the value functions at the adoption threshold: $v(0) \equiv v_i(0)$.

Equations (12) to (15) are the stationary KFEs with boundary values. Recall that g is the growth rate of the minimum of the support and γ is the innovation growth rate. In the normalized setup, firms are moving backwards toward the constant minimum of the support and their growth rate determines the speed at which they are falling back.

Equations (18) and (19) are the Bellman equations in the continuation region and (20) is the value-matching condition that links the value of continuing to the value of technology adoption. The smooth-pasting conditions given in equations (21) and (22) are necessary only if the firms of that particular type are drifting backwards relative to the adoption threshold.

See Figure 2 for a visualization of the normalized Bellman equations.

Definition 1 (Recursive Competitive Equilibrium with Exogenous Innovation)**.** *A recursive competitive equilibrium with exogenous innovation consists of initial distributions $\Phi_i(0, z)$, adoption reservation productivity functions $M_i(t)$, value functions $V_i(t, z)$, interest rates $r(t)$, and sequences of productivity distributions $\Phi_i(t, z)$ such that the following hold:*

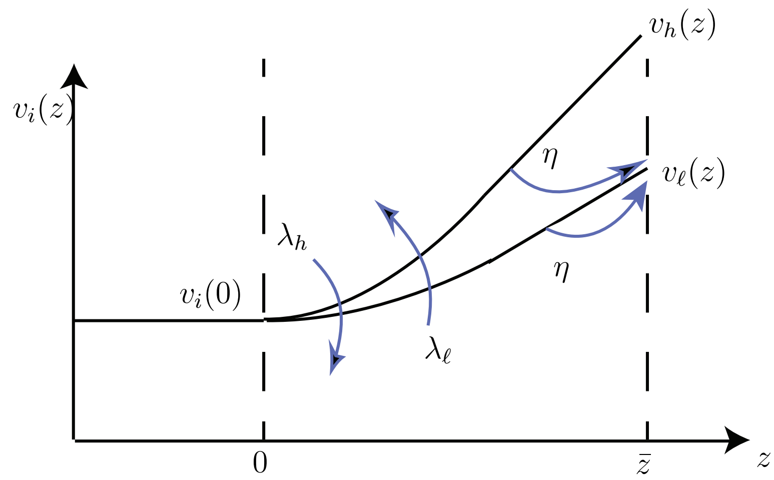

Figure 2: Normalized Stationary Value Functions

1. *Given* $r(t)$ *and* $\Phi_i(t,z)$, $M_i(t)$ *are the optimal adoption reservation productivities, with* $V_i(t,z)$ *the associated value functions.*

2. *Given* $M_i(t)$ *and* $\Phi_i(t,z)$, $r(t)$ *is consistent with the consumer's intertemporal marginal rate of substitution.*

3. *Given* $M_i(t)$, $\Phi_i(t,z)$ *fulfill the laws of motion in* (7) *and* (8) *subject to the initial condition* $\Phi_i(0,z)$.

We restrict our interest to equilibria that are balanced growth paths. Define the growth rate of aggregate output to be $g_Y(t) \equiv \boldsymbol{\partial}_t \mathbb{E}_t[Z]/\mathbb{E}_t[Z]$.

Definition 2 (Balanced Growth Path Equilibrium with Exogenous Innovation)**.** *A balanced growth path equilibrium with exogenous innovation is a recursive competitive equilibrium such that the growth rate of aggregate output is constant and the normalized productivity distributions are stationary. That is,* $g_Y(t) = g_Y$ *and* $F_i(t,z) = F_i(z)\ \forall\ t$.

Lemma 2 shows that on a BGP, there is a tight relationship between the growth rate of the minimum of the support of the distribution, which is driven by firms' adoption decisions, and the growth rate of aggregate output.

Lemma 2 (Growth of the Endogenous Adoption Threshold and Aggregate Output)**.** *On a balanced growth path, the growth rate of the endogenous adoption threshold equals the growth rate of aggregate output. That is,* $g = g_Y$.

Aggregate output is defined as $Y(t) \equiv \mathbb{E}_t[Z]$. Mean productivity written as a function of the productivity CDF is $\mathbb{E}_t[Z] = \int_{M(t)}^{\bar{Z}(t)} [1 - \Phi(t,Z)]\,\mathrm{d}Z$. Using (9) and (10) with a change of variables shows that $Y(t) = M(t)\int_0^{\bar{z}(t)} [1 - F(t,z)]\, e^z \mathrm{d}z$. On a BGP the normalized productivity distributions are constants, i.e., $F_i(z,t) = F_i(z)$ with $\bar{z}$ constant. Hence $g_Y \equiv Y'(t)/Y(t) = M'(t)/M(t) = g$.

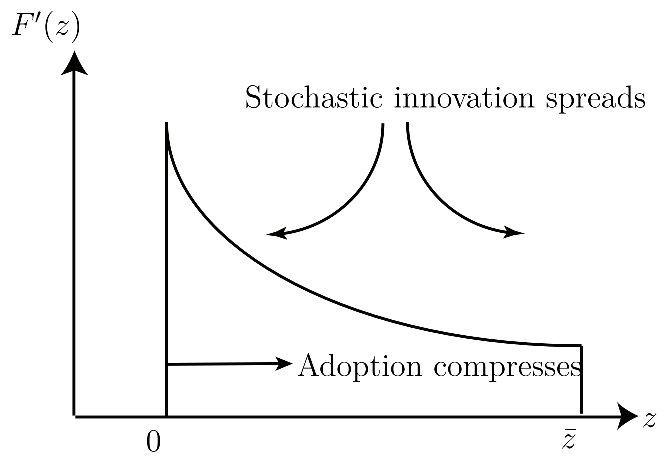

Figure 3: Tension between Stochastic Innovation and Adoption

How Adoption and Innovation Generate a Stationary Normalized Distribution. Figure 3 provides some intuition on how proportional growth and adoption can create a stationary distribution. Without endogenous adoption, nothing prevents the proportional random shocks from spreading out the distribution, driving the variance to infinity.[14] However, when adoption is endogenous, as the distribution spreads, the incentives to adopt a new technology increase, and the adoption decisions of low-productivity agents then act to compress the distribution. In a BGP equilibrium, technology diffusion can balance innovation, thus allowing for a stationary normalized distribution. Note, there are two possible types of normalized stationary distributions: either $\bar{z}$ is a finite constant or the normalized stationary distribution has infinite support. As we show later, these two types of stationary distributions have different implications for how adoption and innovation determine the equilibrium aggregate growth rate.

3 How Adoption and Innovation Interact to Shape the Productivity Distribution

In this section, we compute BGP equilibria for economies with finite-support productivity distributions. There are two main questions that motivate our analysis. First, how do

[14] König et al. (2016) provide a similar intuition in their Proposition 2, which shows the expansion of the distribution in the absence of imitation.

adoption and innovation determine the shape of the productivity distribution? Second, what is the aggregate growth rate, and how is it affected by parameters related to adoption and innovation?

To set the stage, consider Perla and Tonetti (2014), which is essentially a discrete-time version of the economy in this paper with $\eta = \gamma = 0$ and is similar to other papers in the literature, such as Lucas and Moll (2014), as discussed in Buera and Lucas (2018). In Perla and Tonetti (2014) a BGP equilibrium with strictly positive growth exists only if the initial distribution has a power-law tail. Furthermore, the shape of the long-run distribution is given by the shape of the tail of the initial distribution. Additionally, the long-run growth rate is a function of the shape of the initial distribution and the cost of adoption.

Previewing some results, in contrast to Perla and Tonetti (2014), here the shape of the stationary distribution is endogenous, a function of parameters related to innovation and adoption. Furthermore, because the productivity distribution has finite support, the long-run aggregate growth rate equals the growth rate of innovators and is independent of initial conditions and the cost of adoption. In this section, the growth rate of innovators is exogenous (γ), so the aggregate growth rate is exogenous, but Section 4 develops the endogenous growth version of the model in which innovation investment is a choice made by firms.

Proposition 1 characterizes the BGP equilibrium in the case of exogenous innovation and a finite support distribution. Define the constants $\hat{\lambda} \equiv \frac{\lambda_\ell}{\eta+\lambda_h}$, $\bar{\lambda} \equiv \frac{\rho+\lambda_\ell+\lambda_h}{\rho+\lambda_h}$, and $\nu = \frac{\rho+\eta}{\gamma}\bar{\lambda}$.

Proposition 1 (BGP Equilibrium with Exogenous Innovation and Finite Support)**.** *Let $\bar{Z}(0) < \infty$. Then, a unique equilibrium exists with $\bar{z} < \infty$ and $g = \gamma$. The stationary distribution is*

$$F_\ell(z) = \frac{F_\ell'(0)}{(F_\ell'(0) - \eta/\gamma)(1+\hat{\lambda})}(1 - e^{-\alpha z}) \ , \ z \in [0, \bar{z}] \tag{23}$$

$$F_h(z) = \hat{\lambda} F_\ell(z) \ , \ z \in [0, \bar{z}] \tag{24}$$

with

$$\alpha \equiv (1+\hat{\lambda})(F_\ell'(0) - \eta/\gamma), \tag{25}$$

$$\bar{z} = \frac{\log(\gamma F_\ell'(0)/\eta)}{\alpha}. \tag{26}$$

The equilibrium $F_\ell'(0)$ solves the following implicit equation (substituting for α and $\bar{z}$),

$$\zeta + \frac{1}{\rho} = \frac{\gamma F_\ell'(0)\alpha\bar{\lambda}\left(-\frac{e^{-\nu\bar{z}}\left(-1+e^{-\alpha\bar{z}}\right)\eta}{\rho\alpha\nu} + \frac{e^{\bar{z}}\eta\left(e^{-\alpha\bar{z}}-1\right)}{-\alpha\rho} + \frac{-e^{-(\alpha+\nu)\bar{z}}+1}{\nu(\alpha+\nu)} + \frac{-e^{\bar{z}-\alpha\bar{z}}+1}{\alpha-1}\right)}{\gamma(\gamma F_\ell'(0)-\eta)(\nu+1)}. \tag{27}$$

The amount of adopters are

$$S_\ell = \gamma F_\ell'(0), \tag{28}$$

$$S_h = 0. \tag{29}$$

The firm value functions are

$$v_\ell(z) = \frac{\bar{\lambda}}{\gamma(1+\nu)}\left(e^z + \frac{1}{\nu}e^{-\nu z} + \frac{\eta}{\rho}\left(e^{\bar{z}} + \frac{1}{\nu}e^{-\nu\bar{z}}\right)\right), \tag{30}$$

$$v_h(z) = \frac{e^z + (\lambda_h - \eta)v_\ell(z) + \eta v_\ell(\bar{z})}{\rho + \lambda_h}. \tag{31}$$

Proof. See Appendix A.2. A key condition to check when evaluating a candidate equilibrium is that the $F_\ell'(0)$ that solves equation (27) is strictly larger than η/γ so that the relative frontier given in equation (26) is well defined. A proof of uniqueness, showing that $\bar{z} < \infty$ and that $g = \gamma$ is the unique equilibrium growth rate, is in Appendix A.4. □

Since $g = \gamma$, there are no h-type agents that cross the adoption boundary and the total amount of adopters is the flow of ℓ-type agents moving backwards at a relative speed of g across the adoption barrier.

Existence, Uniqueness, and Aggregate Growth. The first main result is that the equilibrium is unique. There is a unique stationary distribution that is independent of initial conditions. Furthermore, the aggregate growth rate equals the growth rate of innovators: $g = \gamma$.

Recall from Lemma 1 that if the initial distribution has finite support, then the stationary distribution will have finite support because the maximum growth rate of a firm is finite. Additionally the growth rate of the frontier equals the growth rate of innovators, i.e., $g_{\bar{Z}} = \gamma$. Furthermore, Lemma 2 stated that the growth rate of aggregate output equals the growth rate of the minimum of support of the distribution, i.e., $g_Y = g$. The minimum of support of the distribution evolves as firms at the minimum choose to adopt better technologies.

Proposition 1 links together the results of these two lemmas. If the productivity distribution has finite support (i.e., if $\bar{Z}(0) < \infty$), then clearly $g \leq g_{\bar{Z}}$, otherwise eventually the minimum of support would be larger than the maximum of support. In addition, if $\bar{z} < \infty$, for the normalized distribution to be stationary, the growth of the minimum and maximum of support must be equal. That is g must equal $g_{\bar{Z}}$. Proposition 1 states that $\bar{z} < \infty$ in equilibrium. Thus, if $\bar{Z}(0) < \infty$, then on a BGP $g = g_Y = g_{\bar{Z}} = \gamma$.

How could the aggregate growth rate be driven by adoption but also equal the growth rate of innovators? Lemma 2 may make it seem as though the behavior of adopters determines the long-run growth rate, but instead of this causal interpretation, it should be interpreted as an equilibrium relationship between adoption and innovation. Adoption is endogenous, and firms invest in adoption to keep up with the frontier. If the frontier were to grow faster, there would be more good new ideas arriving faster, which would make it worthwhile for firms to invest more in adopting those ideas faster.

Economic Intuition for Firm Behavior that Generates a Stationary Distribution. Leapfrogging to the frontier ($\eta > 0$) plays an essential role in delivering a stationary distribution with $\bar{z} < \infty$. Leapfrogging prevents the frontier from escaping from the rest of the distribution by ensuring that frontier technologies never have zero probability of being adopted.[15]

Leapfrogging to the frontier by a positive mass of agents can contain the escape in relative productivities by lucky firms that get streaks of long sojourns in the high-growth state h. As they eventually lose their innovative ability and become ℓ-types, they will be overtaken by others that leapfrog to the frontier from within the productivity distribution and replenish it. This leapfrogging/quality ladder process prevents laggards from remaining laggards forever.

To understand the importance of leapfrogging in generating a finite support, consider a simple modification to the adoption process that allows for higher chances of adopting better technologies. Specifically, let z be drawn from a distortion of the unconditional distribution, $F(Z)^\kappa$. Depending on the value of κ, the distribution is twisted so that adopters are either more or less likely to get better technologies compared to drawing from the unconditional distribution $F(Z)$. For any $\kappa \in (0, \infty)$—even those representing a very-high probability of adopting a very-high productivity—without leapfrogging (i.e., if $\eta = 0$) there does not exist an equilibrium with finite support of the normalized productivity distribution.[16]

The Bellman equations (30) and (31) are the sum of three components: the value of production, the option value of adoption, and the value of jumping to the frontier. That is, in addition to the value of production with the current z modified by time discounting and the probabilities of switching i-type, the value function accounts for changing z through adoption and accounts for the chance of jumping to the frontier. The option value of adoption is decreasing in a firm's productivity level; a firm with a high relative productivity has a long expected time until its relative productivity falls far enough to the point where it chooses to exercise the adoption option.

Whether the support of the distribution is finite has important economic implications. Because the ratio of the frontier to the minimum of the support (the relative frontier, $\bar{z}(t)$)

[15] If η were to equal 0 so there were no leapfrogging, then the distribution would be stationary only asymptotically. That is, even though $\bar{z}(t) < \infty$ for all $t < \infty$, $\lim_{t\to\infty} \bar{z}(t) = \infty$. In the case of $\eta = 0$, there is hysteresis and there is no longer a unique aggregate growth rate of $g = \gamma$.

[16] A simple further extension is to let firms choose κ at some cost. We interpret this as endogenous search intensity, with firms exerting effort to try to adopt better technologies with higher probability.

is finite, in the long run frontier firms still place positive value on the option to adopt. This means that increases in the value of adoption, whether associated with lower costs or higher benefits of adoption, will affect the value of frontier firms. Foreshadowing: In the endogenous growth environment, when γ is a choice, changes in the value of adoption will influence innovation behavior at the frontier, which will affect aggregate growth rates.

Shape. The productivity distribution has an endogenous truncated-tail index, α, that represents the shape of the productivity distribution. Furthermore, there is an additional shape parameter of the distribution: the range of the productivity distribution, given by the max-min ratio $\bar{z}$. Model primitives such as the cost of adoption and the rate of innovation affect both the tail index (α) and how much better the best firm can be relative to the worst firm in the economy ($\bar{z}$). Because $\bar{z}$ is a finite constant, meaning that the $F_i(z)$ have finite support, α is best interpreted as the shape parameter of a right-truncated power law.

Before moving to the endogenous growth case in the next section, we compute an equilibrium with roughly calibrated parameter values and use comparative statics to illustrate properties of the economy. This analysis will show how the cost of adoption and the innovation growth rate affect the shape of the distribution. We choose parameter values to demonstrate model forces in the relevant region of the parameter space.[17] The resulting parameterization is $\gamma = 0.02$, $\rho = 0.01$, $\lambda_\ell = 0.533, \lambda_h = 1.128$, $\zeta = 25.18$, and $\eta = 0.00098$.

First, as shown in Figure 4, v_ℓ and v_h are very similar because the calibrated λ_i are large, and thus the extra benefit of being in the high state or the relative pain from being in the low state does not last very long. Second, the distributions F_i are power-law shaped, with many low-productivity firms and few high-productivity firms, but they are truncated at $\bar{z}$, the relative frontier. The large measure of agents bunched up close to the minimum of support in our model comes from the immediacy of technology adoption; with a stochastic arrival as in König et al. (2016) we would also have an asymptoptically power-law left tail. In this calibration, there are also fewer h firms than ℓ firms at all productivities; those measures are determined largely by the ratio λ_h/λ_ℓ.

Comparative statics on how changes in η, γ, ζ, and λ_h affect $\bar{z}$ and α are shown in Figure 5. Easier innovation, in the sense of a higher growth rate for innovators, spreads out the distribution, creating a more distant technology frontier and a thicker tail. Easier leapfrogging, in the sense of a higher probability of jumps to the frontier, also generates thicker tails but generates less of a productivity gap between the best and worst firms. Easier adoption, captured by lower costs ζ, compresses the distribution, shrinking the relative frontier and thinning the tail.

[17] Details of the simple "calibration" used in our numerical exercises are given in Appendix C. We pick parameters to match firm growth rates, the firm size distribution, the aggregate growth rate, and the risk-free interest rate.

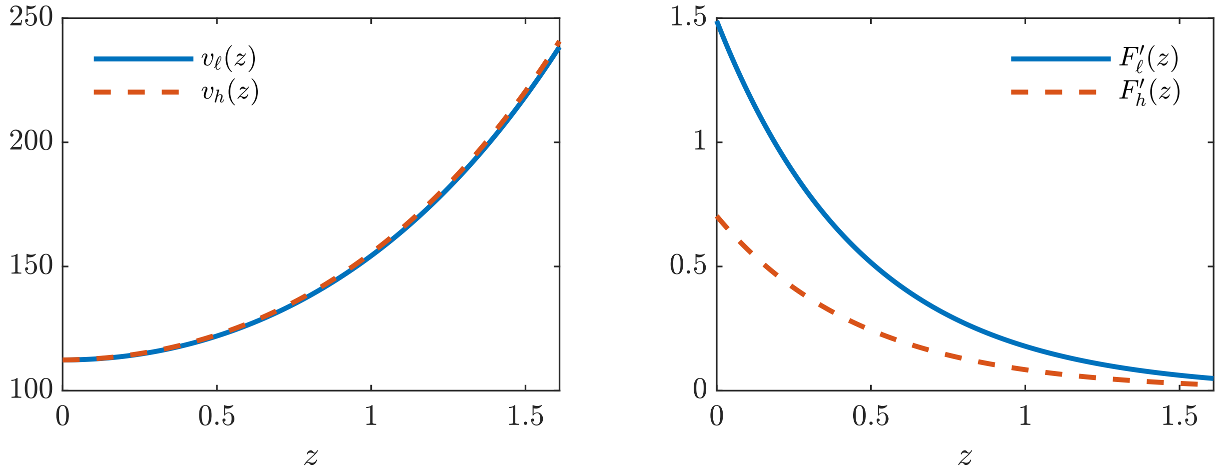

Figure 4: Exogenous $v_i(z)$ and $F'_i(z)$

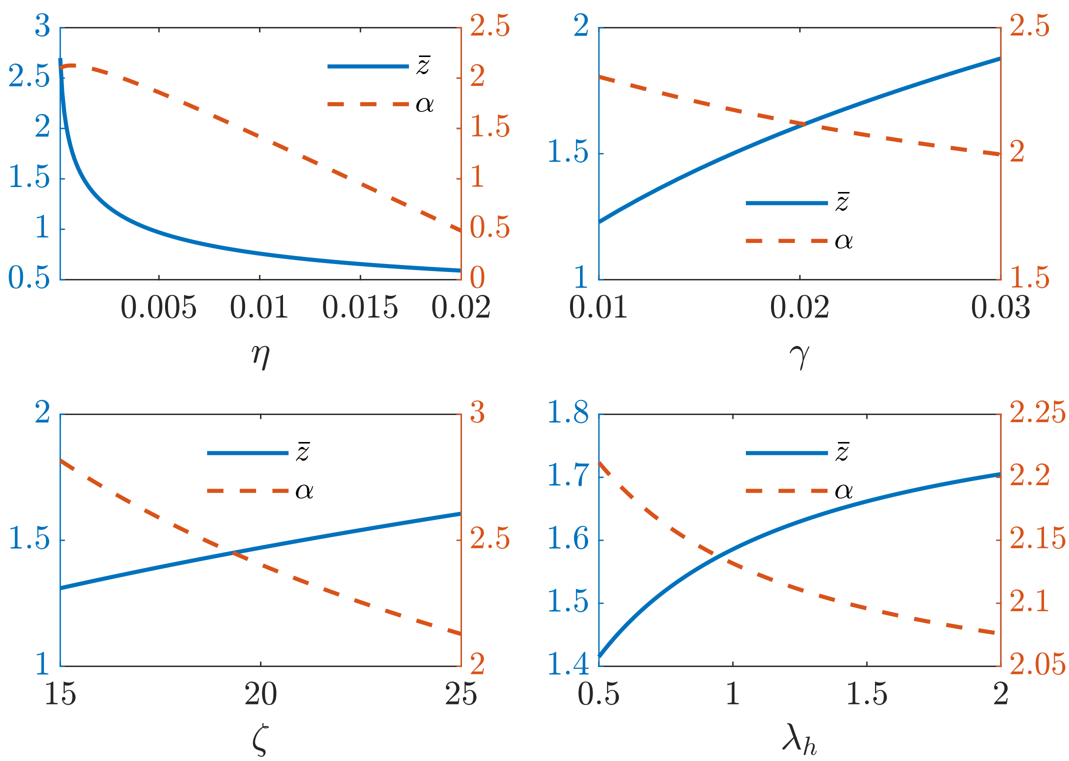

Figure 5: Comparative Statics for Exogenous Innovation

4 How Adoption and Innovation Interact to Determine the Long-run Aggregate Growth Rate

This section introduces endogenous investments in innovation. In the exogenous-innovation model of the previous section, the growth rate of the aggregate economy equaled the growth rate of innovation at the finite frontier, i.e., $g = \gamma$, but the growth rate of innovation was exogenous. With endogenous innovation, there will be an analogous result ($g = \gamma(\bar{z})$), where both the innovation rate $\gamma(\cdot)$ and the frontier $\bar{z}$ are endogenously (and jointly) determined. Hence, we are now in position to discuss how adoption affects the choice of innovation at the frontier, and thus the aggregate growth rate.

4.1 Model with Endogenous Innovation and Excludability

We model firms that can control the drift of their innovation process, as in Atkeson and Burstein (2010) and Stokey (2014).[18] We will focus on two main cases that highlight how adoption activity can affect long-run growth. In both cases, the long-run growth rate is determined by the growth rate of high-productivity innovators, but changes in the cost of adoption affect the growth rate of high-productivity firms by affecting their incentive to invest in innovation.

The first case highlights that when the distribution has finite support, there is a positive option value of adoption even for frontier firms. When firms decide how much to invest in innovation, they take into account the option value of adoption. On the margin, the more attractive is adoption, the stronger is the temptation to free ride and the weaker is the incentive to innovate.

The second case nests the first case and generalizes the model by adding that ideas are excludable, such that adopters need to pay a licensing fee to the higher-productivity firms whose technology they want to adopt. In this case, more adopters lead to more licensing fees. When profits from licensing are tied to the quality of the technology being licensed (e.g., with bargaining over surplus), there can be a positive link between adoption and innovation. In both cases we emphasize how externalities associated with innovation and adoption affect aggregate growth.

As in the exogenous growth model analyzed in Section 3, in the endogenous growth model the long-run growth rate cannot be larger than the maximum innovation rate of firms. Aggregate growth is therefore driven by innovation (i.e., if the cost of technology diffusion went to 0, the long-run aggregate growth rate would still be bounded by γ).

[18] An alternative approach, as taken in Section 6 of König et al. (2016), would be to have the firm choose between adoption and innovation given an investment capacity.

Licensing. Up to now, the firm providing the underlying technology to the adopter was not able to prevent being imitated—i.e., there was no excludability of the technology or intellectual property protection. To bring excludability to this environment with adopters and innovators we model licensing, in which an adopting firm must pay a fee to the technology holder in order to adopt it.

The licensing fee is a fraction of the present discounted value of adopting the technology, paid up front in a lump sum. Firms bargain to determine the size of the licensing fee. The outside option of the adopting firm (i.e., the licensee) is to reject the bargain, pay no fee, and continue on with its existing technology—i.e., $v(0)$. The outside option of the licensor is simply to reject the offer, receive no fee, and continue on. Negotiations take the form of Nash bargaining, with a bargaining power parameter $\psi \in (0, 1]$ for the adopting firm.[19]

Proposition 2 (Profits and Value Matching with Licensing)**.** *Given firms' equilibrium innovation policy $\gamma(\cdot)$, the aggregate growth rate g, and distributions $F_i(\cdot)$, the flow profit function is*

$$\pi(z) = \underbrace{e^z}_{\textit{Production}} + \underbrace{\left(gF_\ell'(0) + (g - \gamma(0))F_h'(0)\right)}_{\textit{Flow of Licensees}} \underbrace{(1 - \psi)\left(v_\ell(z) - v(0)\right)}_{\textit{Profits per Licensee}}. \tag{32}$$

The value-matching condition for adopting firms is

$$v(0) = \frac{1}{\rho} = \int_0^\infty v_\ell(z)\mathrm{d}F(z) - \frac{\zeta}{\psi}. \tag{33}$$

Proof. See Appendix B.2. □

Full bargaining power to the licensor ($\psi = 1$) nests the baseline case without excludability—i.e., $\pi(z) = e^z$ and the cost of adoption is ζ. The value-matching condition reflects that adopters do not gain the full surplus from the newly adopted technology by increasing the effective cost of search to ζ/ψ.

Thus, from an adopter's perspective, the problems with and without license fees are identical, except for a change in the effective cost of adoption and a modification of the post-adoption continuation value of potentially becoming a licensor in the future. The two environments are quite different for the innovator, however, as license fees provide an extra incentive to innovate.

[19] See Hopenhayn and Shi (2020) for a closely related model of licensing. It provides a richer model of bargaining over technology transfers, including a search-and-matching style congestion and creative destruction. Also related is Luttmer (2015a), which provides a model with assignment between teachers and students decentralized through a price system.

Endogenous Innovation. A firm in the innovative state can choose its own growth rate $\gamma \geq 0$ subject to a convex cost proportional to its current z. Let $\chi > 0$ be the productivity of its innovation technology and the cost be quadratic in the growth rate γ. h-type firms will choose an optimal innovation rate $\gamma(z)$ by considering the effect of innovation on the profits from production and licensing, given by $\pi(z)$ in equation (B.24), and the timing of technology diffusion.

With endogenous innovation and licensing, the Bellman equations (18) and (19) become

$$\rho v_\ell(z) = \pi(z) - g v_\ell'(z) + \lambda_\ell(v_h(z) - v_\ell(z)) + \eta(v_\ell(\bar{z}) - v_\ell(z)) \tag{34}$$

$$\rho v_h(z) = \max_{\gamma \geq 0} \left\{ \pi(z) - \underbrace{(g-\gamma)}_{\text{Innovation Drift}} v_h'(z) - \underbrace{\frac{1}{\chi} e^z \gamma^2}_{\text{R\&D cost}} + \lambda_h(v_\ell(z) - v_h(z)) + \eta(v_\ell(\bar{z}) - v_h(z)) \right\} \tag{35}$$

Previously, the smooth-pasting condition was not a necessary condition for h-type firms because they never crossed the adoption boundary on the BGP ($g \leq \gamma$). Now, given that $\gamma(0) < g$ is possible, the h-type smooth-pasting condition may be necessary (see Appendix A.5 for more on this). Consequently,

$$v_\ell'(0) = 0 \text{ if } g > 0 \tag{36}$$

$$v_h'(0) = 0 \text{ if } g > \gamma(0). \tag{37}$$

The laws of motion in equations (12) and (13) also need to take into account the state-dependent $\gamma(z)$, and the possibility that h-type firms may cross the lower boundary if $g > \gamma(0)$, resulting in

$$0 = gF_\ell'(z) + \lambda_h F_h(z) - \lambda_\ell F_\ell(z) - \eta F_\ell(z) + \eta \mathbb{H}(z - \bar{z}) + (S_\ell + S_h)F(z) - S_\ell \tag{38}$$

$$0 = (g - \gamma(z))F_h'(z) + \lambda_\ell F_\ell(z) - \lambda_h F_h(z) - \eta F_h(z) - S_h \tag{39}$$

$$0 = F_\ell(0) = F_h(0) \tag{40}$$

$$1 = F_\ell(\bar{z}) + F_h(\bar{z}) \tag{41}$$

$$S_\ell = gF_\ell'(0) \text{ if } g > 0 \tag{42}$$

$$S_h = (g - \gamma(0))\, F_h'(0) \text{ if } g > \gamma(0) \tag{43}$$

Summary of Equations and Numerical Methods. The endogenous innovation model is given by equations (34) to (43), which include the Bellman equations, value-matching conditions, smooth-pasting conditions, and KFEs. With endogenous growth, the need to jointly solve the nonlinear Hamilton–Jacobi–Bellman equations and the Kolmogorov forward equations necessitates numerical methods. The problem takes the form of a set of ODEs with parameters constrained by equilibrium conditions that are themselves functions of the solutions to the ODEs. We compute the equilibrium using numerical methods based on spectral collocation and quadrature.[20]

4.2 The Option Value of Adoption Affects Long-run Aggregate Growth

To focus on the first case in which adoption can affect long-run growth rates via the option value, in this section we shut off the second mechanism (licensing). That is, firms' only source of profits is production (i.e., $\pi(z) = e^z$) and there is no licensing cost paid by adopters (i.e., $\psi = 1$).

Compared to the exogenous innovation case, the key additional necessary equilibrium condition in the endogenous growth model is the first-order condition of the value function equation (35) with respect to $\gamma(z)$, using $\pi(z) = e^z$ in this no-licensing case. The FOC is

$$\gamma(z) = \tfrac{\chi}{2} e^{-z} v_h'(z). \tag{44}$$

With equation (44), it can be shown that the innovation rate is increasing in productivity—i.e., $\gamma'(z) > 0$. Consider this rate at the adoption boundary $z = 0$ to see that $\gamma(0) = v_h'(0)$. Then, $\gamma(0) = 0$ according to the smooth-pasting equation (37). The intuition is that since the firm is right next to the adoption barrier, there is no additional value in increasing its productivity marginally because it will adopt a new technology immediately.

This demonstrates a tradeoff in firms' innovation decisions: Investing more in innovation grows their productivity and increases their profits, but firms with higher productivity are further from the adoption threshold, and thus innovation decreases the option value of adoption. Since the option value of adoption is a larger component of total value for lower-productivity firms, the lower-productivity firms invest less in innovation. Intuitively, for a firm just above the adoption threshold, why invest in innovation to get an incremental improvement when it can save the cost of innovating and, instead, adopt a technology that is discretely better in expectation than the one it is currently using? Of course, the cost of adopting and innovating will jointly determine this adoption threshold.

[20] The numerical solution technique uses a simple trick: line up the collocation nodes for the function approximation with those of the quadrature nodes for calculating expectations and equilibrium conditions. After everything is lined up, every equation in the model can be naively stacked, including the Bellman equations, KFEs, equilibrium conditions, etc., into a single nonlinear system of equations, and solved without any nested fixed points. In practice, this requires using a high-performance solver and auto-differentiation, but it is easy to implement and reasonably fast.

Proposition 3 (BGP Equilibrium with Endogenous Innovation and Finite Support)**.** *The endogenous innovation choice is such that $\gamma(0) = 0$ and $\gamma(\bar{z}) = g$. A continuum of equilibria exist, parameterized by $\bar{z}$.*

Proof. See Appendix B.1. □

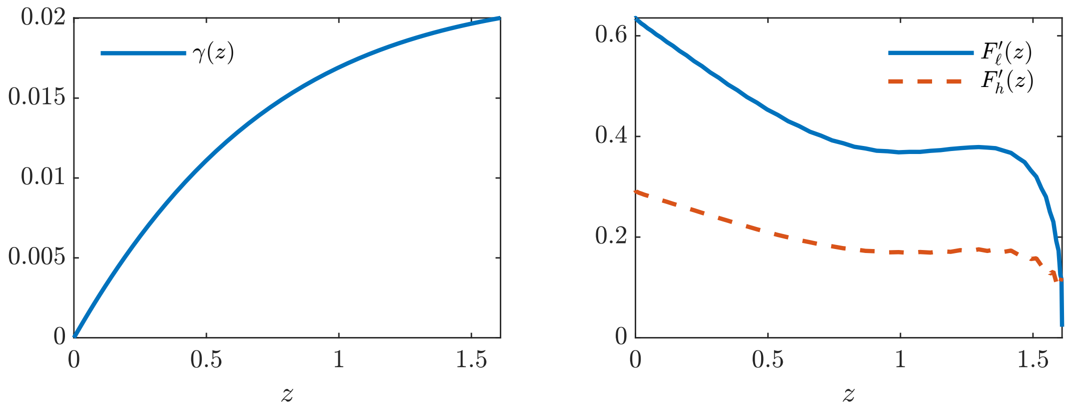

Figure 6: Endogenous $\gamma(z)$ and $F'_i(z)$

An example of the optimal innovation policy and productivity distributions for an endogenous growth BGP is shown in Figure 6.

In the endogenous innovation case, there are a continuum of equilibria indexed by the frontier $\bar{z}$, each with an associated aggregate growth rate $g(\bar{z})$. Compared to Stokey (2014), who features a similar innovation process but differs in the treatment of adoption, here the endogenous choice of γ is complicated by the option value of adoption. Different distributions and associated $\bar{z}$ induce different option values and allow for a continuum of self-fulfilling $\gamma(\bar{z})$. That is, a smaller $\bar{z}$ increases the option value of adoption for innovators at the frontier, which is a disincentive to innovate; this leads to less innovation at the frontier, which, consistently, generates a smaller $\bar{z}$. The hysteresis comes from economic forces (rather than technical properties of stochastic processes or power-law initial conditions) and is due to a complementarity between firms' decisions and the shape of the distribution.

Figure 7 plots the growth rate as a function of the frontier. This figure illustrates the intuition that, because of the self-fulfilling balancing of innovation incentives and frontier location, lower equilibrium values of $\bar{z}$ are associated with lower equilibrium aggregate growth rates. The smaller the relative frontier, the larger the option value of adoption at the frontier, and the lower the incentive to push out the frontier by innovation.

Figure 8 plots the maximum aggregate growth rate from the set of equilibrium g as a function of η. The maximum possible growth rate is a decreasing function of η. With more jumps to the frontier, the distribution becomes more compressed, as discussed in the exogenous innovation case in Section 3 and depicted in Figure 5. As the growth rate of the frontier is determined by the innovation decision of firms with productivity $\bar{z}$, the more

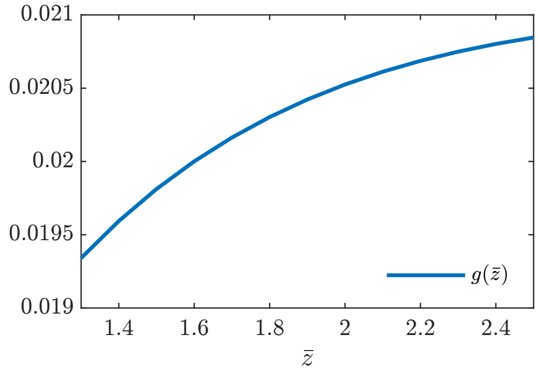

Figure 7: Equilibrium g as a Function of $\bar{z}$

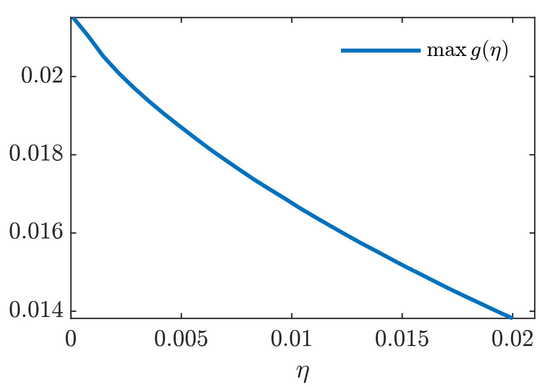

Figure 8: Maximum Equilibrium $g(\eta)$

compressed is the distribution, the lower is the innovation rate, for the same reasons that $g(\bar{z})$ is increasing in Figure 7.

Because of this relationship between the location of the frontier and aggregate growth, whenever $\bar{z} < \infty$, (i.e., when $\eta > 0$), differences in the cost of technology adoption lead to differences in the aggregate growth rate. Recall from Figure 5 that in the exogenous innovation case, a lower adoption cost results in a smaller $\bar{z}$. Easier adoption compresses the distribution by making it easier for laggard firms to keep up with fast-growing innovators. In the endogenous innovation case, this change in the location of the frontier affects innovation at the frontier, and thus the aggregate growth rate.

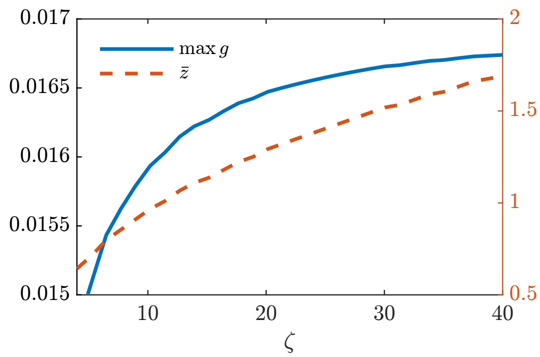

Figure 9: How Aggregate Growth Varies with Adoption Costs Leapfrogging (no licensing $\psi = 1$)

Figure 9 plots comparative statics on the adoption cost ζ in the model with $\bar{z} < \infty$ (i.e., $\eta > 0$). This figure shows that an increase in the adoption cost can increase the aggregate growth rate, because it decreases the free-riding incentive for innovators. A government policy interpretation of this comparative static is that a subsidy to technology adoption financed by lump-sum taxes would decrease the aggregate growth rate. The effect of the subsidy on aggregate growth is stronger when frontier firms are closer to the least-productive firms in the economy.

Comparison to Models with Exogenous Innovation. The results in this section are in contrast to Perla and Tonetti (2014) and the baseline model in Lucas and Moll (2014), where long-run growth is determined by initial conditions. Those papers can be interpreted as modeling how adoption can generate growth in the medium run.

It is also distinct from models with exogenous innovation and diffusion modeled as geometric Brownian motion, such as Staley (2011) and Luttmer (2012b). In those models, for the case of a large number of agents, taking the innovation meeting rate to infinity leads to unbounded idiosyncratic growth rates. Here we get the opposite relationship between adoption rates and growth, in part due to adoption and innovation being choices; decreasing the cost of technology diffusion can decrease the growth rate, as it induces more free-riding for innovators. Another key difference in our setup is that the finite-state Markov process and endogenous investment in innovation at convex cost conspire to yield a finite upper bound on firm-level growth rates.

Endogenous Innovation with $\boldsymbol{\eta \approx 0}$. To emphasize that it is the option value of adoption that generates the link between adoption, innovation, and aggregate growth, we take a short detour to study the case in which $\bar{z} = \infty$ asymptotically, which occurs in the limit as $\eta \to 0$.

Proposition 4 (Endogenous Innovation with $\eta \approx 0$)**.** *For $\eta \to 0$, a BGP equilibrium exists such that $\lim_{t\to\infty} \bar{z}(t) = \infty$ and the unique long-run aggregate growth rate is the solution to the cubic equation*

$$g\left(g^2 + g(2\lambda_h + \lambda_\ell + 3\rho) + 2\rho(\lambda_h + \lambda_\ell + \rho)\right) = \chi(g + \lambda_h + \lambda_\ell + \rho). \tag{45}$$

The endogenous innovation choice is such that $\gamma(0) = 0$ and $\lim_{\bar{z}\to\infty} \gamma(\bar{z}) = g$.

Proof. See Appendix B.1. ☐

As $\eta \to 0$, the number of jumps to the frontier approaches 0, $\bar{z}$ approaches ∞, and the model studied in Proposition 3 converges to the model studied in Proposition 4.

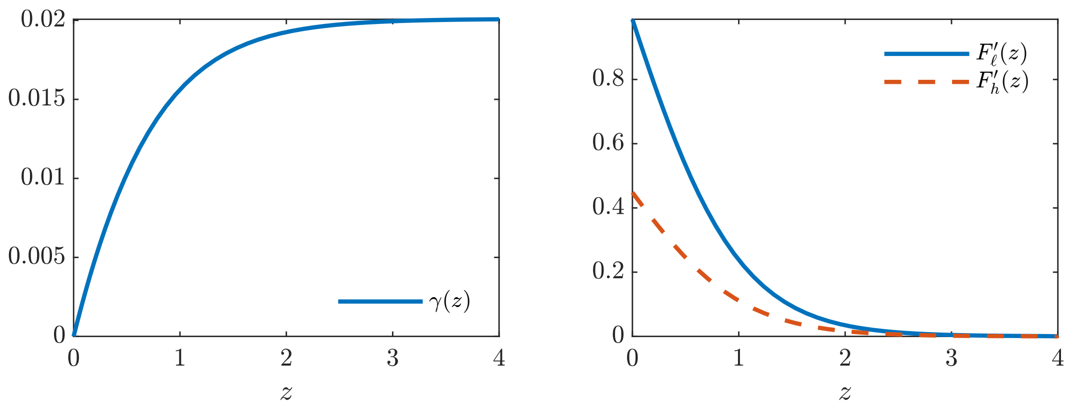

Figure 10: Endogenous $\gamma(z)$, $F_i'(z)$, and $\alpha(z)$ with $\eta \approx 0$

Whenever innovation is a choice, the endogenous aggregate growth rate is the growth rate chosen by innovators at the frontier. In contrast to the $\bar{z} < \infty$ case, when $\eta \to 0$ so that $\bar{z} = \infty$, the aggregate growth rate is independent of the cost of adoption. Note that the cost of adoption does not appear in equation (45). This is because innovating firms at the frontier have zero option value from adoption; therefore, changes in the cost or benefits of adoption do not alter their innovation behavior. For intuition, see that the option value term vanishes for large z in equation (31) (i.e., $v_l(z) \propto e^z$ for $z \to \infty$). The key lesson is that when frontier innovators internalize that they may one day be adopters, the costs and benefits of adoption affect their investments in innovation, which in turn affect aggregate growth.

This result that the cost of adoption affects the aggregate growth rate only through affecting the option value of adoption at the frontier holds when technology is not excludable ($\psi = 1$). We now turn to analyzing the case of excludability and licensing, in which a subsidy to adoption can have a very different effect on innovation and aggregate growth.

4.3 Licensing and Partial Excludability

In order to isolate the effect of licensing on aggregate growth, we focus on the $\eta \approx 0$ case, where Proposition 4 stated that, absent licensing, changes in the cost of adoption do not affect the aggregate growth rate.

Analysis of Flow Profits. To study how licensing affects innovation incentives we differentiate the profit function in equation (B.24):

$$\pi'(z) = e^z + (1-\psi)gF'(0)v_\ell'(z). \tag{46}$$

On the margin, increasing firm productivity z increases profits for two reasons. The first term e^z is the marginal increase in profits from an increase in production. The second term is the increase in profits from an increase in licensing revenue. Since the value function is increasing in productivity, the second term is positive, so licensing provides a positive incentive to innovate. Furthermore, since the value function is convex, licensing provides stronger incentives to innovate for higher productivity firms. The profits from licensing disappear as $z \to 0$ because the surplus from adopting a technology close to the adoption boundary goes to 0. Consequently, $\pi'(0) = 1$. Finally, with licensing, profits become a function of g. Faster growth means more adopters given a fixed $F(z)$, and $F(z)$ and $v_\ell(z)$ are also themselves functions of g.

Role of Excludability. Figure 11 plots the aggregate growth rate as a function of the excludability parameter ψ. When excludability is not too strong, the aggregate growth rate is increasing in the degree of excludability (1-ψ) (i.e., growth increases with weaker bargaining power for the adopter). The increase in the aggregate growth rate is due to the added incentive to invest in growing via innovation, as higher-productivity firms gain extra profits from licensing the better technology to adopting firms.

There is, however, a countervailing force that dominates when excludability is already strong. If the licensor's bargaining power is too strong, the incentive to adopt technologies becomes too small. Consequently, fewer firms adopt new technologies, ultimately generating less licensing revenue. Lower licensing revenue decreases the returns to innovation for all firms, including those near the frontier that determine the aggregate growth rate.

To give a sense of the shape of the distribution, we plot the Gini index. For a wide range of the parameter values, increasing excludability increases innovation activity and generates a more unequal distribution. This shows a trade-off between productivity inequality and aggregate growth rates. This positive association between productivity dispersion and the aggregate growth rate operates through innovation activity, compared to the typical link in the idea diffusion literature (cf., Perla and Tonetti (2014)) that is driven by adoption incentives.

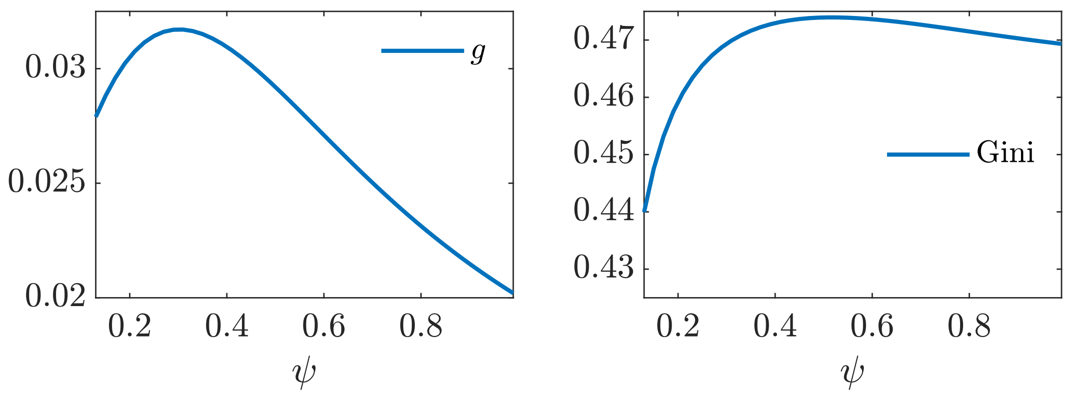

Figure 11: Growth and Distribution Shape under Excludability and Licensing

Figure 12 presents another perspective on the role of excludability by plotting the aggregate growth rate as a function of the adoption cost for various values of ψ. The model used to generate Figure 12 is identical to the one discussed in Proposition 4 ($\eta \approx 0, \bar{z} = \infty$), except ψ is no longer equal to 1. In the absence of excludability (i.e., $\psi = 1$), adoption costs can change the shape of the distribution, but they have no impact on the aggregate growth rate. When $\psi = 1$, the option value of adoption is infinitesimal for the highest-productivity firms that make the innovation decision at the frontier, so changes in the cost of adoption do not affect aggregate growth. With a strong degree of excludability, however, lower adoption costs drive higher aggregate growth, even in this case with $\eta \approx 0$. While the option value of adoption for firms at the frontier is still infinitesimal, an innovating firm gains extra profits by licensing to adopting firms, and there are more adopting firms when adoption costs are lower.

The Gini coefficient is decreasing in the cost of adoption, but only modestly. The reason is that the shape of the distribution near the adoption threshold is impacted by the large mass of agents there, which is mostly determined by the innovation decisions of those lower-productivity firms rather than by the frontier innovation rates.

Finally, we consider a variation on the experiment in Figure 9, which showed that when ($\eta > 0$, $\bar{z} < \infty$), without licensing ($\psi = 1$) a decrease in the cost of technology adoption de-

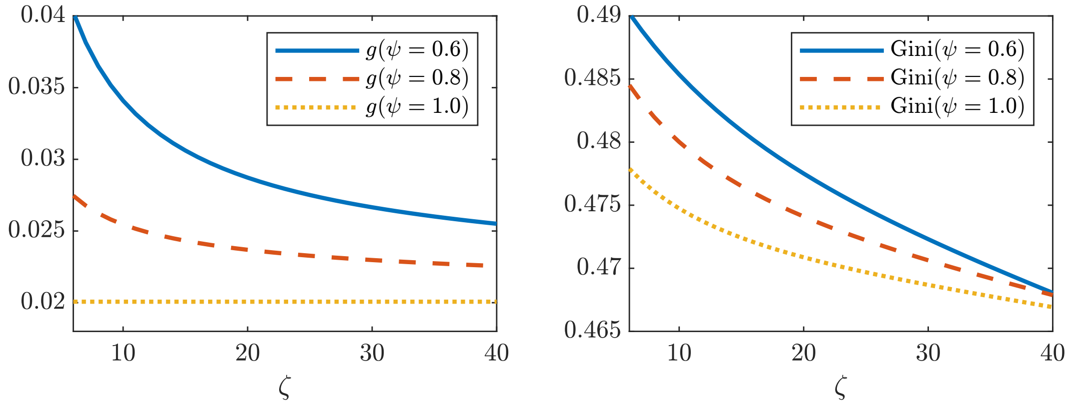

Figure 12: Interaction of Licensing and Adoption Costs

creased the aggregate growth rate (by strengthening the free-riding incentive for innovators). In Figure 13 we plot the same reduction in adoption costs, but in a model with licensing ($\psi = 0.5$) and with positive option value of adoption at the frontier. Figure 13 shows that licensing can overturn the negative relationship between adoption costs and aggregate growth rates, such that subsidizing technology adoption increases the aggregate growth rate.

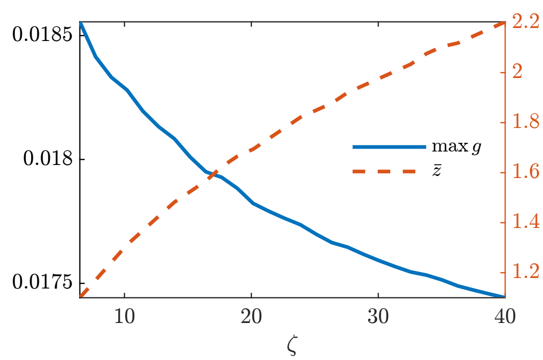

Figure 13: Comparative Statics on Adoption Cost with Option Value of Adoption and Licensing ($\eta = 0.01$ and $\psi = 0.5$)

5 Conclusion

This paper develops a theory of the shape of the productivity distribution and of how productivity improves over time, which generates long-run aggregate growth. Firms make choices to invest in adoption and innovation, and the balance of these two activities across heterogeneous firms determines aggregate outcomes. Adoption is a force that compresses the distribution, helping laggards keep up with an expanding frontier. Innovation is a force that stretches the distribution, pushing out the frontier. Balanced growth path equilibria with stationary normalized distributions exist. On a BGP, a lower cost of adoption creates a thinner-tailed distribution with a smaller distance between best and worst firms. Easier innovation has an opposite effect, generating thicker tails and a larger range in productivity. More firms leapfrogging to the frontier generates thicker tails, but, because it helps laggards keep up with innovators at the frontier, it shrinks the distance from the bottom to the top of the distribution.

In addition to interacting to determine the shape of the distribution, adoption and innovation combine to generate aggregate growth. On a BGP, the long-run aggregate growth rate is the maximum growth rate chosen by innovating firms. If there were no innovation, there would be no long-run growth. In this sense, innovation is the driver of growth. Adoption, however, affects long-run growth by affecting the incentives to innovate. In equilibrium, low-productivity firms choose to adopt and high-productivity firms focus on innovation. Thus, growing through innovation increases the expected time until a firm becomes an adopter. Because innovators may one day become adopters there is an option value of adoption. If this option value is large, e.g., because adoption is low cost, then innovators may be tempted to free ride by investing less in pushing out the frontier, content to fall back to the adoption threshold faster. This complementarity between the distance to the frontier and the incentive to innovate generates multiplicity of BGP. If the initial distribution has a small distance to the frontier, the option value of adoption is high, and the incentives to push out the frontier are low. Thus, the small distance to the frontier is self-reinforcing. In this sense, a subsidy to adoption can reduce aggregate growth.

Adoption, on the other hand, may be a force to increase aggregate growth. When adopting firms must pay a licensing fee to the higher-productivity firm from which they are adopting, there is an extra incentive to innovate. Innovation increases productivity, and profits from licensing a high-productivity technology are larger than profits from licensing a mediocre technology. Thus, more adopters induces more innovation, increasing aggregate growth. Furthermore, there is a growth-maximizing level of excludability, parameterized by the bargaining power of adopters, that balances the amount of adopters and the profits per adopter. Too strong of a bargaining power for adopters limits the innovation incentive provided by the profits from licensing. Too weak of a bargaining power for adopters raises the effective cost of adoption to the point that there are too few adopters.

In sum, the model illustrates how firm choices to adopt and innovate intertwine to generate aggregate and cross-sectional effects on productivity in equilibrium in a growing economy.

References

Acemoglu, D., P. Aghion, C. Lelarge, J. Van Reenen, and F. Zilibotti (2007): “Technology, Information, and the Decentralization of the Firm,” *The Quarterly Journal of Economics*, 122, 1759–1799.

Acemoglu, D., P. Aghion, and F. Zilibotti (2006): “Distance to Frontier, Selection and Economic Growth,” *Journal of the European Economic Association*, 4, 37–74.

Acemoglu, D. and D. Cao (2015): “Innovation by Entrants and Incumbents,” *Journal of Economic Theory*, 157, 255–294.

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2017): “Income and wealth distribution in macroeconomics: A continuous-time approach,” Tech. rep., National Bureau of Economic Research.

Alvarez, F. E., F. J. Buera, and R. E. Lucas (2008): “Models of Idea Flows,” *NBER WP 14135*, 1–12.

——— (2013): “Idea Flows, Economic Growth, and Trade,” Working Paper 19667, National Bureau of Economic Research.

Atkeson, A. and A. T. Burstein (2010): “Innovation, Firm Dynamics, and International Trade,” *Journal of Political Economy*, 118, 433–484.

Benhabib, J., J. Perla, and C. Tonetti (2014): “Catch-up and fall-back through innovation and imitation,” *Journal of Economic Growth*, 19, 1–35.

Buera, F. J. and R. E. Lucas (2018): “Idea Flows and Economic Growth,” *Annual Review of Economics*, 10, 315–345.

Buera, F. J. and E. Oberfield (2019): “The Global Diffusion of Ideas,” *Econometrica (Forthcoming)*.

Chu, A. C., G. Cozzi, and S. Galli (2014): “Stage-dependent intellectual property rights,” *Journal of Development Economics*, 106, 239–249.

Eeckhout, J. and B. Jovanovic (2002): “Knowledge Spillovers and Inequality,” *American Economic Review*, 92, 1290–1307.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016): “The dynamics of inequality,” *Econometrica*, 84, 2071–2111.

Hopenhayn, H. and L. Shi (2020): “Knowledge Creation and Diffusion with Limited Appropriation,” *working paper*.

Hsieh, C.-T. and P. J. Klenow (2009): “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 124, 1403–1448.

Jovanovic, B. and G. M. MacDonald (1994): “Competitive Diffusion,” *Journal of Political Economy*, 102, 24–52.

Jovanovic, B. and R. Rob (1989): “The Growth and Diffusion of Knowledge,” *Review of Economic Studies*, 56, 569–582.

Jovanovic, B. and Z. Wang (2020): “Idea Diffusion and Property Rights,” *working paper*.

Kemmerer, J. E. and J. Lu (2012): “Profitability and royalty rates across industries: Some preliminary evidence,” Tech. rep., KPMG Global Valuation Institute.

König, M., J. Lorenz, and F. Zilibotti (2016): “Innovation vs imitation and the evolution of productivity distributions,” *Theoretical Economics*, 1053–1102.

König, M., K. Storesletten, Z. Song, and F. Zilibotti (2020): “From Imitation to Innovation: Where Is All That Chinese R&D Going?” *working paper*.

Lashkari, D. (2016): “Innovation, Knowledge Diffusion, and Selection,” *Working Paper*.

Lucas, R. E. (2009): “Ideas and Growth,” *Economica*, 76, 1–19.

Lucas, R. E. and B. Moll (2014): “Knowledge Growth and the Allocation of Time,” *Journal of Political Economy*, 122, pp. 1–51.

Luttmer, E. G. (2007): “Selection, Growth, and the Size Distribution of Firms,” *The Quarterly Journal of Economics*, 122, 1103–1144.

——— (2012a): “Eventually, Noise and Imitation Implies Balanced Growth,” *Federal Reserve Bank of Minneapolis Working Paper 699*.

——— (2012b): “Technology diffusion and growth,” *Journal of Economic Theory*, 147, 602 – 622.

——— (2015a): “An Assignment Model of Knowledge Diffusion and Income Inequality,” Tech. Rep. 509, Federal Reserve Bank of Minneapolis.

——— (2015b): “Four Models of Knowledge Diffusion and Growth,” Tech. rep., Federal Reserve Bank of Minneapolis.

Luttmer, E. G. J. (2020): "Bounded Learning from Incumbent Firms," *working paper.*

Nelson, R. R. and E. S. Phelps (1966): "Investment in Humans, Technological Diffusion, and Economic Growth," *American Economic Review*, 56, 69–75.

Perla, J. and C. Tonetti (2014): "Equilibrium Imitation and Growth," *Journal of Political Economy*, 122, pp. 52–76.

Perla, J., C. Tonetti, and M. E. Waugh (2021): "Equilibrium Technology Diffusion, Trade, and Growth," *American Economic Review*, 111, 73–128.

Sampson, T. (2015): "Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth," *The Quarterly Journal of Economics.*

Staley, M. (2011): "Growth and the diffusion of ideas," *Journal of Mathematical Economics*, 47, 470–478.

Stokey, N. L. (2014): "The Race Between Technology and Human Capital," *Working Paper.*

——— (2017): "Technology, Skill, and Long-Run Growth," *Working Paper.*

Syverson, C. (2011): "What Determines Productivity?" *Journal of Economic Literature*, 49, 326–65.

A Exogenous Innovation

This section derives the equilibrium conditions for the model in which innovation is an exogenous process.

A.1 Normalization

Define the normalized productivity distribution, as the productivity distribution relative to the endogenous adoption threshold $M(t)$:

$$\Phi_i(t, Z) \equiv F_i(t, \log(Z/M(t))). \tag{A.1}$$

Differentiate to obtain the pdf

$$\boldsymbol{\partial}_Z \Phi_i(t, Z) = \frac{1}{Z}\frac{\partial F_i(t, \log(Z/M(t)))}{\partial z} = \frac{1}{Z}\boldsymbol{\partial}_z F_i(t, z). \tag{A.2}$$

Differentiate (A.1) with respect to t and use the chain rule to obtain the transformation of the time derivative

$$\boldsymbol{\partial}_t \Phi_i(t, Z) = \frac{\partial F_i(t, \log(Z/M(t)))}{\partial t} - \frac{M'(t)}{M(t)} \frac{\partial_i F(t, \log(Z/M(t)))}{\partial z}. \tag{A.3}$$

Use the definition $g(t) \equiv M'(t)/M(t)$ and the definition of z,

$$\boldsymbol{\partial}_t \Phi_i(t, Z) = \boldsymbol{\partial}_t F_i(t, z) - g(t) \boldsymbol{\partial}_z F_i(t, z). \tag{A.4}$$

Normalizing the Law of Motion This is derived with a more general adoption process, where $\hat{F}_i(t, z)$ is the mass of agents who draw a technology below z with the innovation state i. In our baseline case, $\hat{F}_\ell(t, z) = F(t, z)$ and $\hat{F}_h(t, z) = 0$.

Substitute (A.2) and (A.4) into (7) and (8),

$$\begin{aligned}\frac{\partial F_\ell(t, \log(Z/M(t)))}{\partial t} &= g(t) \frac{\partial F_\ell(t, \log(Z/M(t)))}{\partial z} - \lambda_\ell F_\ell(t, \log(Z/M(t))) + \lambda_h F_h(t, \log(Z/M(t))) \\ &+ (S_\ell(t) + S_h(t))\, \hat{F}_\ell(t, \log(Z/M(t))) - S_\ell(t) \\ &- \eta F_\ell(t, \log(Z/M(t))) + \eta \mathbb{H} \left(\log(Z/M(t)) - \log(\bar{Z}(t)/M(t))\right),\end{aligned} \tag{A.5}$$

$$\begin{aligned}\frac{\partial F_h(t, \log(Z/M(t)))}{\partial t} &= g(t) \frac{\partial F_h(t, \log(Z/M(t)))}{\partial z} - \lambda_h F_h(t, \log(Z/M(t))) + \lambda_\ell F_\ell(t, \log(Z/M(t))) \\ &- \gamma \frac{Z}{Z} \frac{\partial F_h(t, \log(Z/M(t)))}{\partial z} + (S_\ell(t) + S_h(t))\, \hat{F}_h(t, \log(Z/M(t))) - S_h(t) \\ &- \eta F_h(t, \log(Z/M(t))).\end{aligned} \tag{A.6}$$

Use the definition of z and reorganize to find the normalized KFEs,

$$\begin{aligned}\boldsymbol{\partial}_t F_\ell(t, z) &= -\lambda_\ell F_\ell(t, z) + \lambda_h F_h(t, z) + g(t) \boldsymbol{\partial}_z F_\ell(t, z) + S(t) \hat{F}_\ell(t, z) - S_\ell(t) \\ &\quad - \eta F_\ell(t, z) + \eta \mathbb{H}\left(z - \bar{z}(t)\right),\end{aligned} \tag{A.7}$$

$$\boldsymbol{\partial}_t F_h(t, z) = \lambda_\ell F_\ell(t, z) - \lambda_h F_h(t, z) + (g(t) - \gamma) \boldsymbol{\partial}_z F_h(t, z) + S(t) \hat{F}_h(t, z) - S_h(t) - \eta F_h(t, z). \tag{A.8}$$

where the domain of the normalized KFE at time t is $[0, \bar{z}(t)]$. Recall that the unnormalized flux (assuming $M'(t) \geq \gamma M(t)$) is

$$S_\ell(t) \equiv M'(t) \boldsymbol{\partial}_Z \Phi_\ell(t, M(t)), \tag{A.9}$$

$$S_h(t) \equiv \underbrace{(M'(t) - \gamma M(t))}_{\text{Relative Speed of Boundary}} \underbrace{\boldsymbol{\partial}_Z \Phi_h(t, M(t))}_{\text{PDF at boundary}}. \tag{A.10}$$

This is consistent with the solution to the ODEs in equations (7) and (8) at $Z = M(t)$. Normalizing and substituting from (A.2),

$$S_\ell(t) = g(t)\boldsymbol{\partial}_z F_\ell(t,0), \tag{A.11}$$

$$S_h(t) = (g(t)-\gamma)\boldsymbol{\partial}_z F_h(t,0). \tag{A.12}$$

Normalizing the Value Function Define the normalized value of the firm as,

$$v_i(t, \log(Z/M(t))) \equiv \frac{V_i(t,Z)}{M(t)}. \tag{A.13}$$

Rearrange and differentiate (A.13) with respect to t

$$\begin{aligned}\boldsymbol{\partial}_t V_i(t,Z) = M'(t)v_i(t,\log(Z/M(t))) - M'(t)\frac{\partial v_i(t,\log(Z/M(t))}{\partial z} \\ + M(t)\frac{\partial v_i(t,\log(Z/M(t))}{\partial t}.\end{aligned} \tag{A.14}$$

Divide by $M(t)$ and use the definition of $g(t)$

$$\frac{1}{M(t)}\boldsymbol{\partial}_t V_i(t,Z) = g(t)v_i(t,z) - g(t)\boldsymbol{\partial}_z v_i(t,z) + \boldsymbol{\partial}_t v_i(t,z). \tag{A.15}$$

Differentiate (A.13) with respect to Z and rearrange

$$\frac{1}{M(t)}\boldsymbol{\partial}_Z V_i(t,Z) = \frac{1}{Z}\boldsymbol{\partial}_z v_i(t,z). \tag{A.16}$$

Divide (2) by $M(t)$ and then substitute from (A.15) and (A.16),

$$\begin{aligned}r\frac{1}{M(t)}V_h(t,Z) = \frac{Z}{M(t)} + \gamma\frac{M(t)}{M(t)}\frac{Z}{Z}\boldsymbol{\partial}_z v_h(t,z) + g(t)v_h(t,z) - g(t)\boldsymbol{\partial}_z v_h(t,z) \\ + \lambda_h(v_\ell(t,z) - v_h(t,z)) + \frac{\eta}{M(t)}(V_\ell(t,\bar{Z}(t)) - V_h(t,Z)) + \boldsymbol{\partial}_t v_h(t,z).\end{aligned} \tag{A.17}$$

Use (A.13) and the definition of z and rearrange,

$$\begin{aligned}(r - g(t))v_h(t,z) = e^z + (\gamma - g(t))\boldsymbol{\partial}_z v_h(t,z) + \lambda_h(v_\ell(t,z) - v_h(t,z)) \\ + \eta(v_\ell(t,\bar{z}(t)) - v_h(t,z)) + \boldsymbol{\partial}_t v_h(t,z).\end{aligned} \tag{A.18}$$

Similarly, for (1)

$$
\begin{aligned}
(r - g(t))v_\ell(t,z) = e^z - g(t)\boldsymbol{\partial}_z v_\ell(t,z) + \lambda_\ell(v_h(t,z) - v_\ell(t,z)) \\
+ \eta(v_\ell(t,\bar{z}(t)) - v_\ell(t,z)) + \boldsymbol{\partial}_t v_\ell(t,z).
\end{aligned}
\tag{A.19}
$$

Optimal Stopping Conditions Recall that we are solving the general problem with draws to $\hat{\Phi}_i(t,Z)$ which nests our benchmark model. Divide the value-matching condition in (4) by $M(t)$,

$$
\frac{V_i(t,M(t))}{M(t)} = \int_{M(t)}^{\bar{Z}(t)} \frac{V_\ell(t,Z)}{M(t)} \boldsymbol{\partial}_Z \hat{\Phi}_\ell(t,Z)\mathrm{d}Z + \int_{M(t)}^{\bar{Z}(t)} \frac{V_h(t,Z)}{M(t)} \boldsymbol{\partial}_Z \hat{\Phi}_h(t,Z)\mathrm{d}Z - \frac{M(t)}{M(t)}\zeta.
\tag{A.20}
$$

Use the substitutions in (A.2) and (A.13), and a change of variable $z = \log(Z/M(t))$ in the integral, which implies that $\mathrm{d}z = \frac{1}{Z}\mathrm{d}Z$. Note that the bounds of integration change to $[\log(M(t)/M(t)), \log(\bar{Z}(t)/M(t))] = [0, \bar{z}(t)]$

$$
v_i(t,0) = \int_0^{\bar{z}(t)} v_\ell(t,z)\mathrm{d}\hat{F}_\ell(t,z) + \int_0^{\bar{z}(t)} v_h(t,z)\mathrm{d}\hat{F}_h(t,z) - \zeta.
\tag{A.21}
$$

Evaluate (A.16) at $Z = M(t)$, and substitute this into (6) to give the smooth-pasting condition

$$
\boldsymbol{\partial}_z v_i(t,0) = 0.
\tag{A.22}
$$

A.2 Proposition 1

Rather than use the interest rate that is consistent with a representative consumer who has log utility, as in the body of the paper, here in the appendix we allow for a general firm discount rate r.

Proof of Proposition 1. Define the following to simplify notation,

$$
\alpha \equiv (1+\hat{\lambda})\frac{S-\eta}{g},
\tag{A.23}
$$

$$
\hat{\lambda} \equiv \frac{\lambda_\ell}{\eta + \lambda_h},
\tag{A.24}
$$

$$
\bar{\lambda} \equiv \frac{r - g + \lambda_\ell + \lambda_h}{r - g + \lambda_h},
\tag{A.25}
$$

$$
\nu = \frac{r - g + \eta}{g}\bar{\lambda}.
\tag{A.26}
$$

Solve for $F_h(z)$ in (13),

$$F_h(z) = \hat{\lambda} F_\ell(z). \tag{A.27}$$

Substitute this back into (12) to get an ODE in F_ℓ

$$0 = gF'_\ell(z) + (S - \eta)(1 + \hat{\lambda})F_\ell(z) + \eta \mathbb{H}\left(z - \bar{z}\right) - S. \tag{A.28}$$

Solve this ODE using $F_\ell(0) = 0$

$$F_\ell(z) = \begin{cases} \frac{S}{(S-\eta)(1+\hat{\lambda})}(1 - e^{-\alpha z}) & 0 \leq z < \bar{z} \\ \frac{S}{(S-\eta)(1+\hat{\lambda})}(1 - e^{-\alpha \bar{z}}) & z = \bar{z} \end{cases}. \tag{A.29}$$

This function is continuous at $z = \bar{z}$, and therefore so is $F_h(z)$. The unconditional distribution is,

$$F(z) = (1 + \hat{\lambda})F_\ell(\bar{z}), \tag{A.30}$$

$$= \frac{S}{S - \eta}(1 - e^{-\alpha z}). \tag{A.31}$$

Use the boundary condition that $F(\bar{z}) = 1$, and solve for $\bar{z}$ with the assumption that $S > \eta$,

$$\bar{z} = \frac{\log(S/\eta)}{\alpha}. \tag{A.32}$$

The pdf of the unconditional distribution is,

$$F'(z) = \frac{\alpha S}{S - \eta} e^{-\alpha z}. \tag{A.33}$$

To solve for the value, solve (19) for $v_h(z)$,

$$v_h(z) = \frac{e^z + (\lambda_h - \eta)v_\ell(z) + \eta v_\ell(\bar{z})}{r - g + \lambda_h}. \tag{A.34}$$

Substitute into (18) and simplify

$$(r - g + \eta)v_\ell(z) = e^z + \eta v_\ell(\bar{z}) - \frac{g}{\bar{\lambda}} v'_\ell(z). \tag{A.35}$$

Solve (21) with (A.35) and simplify,

$$v_\ell(z) = \frac{\bar{\lambda}}{g + (r + \eta - g)\bar{\lambda}} e^z + \frac{\eta}{r - g + \eta} v_\ell(\bar{z}) + \frac{1}{(r + \eta - g)(\nu + 1)} e^{-z\nu}. \tag{A.36}$$

Evaluate (A.36) at $\bar{z}$ and solve for $v_\ell(\bar{z})$,

$$v_\ell(\bar{z}) = \left(-\frac{\eta}{g - r} + 1\right)\left(\frac{e^{\bar{z}}\bar{\lambda}}{g + (\eta + r - g)\bar{\lambda}} + \frac{e^{-\nu\bar{z}}}{(\eta + r - g)(\nu + 1)}\right). \tag{A.37}$$

Subtitute (A.37) into (A.36) to find an expression for $v_\ell(z)$

$$v_\ell(z) = \frac{\bar{\lambda}}{g(1 + \nu)}\left(e^z + \frac{1}{\nu}e^{-\nu z} + \frac{\eta}{r - g}\left(e^{\bar{z}} + \frac{1}{\nu}e^{-\nu\bar{z}}\right)\right). \tag{A.38}$$

Substitute (A.33) and (A.38) into the value-matching condition in (20) and evaluate the integral,

$$\zeta + \frac{1}{r - g} = \frac{S\alpha\bar{\lambda}\left(-\frac{e^{-\nu\bar{z}}\left(-1+e^{-\alpha\bar{z}}\right)\eta}{(-g+r)\alpha\nu} + \frac{e^{\bar{z}}\eta\left(e^{-\alpha\bar{z}}-1\right)}{\alpha(g-r)} + \frac{-e^{-(\alpha+\nu)\bar{z}}+1}{\nu(\alpha+\nu)} + \frac{-e^{\bar{z}-\alpha\bar{z}}+1}{\alpha-1}\right)}{g(S - \eta)(\nu + 1)}. \tag{A.39}$$

To find an implicit equation for the equilibrium S, take (A.39) and substitute for α and $\bar{z}$ from (A.23) and (A.32).

In the case of log utility: substitute $r = \rho + g, S = \gamma F_\ell'(0)$ and $g = \gamma$ into (A.26), (A.27), (A.34), (A.38) and (A.39) to complete Proposition 1. □

A.3 Distorted and Directed Draws

We can modify the adoption process so that the Z is drawn from a distortion of the unconditional distribution, twisting the distribution so that adopters are either more or less likely to get better technologies compared to drawing from the unconditional distribution F.

The distortion, representing the degree of imperfect mobility or beneficial adoption prospects, is indexed by $\kappa \in (0, \infty)$, where the firm draws its Z from the cdf $\Phi(t, Z)^\kappa$, or $F(z)^\kappa$ after normalization. Note that for higher κ, the probability of a better draw increases, and integer values can be interpreted as the number of draws from the distribution.

With this, the value matching condition integrates over the distorted distribution and

(12) and (20) become

$$v(0) = \frac{1}{\rho} = \int_0^{\bar{z}} v_\ell(z) \; \mathrm{d}F(z)^\kappa - \zeta, \tag{A.40}$$

$$0 = gF_\ell'(z) + \lambda_h F_h(z) - \lambda_\ell F_\ell(z) - \eta F_\ell(z) + \eta \mathbb{H}(z - \bar{z}) + SF(z)^\kappa - S. \tag{A.41}$$

The model remains computationally tractable and is amenable to an extension of firm's choosing κ at a cost.

A.4 No Equilibrium Exists with $g < \gamma$ and $\eta > 0$

In the model discussed in Proposition 1, the fact that $g \leq \gamma$ immediately results from the initial distribution having finite support. Here we demonstrate that there do not exist equilibria with $g < \gamma$ for any $\eta > 0$. Thus, the unique equilibrium growth rate is $g = \gamma$.

Proof of $g = \gamma$ in Proposition 1. To show that there is no stationary equilibrium with $g < \gamma$ and $\eta > 0$, we construct a proof by contradiction: (1) assume there is some constant $g < \gamma$ which is the optimal growth rate; (2) use the constant g to calculate the expectation of the stationary productivity distribution; (3) show that the mean cannot be stationary when normalized relative to the adoption costs, and hence the $g(t)$ cannot have been an optimal choice.

Prior to finding the evolution of the moments, we need to be careful of exactly where agents are removed from the distribution, so replace the S in the CDF with the heaviside function removing at the threshold $\log(M(t)/M(t)) = 0$, i.e., $S\mathbb{H}(z)$. Take the normalized version in (A.7) and (A.8) and instead work in pdfs in order to later take a Laplace transformation. We maintain the assumption of a constant g and S with draws from the unconditional $F(t, \cdot)$. Use that the derivative of the Heaviside is the Dirac delta function, $\boldsymbol{\delta}(\cdot)$, to find,

$$\boldsymbol{\partial}_t f_\ell(t,z) = g\boldsymbol{\partial}_z f_\ell(t,z) + (S - \lambda_\ell - \eta) f_\ell(t,z) + (S + \lambda_h) f_h(t,z) - S\boldsymbol{\delta}(z) + \eta\boldsymbol{\delta}(z - \bar{z}), \tag{A.42}$$

$$\boldsymbol{\partial}_t f_h(t,z) = (g - \gamma)\boldsymbol{\partial}_z f_h(t,z) + \lambda_\ell f_\ell(t,z) - (\lambda_h + \eta) f_h(t,z). \tag{A.43}$$

After differentiating, we have the Dirac delta function, $\boldsymbol{\delta}(z)$, in (A.42) to remove those adopting agents at the threshold (which is normalized to $z = 0$), and the insertion of leapfrogging firms at $\bar{z}$ with arrival rate η.

Taking inspiration from Gabaix et al. (2016), use the bilateral Laplace transform on the z variable to the new ξ space, such that $\mathcal{F}_i(t,\xi) \equiv \int_{-\infty}^{\infty} e^{-\xi z} f_i(t,z)\mathrm{d}z$. Applying this transform to the ODEs in (A.42) and (A.43) gives,[21]

$$\boldsymbol{\partial}_t \mathcal{F}_\ell(t,\xi) = g\xi\mathcal{F}_\ell(t,\xi) + (S - \lambda_\ell - \eta)\mathcal{F}_\ell(t,\xi) + (S + \lambda_h)\mathcal{F}_h(t,\xi) - S + \eta e^{-\bar{z}\xi}, \tag{A.44}$$

$$\boldsymbol{\partial}_t \mathcal{F}_h(t,\xi) = (g-\gamma)\xi\mathcal{F}_h(t,\xi) + \lambda_\ell\mathcal{F}_\ell(t,\xi) - (\eta + \lambda_h)\mathcal{F}_h(t,\xi). \tag{A.45}$$

From Gabaix et al. (2016) equation (16), evaluating at $\xi = -1$ are the moments of the $Z/M(t)$ distribution. Hence, to be a stationary first moment (for a given $\bar{z}$), substitute into a time-invariant (A.44) and (A.45) to find,

$$0 = -g\mathcal{F}_\ell(t,-1) + (S - \lambda_\ell - \eta)\mathcal{F}_\ell(t,-1) + (S + \lambda_h)\mathcal{F}_h(t,-1) - S + \eta e^{\bar{z}}, \tag{A.46}$$

$$0 = -(g-\gamma)\mathcal{F}_h(t,-1) + \lambda_\ell\mathcal{F}_\ell(t,-1) - (\eta + \lambda_h)\mathcal{F}_h(t,-1). \tag{A.47}$$

Solve this algebraic system of equations for $\mathcal{F}_\ell(t,-1)$ and $\mathcal{F}_h(t,-1)$ and then use the linearity of the Laplace transform to find $\mathcal{F}(t,-1) = \mathcal{F}_\ell(t,-1) + \mathcal{F}_h(t,-1)$,

$$\mathbb{E}_t\left[Z/M(t)\right] = \mathcal{F}(t,-1) = \frac{g - \gamma + \eta + \lambda_h + \lambda_\ell}{(g - S + \eta)(g - \gamma + \eta + \lambda_h) - \lambda_l(S - g + \gamma - \eta)}\left(\eta e^{\bar{z}} - S\right). \tag{A.48}$$

Since $\bar{z}(t) = \bar{z}(0) + (\gamma - g)t$ and $g < \gamma$ by assumption, $\lim_{t\to\infty}\bar{z}(t) = \infty$. Therefore, (A.48) diverges for any $\eta > 0$, proving that the mean of the distribution cannot be stationary if $\bar{z} \to \infty$.

To finish the proof by contradiction, recall that the change of variables to $z \equiv \log(Z/M(t))$ was already normalized relative to $M(t)$, and hence is normalized relative to the adoption cost, $\zeta M(t)$. Since $v_\ell(z) > z$, (20) cannot hold with equality at all points. Therefore, we have the contradiction that the proposed $M(t)$ leading to the g cannot have been optimal. $\square$

A.5 Common Adoption Threshold for All Idiosyncratic States

This section derives sufficient conditions for heterogeneous firms to choose the same adoption threshold. It is kept as general as possible to nest both the exogenous and endogenous versions of the model.

Proof. Allow for some discrete type i, and augment the state of the firm with an additional state x (which could be a vector or a scalar). Assume that there is some control u that

[21] Most of the transformation comes through using the linearity of the operator, and the general formula that the bilateral Laplace transform of a derivative. That is, using a simple notation: $\mathcal{L}\left\{f'(z)\right\} = \xi\mathcal{F}(\xi)$. The other important formula is that $\mathcal{L}\left\{\boldsymbol{\delta}\left(z - c\right)\right\} = e^{-c\xi}$, and $\mathcal{L}\left\{\boldsymbol{\delta}\left(z\right)\right\} = 1$

controls the infinitesimal generator $\mathbb{Q}_u$ of the Markov process on type i and, potentially, x.[22] Also assume that the agent can control the growth rate $\hat{\gamma}$ at some cost. The feasibility set of the controls is $(u, \hat{\gamma}) \in U(t, i, z, x)$. The cost of the controls for adoption and innovation have several requirements for this general property to hold: (a) the net value of searching, $v_s(t)$, is identical for all types i, productivities z, and additional state x, (b) the minimum of the cost function is 0 and in the interior of the feasibility set: $\min_{(\hat{\gamma},u)\in U(t,i,z,x)} c(t, z, \hat{\gamma}, i, x, u) = 0$, for all t, x, i; and (c) the value of a jump to the frontier, $\bar{v}(t)$, is identical for all agent states (e.g. $\bar{v}(t) = v_\ell(t, \bar{z}(t)) = v_h(t, \bar{z}(t))$).[23]

Let flow profits be potentially type-dependent, $\pi_i(t, z, x)$, but require that $\pi(t, 0) \equiv \pi_i(t, 0, x)$ is identical for all i and x. Then, the normalization of the firm's problem gives the following set of necessary conditions:

$$
\begin{aligned}
(r - g(t))v_i(t, z, x) = \max_{(\hat{\gamma},u)\in U(\cdot)} \Bigg\{ & \pi_i(t, z, x) - c(t, z, \hat{\gamma}, i, x, u) + (\hat{\gamma} - g)\frac{\partial v_i(t, z, x)}{\partial z} \\
& + \frac{\partial v_i(t, z, x)}{\partial t} + \mathbf{e_i} \cdot \mathbb{Q}_u \cdot v(t, z, x) + \eta(\bar{v}(t) - v_i(t, z, x)) \Bigg\}
\end{aligned}
\tag{A.49}
$$

$$
v_i(t, \underline{z}(t, i, x), x) = v_s(t) \tag{A.50}
$$

$$
\frac{\partial v_i(t, \underline{z}(t, i, x), x)}{\partial z} = 0, \tag{A.51}
$$

where $\underline{z}(t, i, x)$ is the normalized search threshold for type i and additional state x. To prove that these must be identical, we will assume that $\underline{z}(t, i, x) = 0$ for all types and additional states, and show that this leads to identical necessary optimal stopping conditions. Evaluating at $z = 0$,

$$
v_i(t, 0, x) = v_s(t) \tag{A.52}
$$

$$
\frac{\partial v_i(t, 0, x)}{\partial z} = 0. \tag{A.53}
$$

[22] Ordering the states as $\{l, h\}$, the infinitesimal generator for this continuous-time Markov chain is $\mathbb{Q} = \begin{bmatrix} -\lambda_\ell & \lambda_\ell \\ \lambda_h & -\lambda_h \end{bmatrix}$, with adjoint operator $\mathbb{Q}^*$. The KFE and Bellman equations can be formally derived using these operators and the drift process.

[23] Without this requirement, firms may have differing incentives to "wait around" for arrival rates of jumps at the adoption threshold. A slightly weaker requirement is if the arrival rates and value are identical only at the threshold: $\eta(t, 0, \cdot)$ and $\bar{v}(t, 0, \cdot)$ are idiosyncratic states.

Note that equations (A.52) and (A.53) are identical for any i and x. Substitute equations (A.52) and (A.53) into equation (A.49) to obtain

$$(r - g(t))v_s(t) = \max_{(\hat{\gamma},u)\in U(\cdot)} \left\{\pi(t,0) - c(t,z,\hat{\gamma},i,x,u) + \mathbf{e_i}\cdot\mathbb{Q}_u\cdot v_s(t) + \eta(\bar{v}(t) - v_s(t)) + v_s'(t)\right\}. \tag{A.54}$$

Since in order to be a valid intensity matrix, all rows in $\mathbb{Q}_u$ add to 0 for any u, the last term is 0 for any i or control u,

$$(r - g(t))v_s(t) = \max_{(\hat{\gamma},u)\in U(\cdot)} \left\{\pi(t,0) - c(t,z,\hat{\gamma},i,x,u) + v_s'(t) + \eta(\bar{v}(t) - v_s(t))\right\}. \tag{A.55}$$

The optimal choice for any i or x is to minimize the costs of the $\hat{\gamma}$ and u choices. Given that $\hat{\gamma}$ only shows up as a cost, and our assumption that the cost at the minimum is 0 and is interior,

$$(r - g(t))v_s(t) = \pi(t,0) + v_s'(t) + \eta(\bar{v}(t) - v_s(t)) \quad \text{for all } i. \tag{A.56}$$

Therefore, the necessary conditions for optimal stopping are identical for all i, x, z, confirming our guess. Furthermore, equation (A.56) provides an ODE for $v_s(t)$ based on aggregate $g(t)$ and $\bar{v}(t)$ changes. Solving this in a stationary environment gives an expression for v_s in terms of equilibrium g, $\bar{v}$ and the common $\pi(0)$,

$$v(0) \equiv v_s = \frac{\pi(0) + \eta\bar{v}}{r - g + \eta}. \tag{A.57}$$

Furthermore, note from equation (B.24) that $\pi(0) = 1$ for all variations of the model in the body of the paper. □

A.6 Limit of Draw Arrival Process

The following is a heuristic derivation of the law of motion and cost function for adoption which yields a conditional draw above the firm's adoption threshold.

Instead of instantaneous draws, assume a firm choosing to adopt has an arrival rate of $\bar{\lambda} > 0$ of opportunities. While attempting to adopt, they pay a normalized flow cost of $\zeta\bar{\lambda}$. Note that we are scaling the flow cost by the arrival rate in order to take the limit and have a finite expected value of search costs.

Furthermore, assume that the firm draws *unconditionally* from the z distribution in the economy (rather than simply those above their current threshold), and starts with a ℓ type. In the stationary equilibrium, as all agents start low after adoption, searching firms accept a draw if they get above the normalized cutoff of 0. The proof will construct a limit where agents get a successful draw above 0 in any infinitesimal time period, and hence draw from the conditional distribution of $z \geq 0$.

Define $F_{\ell\lambda}(z)$ and $F_{h\lambda}(z)$ to be the cdfs of agents in the $z < 0$ region. As firms in the region $F_{\ell\lambda}(z)$ are otherwise identical, define the mass of searching agents as $F_{\ell\lambda}(0)$. Assume that agents have an unconditional draw of all z within the economy, then conditional on a draw, the probability of escaping the $F_{\ell\lambda}(0)$ mass is $(1 - F_{\ell\lambda}(0) - F_{h\lambda}(0))$. It is easily shown that the the arrival rate of *successful* draws is then $\bar{\lambda}(1 - F_{\ell\lambda}(0) - F_{h\lambda}(0))$. The distribution of waiting times until the first success is an exponential distribution with this parameter. The survivor function is therefore: $e^{-\bar{\lambda}(1-F_{\ell\lambda}(0)-F_{h\lambda}(0))t}$. Due to the total mass of one, $F_{\ell\lambda}(0) + F_{h\lambda}(0)) \in (0,1)$, so the survivor function is decreasing in t. $F_{\ell\lambda}(0)$ is independent of the $\bar{\lambda}$ arrival rate when taking limits as no agents enter this region from successful searches. Taking the limit for any t, $\lim_{\bar{\lambda}\to\infty} e^{-\bar{\lambda}(1-F_{\ell\lambda}(0)-F_{h\lambda}(0))t} = 0$. Therefore, in the limit in equilibrium, $F_{\ell\lambda}(0) = 0$ as measure 1 agents get a successful draw in any strictly positive interval. The same arguments can be used to explain why $F_{h\lambda}(0) = 0$.

To ensure that the expected search costs in this limit are finite, calculate the present discounted value of flow payments until the first success. This is the exponential distribution with parameter $\bar{\lambda}(1 - F_{\ell\lambda}(0) - F_{h\lambda}(0))$ and flow cost $\zeta\bar{\lambda}$

$$\mathbb{E}\left[\text{search costs}\right] = \int_0^\infty \left(\int_0^t \zeta\bar{\lambda}e^{-rs}\mathrm{d}s\right)\bar{\lambda}(1 - F_{\ell\lambda}(0) - F_{h\lambda}(0))e^{-\bar{\lambda}(1-F_{\ell\lambda}(0)-F_{h\lambda}(0))t}\mathrm{d}t, \tag{A.58}$$

$$= \frac{\bar{\lambda}\zeta}{r + \bar{\lambda}(1 - F_{\ell\lambda}(0) - F_{h\lambda}(0))}. \tag{A.59}$$

Take the limit and use the result that $F_{\ell\lambda}(0)$ and $F_{h\lambda}(0)$ converge to show

$$\lim_{\bar{\lambda}\to\infty} \mathbb{E}\left[\text{search costs}\right] = \zeta. \tag{A.60}$$

Therefore, in the limit the model can have draws directly from above the current threshold, with measure 0 remaining behind, and a cost for an instantaneous adoption of ζ. Since the flow of adopters is of measure 0, whether the draw is from the conditional or unconditional distribution is irrelevant.

B Endogenous Markov Innovation

This builds on the previous section to add additional derivations for the case in which firms choose to invest in innovation. For brevity, any equations that remain identical will be left out of the discussion.

Keep in mind that two things that change with endogenous innovation are the growth rate in the innovation state, i.e., $\gamma \to \gamma(\cdot)$, and the profit function due to licensing, i.e., $e^z \to \pi(z)$.

B.1 Stationary BGP with Endogenous Innovation

Proof of Propositions 3 and 4 . To create a stationary solution for the value function define a change of variables,[24]

$$w_i(z) \equiv e^{-z} v_i'(z) \tag{B.1}$$

From (36) and (37),

$$w_\ell(0) = w_h(0) = 0. \tag{B.2}$$

Differentiate (B.1) and reorganize ,

$$e^{-z} v_i''(z) = w_i'(z) + w_i(z). \tag{B.3}$$

Take the first order necessary condition of the Hamilton-Jacobi-Bellman equation in (35), and reorganize

$$\gamma(z) = \tfrac{\chi}{2} e^{-z} v_h'(z). \tag{B.4}$$

Substitute this back into (35) to get a non-linear ODE,

$$(r-g)v_h(z) = \pi(z) - g v_h'(z) + \tfrac{\chi}{4} e^{-z} v_h'(z)^2 + \lambda_h(v_\ell(z) - v_h(z)) + \eta\left(v_\ell(\bar{z}) - v_h(z)\right). \tag{B.5}$$

Differentiate (34),

$$(r-g)v_\ell'(z) = \pi'(z) - g v_\ell''(z) + \lambda_\ell(v_h'(z) - v_\ell'(z)) - \eta v_\ell'(z). \tag{B.6}$$

Multiply (B.6) by e^{-z} and use (B.1) and (B.3) and (46).

$$(r + \lambda_\ell + \eta - (1-\psi) g F'(0)) w_\ell(z) = 1 - g w_\ell'(z) + \lambda_\ell w_h(z). \tag{B.7}$$

[24] Our approach is to normalize and then substitute the FOC of the HJBE into the Bellman equation to form a nonlinear ODE, which we can solve numerically using collocation methods. An alternative approach to solving the HJBE numerically might be to use upwind finite difference methods as in Achdou et al. (2017) or Perla et al. (2021).

Note that using (B.3),

$$e^{-z}\boldsymbol{\partial}_z\left(e^{-z}v_h'(z)^2\right) = 2e^{-z}v_h''(z)e^{-z}v_h'(z) - \left(e^{-z}v_h'(z)\right)^2, \tag{B.8}$$

$$= 2w_h(z)w_h'(z) + w_h(z)^2. \tag{B.9}$$

Differentiate (B.5), multiply by e^{-z}, and use (B.1), (B.3) and (B.9) and (46)

$$(r+\lambda_h+\eta)w_h(z) = 1 - \left(g - \tfrac{\chi}{2}w_h(z)\right)w_h'(z) + (\lambda_h + (1-\psi)gF'(0))\,w_\ell(z) + \tfrac{\chi}{4}w_h(z)^2. \tag{B.10}$$

From (B.4),

$$\gamma(z) = \tfrac{\chi}{2}w_h(z), \tag{B.11}$$

$$g \equiv \tfrac{\chi}{2}w_h(\bar{z}). \tag{B.12}$$

□

Proof of Proposition 4 . For the case where $\eta \to 0$, and $\bar{z} \to \infty$, we can check the asymptotic value in (B.1)

$$\lim_{z\to\infty} w_i(z) = c_i. \tag{B.13}$$

To find an upper bound on g, note that as $w_i(z)$ is increasing, the maximum growth rate is as $\bar{z} \to \infty$. In the limit, $\lim\limits_{z\to\infty} w_i'(z) = 0$ as $w_i(z)$ have been constructed to be stationary. Furthermore, note that the maximum g from (B.12) is,

$$g = \lim_{\bar{z}\to\infty} \tfrac{\chi}{2}w_h(\bar{z}) = \tfrac{\chi}{2}c_h. \tag{B.14}$$

Therefore, looking at the asymptotic limit of (B.7) and (B.10),

$$(r+\lambda_\ell+\eta-(1-\psi)\tfrac{\chi}{2}c_hF'(0))c_\ell = 1+\lambda_\ell c_h, \tag{B.15}$$

$$(r+\lambda_h+\eta)c_h = 1+\left(\lambda_h+(1-\psi)\tfrac{\chi}{2}c_hF'(0)\right)c_\ell + \tfrac{\chi}{4}c_h^2. \tag{B.16}$$

Given an $F'(0)$, (B.15) and (B.16) provide a quadratic system of equations in c_l and c_h—and ultimately g through (B.14). While analytically tractable given an $F'(0)$, this quadratic has a complicated solution—except if $\psi = 0$. For that case, define

$$\bar{\lambda} \equiv \frac{r+\eta+\lambda_\ell+\lambda_h}{r+\eta+\lambda_\ell}. \tag{B.17}$$

Then, an upper bound on the growth rate with $\psi = 1$ and $\eta > 0$ is

$$g < \bar{\lambda}(r+\eta)\left[1 - \sqrt{1 - \frac{\chi}{\bar{\lambda}(r+\eta)^2}}\right], \tag{B.18}$$

where if $\eta = 0$, the unique solution is,

$$g = \bar{\lambda}r\left[1 - \sqrt{1 - \frac{\chi}{\bar{\lambda}r^2}}\right], \tag{B.19}$$

where a necessary condition for an interior equilibrium is

$$r > \sqrt{\frac{\chi}{\bar{\lambda}}}. \tag{B.20}$$

For the log utility case, substitute $r = \rho + g$ and (B.17) into (B.19) to yield (45).

□

B.2 Bargaining Derivation

Proof of Proposition 2 . From standard Nash bargaining, with a total surplus value of $v_\ell(z)$, let $\hat{v}$ be the proportion of the surplus obtained by the licensee and $v_\ell(z) - \hat{v}$ be the proportion obtained by the licensor. As is apparent in (B.22), if $\psi = 1$, then the technology is adopted for free and the licensee gains the entire value such that $\hat{v}(z) = v_\ell(z)$.

The timing is that the adopting firm first pays the adoption cost and then, upon the realization of the match, negotiations over licensing commence.

$$\arg\max_{\hat{v}} \left\{(\hat{v} - v(0))^{\psi}(v_\ell(z) - \hat{v})^{1-\psi}\right\}. \tag{B.21}$$

Solving for the surplus split, the value for a licensee that matches a firm with productivity z is,

$$\hat{v}(z) = (1-\psi)v(0) + \psi v_\ell(z), \tag{B.22}$$

while the licensor gains

$$v_\ell(z) - \hat{v}(z) = (1-\psi)(v_\ell(z) - v(0)). \tag{B.23}$$

There is an equal probability of adopting from any licensor and a unit measure of firms.

Thus, the flow of adopters engaging any licensing firm is just the flow of adopters $S = S_\ell + S_h$,

$$\pi(z) = e^z + \underbrace{gF_\ell'(0) + (g - \gamma(0))F_h'(0)}_{\text{Amount of Licensees}} \underbrace{(1-\psi)\left(v_\ell(z) - v(0)\right)}_{\text{Profits per Licensee}}. \tag{B.24}$$

Since the surplus split does not introduce state dependence to the cost of adopting, the smooth-pasting condition is unchanged. However, since adopters may not gain the full surplus from the newly adopted technology due to licensing costs, the value-matching condition takes bargaining into account. Adapting (20), the value-matching condition is

$$v(0) = \frac{1}{\rho} = \int_0^\infty \underbrace{\left[\psi v_\ell(z) + (1-\psi)v(0)\right]}_{\text{Surplus with licensing}} \mathrm{d}F(z) - \zeta. \tag{B.25}$$

Rearranging, the value-matching condition is identical to that previously derived in (20), but with a proportional increase in the adoption cost,

$$v(0) = \frac{1}{\rho} = \int_0^\infty v_\ell(z)\mathrm{d}F(z) - \frac{\zeta}{\psi}. \tag{B.26}$$

□

C Rough Calibration Parameter Values

See Table 1 for a summary of the parameter values used in the numerical examples.

We set $\gamma = 0.02$ to target a 2 percent growth rate and $\rho = 0.01$ to generate a real interest rate of 3 percent.

Transition rates λ_h and λ_ℓ are chosen to roughly match firms' growth-rates, with firms growing faster than 5 percent annually labeled h types and all other firms labeled ℓ-types. While the transition rates are sensitive to the h-threshold growth rate, all resulting numerical analysis is unchanged by wide variation in this threshold, as all calibrated transition rates are high enough to suggest that there is little persistence in either state and that the process essentially acts like iid growth rates.

Where appropriate, we target an approximate tail index of $\alpha = 2.12$, which corresponds to a tail parameter of 1.06 in the size or profits distribution used in Luttmer (2007) with a typical markup. Since the model is not sensitive to the tails, this is not a sensitive parameter but provides some discipline when jointly choosing the adoption and innovation costs.

We use $\bar{z} = 1.61$ (i.e., the frontier to minimum productivity ratio is 5). This ratio is larger than the $\bar{z} = 0.651$ (1.92 ratio in levels) documented by Syverson (2011) between the top and bottom decile within narrowly defined industries, and is closer to the 5:1 ratio found in Hsieh and Klenow (2009) in India and China. The qualitative results are not strongly

Parameters	Value/Target	Calibration
$\{\lambda_\ell, \lambda_h\}$	$\{0.533074, 1.12766\}$	Matches estimation of 2 state Markov transition matrix for firm growth rates using Compustat with firms in SIC 2000-3999. Model is not sensitive to these parameters, as long as they are not too low.
ψ	$[0.5, 1.0]$ with 0.95 baseline	Matches median 5% royalties of large firms reported from RoyaltySource in Kemmerer and Lu (2012).
ρ	0.01	Target interest rate $r = .03$ when $g = 0.02$.
$\{\chi, \zeta\}$	$g = 0.02$ and $\alpha = 2.12$	Targets 2% growth rate, and an underlying tail parameter of the firm size distribution of 1.06 (which translates to $\alpha = 2.12$ using the rough adjustment implied by monopolistic competition). Note: The growth rates are a function of ψ and other parameters which are calibrated separately.
$\{\bar{z}, \eta\}$	$\bar{z} \in [0.651, \infty]$	If $\eta \approx 0$, then $\bar{z}$ is set large enough for numerical stability to approximate ∞ (keeping in mind that $e^{\bar{z}}$ is the actual multiplier on productivity of the frontier, so $\bar{z} = 3.0$ translates to a ratio of productivity of the frontier to the threshold of 20.1.)

Table 1: Summary of Parameter Values

driven by this fact, as we look across various $\bar{z}$ level.