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

Jess Benhabib
NYU

Jesse Perla
UBC

Christopher Tonetti
Stanford GSB

January 4, 2017
Draft Version: 329

Active Work in Progress[0]
[Download Newest Version]
[Download Technical Appendix]

Abstract

We study how innovation and technology diffusion interact to endogenously determine the productivity distribution and generate aggregate growth. We develop a model in which firms choose to innovate, adopt technology, or keep producing with their existing technology. We show how innovation tends to stretch the distribution while adoption compresses it. These forces can balance, generating stationary distributions, with the relative strength of the forces determining the shape of the distribution. We analyze the degree to which innovation and technology diffusion at the firm level contribute to aggregate economic growth and can lead to hysteresis, which depends crucially on the cardinality of the support of the productivity distribution. With finite support productivity distributions, the aggregate growth rate equals the growth rate of innovators at the frontier. Changes in the costs or benefits of adoption, however, can influence the aggregate growth rate by affecting the incentives to innovate, either directly through licensing arrangements when technologies are excludable or indirectly via the option value of adoption.

Keywords: Endogenous Growth, Technology Diffusion, Innovation, Imitation, R&D, Technology Frontier

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

[0]An earlier version of this paper circulated under the title "The Growth Dynamics of Innovation, Diffusion, and the Technology Frontier."

> *"At any moment, there is a large gap between average and best practice technology; reducing this gap by disseminating the techniques used by producers at the cutting edge of knowledge is technological progress without inventions. Any discussion of the gap between average and best practice techniques makes little sense unless we have some notion of where the best practice technique came from in the first place. Without further increments in knowledge, technological diffusion and the closing of the gap between practices will run into diminishing returns and eventually exhaust itself."*
>
> —Joel Mokyr, "The Lever of Riches"

1 Introduction

This paper studies how the interaction between adoption and innovation jointly determine 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 even within 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 who purposefully choose to avoid innovating and, instead, adopt already invented ideas.

New ideas/technologies are invented and adopted frequently. For example, taking the notion of ideas-as-recipes literally, *"Now-ubiquitous dishes, such as molten chocolate cake or miso-glazed black cod, did not just pop up like mushrooms after a storm. Each debuted in a specific restaurant but soon migrated outward in slightly altered form. The putative inventors (Jean-Georges Vongerichten in the case of molten chocolate cake, Nobu Matsuhisa for miso black cod) can claim no royalties on their creation."* (Raustiala and Sprigman, 2012, p.63). Miso black cod eventually diffused from Nobu in New York to Sakura Sushi in Whitehorse, Yukon, Canada. These examples from the restaurant industry motivate two key building blocks of our theory. First, it is useful to distinguish between innovation and adoption activity, as the economic incentives, costs, risks, abilities, and players involved in the two activities are very different. Second, there are often a wide range of very different firms concurrently producing closely related varieties, without the subsequent producer pushing the original inventor out of business due to some winner-take-all force like creative destruction.

The huge spread in productivity within narrowly defined industries, the importance of technology diffusion as a key source of growth for the less-productive, and the importance of innovation in generating long-run growth are well established, yet there are few theories with which to study these linked phenomena. The main contribution of this paper is to develop a model that provides tools to inspect data with these forces in mind. Crucially, the model delivers a finite endogenously-expanding frontier with wide productivity dispersion as the result of optimal firm behavior.

We build a model that avoids the puzzle of collapse at the frontier and the associated need for infinite support productivity distributions, while reconciling models of innovation and idea diffusion. A finite frontier is *prima facie* supported by the data, and turns out to be a useful and consequential model feature. If the frontier were finite and constant, there would be a puzzle of collapse to the frontier with no long-run growth; as Mokyr points out "without further increments in knowledge, technological diffusion. . . eventually exhaust[s] itself." To address this, previous models of idea diffusion such as Lucas and Moll (2014) and Perla and Tonetti (2014) have either assumed an infinite support distribution or some exogenous expansion of the frontier, so that there is no exhaustion of ideas. In a sense, this phenomena represents "latent growth", i.e., growth that is inherent in the interplay of initial conditions and exogenous stochastic processes and is foreign to the technology diffusion mechanism at the heart of the model. The implicit assumption is that some process outside of the model is generating an expansion of the frontier and the model is focusing on the change in the productivity distribution generated by the diffusion process. For some purposes, like studying medium-run growth rates or examining the range of aggregate growth rates consistent with exogenously given productivity distributions, this may be a very useful assumption.[1] For other purposes, like explaining the sources of long-term growth and the role of technology diffusion in determining growth rates, is necessary to close the model with a joint theory of innovation and diffusion. Instead of using the infinite support assumption, we build such a model of an endogenous expanding finite frontier with innovation and adoption in the long-run determining the aggregate growth rate and shape of the productivity distribution. Furthermore, we show that the finite support of the distribution has critical implications for key model properties concerning latent growth, hysteresis and multiplicity, and how adoption and innovation interact.

While diffusion models typically use infinite support or exogenous innovation processes to avoid collapse at the frontier, Schumpeterian models are designed expressly to study the endogenous expansion of the frontier. These models of creative destruction, however, typically model the frontier and near-frontier firms, with the many low-productivity firms and associated adoption activity absent. By combining adoption, innovation, and quality-ladder like jumps to the frontier, our model generates substantial, but bounded, productivity dispersion consistent with the firm distribution data. Innovation pushes out the frontier and creates the technologies that will eventually be adopted, while adoption helps compress the distribution keeping the laggards from falling too far behind. Furthermore, innovation activity affects adoption incentives and adoption can affect innovation incentives. Thus, it is the interaction between these two forces that determines the shape of the productivity distribution and the aggregate growth rate. Since optimal adoption and innovation behaviors generate the shape of the productivity distribution, including the spread between best and worst firms, the model is well-suited to analyze the determination of the full productivity distribution, not just the few firms at the frontier. Long-run growth is driven by innovation, but that does not necessarily mean that adoption of already discovered ideas can not affect long-run growth rates. Rather, it means that adoption affects growth rates by affecting the incentives to innovate. This paper studies how innovation and adoption jointly generate the cross-section of firm productivity and its movement over time.

Model Overview and Main Results. We first build a simple model of exogenous innovation and growth to focus on how innovation and adoption affect the shape of the productivity distribution. We then add an innovation decision in which aggregate growth is endogenously driven in the long run by the innovation activity of high productivity firms. At the core of the model are the costs and benefits of adoption and innovation. Section 2.1 discusses how we model innovation and adoption and why. 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 productivity is drawn from a distribution related to the existing distribution of technologies currently in use for production. To represent innovation, we model firms as being in either a creative or stagnant innovation state, and when creative, innovation generates geometric growth in productivity at a rate increasing in firm-specific innovation expenditure. A firm's innovation state evolves according to a two-state Markov process, and this style of stochastic model of innovation 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 that there are better technologies to be invented,

[1] In particular, these diffusion models isolate the impact of the shape of the productivity distribution on determining the incentives to adopt technology—which may be the dominant force in short- and medium-term growth for most of the world.

in contrast to all knowledge that will ever be known being in use for production at time zero.[2] At each point in time any firm has the ability to innovate or adopt, and firms' optimally choose if 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, there will be low productivity firms investing in adopting technologies, stagnant firms falling back relative to creative firms, medium productivity creative firms investing small amounts to grow a bit through innovation, and higher productivity creative firms investing a lot in R&D to grow fast, create new knowledge, and push out the productivity frontier. 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 behind too far. A low cost of innovation tends to spread the distribution, as the high productivity firms can more easily escape from the pack. The stochastic innovation state ensures that some firms who have bad luck and stay uncreative for a stretch of time fall back relative to adopting and innovating firms, generating non-degenerate normalized distributions with adopting activity existing in the long run. Thus, the shape of the distribution, which typically looks like a truncated Pareto with finite support, is determined by the relative ease of adoption and innovation through the differing rates at which high and low productivity firms grow.

Adoption and innovation are not two completely independent processes with some firms perpetual adopters and some perpetual innovators. Rather, that all firms have the ability to invest in both activities generates general equilibrium interactions between actions. The key spillover between adoption and innovation can be seen in the option value of adoption. For high productivity firms far away from being a low productivity adopter, the value of having the option to adopt is small. The lower a firm's productivity, the closer they are to being an adopter and 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. Thus, the value of adoption, which is determined by the cost of acquiring a new technology and the probability of adopting a good technology, affects incentives to innovate.

In addition to the baseline model, we introduce a sequence of extensions designed to enrich the model to capture more ways in which innovation and adoption might interact and to relax some of the stark assumptions prevalent in the literature. We introduce a version of quality ladders by including a probability of leap-frogging to the frontier technology. In an extension in which technologies are partially excludable and there is licensing, because adopters pay a fee to the firm whose technology they adopt there is an extra direct link between adoption behavior and innovation incentives affecting the shape of the distribution and aggregate growth rates. The baseline model has undirected search for a new technology in that a draw is from the unconditional distribution of technologies and there is no action a firm can take to influence the source distribution. In an extension we model "directed" adoption in that firms can obtain a draw from a skewed distribution in which they can increase the probability of adopting better technologies at a cost. In the baseline model firms exist for all time, their output and profits equal their productivity, there is no explicit cost of production, and there is a single market for the common good all firms produce. While this delivers the cleanest framework for analyzing the key forces, the model is extended to include endogenous entry, exogenous exit and firms who hire labor to produce a unique variety sold via monopolistic competition to a CES final good producer. For each extension we examine properties of the BGP productivity distribution, such as the tail index and the ratio of the frontier to the minimum productivity, and whether the equilibrium is unique or if there is hysteresis in the sense that the long run distribution and growth rate depends on initial conditions.

[2] Given a continuum of firms, modeling stochastic innovation using geometric Brownian motion, as is common in the literature, would generate infinite support instantly, while the finite-state Markov process allows for finite support for all time.

Through the baseline model and extensions we show what types of stochastic processes can generate data consistent with the empirical evidence: balanced growth with a nearly Pareto firm size distribution in the right-tail and finite support when normalized by aggregates. We show which features are necessary to have both innovation and adoption activity exist in the long run and when and how adoption affects the aggregate growth rate. Finally, we also show that assumptions such as infinite initial support are not innocuous, in that the obviously counterfactual infinite support initial condition implies very different important model properties than the finite support initial condition. An important distinction between BGPs with infinite and finite support is whether there is latent growth in that the aggregate growth rate can be greater than the rate of innovation, or more starkly, whether long-run growth can exist without innovation.

While many versions of the model can be studied analytically, the endogenous growth cases must be solved numerically. A final contribution of this paper is the development of a generally applicable numerical technique—based on spectral collocation and quadrature—to solve continuous time models with heterogeneous agents that take the form of coupled Hamilton-Jacobi-Bellman equations, Kolmogorov forward equations, and integral constraints.

Finally, since many of the results depend on the technology frontier, in Section 6 we provide some exploratory empirical analysis on the relative frontier using Compustat data. We show that the relative frontier, as proxied by the ratio of the 90th to the 10th percentile of the firm size distribution varies significantly across industries, but has been relatively stable within industry over the past few decades.

1.1 Recent Literature

Our literature connects to the four major lines of the literature on economic growth: 1) Expanding ideas/varieties as exemplified by Romer (1990) and Jones (1995); 2) Schumpeterian creative-destruction, as in Aghion and Howitt (1992), Grossman and Helpman (1991), and Klette and Kortum (2004); 3) Variety improvement and human capital, as in Uzawa (1961), Arrow (1962), and Lucas (1988); and 4) Technology/idea diffusion, as in Luttmer (2007), Perla and Tonetti (2014), and Lucas and Moll (2014). This section will first briefly outline the relation to the major strands of growth theory, and then focus on a few recent papers that are most closely related.

While closely linked to many of the concepts in the literature, the model developed in this paper is designed to address phenomena not jointly captured by most of the existing literature, namely within firm productivity improvement via adoption and innovation with large productivity dispersion among actively producing incumbents. To focus on the interaction between adoption and innovation in a simple environment, we abstract from some features prominent in other models of growth. Specifically, we develop a one-product firm model and focus on within-firm growth. Thus, our baseline model omits the creation of new varieties (although we add endogenous entry in an extension) and we omit selection into exit. All of these features can be merged into one large model, but for exposition and clarity we introduce our model of adoption and innovation in a minimalist environment without these extra features.

Expanding Varieties. The expanding varieties literature develops a theory of TFP growth by modeling the process of discovering completely new products/ideas. The key concept in Romer (1990) is that ideas are non-rival. Thus, when combined with a production function that features constant returns to scale in physical (rival) inputs, the production function has increasing returns to scale in ideas and physical inputs. Essentially, since an idea improves productivity and is non-rival, it only needs to be invented once and can be applied without exhaustion to improve output from the stock of physical inputs. Because ideas are non-rival, if they could be copied by a competitor instantly, there would be no incentive to invent a new idea. Thus, excludability plays a central role in determining the incentives to innovate. Given the setup in our model, there are incentives to innovate even absent excludability since firms can coexist producing similar products, just like Nobu inventing miso black cod and many other firms later adopting the recipe. In an extension in this paper, we introduce excludability into our model of innovation and adoption, bringing this concept that is crucial in this literature to the diffusion literature, capturing similar incentives, but delivering new economics via the interaction between adoption, innovation, and the shape of the distribution.

Typically, in this literature, there is zero within variety improvements and all varieties are symmetric, such that the model has no concept of heterogeneous firms. Thus, the literature is not able to speak to the wide dispersion in the firm size distribution, incumbents improving over time, or different firms undertaking different types of activities to improve productivity, like adoption versus innovation. Another related issue in the literature is how to model the idea production function, in which the productivity of discovering new ideas is assumed to depend on the stock of already discovered ideas. In some sense, our model in which the aggregate growth rate depends on the entire productivity distribution is a micro-foundation for the dependence of the idea production function on past ideas.

Creative Destruction. Shumpeterian models tend to have a single leader producing each variety and an inactive fringe, rather than a distribution of firms producing similar products with different levels of productivity. For example, in Klette and Kortum (2004) and Acemoglu, Akcigit, Bloom, and Kerr (2013), each good is produced by a single firm who may lose their exclusivity to an innovative competitor creating a new leading-edge version of the product, taking over the entire product line. Crucially, all growth occurs through creative destruction rather than from innovations within the firm.[3] See Aghion, Akcigit, and Howitt (2014) for a survey of the creative destruction literature.

Creative destruction models are expressly developed to model the expansion of the technology frontier, and capture the cutthroat competitive environment associated with innovating at the frontier. They do not, however, have a wide dispersion in firm productivity within very narrow products, as is apparent in the data. In this sense, our paper contains a less sophisticated model of innovation at the frontier, but it is well suited to model the determination of the entire productivity distribution, not just the frontier firms. Creative destruction models cannot, however, provide guidance on the evolution of the productivity and firm size distribution for those producing well below the frontier technology. In this paper, we have many lower productivity firms and we have separate processes by which high and low productivity firms improve. That is, we can capture the increases in productivity for the large mass of firms that do not improve through innovation that is at the heart of the creative destruction literature. In future research it could be fruitful to combine a rich model of innovation at the frontier as is typically modeled in the creative destruction literature with the adoption and dispersion features developed in this paper.

In some sense, one could view the quality ladder model, in which innovation for a given variety is on top of the frontier production technology, as a form of technology diffusion. That is, the productivity of the new leader is typically a function of the old leaders productivity (e.g., a given multiplicative step size). However, even this interpretation is very different from the adoption we model, as it is focused on diffusion from frontier firms to new frontier firms, not firms spread throughout the productivity distribution who typically operate with low productivity technologies even post-adoption. To extend the ladder analogy to diffusion models, there may be a large number of followers on the same ladder distributed across previously invented steps. These followers improve over time by taking modest steps up the ladder that other frontier firms built, but they rarely get close to the fiercely competitive top step.

[3]Differing from the expanding variety models, creative destruction models generate growth via within variety improvements, however both tend to feature one firm producing each variety.

Human Capital and Variety Improvement. In contrast to the variety expansion models that focus on the creation of new varieties, this line of literature focuses on the creation of better ideas, either within-variety or within-individual. In the baseline model of this paper we develop all growth comes from within firm productivity improvements, just as in the human capital and variety improvement literature. A key distinction is that we model multiple ways for firms to improve their variety, via either adoption or innovation, and study how the interaction between these two actions generates growth and the productivity distribution. See Grossman, Helpman, Oberfield, and Sampson (2016) for a recent example in this line of research.

Technology Diffusion. This paper is closely related conceptually and technically to the idea diffusion literature, including Kortum (1997), Alvarez, Buera, and Lucas (2008), Lucas (2009), Alvarez, Buera, and Lucas (2013), Perla and Tonetti (2014), and Lucas and Moll (2014). This technology/idea diffusion literature focuses on how existing production technologies are distributed among agents throughout the entire distribution and how the better technologies diffuse to low productivity agents.[4] The literature does not emphasize, however, how ideas are created in the first place nor the role of the technology frontier.[5] These papers focus on adoption, implicitly assuming some process that generates the ideas to be copied in the long run without explicitly modeling that innovative behavior.[6] In these models, while growth in the short- or medium-run can be dominated by technology adoption, long-run growth is only possible if the productivity distribution has infinite support with fat tails (i.e., a power-law), with a direct relationship between the power-law's tail parameter and the growth rate.[7]

The interaction between innovation and technology diffusion explored in this paper also appears in Luttmer (2007, 2012, 2015a) and Sampson (2015), although with a different emphasis and mechanism. The main similarity with this paper is that in all cases diffusion is modeled as some firms drawing a new productivity from the productivity distribution of incumbents. The big difference is that in those papers, diffusion is generated from incumbents to new entrants and it is the equilibrium selection that the worst firms exit that drives diffusion and growth. In contrast, this paper features incumbents adopting better technologies in equilibrium, with different implications for policy counterfactuals and mapping the model to data. Another common modeling feature is that Luttmer (2011) also features fast and slow growing incumbent firms—driven by differences in the quality of blueprints for size expansion.[8] Luttmer (2015a) emphasizes the role of risky "experimentation," modeled as a stochastic process distinct from deterministic innovation, as important in the generation of endogenous tail parameters. Our model also has a stochastic and deterministic component of innovation, and the risky part of innovation makes sure there is mixing in the distribution generating adoption in the long-run. Luttmer (2015b) describes a continuum of long-run growth rates possible in a class of diffusion models. We add to this by showing that the "latent growth" and hysteresis in these models comes from the infinite tail (either from an initial condition or unbounded geometric random shocks). While Luttmer (2015b) introduces a selection device,

[4]Some of the key empirical papers on technology diffusion examine cross-country adoption patterns, such as Comin and Hobijn (2004, 2010); Comin, Hobijn, and Rovito (2008).

[5]Grassi and Imbs (2016) is a recent model of technology diffusion with a finite number of firms, and thus finite frontier, that focuses on the role of granularity in generating the measured increasingly positive correlation between growth and volatility as the share of large firms in a sector rises. Luttmer (2015b) also discusses the role of a finite numbers of agents.

[6]Lucas and Moll (2014) provides an extension of their baseline diffusion model with the addition of exogenous innovators in order to discuss finite support of the initial distribution.

[7]See the transition dynamics in Figure 2 of Perla and Tonetti (2014) for an example with a finite support productivity distribution without the creation of new ideas. In the simple calibration, growth exists for decades purely driven by diffusion, before leveling off to zero growth eventually.

[8]In his model, firms stochastically enter the absorbing slow growth-state, where in this paper firms can jump back and forth between the states.

which is to pick the smallest growth rate as it is the one that would arise from a finite support initial distribution when paired with GBM shocks, we directly concentrate on the role of finiteness and its interaction with innovation decision.

Recent Papers Combining Mechanisms. Our paper is most closely related in spirit to recent papers that combine different productivity growth mechanisms in one model. Coming from the creative destruction literature, Acemoglu and Cao (2010), Akcigit and Kerr (2016), and König, Lorenz, and Zilibotti (2016) model own-variety improvement within a Schumpeterian framework. König, Lorenz, and Zilibotti (2016) has the same structure as Klette and Kortum (2004) with a single firm using the best-practice technology to produce a variety subject to an inactive competitive fringe, but allows both an innovation choice in the spirit of Akcigit and Kerr (2016) and imitation of a firm's quality from a different product-line. In this sense, it is loosening the strict separation across product lines and departing from a strictly Schumpeterian interpretation of productivity growth.[9] König and Rogers (2016) combines Klette and Kortum (2004) with an expanding variety model using collaborative networks. In each case, there is still only a single firm producing a given product line at any time.[10] Coming from the technology diffusion literature, Buera and Oberfield (2015) is a related semi-endogenous growth model of the international diffusion of technology and its connection to trade in goods. Buera and Oberfield (2015) combines the process of idea diffusion with innovation. This is in the spirit of Jovanovic and Rob (1989), in which there is a positive stochastic spillover from the diffusion process itself. In contrast, this paper models these as distinct actions potentially undertaken by different firms, they model productivity upgrading according to one joint process that mixes innovation and adoption. 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.

Acemoglu, Aghion, and Zilibotti (2006), Chu, Cozzi, and Galli (2014), Stokey (2014), and Benhabib, Perla, and Tonetti (2014) also explore the relationship between innovation and diffusion from different perspectives. Similar to this paper, 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, distorting their innovation choices. That is, incumbent firms not adopting today realize they may adopt in the future, and derive positive value from this option to adopt.[11]

In our extensions, we take small steps towards reconciling our model with creative destruction and expanding variety models. For example, the important role of excludability (often a precondition for models of variety expansion) is explored in Section 5, and leap-frogging to the frontier as developed in Section 3.3 is a step towards Schumpeterian forces—albeit without the crucial force of creative-destruction upon the jump.

[9] In contrast to König, Lorenz, and Zilibotti (2016), we abstain from modeling Schumpeterian forces and keep both the innovation and adoption technology fairly simple in order to focus on the economics of their interaction. This comes at a loss of our ability to model the richness of the innovation process for which Schumpeterian models excel, but we gain tractability which enables us to better study the non-Schumpeterian forces and consider the role of infinite support, a finite technology frontier, and "latent growth." A key distinction is that the finite state Markov innovation process that we model allows for a finite frontier, while in models with a continuum of firms on quality ladders and Poisson arrivals of multiplicative jumps the frontier quality diverges to infinity from any initial condition in the same way that it does for Brownian motion.

[10] Lashkari (2016) also models the role of innovation and technology diffusion—but emphasizes the interaction with selection in a model of creative destruction.

[11] Similar to our paper, Eeckhout and Jovanovic (2002) model technological spillovers that are a function of the distribution of firm productivity. Acemoglu, Aghion, Lelarge, Van Reenen, and Zilibotti (2007) also model spillovers across firms in innovation, captured by the number of firms that have attempted to implement a technology before.

Interpreting Productivity Dispersion and Firm Growth. Papers that closely examine productivity find a high degree of dispersion at *every level of aggregation*—even for narrowly defined industries where we would expect firms to be producing goods with similar characteristics.[12] For example, Syverson (2011) surveys the evidence on productivity dispersion and finds that within the US, the ratio of the top to the bottom decile is approximately 1.92:1. In places such as China and India, Hsieh and Klenow (2009) finds the ratio is closer to 5:1.[13]

Absent measures of physical productivity, either revenue or firm size are commonly used to proxy for productivity or quality—which are difficult to separately identified and are typically combined into one dimension of firm profitability. As only one or two firms produce each variety within a creative destruction model, all productivity or quality differences across firms are interpreted as firms having fewer product lines that are at the frontier or having frontier products with lower quality/productivity relative to other frontier products. The typical models do not allow any interpretation as firms producing the same variety another firm is producing, but with below frontier quality/productivity.

With this in mind, Garcia-Macia, Hsieh, and Klenow (2016) decomposes changes in employment shares of firms as a proxy for quality into multiple possible growth channels. They find that 48 percent of such changes are attributable to own-variety improvements by incumbents, 23 percent comes from creative destruction (from both incumbents and entrants), and the expansion of varieties accounts for 29 percent. This confirms the important role of incumbent improvements that is highlighted in this paper. From the perspective of our paper, however, not all such improvements are attributed to innovation activity as many of the firms improve via adoption.

While the stationary productivity distribution gives a sense of dispersion, dynamics of the distribution gives a sense of whether the creative destruction dynamics are active throughout the whole distribution or only active for firms near the front. A classic empirical example analyzing productivity dynamics is Baily, Hulten, Campbell, Bresnahan, and Caves (1992). In their Figure 4, they show that of the all plants in the lowest (5th) quintile of the productivity distribution in 1972, 54 percent are in the 4th or 5th quintile in 1982. Of those in the 1st quintile in 1972, 42 percent remain in the top quintile in 1982, while only 6.5 percent exit during that period. This suggests typical changes to plants are modest in size and occur slowly, inconsistent with Schumpeterian models of leap-frogging (i.e., not too many changes in leadership). Creative destruction models are designed to capture frontier firms, so it is very possible Schumpeterian forces are very strong inside the top quintile. This suggests, that while very relevant for the expansion of the frontier, creative destruction may have more modest effects on the bottom 4 quintiles. See Section 6 for basic empirics of the technology frontier using ratios of moments in the distribution, in a sense similar to Hsieh and Klenow (2009) and Acemoglu, Aghion, Lelarge, Van Reenen, and Zilibotti (2007).

2 Baseline Model with Exogenous Stochastic Innovation

We first analyze an exogenous growth model to simplify the introduction of the environment; to focus on the economic forces that determine the shape of the stationary normalized productivity distribution; and to highlight that key properties of BGP equilibria depend on the cardinality of the support of the productivity distribution.

The only choice a firm makes in this version of the model is whether to adopt a new technology or to continue producing with its existing technology. In Section 4, we develop an extension of

[12] More aggregated approaches, such as Acemoglu and Dell (2010) find considerable dispersion both between and within countries, and attribute most of it to differences in technological know-how.

[13] Furthermore, if productivity is partially embodied in management practices, as described in Bloom and Van Reenen (2010) and Bloom, Sadun, and Reenen (2014), then the dispersion of management practices between and within countries also provides strong evidence for a high degree of productivity dispersion for similar products.

this model—in which a firm chooses its innovation rate—to study how adoption and innovation activities interact to jointly determine the shape of the productivity distribution and the aggregate growth rate.

Throughout the paper, we present stark models in the interest of parsimony. In Technical Appendix G, we develop a more elaborate general equilibrium model of monopolistically competitive firms facing CES demand that hire labor for production and productivity-improving activities to maximize profits, with all costs denominated in units of labor at a market wage and free entry determining the endogenous mass of firms. Results are qualitatively equivalent, so we focus on the simpler model for clarity of exposition.

2.1 How we Model Innovation and Adoption and Why

We model adoption and innovation as two separate methods for generating within-firm productivity gains. It could be that there is no such activity as "pure" adoption and that each act of adoption has the chance to create something new (as in Jovanovic and Rob (1989) and Buera and Oberfield (2015)) or requires some innovation to adapt the adopted technology to local uses. Likewise, perhaps every act of innovation requires the concurrent adoption of some external knowledge. There is, however, a sense in which some improvements fall closer to the concept of pure innovation and some closer to pure adoption on the spectrum. Investments in adoption and innovation activities have different costs, benefits, and risks, and are done differentially by different types of firms. Thus, we believe that it is fruitful to distinguish between these activities. It is useful to model them as completely separate actions, to better understand the incentives driving firms to align themselves closer to primarily adoptive or innovative activity and to see how these two activities interact to generate dynamics of the productivity distribution. The key characteristic of innovation is that it is the process that creates new knowledge. There are two essential characteristics of adoption. First, adoption does not create new knowledge, but, rather, transfers existing knowledge across agents. Second, the opportunities available to an adopting firm necessarily depend on the existing technologies that are available to adopt. To make sure we can separately capture the incentives and effects of these two types of activities, the innovation process in our model does not share these two essential properties of adoption and the adoption process does not share this key characteristic of innovation.

To capture these features, we model adoption as a meet and copy process by which an adopter draws a new technology with some randomness that is a function of the distribution of technologies actively in use. This is similar to the diffusion models in Perla and Tonetti (2014) and Lucas and Moll (2014). Adopters would prefer to copy the frontier technology, but due to the difficulty of finding and implementing the best technologies, they will typically end up with technology that is better than their own but not at the frontier. Innovation is important in this paper because it generates an endogenously expanding productivity frontier. We model innovation as autarkic geometric growth (as in Atkeson and Burstein (2010) and Stokey (2014)), in the sense that the returns are only a function of the innovating firm's productivity level and not the productivity distribution. This model of innovation delivers an endogenously expanding frontier in the simplest way, so that we can focus on the interaction between the distinct activities of innovation and adoption.

Since the key property of innovation is the creation of new technologies, which we capture via the expansion of the frontier, it is essential that our model of innovation permits a productivity distribution that has finite support for all finite times. As we will show, it turns out that a key distinction that determines the fundamental properties of BGP equilibria is whether the long-run productivity distribution has finite or infinite support. The important difference is between infinite and finite support in the long run, independent of whether infinite support occurs due to initial conditions with infinite support or to an exogenous process that generates infinite support.[14] This motivates our use of a finite-state Markov process for innovation growth rates, with which growth rates are always bounded for all firms over any positive time interval, and infinite support in the stationary distribution arises only from initial conditions with infinite support.

There is an additional advantage to modeling an innovation process with a bounded maximum growth rate (such as our finite-state Markov chain) beyond just allowing for a finite technology frontier. Since the process is bounded, the innovation rate in a fully-specified growth model must be less than or equal to the maximum growth rate—whether that rate be determined endogenously or exogenously. Hence, if the growth rate of the aggregate economy is greater than the maximum innovation rate—as it is in the infinite support version of our model—then the degree of "latent growth" in the economy is clearly defined. That is, if the aggregate growth rate is greater than the maximum idiosyncratic growth rate, there must be latent growth. With unbounded innovation rates, such as GBM or multiplicative Poisson arrivals, there is no maximum rate of innovation or zero-latent-growth benchmark for the economy, and thus it is difficult to determine what the sign of the latent growth may be.

See Technical Appendix A for more discussion of the strengths, weaknesses, and failures of this and many alternative innovation processes.

2.2 The Baseline Model

Firm Heterogeneity and Choices. A continuum of firms produce a homogeneous product and are heterogeneous over their productivity, Z, and innovation ability, $i \in \{\ell, h\}$. For simplicity, firm output equals firm profits equals firm productivity, Z. The mass of firms of productivity less than Z in innovation state i at time t is defined as $\Phi_i(t, Z)$ (i.e., an unnormalized CDF). Define the technology frontier as the maximum productivity of any firm, $\bar{Z}(t) \equiv \sup\{\text{support}\{\Phi_\ell(t,\cdot)\}, \text{support}\{\Phi_h(t,\cdot)\}\} \leq \infty$, and normalize the mass of firms to 1 so that $\Phi_\ell(t, \bar{Z}(t)) + \Phi_h(t, \bar{Z}(t)) = 1$. At any point in time, the minimum of the support of the distribution will be an endogenously determined $M_i(t)$, so that $\Phi_i(t, M_i(t)) = 0$. 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 a deterministic rate γ.[15] In the low innovation ability state (ℓ), it has zero productivity growth from innovation (without loss of generality).[16] Sometimes, firms have good ideas or projects that generate growth, and sometimes firms are just producing using their existing technology. Innovation ability evolves according to a continuous-time two-state Markov process that drives the exogenous growth rate of an operating firm. This process allows for stochastic innovation with a finite frontier and will permit equilibria in which adoption persists in the long

[14] Infinite support can be present from the beginning due to an initial condition (e.g., Perla and Tonetti (2014), Lucas and Moll (2014)). Alternatively, infinite support can arise from an innovation process with geometric stochastic shocks generating an infinite-support power law as its stationary distribution from any initial condition (e.g., Luttmer (2007), Perla, Tonetti, and Waugh (2015) Appendix, and the GBM example in Technical Appendix E). For example, in the case of processes using Brownian Motion or a Poisson arrival of geometric jumps, even an initial condition with finite support immediately becomes infinite with a continuum of agents. The issue in either of those two cases is that the growth rates in any strictly positive amount of time are unbounded. The properties of the equilibria are largely the same if the infinite support comes from geometric stochastic shocks or initial conditions, as is seen in the nested solution of Perla, Tonetti, and Waugh (2015).

[15] In Section 4, the growth rate γ will become a control variable for a firm, with the choice subject to a convex cost.

[16] The different innovation types are akin to Acemoglu, Akcigit, Bloom, and Kerr (2013)—albeit without the selection of entry inherent in that model.

run and growth is driven by the innovation choices of frontier firms. See Technical Appendix A for a discussion of the strengths, weaknesses, and failures of this and many alternative innovation processes (e.g., Poisson arrival of jumps or drifts; IID growth rates and singular perturbation methods; reflected geometric Brownian motion (GBM); deterministic heterogeneous growth rates, etc.). The innovation states are modeled primarily for technical reasons related to continuous-time, combined with the desire to model stochastic innovation and finite-support distributions, rather than being of first order interest per se. In some loose sense, with high switching rates this process is quantitatively similar to IID growth rate shocks without the associated continuous-time measurability issues.

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 (except, perhaps, exactly at an initial condition). Recall that support $\{\Phi(t,\cdot)\} \equiv [M(t), \bar{Z}(t))$. The growth rate of the upper and lower bounds of the support are defined as $g(t) \equiv M'(t)/M(t)$ and $g_{\bar{Z}}(t) \equiv \bar{Z}'(t)/\bar{Z}(t)$ if $\bar{Z}(t) < \infty$.

For notational simplicity, 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 will be denoted as $M'(z) \equiv \boldsymbol{\partial}_z M(z) \equiv \frac{\mathrm{d}M(z)}{\mathrm{d}z}$.

Adoption and Technology Diffusion. A firm has the option to adopt a new technology by paying a cost. Adoption means switching production practice by changing to a technology that some other firm is using. We model this adoption process as undirected search across firms.[17] That is, when a firm chooses to adopt, it immediately switches its productivity to a draw from a function of the existing productivity distribution $\Phi_i(t,Z)$.[18] Assume that an adopting firm draws an (i,Z) from distributions $\hat{\Phi}_\ell(t,Z)$ and $\hat{\Phi}_h(t,Z)$, where $\hat{\Phi}_i(t,Z)$ is the probability that an adopting firm becomes type i and has productivity less than Z. Let $\hat{\Phi}_i$ be such that $\hat{\Phi}_\ell(t,0) = \hat{\Phi}_h(t,0) = 0$ and $\hat{\Phi}_\ell(t,\bar{Z}(t)) + \hat{\Phi}_h(t,\bar{Z}(t)) = 1$. $\hat{\Phi}_i$ will be determined by the equilibrium $\Phi_\ell(t,Z)$ and $\Phi_h(t,Z)$. The exact specification of $\hat{\Phi}_i(t,Z)$ typically does not affect the qualitative results, so we will write the process fairly generally and then analyze specific cases. One simplification that we maintain is that the draw of a new (i,Z) is independent of a firm's current type. A version of the model in which the firm can direct its search towards better technologies at some cost is nested in Section 5.

The cost of adoption grows as the economy grows. The endogenous scale of the economy is summarized by the minimum of support of the productivity distribution $M(t)$. For simplicity, the cost of adoption is proportional to the scale of the aggregate economy: $\zeta M(t)$.

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

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

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

[17] Instead of random matching, Luttmer (2015a) uses matching of teachers and students. The sorts of learning delays embedded in Luttmer (2015a,b) give a qualitatively similar foundation to random matching from some distortion of the productivity distribution. Furthermore, If the discrete choice of an immediate innovation arrival seems stark, an alternative is to use choice of an innovation intensity, as in Lucas and Moll (2014). Since the optimal innovation intensity is decreasing in relative productivity in that model, it would deliver qualitatively similar results at the aggregate level.

[18] See Technical Appendix C.2 for a proof showing that a firm's ability to recall its last productivity doesn't change the equilibrium conditions; and Technical Appendix C.1 for a derivation when adoption is not instantaneous—where our main specification is the limit of the adoption arrival rate going to infinity.

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.

The value of adopting is the continuation value of having a new productivity and innovation type drawn from $\hat{\Phi}_\ell$ and $\hat{\Phi}_h$ minus the cost of adoption:

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

Optimal Adoption Policy and the Minimum of the Productivity Distribution Support. 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. All firms choose an identical threshold, $M_a(t)$, above which they will continue operating with their existing technology and, otherwise, will adopt a new technology (i.e., a firm with $Z \leq M_a(t)$ is in the adoption region of the productivity space). While the adoption threshold could depend on the type i, see Appendix A.2 for a proof showing that ℓ and h firms choose the same threshold, $M_a(t)$, if 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 $\Phi_i(t,Z)$ distribution, $M(t)$, and in an abuse of notation, we will refer to both the minimum of support and the firm adoption policy as $M(t)$ going forward.[19]

In principle, there may be adopters of either innovation type with productivity in the common adoption region with $Z \leq M(t)$. Define $S_i(t) \geq 0$ as the flow of i-type firms entering the adoption region at time t. 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. Necessary conditions for the optimal stopping problem include the continuation value functions and, at the endogenously chosen adoption boundary, $M(t)$, value matching,

$$\underbrace{V_i(t,M(t))}_{\text{Value at Threshold}} = \underbrace{\int_{M(t)}^{\bar{Z}(t)} V_\ell(t,Z)\mathrm{d}\hat{\Phi}_\ell(t,Z) + \int_{M(t)}^{\bar{Z}(t)} V_h(t,Z)\mathrm{d}\hat{\Phi}_h(t,Z)}_{\text{Gross Adoption Value}} - \underbrace{\zeta M(t)}_{\text{Adoption Cost}}, \tag{4}$$

as well as 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, a firm must be indifferent between producing with the reservation productivity or adopting a new productivity. Smooth-pasting is a technical requirement that can be interpreted as an intertemporal no-arbitrage condition that ensures that the recursive system of equations is equivalent to the fundamental optimal stopping problem. Thus, the smooth-pasting conditions are necessary only if firms at the boundary are moving backwards relative to the boundary over time.

[19] To see why the minimum support is the endogenous threshold, consider instantaneous adoption as the limit of the 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. Because of the immediacy of draws, the stationary equilibrium does not depend on whether draws are from the unconditional distribution or from the distribution conditional on being above the current adoption threshold. This is the same as the small time limit of Perla and Tonetti (2014), which solves both versions of the model. A more formal derivation of this cost function as the limit of the arrival rate of unconditional draws is in Technical Appendix C.1.

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 $\bar{Z}(t)$; that is, $\bar{Z}'(t)/\bar{Z}(t) = \gamma$ if $\Phi_h(t, \bar{Z}(t)) - \Phi_h(t, \bar{Z}(t) - \epsilon) > 0$, for all $\epsilon > 0$. With the continuum of firms and the memoryless Poisson arrival of changes in i, there will always be some h firms that have not jumped to the low state for any t, 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 CDFs for ℓ and h type firms are

$$\boldsymbol{\partial}_t \Phi_\ell(t, Z) = \underbrace{-\lambda_\ell \Phi_\ell(t, Z) + \lambda_h \Phi_h(t, Z)}_{\text{Net Flow from Jumps}} + \underbrace{(S_\ell(t) + S_h(t))}_{\text{Flow of Adopters}} \underbrace{\hat{\Phi}_\ell(t, Z)}_{\text{Draw} \leq Z} - \underbrace{S_\ell(t)}_{\ell\text{-Adopters}} \tag{7}$$

$$\boldsymbol{\partial}_t \Phi_h(t, Z) = \underbrace{-\gamma Z \boldsymbol{\partial}_Z \Phi_h(t, Z)}_{\text{Innovation}} - \lambda_h \Phi_h(t, Z) + \lambda_\ell \Phi_\ell(t, Z) + (S_\ell(t) + S_h(t)) \hat{\Phi}_h(t, Z) - S_h(t). \tag{8}$$

For each type i, the KFEs keep track of inflows and outflows of firms with a productivity level below Z. An i-type firm with productivity less than Z stops being in the i-distribution 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 two ways: adoption or innovation. Because an adopting firm has probability $\hat{\Phi}_i(t, Z)$ of becoming type i and drawing a productivity less than Z, and there are $(S_\ell(t) + S_h(t))$ number of firms adopting, $(S_\ell(t) + S_h(t))\hat{\Phi}_i(t, Z)$ is added to the i-distribution. Additionally, the flow of adopters of each type, $S_i(t)$, is subtracted from the corresponding distribution (this term appears conditional on any Z because adoption occurs at the minimum of 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 support, removes them from the distribution, and inserts them back into the distribution according to $\hat{\Phi}$.[20]

The KFE for the h-types has a term that subtracts the firms that grow above Z through innovation: there are $\boldsymbol{\partial}_Z \Phi_h(t, Z)$ number of h-type firms at productivity Z, and because innovation is geometric, they grow above Z at rate γ.[21]

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, there are $\Phi_\ell(t, Z)$ many ℓ-type firms with productivity less than Z, and they leave the ℓ distribution at rate λ_ℓ and enter the h-distribution at the same rate.

[20] Recognizing that the λ_i jumps are of measure 0 when calculating how many firms cross the boundary in any infinitesimal time period, the flow of adopters comes from the flux across the moving $M(t)$ boundary.

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

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

This is consistent with the solution to the ODEs in equations (7) and (8) at $Z = M(t)$, as is apparent in the normalized equations (21) and (22).

[21] To account for the drift component, assume a stochastic process with a drift $\mu(z)$, PDF $f(t, z)$, and CDF $F(t, z)$. The KFE comes from the adjoint to infinitesimal generator:

$$\boldsymbol{\partial}_t f(t, z) = -\boldsymbol{\partial}_z \left[\mu(z) f(t, z)\right] + \ldots \tag{11}$$

Integrating to get the CDF $F(t, z)$, either using the fundamental theorem of calculus or interchanging the order of differentiation and integration, yields

$$\boldsymbol{\partial}_t F(t, z) = -\mu(z) f(t, z) + \ldots \tag{12}$$

$$= -\mu(z) \boldsymbol{\partial}_z F(t, z) + \ldots \tag{13}$$

Consumer Welfare and the Interest Rate The firms are owned by a representative consumer who values the undifferentiated good with log utility and a discount rate $\rho > 0$.[22] Given the productivity distribution $\Phi(t, Z)$, welfare is

$$U(t) = \int_0^\infty e^{-\rho\tau} \log\left(\underbrace{\int_{M(t)}^\infty Z \boldsymbol{\partial}_Z \Phi(t+\tau, Z) \mathrm{d}Z}_{\text{Aggregate Output}}\right) \mathrm{d}\tau. \tag{14}$$

As is standard, along a balanced growth path with aggregate output growing at rate g, the discount rate that a firm faces is

$$r = \rho + g. \tag{15}$$

2.3 Normalization, Stationarity, and Balanced Growth Paths

In this paper, we study economies 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 growing at the same rate as the economy, it is most convenient to normalize variables relative to the endogenous boundary $M(t)$. Define the change of variables, normalized distribution, and normalized value functions as,

$$z \equiv \log(Z/M(t)) \tag{16}$$

$$F_i(t, z) = F_i(t, \log(Z/M(t))) \equiv \Phi_i(t, Z) \tag{17}$$

$$v_i(t, z) = v_i(t, \log(Z/M(t))) \equiv \frac{V_i(t, Z)}{M(t)}. \tag{18}$$

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

With the above normalizations, it is possible that the value function, productivity distribution, and growth rates can be stationary—i.e., independent of time.[23]

Summary of Stationary KFEs and Firm Problem. A full derivation of the normalized system is detailed in Appendix A.1. Here, we summarize the resulting equations that characterize

[22] The appendix uses a general firm discount rate r, and the numerical algorithm in Technical Appendix B is implemented for an arbitrary CRRA utility function. Using log utility here simplifies a few of the parameter restrictions and expressions in the algebra compared to linear utility or general CRRA.

[23] An important example is when $\Phi(t, Z)$ is Pareto with the minimum of support $M(t)$ and a tail parameter α:

$$\Phi(t, Z) = 1 - \left(\frac{M(t)}{Z}\right)^\alpha, \text{ for } M(t) \leq Z. \tag{19}$$

Then $F(t, z)$ is independent of M and t:

$$F(t, z) = 1 - e^{-\alpha z}, \text{ for } 0 \leq z < \infty. \tag{20}$$

This is the CDF of an exponential distribution, with parameter $\alpha > 1$. From a change of variables, if $X \sim Exp(\alpha)$, then $e^X \sim Pareto(1, \alpha)$. Hence, α is the tail index of the unnormalized Pareto distribution for Z.

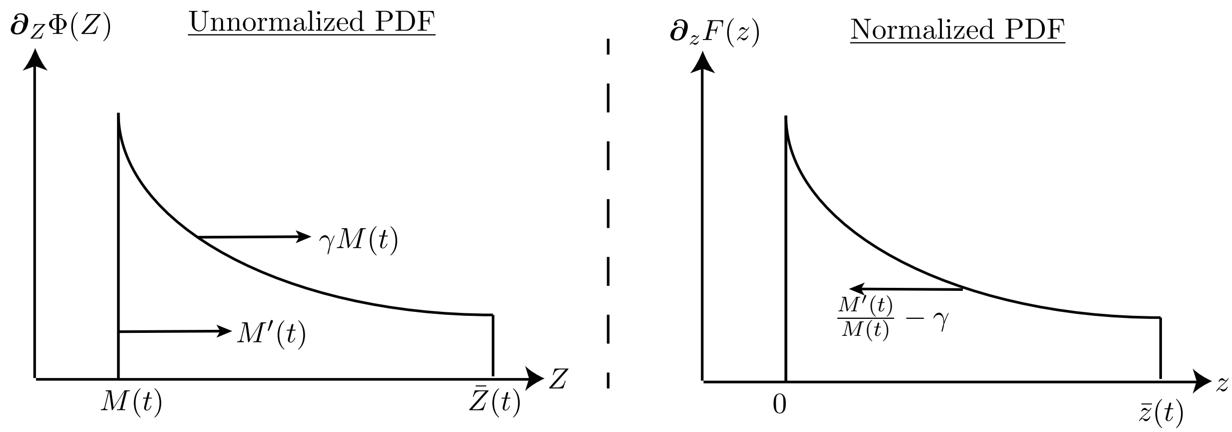

Figure 1: Normalized vs. Unnormalized Distributions

the law of motion for the normalized productivity distribution and the normalized firm problem. The equations that determine the stationary productivity distributions are:

$$0 = gF_\ell'(z) - \lambda_\ell F_\ell(z) + \lambda_h F_h(z) + (S_\ell + S_h)\hat{F}_\ell(z) - S_\ell \tag{21}$$
$$0 = (g-\gamma)F_h'(z) - \lambda_h F_h(z) + \lambda_\ell F_\ell(z) + (S_\ell + S_h)\hat{F}_h(z) - S_h \tag{22}$$
$$0 = F_\ell(0) = F_h(0) \tag{23}$$
$$1 = F_\ell(\bar{z}) + F_h(\bar{z}) \tag{24}$$
$$S_\ell = gF_\ell'(0), \text{ if } g > 0 \tag{25}$$
$$S_h = (g-\gamma)\,F_h'(0), \text{ if } g > \gamma. \tag{26}$$

The necessary conditions of the normalized firm problem are:

$$\rho v_\ell(z) = e^z - gv_\ell'(z) + \lambda_\ell\,(v_h(z) - v_\ell(z)) \tag{27}$$
$$\rho v_h(z) = e^z - (g-\gamma)v_h'(z) + \lambda_h\,(v_\ell(z) - v_h(z)) \tag{28}$$
$$v(0) = \frac{1}{\rho} = \int_0^{\bar{z}} v_\ell(z)\mathrm{d}\hat{F}_\ell(z) + \int_0^{\bar{z}} v_h(z)\mathrm{d}\hat{F}_h(z) - \zeta \tag{29}$$
$$v_\ell'(0) = 0, \text{ if } g > 0 \tag{30}$$
$$v_h'(0) = 0, \text{ if } g > \gamma. \tag{31}$$

Given that both firm types 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 (21) to (24) are the stationary KFEs with initial conditions and boundary values. Recall that g is the growth rate of the minimum of support and γ the innovation growth rate. In the normalized setup, firms are moving backwards towards the constant minimum of support, and their growth rate determines the speed at which they are falling back. S_ℓ in equation (25) is the flow of ℓ agents moving backwards at a relative speed of g across the adoption barrier, while S_h in equation (26) is the flow of h agents moving backwards at the slower relative speed of $g - \gamma$ across the barrier. The $\hat{F}_i(z)$ specification is some function of the equilibrium $F_i(z)$, and will be analyzed further in Sections 3.1 to 3.3.

Equations (27) and (28) are the Bellman Equations in the continuation region, and (29) is the value-matching condition between the continuation and technology adoption regions. The smooth-pasting conditions in (30) and (31) are necessary only if the firms of that particular i are drifting backwards relative to the adoption threshold. See Figure 2 for a visualization of the normalized Bellman equations.

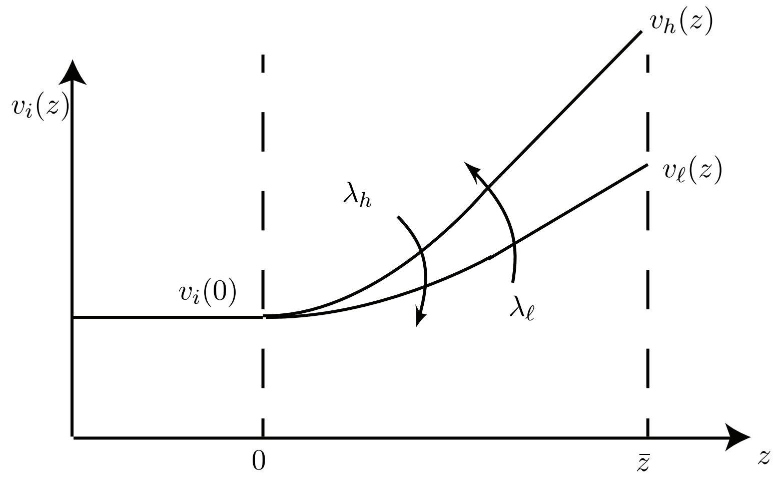

Figure 2: Normalized, Stationary Value Functions

Equilibrium Definitions. This paper studies balanced growth path equilibria, defined below.

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*

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 $\Phi_i(0, z)$.*

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. This is equivalent to requiring that $F_i(t, z) = F_i(z)$ and $g(t) \equiv M'(t)/M(t) = g$.*

Define the growth rate of aggregate output as $g_E(t) \equiv \boldsymbol{\partial}_t \mathbb{E}_t [Z] / \mathbb{E}_t [Z]$; then,

Lemma 1 (Growth of the Endogenous Adoption Threshold and Aggregate Output)**.** *On a balanced growth path, the growth rate of the endogenous threshold, $M(t)$, must be the same as the growth rate of aggregate output. That is, $g = g_E$.*

Proof. The value-matching condition in (29) is normalized to the endogenous threshold, and, hence, adoption has a constant and strictly positive cost. Thus, as $v(z) \geq z$, if the expected value of a draw from the technology distribution were not stationary relative to the adoption cost, then the value-matching could not hold with equality for all t. See Appendix A.7 for a similar, and more formal, case. $\square$

How Adoption and Innovation Generate a Stationary Normalized Distribution. As we will show, with geometric growth, the stationary solutions will endogenously become power-law distributions asymptotically, as discussed with generality in Gabaix (2009). Figure 3 provides some intuition on how proportional growth and adoption can create a stationary distribution. Stochastic innovation spreads out the distribution and, in the absence of endogenous adoption, this would prevent the existence of a stationary distribution. Without endogenous adoption, there is no "absorbing" or "reflecting" barrier, and proportional random shocks generate a variance diverging to infinity. 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.

Much of the intuition for how innovation and adoption can generate a stationary normalized power-law productivity distribution can be seen in the same model, except with innovation that follows geometric Brownian motion. Technical Appendix E solves this modified model in closed form. In the GBM case, however, even with a finite $\bar{Z}(0)$ initial condition, the support of a stationary $F(z)$ will be $[0, \infty)$ since, with a continuum of agents, Brownian motion instantaneously increases the support of the distribution to infinity. Thus, this GBM model is well-suited to gaining intuition for some features of our baseline model, but it is ill-suited to study the main questions surrounding the expansion of the frontier, which drive the analysis in this paper.

Another advantage of our specification compared to GBM is that any firm's maximum growth rate is bounded by the innovation rate γ. Hence, if the aggregate growth rate $g > \gamma$, then we can easily identify conditions under which extra growth beyond innovation is possible—i.e., latent growth. Alternatively, if $g < \gamma$, it means that the growth rate of the distribution as a whole is unable to keep up with the innovation rate—i.e., latent growth with a negative sign.

The existence of a stationary normalized distribution immediately places restrictions on the relationship between the growth of the frontier and adoption behavior. Recall that g is the growth rate of the reservation adoption productivity and, thus, the growth rate of the minimum of support of the unnormalized distribution. A necessary condition for the existence of a stationary normalized distribution with a finite frontier (i.e., if $\bar{Z}(t) < \infty \ \forall\ t$) is that $g = g_{\bar{Z}} \leq \gamma$. That is, the minimum of support must grow at the same rate as the aggregate economy, and cannot grow faster than the frontier. The requirement that $g = g_{\bar{Z}}$ occurs since, otherwise, the returns to adoption would grow faster than the cost, contradicting the agent's endogenous adoption choice.

This may seem as though the growth rate of the minimum of support, which is determined by adopters, determines the long-run growth rate, but it should be read as an equilibrium relationship between adoption and innovation. The flow of adopters endogenously increases or decreases—adjusting the growth rate $g(t)$ until it is in balance with the growth rate of the frontier, γ.

Three Types of Stationary Distributions. There are three possibilities for the stationary normalized frontier, $\bar{z}$, that we will analyze separately. These are not mere technicalities, but, rather, each type of distribution is associated with very different firm behavior in a manner that is informative of the economic relationship among adoption, innovation, aggregate growth, and the shape of the productivity distribution.

The first case is if $\bar{z} = \infty$, which we call "infinite-support." Because we model innovation as a stochastic process with bounded growth rates, the infinite-support case in this model occurs if and only if the initial productivity distribution has infinite-support (i.e., $\bar{z}(0) = \infty$ or, equivalently, $\sup\{\text{ support }\{F(0, \cdot)\}\} = \infty$). Since essential properties of the equilibria are determined by whether the long-run productivity distribution has infinite or finite-support, this case also corresponds to economies like those studied in Luttmer (2007) (in which GBM generates infinite-support for any initial condition) and the infinite-support examples of Perla and Tonetti (2014) and Lucas and Moll (2014).

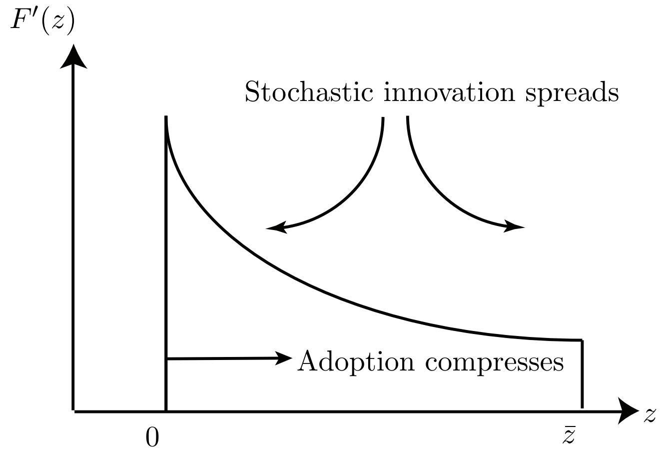

Figure 3: Tension between Stochastic Innovation and Adoption

The other two cases both have finite-support distributions ($\bar{Z}(t) < \infty$ at all time periods) but different properties of the normalized productivity frontier. The second case is when $\bar{z}(0) < \infty$ (which implies $\bar{z}(t) < \infty \; \forall \; t$), but where $\lim_{t\to\infty} \bar{z}(t) = \infty$. We label this case "finite unbounded support." The final case is when the initial condition has finite-support and $\lim_{t\to\infty} \bar{z}(t) < \infty$, which we refer to as "finite bounded support."

In Section 3, we provide the conditions necessary for the existence of each type of stationary distribution; show when there is hysteresis in which the stationary distribution depends on the initial conditions; derive properties of the stationary distribution, including the shape as a function of primitive adoption and innovation parameters; and study how adoption and innovation activity determine the aggregate growth rate. Finally, as the empirics in Section 6 show, we need to understand both the unbounded and bounded cases as there is no conclusive answer on whether the frontier is bounded—for most industries, at least.

3 Stationary BGP Equilibria with Exogenous Innovation

In this section, we compute BGP equilibria for economies with an exogenous innovation process and study their properties. There are three main questions that motivate this analysis. First, does adoption activity affect long-run growth rates and, if so, how and why? Second, how do adoption and innovation determine the shape of the productivity distribution? Third, under what conditions is the aggregate growth rate not equal to the innovation rate.

We show that the relationship of adoption to long-run aggregate growth depends crucially on whether the productivity distribution has finite-support. Model assumptions that generate infinite-support power-law distributions (whether assumed as an initial condition or as properties of the stochastic process) are not innocuous, in the sense that economies leading to distributions with finite-support have very different properties than those leading to infinite-support power-law economies.

First, whether the support is finite or infinite determines whether the aggregate growth rate can be larger than the growth rate of innovators. In finite-support BGP equilibria, the aggregate growth rate cannot exceed the growth rate of innovators (a parameter in this exogenous innovation section). In infinite-support BGPs, it is possible for aggregate growth to be faster than growth from innovation. For example, even if there is zero innovation, positive long-run aggregate growth can be sustained via adoption if the productivity distribution has infinite-support. Furthermore, conditions generating bounded versus unbounded stationary distributions will determine whether the growth rate can be strictly less than the innovation rate (i.e., adopters and aggregate output are not keeping up with the innovation rate).

Second, these different types of stationary distributions imply a different relationship between adoption activity and innovation incentives. Beyond determining if the growth rate can be less than the innovation rate, the distinction between bounded and unbounded stationary distributions becomes most important when innovation is endogenous. Although in this section we maintain exogenous innovation rates, by studying the firm value function, we can already see under which assumptions adoption impacts the long-run growth rate and why. Whether adoption influences long-run aggregate growth depends on how it impacts innovation incentives, which in turn depends on whether innovators at the frontier think they will become adopters in any reasonable time frame. Since adoption happens at the bottom of the distribution and innovation pushes out the frontier, the distance between the frontier and the minimum of support will partially determine the influence of adoption on long-run growth. In this respect, the unbounded support BGPs are closer to the infinite-support case, in that the firms at the frontier have an arbitrarily large $\bar{z}$ relative to the adopters, and thus have an infinitesimal option value of adoption. With finite bounded support distributions, innovators that are pushing out the frontier have positive value from the option to adopt and adoption can affect long-run growth by affecting innovation incentives.

Third, in addition to determining the aggregate growth rate, adoption and innovation determine the shape of the productivity distribution. Adoption is a force that compresses the distribution while innovation stretches it. Geometric growth through innovation combined with adoption generates, asymptotically, a power-law productivity distribution, roughly consistent with firm size data. In this section we detail how innovation and adoption determine the ratio of best to worst technologies and the shape (tail-index) of the productivity distribution.

Fourth, we also show under which conditions hysteresis exists, in which the stationary distribution is a function of the initial distribution. While there can be a unique equilibrium (and aggregate growth rate) associated with finite BGPs, infinite-support allows for a continuum of possible equilibria (and growth rates) that depend on the exact initial condition.

All of these results suggest that caution should be exercised when using an infinite-support distribution as an approximation of a finite, even if ultimately unbounded, empirical distribution.

3.1 Stationary BGP: Infinite Support

To set the stage, we first study the economy with infinite-support, in which the maximum growth rate of innovations is γ. The infinite-support case is useful as a foil to the finite-support BGPs. For reference, the characteristics of this model with infinite-support are qualitatively similar to the model with innovation driven by GBM that is solved in closed form in Technical Appendix E. The infinite-support case in this economy with the two-state Markov innovation process, however, is developed here for better comparison to the finite-support cases.

Since we are most interested in when growth rates can exceed innovation rates, we will concentrate on cases in which $g \geq \gamma$. While in the finite-support case it must be that $g \leq \gamma$, in the infinite-support case there is no such requirement. An important property of the infinite-support economy is the possibility that $g > \gamma$, and we will focus on these equilibria of interest.[24]

For ease of exposition, in this section, we model the adoption technology as firms copying both the type and productivity of their draw from the unconditional distribution.[25] That is, the normalized adoption measures are $\hat{F}_\ell(z) \equiv F_\ell(z)$ and $\hat{F}_h(z) \equiv F_h(z)$. Together with equations (23)

[24]In cases in which $g < \gamma$, the infinite- and finite-support distributions are equivalent. See Appendix A.6.

[25]While an exactly correlated draw of the type and the productivity is not necessary here, see Technical Appendix C.3 for a proof showing that independent draws for adopters of Z and the innovation type i have infinite-support equilibria only with degenerate stationary distributions. The finite-support cases do not impose the same requirements.

to (28), (30) and (31), the following value-matching and KFEs are the necessary conditions for a BGP equilibrium.

$$v(0) = \frac{1}{\rho} = \underbrace{\int_0^\infty v_\ell(z)\mathrm{d}F_\ell(z) + \int_0^\infty v_h(z)\mathrm{d}F_h(z)}_{\text{Adopt both } i \text{ and } Z \text{ of draw}} - \zeta \tag{32}$$

$$0 = gF_\ell'(z) - \lambda_\ell F_\ell(z) + \lambda_h F_h(z) + (S_\ell + S_h)F_\ell(z) - S_\ell \tag{33}$$

$$0 = (g-\gamma)F_h'(z) - \lambda_h F_h(z) + \lambda_\ell F_\ell(z) + (S_\ell + S_h)F_h(z) - S_h. \tag{34}$$

If $\Phi(0, Z)$ has infinite-support, the normalized $F(t, Z)$ will converge to a stationary distribution as $t \to \infty$. A continuum of stationary distributions exists; they are determined by initial conditions, and each is associated with its own aggregate growth rate. To characterize the continuum of stationary distributions, parameterize the set of solutions by a scalar α. By construction, α will be the tail index of the unconditional distribution $F(z)$. Define the following as a function of the parameter α with an accompanying growth rate, $g(\alpha)$,

$$\vec{F}(z) \equiv \begin{bmatrix} F_\ell(z) \\ F_h(z) \end{bmatrix} \qquad \vec{v}(z) \equiv \begin{bmatrix} v_\ell(z) \\ v_h(z) \end{bmatrix} \tag{35}$$

$$A \equiv \begin{bmatrix} \frac{1}{g} \\ \frac{1}{g-\gamma} \end{bmatrix} \qquad B \equiv \begin{bmatrix} \frac{\rho+\lambda_\ell}{g} & -\frac{\lambda_\ell}{g} \\ -\frac{\lambda_h}{g-\gamma} & \frac{\rho+\lambda_h}{g-\gamma} \end{bmatrix} \tag{36}$$

$$\varphi \equiv \sqrt{(\lambda_h - \alpha\gamma)^2 + 2\lambda_l(\alpha\gamma + \lambda_h) + \lambda_l^2} \tag{37}$$

$$C \equiv \begin{bmatrix} \frac{-\alpha\gamma + 2\alpha g + \lambda_h - \lambda_l + \varphi}{2g} & \frac{\lambda_h}{g} \\ \frac{\lambda_l}{g-\gamma} & \frac{-\alpha\gamma + 2\alpha g - \lambda_h + \lambda_l + \varphi}{2(g-\gamma)} \end{bmatrix} \tag{38}$$

$$D \equiv \begin{bmatrix} \frac{\lambda_h(g(\alpha\gamma - \lambda_h + \varphi - \lambda_l) + 2\gamma\lambda_l)}{\gamma g(\alpha\gamma - \lambda_h + \varphi + \lambda_l)} \\ \frac{g(\alpha\gamma - \lambda_h + \varphi - \lambda_l) + 2\gamma\lambda_l}{2\gamma(g-\gamma)} \end{bmatrix} \tag{39}$$

Proposition 1 (Stationary Equilibrium with Infinite Support)**.** *There exists a continuum of equilibria parameterized by $\alpha > 1$, which, by construction, is the tail index of F. An equilibrium is determined by the $g(\alpha)$ that satisfies*

$$\frac{1}{\rho} + \zeta = \int_0^\infty \left[\left[(\mathbf{I} + B)^{-1} \left(e^{\mathbf{I}z} + e^{-Bz} B^{-1} \right) A \right]^T e^{-Cz} D \right] \mathrm{d}z \tag{40}$$

and the parameter restrictions given in (A.83) *to* (A.86)*. The stationary distributions and the value functions are given by:*

$$\vec{F}(z) = \left(\mathbf{I} - e^{-Cz} \right) C^{-1} D \tag{41}$$

$$\vec{F}'(z) = e^{-Cz} D \tag{42}$$

$$\vec{v}(z) = (\mathbf{I} + B)^{-1} \left(e^{\mathbf{I}z} + e^{-Bz} B^{-1} \right) A. \tag{43}$$

The aggregate growth rate is $g(\alpha)$.

Proof. See Appendix A.5 and (A.83) to (A.86) for parameter restrictions that ensure that $r > g$ for a general CRRA and positive eigenvalues of B and C. Positive eigenvalues of B ensure that value-matching is defined, and the option value of diffusion asymptotically goes to 0 for large z. Positive eigenvalues of C are equivalent to $\alpha > 0$. □

With any infinite-support initial condition, the BGP equilibrium maintains the infinite support, and the stationary productivity distribution F is a power-law with a particular tail index and accompanying aggregate growth rate. Proposition 1 states that there is a continuum of combinations of growth rates and shapes of the distribution (tail index α) that are equilibria.[26] Essentially, a fatter-tailed stationary distribution provides stronger incentives to adopt, which generates faster growth. Thus, generically, the aggregate growth rate is not equal to γ, but to $g > \gamma$, and long-run growth is affected by adoption per se, independent of adoption's affect on innovation activity. Of course, g is a function of γ, so aggregate growth depends on innovation, but even if $\gamma = 0$, growth can be positive.[27] This hysteresis and the possibility of long-run growth without innovation are reminiscent of the equilibria computed in Perla and Tonetti (2014) and Lucas and Moll (2014). Furthermore, adoption and aggregate growth are a function of the shape of the distribution, but the shape of the distribution is determined by the initial distribution or exogenous stochastic parameters, not endogenously by the interaction of adoption and innovation activity. In the next section, we compute the finite-support BGP—which comes about from finite initial conditions—and compare its properties with those of the infinite-support case.

3.2 Stationary BGP: Finite Unbounded Relative Support

In this section, we study the BGP equilibrium in the case when the initial distribution $\Phi(0, Z)$ has finite support. Due to the bounded growth rates of the Markov process, if the support of $\Phi(0, z)$ is finite, then it remains finite as it converges to a stationary distribution. Recall that, with a finite frontier, $g \leq \gamma$ to have a BGP with non-degenerate $F_i(z)$. Hence, in the stationary equilibrium, there are no h-type agents hitting the adoption threshold (and the smooth-pasting condition for h firms is not a necessary condition). As will become clear when we contrast bounded and unbounded support examples in Section 3.3, the $g = \gamma$ case is the most natural, and the $g < \gamma$ cases occur only as knife-edge results.

In this section and the rest of the paper, for simplicity we consider the process of adopting new technologies disruptive to R&D, in that an adopting firm becomes an ℓ type regardless of its former type.[28] That is,

$$\hat{F}_\ell(z) = F_\ell(z) + F_h(z) = F(z) \tag{44}$$
$$\hat{F}_h(z) = 0. \tag{45}$$

Necessary conditions for a stationary equilibrium with a finite initial frontier are $v_\ell(z)$, $v_h(z)$, $F_\ell(z)$, $F_h(z)$, and S—such that (23) to (25), (27) and (30), and

$$\rho v_h(z) = e^z + \lambda_h(v_\ell(z) - v_h(z)) \tag{46}$$
$$v(0) = \frac{1}{\rho} = \int_0^{\bar{z}} v_\ell(z) \mathrm{d}F(z) - \zeta \tag{47}$$
$$0 = gF_\ell'(z) + SF(z) + \lambda_h F_h(z) - \lambda_\ell F_\ell(z) - S \tag{48}$$
$$0 = \lambda_\ell F_\ell(z) - \lambda_h F_h(z) \tag{49}$$

Define the constants, $\hat{\lambda} \equiv \frac{\lambda_\ell}{\lambda_h}$ and $\bar{\lambda} \equiv \frac{\lambda_\ell}{\rho+\lambda_h} + 1$.

[26] Luttmer (2012) discusses sources of hysteresis in related models. He shows that if the initial distribution is thin-tailed and innovation occurs via a GBM, then the unique stationary distribution is that associated with the thinest power-law initial condition. We would expect similar results with our model.

[27] In the case studied in Technical Appendix E, in which innovation follows GBM instead of the finite-state Markov process, g can be computed in closed form and can be decomposed into growth that would occur absent any innovation plus growth from the drift in innovation plus growth from the randomness in innovation that induces adoption.

[28] Unlike the infinite-support case in Section 3.2, important properties of the finite-support equilibria are not sensitive to the degree of correlation in the draws, and we have simply chosen the most analytically convenient.

Proposition 2 (Stationary Equilibrium with Finite Unbounded Support)**.** *There exists a unique maximum growth equilibrium with $g = \gamma$ and $\bar{z} \to \infty$. There does not exist an equilibrium with finite and bounded support.*

The unique stationary distribution is,

$$F_\ell(z) = \tfrac{1}{1+\hat{\lambda}} e^{-\alpha z} \tag{50}$$

$$F_h(z) = \hat{\lambda} F_\ell(z), \tag{51}$$

where α is the tail index of the power-law distribution:

$$\alpha \equiv (1 + \hat{\lambda}) F_\ell'(0), \tag{52}$$

and $F_\ell'(0)$ is determined by model parameters:

$$\tag{53}$$

$$F_\ell'(0) = \frac{\lambda_h \Big(\zeta(\rho+\gamma)((\rho+\gamma)+\lambda_h+\lambda_\ell) - \sqrt{\zeta\Big(\Big(4\gamma+(\rho+\gamma)^2\zeta\Big)(-\gamma+(\rho+\gamma)+\lambda_h)^2 + 2(-2\gamma+(\gamma-(\rho+\gamma))(\rho+\gamma)\zeta)(\gamma-(\rho+\gamma)-\lambda_h)\lambda_\ell + (\gamma-(\rho+\gamma))^2\zeta{\lambda_\ell}^2\Big)} + \zeta\gamma^2 2 + \zeta(-\gamma)(3(\rho+\gamma)+2\lambda_h+\lambda_\ell)\Big)}{\gamma 2\zeta(\lambda_h+\lambda_\ell)(\gamma-(\rho+\gamma)-\lambda_h)} \tag{54}$$

The firm value functions are,

$$v_\ell(z) = \frac{\bar{\lambda}}{\gamma + \rho\bar{\lambda}} e^z + \frac{1}{\rho(\nu+1)} e^{-\nu z} \tag{55}$$

$$v_h(z) = \frac{e^z + \lambda_h v_\ell(z)}{\rho + \lambda_h}, \tag{56}$$

where ν is defined as

$$\nu \equiv \frac{\rho\bar{\lambda}}{\gamma} > 0. \tag{57}$$

Proof. See Appendix A.3. □

Proposition 2 shows that while, initially, finite-support distributions maintain finite support for all $t < \infty$, stationary distributions all have asymptotically infinite relative support. That is, the ratio of frontier to lowest productivity goes to infinity: $\bar{z} \to \infty$. Given the Markov process for ℓ and h, there will be some agents who hit lucky streaks more than others, escape from the pack, and break away. There is no finite bounded BGP because a diminishing, yet strictly positive number of firms at the frontier keep getting lucky and grow at γ forever.[29]

Why don't the low-productivity firms simply adopt more rapidly to maintain a compact distribution? The reason is that although the mass of firms near the frontier is always positive, the number of such firms is vanishingly small. Consequently, the probability an adopting firm will adopt technologies near the frontier becomes arbitrarily close to zero.

The value functions given by equations (55) and (56) have an intuitive interpretation that is informative about firms' adoption incentives and behavior. The stationary value of an h firm is that it will continue to produce with its current z continually earning flow profits e^z, discounted for time and accounting for the probability that it will switch to being an ℓ type at rate λ_h. Similarly, an ℓ firm values producing with its current z earning flow profits e^z, discounting for time and switches to h, but also considers that its productivity is falling behind relative to innovators who are growing

[29]This result is robust to variations in the diffusion specification, including assuming that adopting agents draw from the $F_h(z)$ distribution and start with an h type, which adds the strongest incentives to increase diffusion and compress the distribution.

at rate γ. Furthermore, since it is the ℓ firms that adopt in equilibrium, they have an extra term that captures the option value of being able to adopt. The elasticity parameter ν measures how the adoption option value decreases with higher productivity. A firm with higher productivity will take longer in expectation to fall back to the adoption threshold, and, thus, higher-productivity firms place less value on the option to adopt. Note that $\nu > 0$ shows that in the unbounded case with $\bar{z} \to \infty$, in the limit, firms at the frontier place no value on the option to adopt.[30]

By comparing Propositions 1 and 2, we see that the assumption that the distribution has infinite-support at time zero is not innocuous. When the initial distribution has infinite-support, there is a continuum of BGP equilibria indexed by the tail parameter of the stationary distribution (α), with each initial distribution mapping to a particular α. In this infinite-support case, aggregate growth can be faster than the rate of innovation—i.e., $g(\alpha) > \gamma$ or latent growth. In contrast, the BGP in the finite unbounded case—which *asymptotically* features a power-law distribution with an unbounded relative support ($\bar{z} \to \infty$)—has a unique maximum-growth induced stationary distribution that is independent of the initial distribution and has $g = \gamma$.

Thus, in the finite unbounded economy with exogenous innovation, adoption affects the shape of the distribution, but the aggregate growth rate is determined solely by the innovation rate parameter γ. When γ is a choice and growth is endogenous, the option value of adoption for frontier firms will influence whether and how adoption affects long-run growth.

3.2.1 Equilibrium with $g < \gamma$

While we have concentrated on equilibria with maximum growth, $g = \gamma$, there may be other equilibria with $g < \gamma$. This is the equivalent to latent growth, but with a negative sign. Intuitively, the initial conditions of the economy are dragging it down so that a particular BGP growth rate is kept below the innovation rate. An initial condition with a different arrangement of productivities that provided better adoption incentives could bring the aggregate growth rate up to the innovation rate.

Proposition 3 (Finite Unbounded Equilibrium with $g < \gamma$)**.** *For both a infinite- and finite-support initial conditions, there exists a continuum of equilibrium parameterized by $\alpha > 1$, which, by construction, is the tail index of the productivity distribution. An equilibrium is determined by the aggregate growth rate $g(\alpha) < \gamma$ that satisfies*

$$\frac{1}{\rho} + \zeta = a_2 b_1 \left(\frac{a_3}{\alpha + \beta_2} - \frac{a_3 b_2}{\bar{\alpha} + \beta_2} - \frac{1}{\alpha + \beta_1} + \frac{b_2}{\bar{\alpha} + \beta_1} \right) + a_1 b_1 \left(\frac{1}{1 - \alpha} + \frac{b_2}{\bar{\alpha} - 1} \right) \tag{58}$$

given $a_1, a_2, a_3, b_1, b_2, \beta_1, \beta_2$, and $\bar{\alpha}$ defined in (A.103)*,* (A.108) *to* (A.110) *and* (A.112) *to* (A.114)*.*

Proof. See Appendix A.6. Parameter restrictions are given in (A.116), (A.120) and (A.121). □

The example of Figure 18 shows continuity from the endogenous α at the $g = \gamma$ solution calculated in Section 3.2 to lower g as the tail thins out. However, as we will see in Section 3.3, the $g < \gamma$ cases disappear whenever there are conditions that generate a finite bounded distribution.

The low-productivity firms don't adopt fast enough to keep up with the innovation rate for reasons similar to why there is no bounded equilibrium. As the mass of the lucky agents with extremely high z thins out, it doesn't strongly affect the diffusion incentives and adoption probabilities of those with a low z. The reason that adopters don't find it worthwhile to keep up is that the mean can expand more slowly than the frontier. The incentive for adoption, which drives the speed of the moving barrier at the minimum of the distribution, is given by the expected draw in

[30]This interpretation applies more generally to other models with knowledge diffusion. For example, a similar result would hold in Lucas and Moll (2014) or Luttmer (2015b), where the "option value" in our model would be the component of the value function that captures the opportunities to learn from agents with higher productivity.

productivity. Therefore, if the frontier diverges to infinity, but the mean doesn't keep growing at the same rate, the frontier technology will diverge.[31]

Possible multiplicity in the $g < \gamma$ direction is less concerning than the $g > \gamma$ case, because it is easy to believe there is a possibility that innovations are inefficiently disseminated across the economy. Essentially, since not all firms are growing at the rate γ, if the tails are thin enough, then the incentives for adopting—conditional on that particular distribution shape—can be insufficient to maintain aggregate growth at the rate of the frontier. However, as Section 3.3 demonstrates, this multiplicity is unstable and is eliminated as soon as there is any chance of jumping to the frontier. Loosely speaking, in models with Schumpeterian forces, $g < \gamma$ equilibria would not exist and latent growth would never have a negative sign.

3.2.2 Extension to Distorted Adoption Distribution

The reason that there is no bounded support BGP is that the adopters don't keep up with the innovators at the frontier. 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.[32] We show that unboundedness is a rather robust result, in the sense that higher chances of adopting better technologies are not enough to generate bounded equilibria.

The distortion, representing the degree of imperfect mobility or beneficial adoption prospects, is indexed by $\kappa > 0$, where the agent draws its Z from the CDF $\Phi(t, Z)^\kappa$. Note that for higher κ, the probability of a better draw increases.The stationary normalized adoption distribution is

$$\hat{F}_\ell(z) = (F_\ell(z) + F_h(z))^\kappa = F(z)^\kappa \tag{59}$$
$$\hat{F}_h(z) = 0. \tag{60}$$

We write $F(z)^\kappa$ for the normalized draw process.
Necessary conditions for a stationary equilibrium with a finite initial frontier are $v_\ell(z)$, $v_h(z)$, $F_\ell(z)$, $F_h(z)$, and S—such that (23) to (25), (27), (30) and (46), and

$$v(0) = \frac{1}{\rho} = \int_0^{\bar{z}} v_\ell(z) \mathrm{d}F(z)^\kappa - \zeta \tag{61}$$
$$0 = gF_\ell'(z) + SF(z)^\kappa + \lambda_h F_h(z) - \lambda_\ell F_\ell(z) - S \tag{62}$$
$$0 = \lambda_\ell F_\ell(z) - \lambda_h F_h(z). \tag{63}$$

Corollary 2. *There does not exist an equilibrium with finite and bounded support for any $\kappa > 0$. Furthermore, κ distorts the unique tail parameter of the (asymptotically) power-law stationary distribution.*

Proof. See Technical Appendix C.4. □

Modifying κ may provide a better description of the adoption process and allow for an improved fit to the data for the shape of the stationary productivity distribution. Nonetheless, smoothly

[31] Given a fixed barrier, the logic is similar to the linear or asymptotically linear Kesten processes bounded away from zero with affine terms, for which, under appropriate conditions, the asymptotic tail index can be explicitly computed in terms of the stationary distribution induced by the Markov process. In our model, however, the adoption process introduces non-linear jumps that are not multiplicative in productivities and that do not permit the simple characterization of the tail index. The endogenous absorbing adoption barrier (which acts like a reflecting barrier with stochastic jumps) complicates the analogy since it moves according to the optimal adoption behavior of firms.

[32] While κ is exogenous here, in Section 5, we solve a version of the model in which κ is endogenous. We interpret the choice of κ as limited directed search, in that, at some cost, firms can focus their attention on adopting better technologies with higher probability.

twisting the adoption distribution so that firms obtain substantially better technologies with high probability is not enough to compress the distribution by allowing adopters to keep up with the frontier. In the next section, we extend the model to obtain existence of BGP with a finite bounded stationary distribution. The bounded and unbounded equilibria are similar in many ways, but comparison of the firm value functions will show how adoption affects innovation once firms choose how fast to innovate.

3.3 Stationary BGP: Finite Bounded Relative Support

For a finite bounded equilibrium to exist, Proposition 2 shows that it is not enough that adopters have a set of adoption possibilities growing with the scale of the economy. Rather, there must be some way that firms not at the frontier can catch up to those few lucky firms that have been h-hypes at the frontier forever. In this section, we extend the model by introducing another type of productivity-enhancing activity to construct finite bounded BGP equilibria.

This section helps us to further understand the hysteresis and deviations from $g = \gamma$. Recall that the multiplicity with $g > \gamma$ inherent in the infinite-support case of Section 3.1 was eliminated because we both bounded the growth rates and assumed a finite initial condition on the distribution. However, this left a possibility of $g < \gamma$. In this section, we will show that conditions generating a finite bounded relative support also ensure a unique equilibrium with $g = \gamma$.

One motivation for modeling a bounded relative support is that, regardless of the commodity space or model of competition, for all practical purposes, a firm grossly less productive than the frontier—whether $100\times$, $1000\times$, or larger, but at some point—likely could not survive in a market populated by firms with such better technology. The more productive firms would move in, undercut, displace, and capture the market, or be imitated. In some sense, this bounded frontier model provides an intermediate case between Schumpeterian models, in which one or two firms at the frontier dominate all production, and the infinite dispersion model, in which firms of radically different productivities coexist. A bounded frontier also seems consistent with the, admittedly, limited evidence on the stability of the right tail of the relative firm size distribution, which Luttmer (2010) shows have been constant since 1977 using the Business Dynamics Statistics data.

In this section, we show that the bounded frontier case that seems reasonable on first principles has important implications for the determination of the aggregate growth rate and for how adoption and innovation interact. In the exogenous growth context, this bounded frontier case is a building block for the model with a frontier that evolves endogenously in response to innovation decisions of high-productivity firms developed in Section 4.2. As will become clear, an important interaction determining the endogenous growth rate appears in the decisions of firms at the frontier. With an unbounded frontier, frontier firms are arbitrarily far away from the adoption threshold, and, thus, the interactions between technology adoption and innovation are limited, unlike in this bounded case.

An alternative way to model within-firm productivity change is to assume that firms have some chance of leapfrogging to the frontier. Such leapfrogging is a continuous-time version of a quality-ladders model.[33] For generality, the jumps can occur either for adopting or for innovating firms. Innovators jumping to the frontier represents some chance that the innovation process generates a big insight, instead of steady incremental progress. Different from autarkic innovation that generates γ proportional growth through process improvement, leapfrogging may be viewed as a melding of innovation and diffusion since the jump is a function of the existing productivity

[33] In a continuous-time model with leapfrogging event arrivals that generate a multiplicative step above the frontier, the frontier would become infinite immediately, as there would be some agent with an arbitrarily large number of jump arrivals in any positive time interval (as in König, Lorenz, and Zilibotti (2016)). An alternative process with similar qualitative implications is to recast leapfrogging as a step-by-step innovation model in the spirit of Aghion, Akcigit, and Howitt (2014). In either case, infinite support allows for the possibility of latent growth.

distribution. Adopters jumping to the frontier captures 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 $\ell-$types and must wait for the Markov transition to h before they become innovators again.[34]

We modify the model that generated finite unbounded BGP, as characterized in Proposition 2, by adding an arrival rate of jumps to the frontier, $\eta \geq 0$, as depicted in Figure 4.

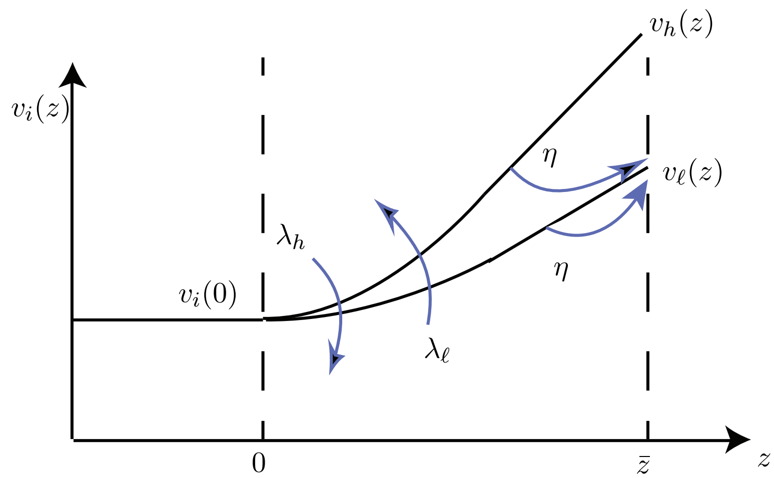

Figure 4: Normalized, Stationary Value Functions with Bounded Support

In principle, there *could* be a jump discontinuity in the CDF $F(z)$ at $\bar{z}$. Due to right continuity of the CDF, the mass at the discontinuity $z = \bar{z}$ is

$$\Delta_i = \lim_{\epsilon \to 0} \left(F_i(\bar{z}) - F_i(\bar{z} - \epsilon) \right). \tag{64}$$

When all leapfrogging firms become type ℓ, we can show that $F_i(z)$ are continuous and $\Delta_\ell = \Delta_h = 0$. This is because type ℓ firms immediately fall back in relative terms and don't pool at the frontier. As those firms falling back also jump back and forth to the h type, the $F_\ell(z)$ and $F_h(z)$ distribution smoothly mix, ensuring continuity.

Define $\mathbb{H}(z)$ as the Heaviside operator. Along with (23) to (25) and (30), the following characterizes the necessary conditions for a stationary equilibrium:

$$\rho v_\ell(z) = e^z - g v_\ell'(z) + \lambda_\ell(v_h(z) - v_\ell(z))) + \eta(v_\ell(\bar{z}) - v_\ell(z)) \tag{65}$$
$$\rho v_h(z) = e^z + \lambda_h(v_\ell(z) - v_h(z)) + \eta(v_\ell(\bar{z}) - v_h(z)) \tag{66}$$
$$0 = g F_\ell'(z) + \lambda_h F_h(z) - \lambda_\ell F_\ell(z) - \eta F_\ell(z) + \eta \mathbb{H}\left(z - \bar{z}\right) + S F(z) - S \tag{67}$$
$$0 = \lambda_\ell F_\ell(z) - \lambda_h F_h(z) - \eta F_h(z). \tag{68}$$

[34]The assumption that all leapfroggers switch to the ℓ state is purely for analytical convenience and can be changed without introducing qualitative differences. If some firms leapfrogged to the h state instead, then a right discontinuity in $F_h(z)$ would exist ($\Delta_h > 0$) and more care would be necessary in solving the KFE and integrating the value-matching condition. Furthermore, this simplification ensures that the values of jumps to the frontier remain identical for both agents, and, hence, both types have the same adoption threshold, as demonstrated in Appendix A.2. With this specification, a possible downside is that $v_\ell(\bar{z}) < v_h(\bar{z} - \epsilon)$ for some set of small ϵ, and those firms would rather keep the lower z rather than innovate. In the calibrated model, the difference between v_ℓ and v_h is tiny due to high switching probabilities, and, thus, this is binding for a very small number of firms.

While the value-matching condition is identical to equation (47) in the unbounded setting, the value of adopting is now dependent on the value at the frontier:

$$v(0) = \frac{1 + \eta v_\ell(\bar{z})}{\rho + \eta} = \int_0^{\bar{z}} v_\ell(z) \mathrm{d}F(z) - \zeta. \tag{69}$$

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 4 (Stationary Equilibrium with a Bounded Frontier)**.** *Given parameter restrictions, a unique equilibrium with $\bar{z} < \infty$ exists with $g = \gamma$. The stationary distribution is*

$$F_\ell(z) = \frac{F_\ell'(0)}{(F_\ell'(0) - \eta/\gamma)(1 + \hat{\lambda})}(1 - e^{-\alpha z}) \tag{70}$$

$$F_h(z) = \hat{\lambda} F_\ell(z), \tag{71}$$

with

$$\alpha \equiv (1 + \hat{\lambda})(F_\ell'(0) - \eta/\gamma) \tag{72}$$

$$\bar{z} = \frac{\log(\gamma F_\ell'(0)/\eta)}{\alpha}. \tag{73}$$

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

$$\zeta + \frac{1}{\rho} = \frac{\gamma F_\ell'(0) \alpha \bar{\lambda} \left(-\frac{e^{-\nu \bar{z}}\left(-1 + e^{-\alpha \bar{z}}\right)\eta}{\rho \alpha \nu} + \frac{e^{\bar{z}}\eta\left(e^{-\alpha \bar{z}} - 1\right)}{-\alpha \rho} + \frac{-e^{-(\alpha+\nu)\bar{z}} + 1}{\nu(\alpha + \nu)} + \frac{-e^{\bar{z} - \alpha \bar{z}} + 1}{\alpha - 1} \right)}{\gamma(\gamma F_\ell'(0) - \eta)(\nu + 1)}. \tag{74}$$

The value functions for the firm are,

$$v_\ell(z) = \frac{\bar{\lambda}}{\gamma(1 + \nu)} \left(e^z + \frac{1}{\nu} e^{-\nu z} + \frac{\eta}{\rho} \left(e^{\bar{z}} + \frac{1}{\nu} e^{-\nu \bar{z}} \right) \right) \tag{75}$$

$$v_h(z) = \frac{e^z + (\lambda_h - \eta) v_\ell(z) + \eta v_\ell(\bar{z})}{\rho + \lambda_h}. \tag{76}$$

Proof. See Appendix A.4. Parameter restrictions include those such that the $F_\ell'(0)$ that solves equation (74) is larger than η/γ. A proof of uniqueness, showing that $g < \gamma$ equilibria do not exist, is in Appendix A.7. □

As before, α represents the shape of the productivity distribution. The α is an empirical "tail index" that can be estimated from a discrete set of data points. Unlike in the unbounded case, $\bar{z}$ is now bounded, and, hence, this tail index is better interpreted as the shape parameter of a right-truncated power-law.

In contrast to the finite unbounded BGP, the Bellman equations (75) and (76) now contain the value of production and the option value of adoption for the frontier $\bar{z}$. That is, in addition to the value of production with current z modified by time discount and the probabilities of switching i-type, the value function accounts for the chance of jumping to the frontier. Thus, movements in the frontier provide enough incentive for adopters at the bottom of the distribution to keep up with the frontier, allowing for a bounded stationary distribution.

As noted at the end of the previous section, 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 group 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 almost every laggard from remaining a laggard forever. The distribution of relative productivities then remains bounded, as the frontier acts as a locomotive in a relay race. Note that this locomotive process is similar to models of technology diffusion in which the growth rate of adopters is an increasing function of the distance to the frontier, unlike this model of innovators with multiplicative growth in their productivity level (see, e.g., Nelson and Phelps (1966) and Benhabib, Perla, and Tonetti (2014)). If adopters were to fall behind, they would adjust their behavior, endogenously increasing their adoption activity to match innovators' growth rates, ensuring that relative productivities would be bounded in equilibrium.

Additionally, now that the relative frontier is bounded, in the long run, frontier firms still place positive value on the option to adopt. This means that increases in the value of adoption, either associated with lower costs or higher benefits of adoption, will affect the value of frontier firms. Depending on how far the frontier is from the adoption threshold, in the endogenous growth environment changes in the value of adoption will influence innovation behavior, which may affect aggregate growth rates.

Before moving to the endogenous growth case in the next section, we compute a finite bounded equilibrium with calibrated parameter values and use comparative statics to illustrate properties of the economy. The calibration is designed to set roughly set parameters in the relevant region of the state space. Details of the calibration are discussed in Appendix D.6. We set $\gamma = .02$ 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. We target a tail index $\alpha = 2.12$, which corresponds to a tail parameter of 1.06 in the size or profits distribution in a monopolisitic competition model, as used in Luttmer (2007). We target $\bar{z} = 1.61$ (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. Alternatively, comparing the largest to smallest firms in the economy yields a much larger 279:1 ratio ($\bar{z} = 5.63$), which is numerically close to ∞ for all intents and purposes in this model. The resulting parameterization is $\gamma = 0.02$, $\rho = 0.01$, $\lambda_\ell = 0.533, \lambda_h = 1.128$, and $\zeta = 25.18$ $\eta = 0.00098$.

First, as shown in Figure 5, 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 truncated power-law shaped, with many low-productivity firms and few high-productivity firms, but is truncated above at the $\bar{z}$ relative frontier. In this calibration, there are also fewer h firms than ℓ firms at all productivities.

Comparative statics on how changes in η, γ, ζ, and λ_h affect $\bar{z}$ and α are shown in Figure 6. 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.

3.4 Stationary BGPs with Exogenous Innovation: Summary and Analysis

An analysis of BGP equilbria with exogenous growth points to economically important differences between economies with infinite-support, finite bounded support, and finite bounded support distributions. There are two major difference between infinite- and finite-support BGPs. First, with

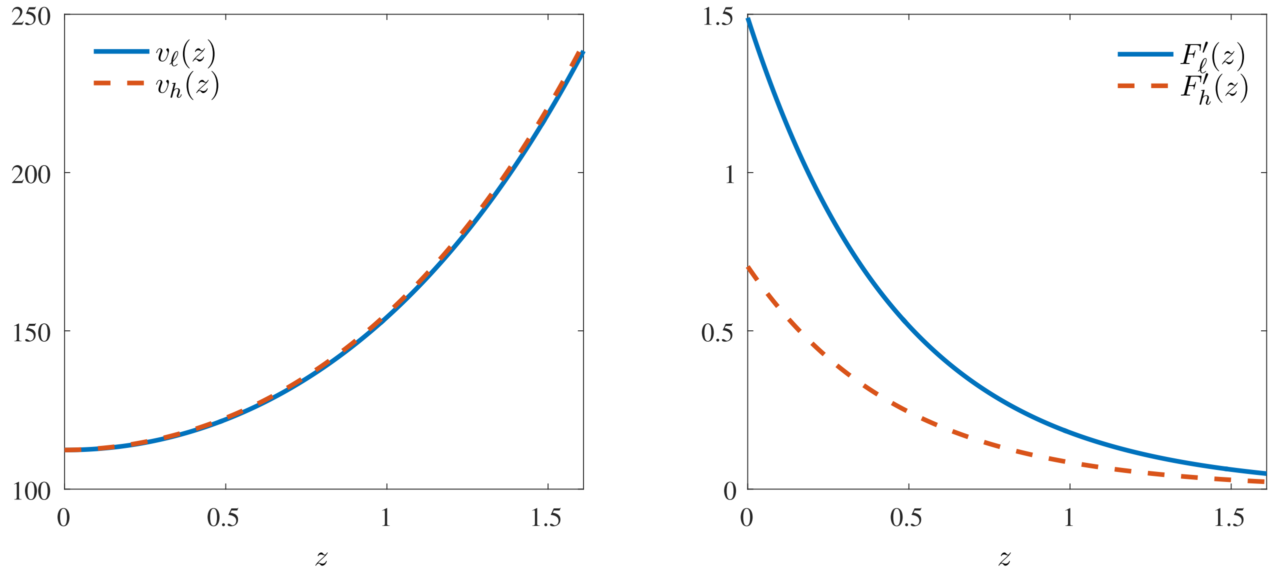

Figure 5: Exogenous $v_i(z)$, and $F'_i(z)$ with a Bounded Frontier

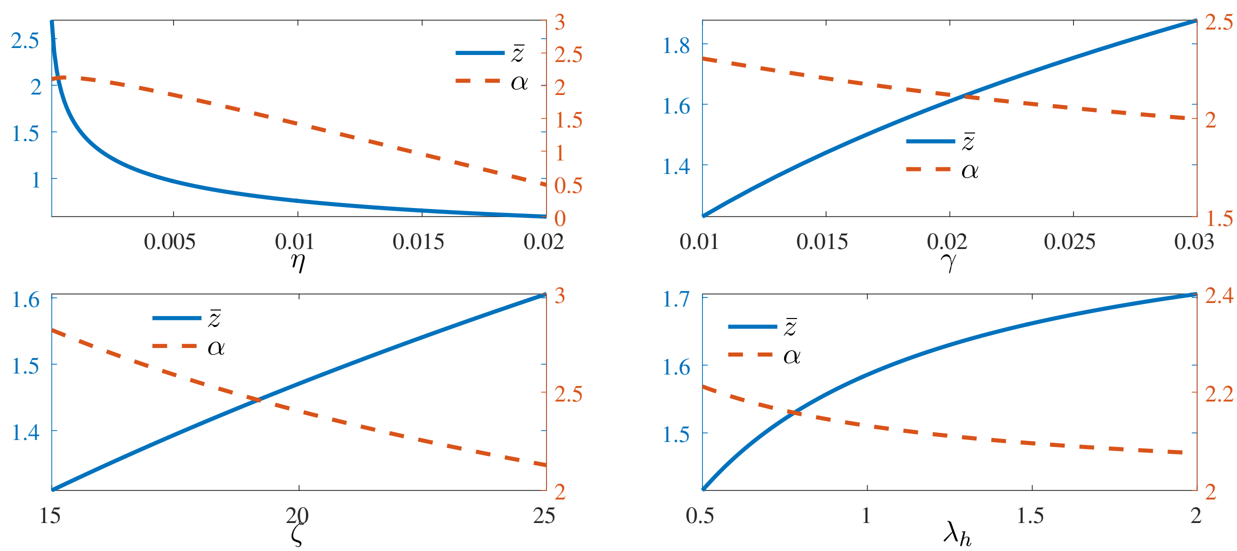

Figure 6: Comparative Statics with a Bounded Frontier

finite-support, the aggregate growth rate must be weakly less than the growth rate of innovators, while in the infinite-support case the aggregate growth rate can exceed the growth rate of innovators. Second, in the finite unbounded case, there is still the possibility of growth rates strictly less than the innovation rate. Finally, in the finite bounded support case, there is no such hysteresis; there is a unique equilibrium stationary distribution and a growth rate $g = \gamma$. Since the finite bounded cases' unique BGP occurs for an arbitrarily small arrival rate of jumps (i.e., $\eta \to 0$), we focus attention on the equilibrium with $g = \gamma$ even in the finite unbounded case.

These results on infinite-support distributions arise when the initial distribution has infinite-support but can be mapped to models that generate infinite-support distributions from arbitrary finite-support initial conditions—such as the GBM models discussed previously. In those cases, the result is stronger: any sort of stochastic shocks with unbounded growth rates generate a continuum of stationary distributions and growth rates, which can have a growth rate greater than the innovation rate.

We live in a finite-support world: there is no firm operating at any point in time with the best technology that could ever be invented and there were no car manufacturers in the Stone Age. While it may be appropriate for certain questions, because the economic properties of infinite- vs. finite-support equilibria are so different, using an infinite-support assumption as an approximation to finite-support requires caution.

Even for finite-support distributions, when normalized, the relative frontier of best to worst firms could be bounded or could be diverging to infinity. We showed that in order to maintain a bounded relative frontier, there needs to be some way for laggards to keep up with the frontier innovators by, for example, leapfrogging. Beyond eliminating the $g < \gamma$ hysteresis, the key difference between the bounded and unbounded cases appears in the option value of adoption for firms at the frontier. In the bounded equilibrium, even the best firm realizes that it may become an adopter in some reasonable time frame and places some positive value on the option to adopt. Since it is the innovative firms with the highest productivity that push out the frontier, this creates a link between their innovation incentives and the value of adoption, through which a change in the cost of the benefits of adoption will affect aggregate growth. In Section 4, we study this interaction between adoption and innovation in an endogenous growth version of the model.

4 Stationary Equilibria with Endogenous Stochastic Innovation

This section introduces endogenous innovation into the stochastic model with finite-support. We assume that firms can control the drift of their innovation process, as in Atkeson and Burstein (2010) and Stokey (2014). In Section 4.1, in order to analyze the unbounded case with $\bar{z} \to \infty$ we first assume that the arrival rate of jumps to the frontier, η, is zero. This allows us to focus on the link between innovation and aggregate growth before adding the complication of the aggregate growth rate being affected by the option value of adoption for innovators at the frontier. Then, in Section 4.2, we move to the model with jumps to the frontier and bounded support and analyze how the innovation choice of frontier firms is influenced by the option value of adoption, which affects the aggregate growth rate. Thus far, the main link between innovation and adoption is through the option value of adoption. In Section 5, we also model technology as excludable, and allow adopting firms to bargain with the owner of the technology that they are trying to adopt. We show how payments from adopters affect innovation investments and aggregate growth.

For simplicity, we model the innovation choice with no direct spillovers from other firms, in order to focus on the new interactions that we are modeling. With this simplification, the interactions here arise via tradeoffs in the firms' choices, rather than via a coupled innovation and adoption technology.[35]

To model 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 R&D technology and the cost be quadratic in the growth rate γ. Thus the h-firm Bellman becomes

$$\rho v_h(z) = \max_{\gamma \geq 0} \left\{ e^z - \underbrace{(g-\gamma)}_{\text{Drift}} v_h'(z) - \underbrace{\tfrac{1}{\chi} e^z \gamma^2}_{\text{R\&D cost}} + \lambda_h(v_\ell(z) - v_h(z)) + \eta(v_\ell(\bar{z}) - v_h(z)) \right\} \tag{77}$$

The key additional necessary equilibrium condition in the endogenous growth model comes from the first-order condition associated with optimal innovation rates:

$$\gamma(z) = \tfrac{\chi}{2} e^{-z} v_h'(z). \tag{78}$$

Thus, the value function of the h-firm in (77) becomes the following non-linear ODE:

$$\rho v_h(z) = e^z - g v_h'(z) + \tfrac{\chi}{4} e^{-z} v_h'(z)^2 + \lambda_h(v_\ell(z) - v_h(z)). \tag{79}$$

Instead of all h-firms growing at rate γ exogenously, h-firms are now choosing a growth rate γ that is a function of their current productivity level, z.[36]

Adapting the equations in Section 3.3, the equations that characterize an equilibrium are (23) to (26), (30), (31) and (69) with the following Bellman and KFE equations:

$$\rho v_\ell(z) = e^z - g v_\ell'(z) + \lambda_\ell(v_h(z) - v_\ell(z)) + \eta(v_\ell(\bar{z}) - v_\ell(z)) \tag{80}$$

$$\rho v_h(z) = e^z - g v_h'(z) + \tfrac{\chi}{4} e^{-z} v_h'(z)^2 + \lambda_h(v_\ell(z) - v_h(z)) \tag{81}$$

$$0 = g F_\ell'(z) + +\lambda_h F_h(z) - \lambda_\ell F_\ell(z) - \eta F_\ell(z) + \eta \mathbb{H}(z - \bar{z}) + S F(z) - S_\ell \tag{82}$$

$$0 = \underbrace{(g - \gamma(z))}_{\text{Drift}} F_h'(z) + \lambda_\ell F_\ell(z) - \lambda_h F_h(z) - \eta F_h(z) - S_h. \tag{83}$$

Numerical Methods. For all endogenous growth cases, the need to jointly solve for the nonlinear Hamilton-Jacobi-Bellman equations and the Kolmogorov forward equations necessitates non-trivial numerical solution 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 a generally applicable numerical technique based on spectral collocation and quadrature, as detailed in Technical Appendix B.[37]

4.1 Endogenous Innovation with a Finite Unbounded Frontier

In this section, we study the finite unbounded BGP, which is the case corresponding to $\eta = 0$.

[35] This is in contrast to Schumpeterian models, where innovations are built upon on diffusion from the frontier, and approaches such as that of Chor and Lai (2013), who are interested in the direct interaction with a dependent innovation process, with aggregate spillovers of knowledge.

[36] As the choice of γ is increasing in z in equilibrium, agents in the h state will end up crossing the endogenous adoption threshold, as shown in Appendix A.2, and, thus, the smooth-pasting condition for h types is now necessary. To generalize the approach, different innovation types are best understood as having different productivity in innovation, χ, with the bellman equation for each type as (79).

[37] The 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, you can naively stack every equation in the model, including the Bellman Equations, KFEs, equilibrium conditions, etc. into a single nonlinear system of equations, and solve without any nested fixed points. In practice, this requires using a high-performance solver and auto-differentiation, but is easy to implement and reasonably fast.

Proposition 5 (Endogenous Innovation and Finite Unbounded Support BGP)**.** *For $\eta \to 0$, a unique equilibrium exists with a growth rate that is the solution to the cubic equation*[38]

$$g\left(g^2 + g(2\lambda_h + \lambda_\ell + 3\rho) + 2\rho(\lambda_h + \lambda_\ell + \rho)\right) = \chi(g + \lambda_h + \lambda_\ell + \rho). \tag{84}$$

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

Proof. See Appendix B. The numerical method to compute the equilibrium $\gamma(z)$ and $F_i(z)$ is described in Technical Appendix B. □

In addition to the existence and uniqueness of the equilibrium, a main result is that the innovation rate is increasing in productivity—i.e., $\gamma'(z) > 0$. There is 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, have a smaller option value of adoption. Since the option value of adoption is a larger component of total value for lower-productivity firms, the low 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 they can save the cost of innovating and, instead, adopt a technology discretely better in expectation than the one it is currently operating? Of course, the cost of adopting and innovating will jointly determine this adoption threshold.[39]

The innovation rate and statistics of the productivity distribution for a computed equilibrium are presented in Figure 7. In addition to the calibration discussed in Section 3.3, the innovation cost parameter $\chi = 0.001074$ is calibrated to match $g = 0.02$ and $\zeta = 27.2$ to match the tail parameter of $\alpha = 2.12$.

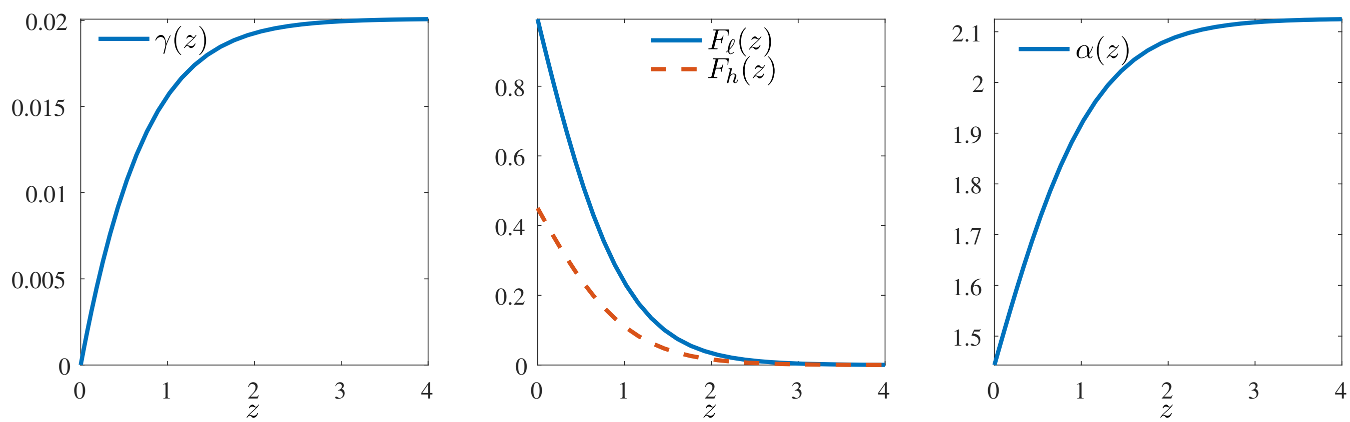

Figure 7: Endogenous $\gamma(z)$, $F_i(z)$, and $\alpha(z)$ with a Finite Unbounded Frontier

In order to get a sense of the shape of the unconditional distribution, we define the "local" tail index

$$\alpha(z) \equiv \frac{F'(z)}{1 - F(z)}. \tag{85}$$

[38]The limit is necessary here for uniqueness, because otherwise a $g < \gamma$ growth rate along the lines of Section 3.2.1 is possible and (84) only determines $\gamma(\infty)$. For the same reasons as discussed in Section 3.2.1 we do not focus on the negative latent growth cases here.

[39]Research that separates internal innovation on existing products from Schumpeterian innovation replacing products, such as Akcigit and Kerr (2016), find that internal innovation scales moderately faster with size. Since all innovation in our model is "internal innovation" of this sort, this is consistent with our result that innovation rates are increasing in productivity.

In the standard log-log plot used for estimating power-laws, as in Gabaix (2009), this $\alpha(z)$ would be the slope of the non-linear equation at z. Note that with this definition, the "local" tail index of a Pareto distribution is constant and equal to its true tail index. Furthermore, for any distribution with infinite-support, the tail index is $\alpha \equiv \lim_{z\to\infty} \alpha(z)$. Figure 7 plots the local tail coefficient, converging to the calibrated value of 2.12. As the tail index is increasing, this shows that there is more productivity dispersion for firms with lower relative productivity.

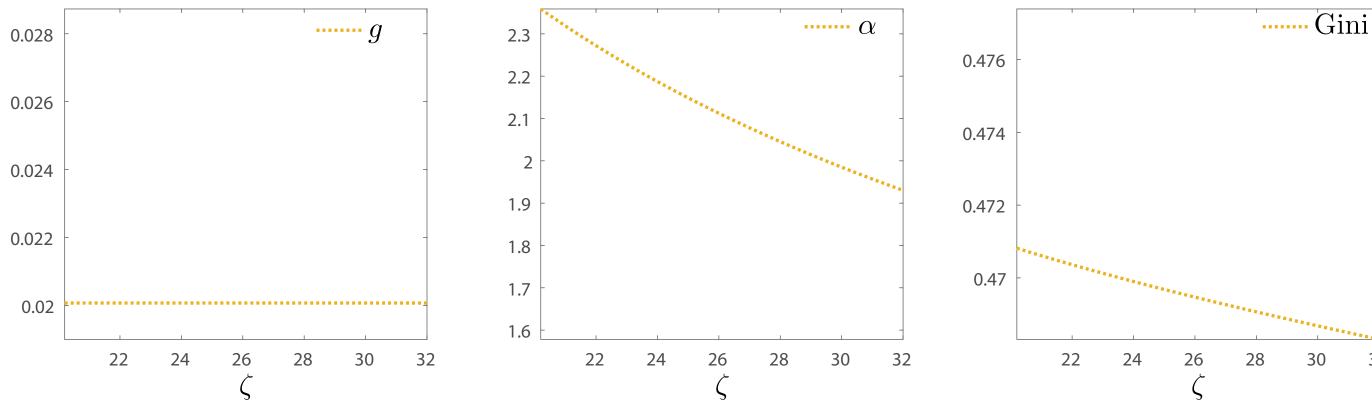

Figure 8: Comparative Statics on Adoption Cost

The endogenous aggregate growth rate is the growth rate chosen by innovators at the frontier. As alluded to in the exogenous growth sections, now that aggregate growth is endogenous, we can see that, in the unbounded case, the aggregate growth rate is independent of the cost of adoption. 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. Figure 8 plots comparative statics for a change in the adoption cost. As the theory shows, the aggregate growth rate is invariant. However, changes in the cost of adoption affect the shape of the distribution. Higher adoption costs induce a significantly thicker tail, but almost no change in the Gini coefficient.

4.2 Endogenous Innovation with a Finite Bounded Frontier

In this section, we study the finite bounded BGP, which is the case corresponding to $\eta > 0$. For the same reasons as discussed in the exogenous innovation case, without leapfrogging, $\lim_{\eta\to 0} \bar{z}(\eta) = \infty$.

Proposition 6 (Stationary Equilibrium with Continuous Endogenous Innovation and Bounded Support)**.** *The endogenous innovation choice is such that $\gamma(0) = 0$ and $\gamma(\bar{z}) \equiv g$. A continuum of equilibria exist, parameterized by a $\bar{z}$.*

Proof. See Appendix B. A numerical method to solve for the continuum of equilibria is described in Technical Appendix B. □

In the limit approaching the unbounded case, as $\eta \to 0$, the upper bound on g equals the unbounded growth rate in Proposition 5. However, in the endogenous growth bounded distribution 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. In the unbounded case, the aggregate growth rate—i.e., $\gamma(z)$ as $z \to \infty$—is unique because the option value disappears, as opposed to when $\bar{z}$ is bounded. Different distributions and associated $\bar{z}$ induce different growth 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 multiplicity here is due to complementarity of firm decisions and is fundamentally different than the $g > \gamma$ latent growth multiplicity driven by initial conditions as discussed previously (see Appendix C.1).

Since the option value of adoption is positive for innovators at the frontier, the aggregate growth rate is affected by adoption costs and benefits.[40]

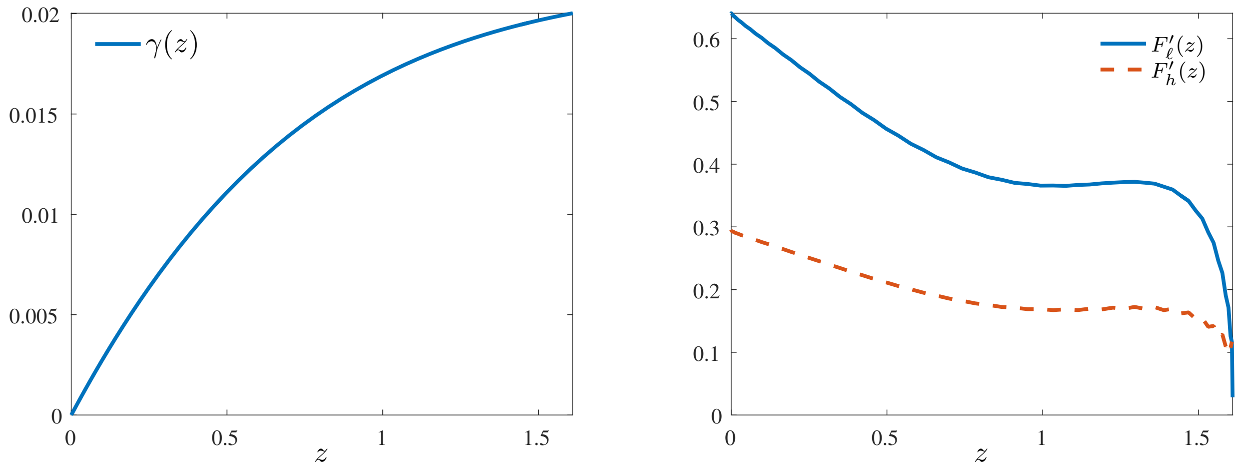

Figure 9: Endogenous $\gamma(z)$ and $F'_i(z)$ with an Bounded Frontier

A calibrated example of the optimal innovation policy and productivity distributions for a bounded endogenous growth BGP is shown in Figure 9. Figure 10 documents the intuition that lower $\bar{z}$ are associated with lower aggregate growth rates, due to the self-fulfilling balancing of innovation incentives. The larger the option value of adoption—which is increasing in the distance of the relative frontier—the lower are the incentives to push out the frontier by innovation.

Figure 11 plots the maximum growth rate of the set of admissible g as a function of η, using equation (84) with the same parameters used in Figure 7 (Endogenous Unbounded BGP). As $\eta \to 0$, the number of jumps to the frontier approaches 0, and the bounded model studied in Section 4.2 asymptotically becomes the unbounded model studied in Section 4.1. The intuition for a decreasing $\max(g(\eta))$ is that with more jumps to the frontier, the distribution becomes more compressed. As the growth rate of the frontier is determined by the innovation decision at $\bar{z}$, which takes into account the option value of diffusion, the more compressed the distribution, the lower the innovation rate, for the same reasons that $g(\bar{z})$ is increasing in Figure 10.

[40]Due to the multiplicity of equilibria, comparative statics of stationary distributions are not well-defined. Numerically, we find that—fixing all parameters—the range of feasible g in these stationary equilibria tends to be fairly small. For example, Figure 10 contains all regions of g where the numerical algorithm converged.

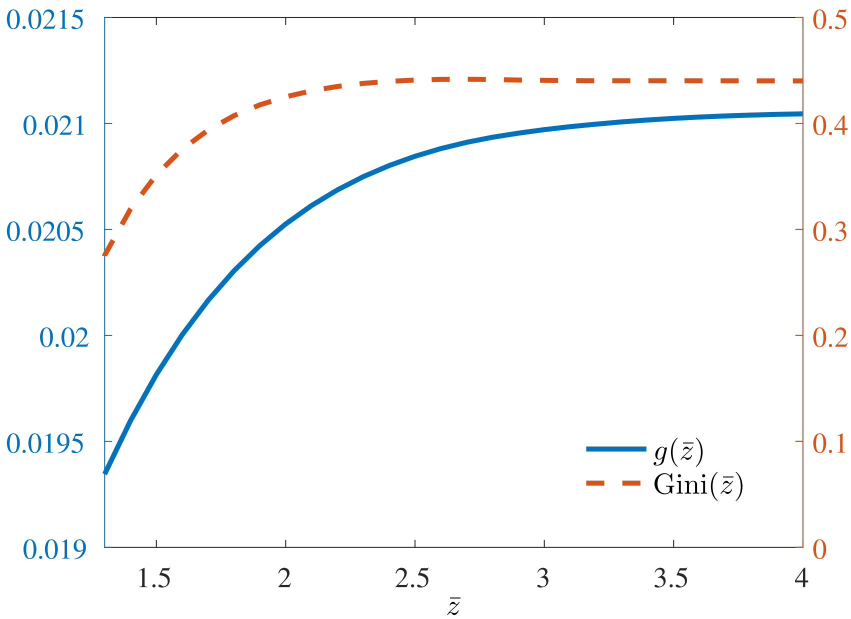

Figure 10: Equilibrium g as a function of $\bar{z}$

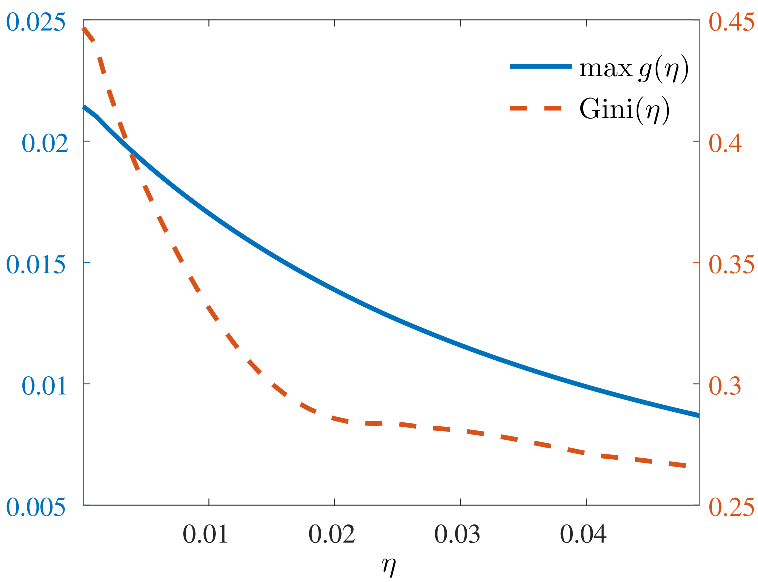

Figure 11: Maximum Equilibrium $g(\eta)$ and associated Gini coefficient

5 Extensions: Licensing and Directed Adoption

In this section, we combine all of the elements from earlier sections, and add additional features to endogenize all of the innovation and technology diffusion choices. These additions show that the model is extensible and further highlight how adoption and innovation interact. The extension to licensing brings the concept of variable degrees of excludability to models of technology diffusion. The extension to directed adoption captures that firms may be able to better target their adoption activities on technologies than if they merely obtain unconditional draws of the existing technologies in use. This also gives the model flexibility to produce firm dynamics for adopters that may be more consistent with the panel data.

Complete Nested Model (with Extensions). Since a firm may receive profits from licensing its technology, its profits no longer necessarily equal its productivity. We define $\pi(z)$ to be the total profits of the firm, with $\pi(z) = e^z$ in the baseline model with no excludability. $\pi(z)$ in the model with licensing will be derived in Section 5.1. To endogenize leapfrogging, we allow firms that are adopting to jump to the frontier with probability $\theta \in [0, 1)$. Just as firms can choose their innovation rate γ subject to a convex cost, firms can also choose the probability of a jump to the frontier with a convex cost. The cost of choosing jump probability θ is $\frac{1}{\vartheta}\theta^2$. Thus, when a firm upgrades its technology through adoption, it has some chance of adopting a state-of-the-art invention and jumping to the frontier, and it can invest on the intensive margin to try to obtain such a large upgrade. Finally, we allow firms to choose the degree of distortion in their draws (i.e., directed adoption) by choosing $\kappa > 0$ at a cost of $\frac{1}{\varsigma}\kappa^2$. In this section, we adapt Sections 3.3 and 4.1 to include the endogenous innovation choice of γ, partial-excludability, a choice of θ, and a choice of κ. The system of equations that characterize the equilibrium are (23) to (28), profits as defined in equation (99), and

$$\rho v_\ell(z) = \pi(z) - g v_\ell'(z) + \lambda_\ell(v_h(z) - v_\ell(z)) + \eta(v_\ell(\bar{z}) - v_\ell(z)) \tag{86}$$

$$\rho v_h(z) = \max_{\gamma \geq 0} \left\{ \pi(z) - \frac{1}{\chi} e^z \gamma^2 - (g - \gamma) v_h'(z) + \lambda_h(v_\ell(z) - v_h(z)) + \eta(v_\ell(\bar{z}) - v_h(z)) \right\} \tag{87}$$

$$v(0) = \frac{1 + \eta v_\ell(\bar{z})}{\rho + \eta} = \max_{\theta \geq 0,\, \kappa > 0} \left\{ (1 - \theta) \int_0^{\bar{z}} v_\ell(z) \mathrm{d}F(z)^\kappa + \theta v_\ell(\bar{z}) - \frac{1}{\psi}\left(\zeta + \frac{1}{\vartheta}\theta^2 + \frac{1}{\varsigma}\kappa^2\right) \right\} \tag{88}$$

$$\begin{aligned} 0 &= g F_\ell'(z) + \lambda_h F_h(z) - \lambda_\ell F_\ell(z) - \eta F_\ell(z) \\ &\quad + (\eta + \theta(S_\ell + S_h))\mathbb{H}(z - \bar{z}) + (1 - \theta)(S_\ell + S_h) F(z)^\kappa - S_\ell \end{aligned} \tag{89}$$

$$0 = (g - \gamma(z)) F_h'(z) + \lambda_\ell F_\ell(z) - \lambda_h F_h(z) - \eta F_h(z) - S_h \tag{90}$$

See Appendix B for a nested derivation of all extensions. Note that when a particular firm chooses θ or κ, it does not influence the θ or κ chosen by the other firms. Firms will take into account the effects of the aggregate "directed adoption" choice on $F_i(z)$, and as all adopting firms are a priori identical, each firm will choose the same θ, which will induce an $F_i(z)$ that is consistent with firms' beliefs about $F_i(z)$ in equilibrium.

In addition to all other equilibrium requirements, the chosen intensity of jumps to the frontier for adopting agents comes from the first-order condition of (88),

$$\theta = \frac{\psi\vartheta}{2} \int_0^{\bar{z}} v_\ell'(z) F(z)^\kappa \mathrm{d}z, \tag{91}$$

and the directed technology diffusion choice from the first-order condition of (88) solves the implicit equation

$$\kappa = \frac{-\varsigma\psi(1 - \theta)}{2} \int_0^{\bar{z}} v_\ell'(z) \log(F(z)) F(z)^\kappa \mathrm{d}z. \tag{92}$$

5.1 Licensing and Partial Excludability

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 (no 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 fee is modeled as a portion of the present discounted value of adopting the technology and, instead of being paid as a sequence of residual payments, the fee is paid up front in one lump sum. Firms bargain to determine the size of the licensing fee, with the fee reflecting the bargaining power of adopters and license issuers. There is no cost to the actual transfer of technology.

Bargaining. The timing is that the adopting firm first pays the adoption cost and then, upon the realization of the match, negotiations over licensing commence. Negotiations take the form of Nash bargaining, with a bargaining power parameter of $\psi \in (0, 1]$. $\psi = 1$ represents no excludability, in which the technology is adopted for free, without profit to the higher-productivity firm. The outside option of the adopting firm (i.e., the licensee) is to reject the offer and continue on with its existing technology—i.e., $v(0)$. The outside option of the licensor is simply to reject the offer and gain nothing. 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. The Nash bargaining problem is

$$\arg\max_{\hat{v}} \left\{ (\hat{v} - v(0))^{\psi} (v_\ell(z) - \hat{v})^{1-\psi} \right\}. \tag{93}$$

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

$$\hat{v}(z) = (1 - \psi)v(0) + \psi v_\ell(z), \tag{94}$$

while the licensor gains

$$v_\ell(z) - \hat{v}(z) = (1 - \psi)(v_\ell(z) - v(0)). \tag{95}$$

As is apparent in equation (94), if $\psi = 1$, then the licensee gains the entire value with $\hat{v}(z) = v_\ell(z)$. Given that firms are risk-neutral, the value-matching condition integrates over the adopter's surplus from the matches across all possible z.

Licensing and Flow Profits If both $\psi < 1$ and $\kappa \neq 1$, then the matching probabilities are distorted, and profits become a direct function of $F(z)$. To simplify the analysis and numerical algorithm, we will restrict our analysis to undirected adoption when there is licensing—i.e., $\kappa = 1$ if $0 < \psi < 1$.[41] In the simple case of $\kappa = 1$ with no distortions in the adoption distribution, there is an equal probability of adopting from any licensor. Thus, the flow of adopters engaging any licensing firm is just the flow of adopters $S = gF'(0)$.

[41] This assumption shuts down the interaction between κ choices and licensing, but is otherwise innocuous. The technical issue is that the marginal profits in (100) becomes a direct function of $F(z)$, which means that the HJBE and the KFE need to be solved concurrently. To find the solution for any κ with an interior ψ, use a conservation of the total surplus from the value-matching condition with draws of $F(z)^\kappa$ distribution, the total surplus flows to the operating firms with distribution $F(z)$, and the flow (95). Then, for some $q(z)$ representing the distortion of the surplus flow to z agents, the flow of adopters licensing the firm's technology is $Sq(z)$, and the conservation of total licensing flows is

$$S \int (v_\ell(z) - v(0))\, \mathrm{d}F(z)^\kappa = S \int (v_\ell(z) - v(0))\, q(z) \mathrm{d}F(z). \tag{96}$$

Since $\gamma(0) = 0$ in equilibrium (as shown in Section 4.1), then,

$$\pi(z) = e^z + \underbrace{gF'(0)}_{\text{Amount of Licensees}} \underbrace{(1-\psi)\left(v_\ell(z) - v(0)\right)}_{\text{Profits per Licensee}}. \tag{99}$$

Differentiate and multiply by e^{-z}, we find a function for marginal profits relative to profits from production,

$$e^{-z}\pi'(z) = 1 + (1-\psi)gF'(0)e^{-z}v_\ell'(z). \tag{100}$$

When $\psi < 1$, profits are a function of g since (1) faster growth means more adopters—given a fixed $F(z)$; and (2) faster growth increases the licensor Nash bargaining threat of $v(0)$ by making it relatively more valuable to wait for the economy to grow before accepting a new technology. Given the equilibrium $F(z)$ and $v_\ell(z)$—which are themselves determined by g—these tensions help determine the relative profits from licensing. Also, as $z \to 0$, the profits from licensing disappear as the surplus of accepting a draw close to the adoption boundary goes to 0. Consequently, $\pi(0) = 1$. Since (100) is increasing in z, the incentives generated by licensing are such that firms with the highest productivity have the strongest incentive to innovate.

Value-Matching Condition. Since the surplus 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 changes. Combining (47), (94) and (A.29), the value-matching condition is

$$v(0) = \frac{1}{\rho} = \int_0^\infty \left[\psi v_\ell(z) + (1-\psi)v(0)\right] \mathrm{d}F(z) - \zeta. \tag{101}$$

Rearranging, the value-matching condition is identical to that previously derived in (47), but with a proportional increase in the adoption cost,

$$v(0) = \frac{1}{\rho} = \int_0^\infty v_\ell(z)\mathrm{d}F(z) - \frac{\zeta}{\psi}. \tag{102}$$

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 being a licensor in the future. The two environments are quite different for the innovator, as license fees provide an extra incentive to innovate.

With (59), we can reverse engineer the distortion as $q(z) \equiv \kappa F(z)^{\kappa-1}$. From (95), we find the flow profits, including the direct value of production and the flow licensing given z as

$$\pi(z) \equiv \underbrace{e^z}_{\text{Production}} + \underbrace{\overbrace{\kappa F(z)^{\kappa-1} S}^{\text{Licensee Flow}}\overbrace{(1-\psi)(v_\ell(z) - v(0))}^{\text{Profits per Licensee}}}_{\text{Total Licensing Profits}} \tag{97}$$

The more general form of (100), using definitions (B.1) and (B.13) is, then,

$$e^{-z}\pi'(z) \equiv 1 + (1-\psi)\kappa g F'(0)F(z)^{\kappa-1}\left(w_\ell(z) + (\kappa-1)F(z)^{-1}F'(z)e^{-z}\int_0^z w_\ell(\tilde{z})\mathrm{d}\tilde{z}\right) \tag{98}$$

Note that after substitution, the HJBE in $w_\ell(z)$ would then be a system of integro-differential equations, significantly increasing the complexity of the problem. As we do not feel the economics of the κ and ψ interaction are sufficiently interesting, we leave these complexities out of the main body, but handle all cases in the numerical solution described in Technical Appendix B.

Role of Excludability. Figure 12 uses the same parameterization as in Figure 7 (Unbounded Endogenous Growth BGP with $\psi = 1$—zero excludability), but varies the degree of excludability parameter ψ. To set a baseline value, according to Kemmerer and Lu (2012), the median fraction of profits that a licensor must pay is 5%, corresponding roughly to $\psi = 0.95$ if most licenses are for high-productivity technologies.

When excludability is not too strong, the aggregate growth rate is increasing in the degree of excludability (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 R&D for all firms, including those near the frontier that determine the aggregate growth rate. This is analogous to the forces at play in the optimal patent length literature.

To give a sense of the distribution shape, the local tail index $\alpha(z)$ is plotted at large $z = 2.4$ (i.e., for firms with approximately 11:1 productivity relative to the minimum).[42] Increasing excludability increases innovation activity, generating a more unequal distribution with a fatter tail parameter. This shows a trade-off between productivity inequality and aggregate growth rates. This independent positive association between productivity dispersion and the aggregate growth rate operates through innovation activity, compared to the typical link (i.e., Perla and Tonetti (2014)) driven by adoption incentives.

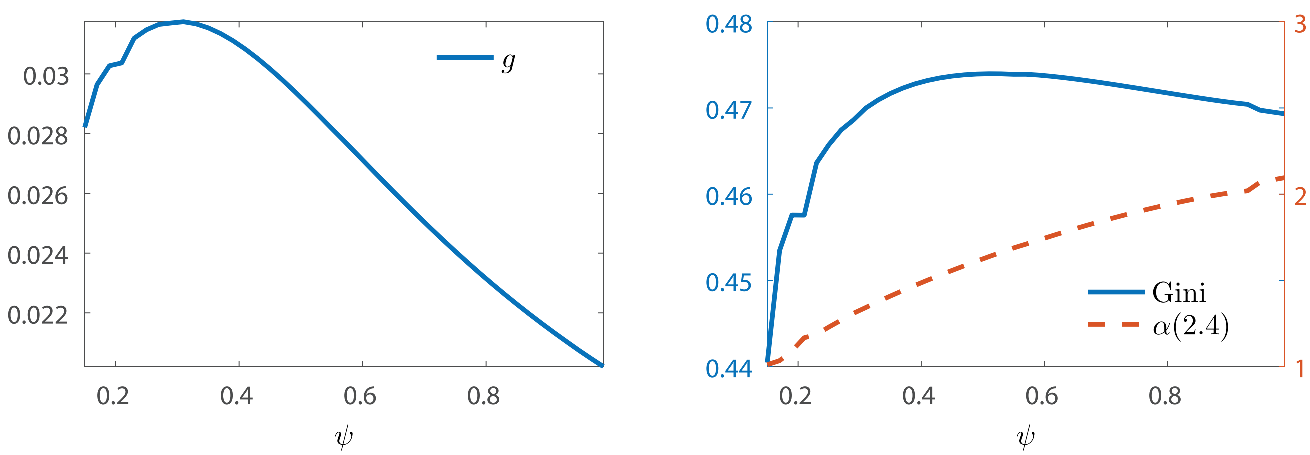

Figure 12: Growth and Distribution Shape under Excludability

Figure 13 presents another perspective on the role of excludability. For several levels of ψ, the growth rate is plotted for various adoption costs, ζ. In the absence of excludability (i.e., $\psi = 1$), adoption costs change the shape of the distribution, but have no impact on the aggregate growth rate—since the option value of adoption is infinitesimal for the highest-productivity agents making the innovation decision. With a strong degree of excludability, however, lower adoption costs drive higher aggregate growth, even in the unbounded case. While the option value of adoption for firms at the frontier is still infinitesimal, an innovating firm gains extra profits from adopting firms, and the number of adopting firms increases when adoption costs are lower.

[42] A level smaller than the asymptotic $z \to \infty$ is chosen due to numerical instabilities of calculating $\alpha(z) \equiv F'(z)/(1 - F(z))$ where $F(z) \approx 1$. The asymptotic tail index α will be only slightly larger than the local tail parameter for this large of a z.

The shape of the distribution is more subtle since the tail parameter is decreasing substantially as adoption costs increase, which means that tail inequality is substantially increasing. If the distribution was Pareto, then the Gini coefficient would be decreasing in the tail parameter (Pareto Gini= $\frac{1}{2\alpha-1}$). However, here, the Gini coefficient decreases modestly, but is nearly flat. The reason is that the shape of the distribution near the adoption threshold is impacted by the large mass of agents there, which is generally determined by the innovation decisions of those firms rather than by the tail innovation rates. Consequently, (1) more excludability leads to a large increase in tail inequality and a modest increase in broader measures of inequality; (2) only looking at changes in tail inequality for comparative statics is deceptive in the case of endogenous innovation; and (3) increasing excludability can lead to substantial increases in the aggregate growth rate, but decreasing adoption costs leads only to a modest increase in aggregate growth when starting around reasonable excludability levels and adoption costs.

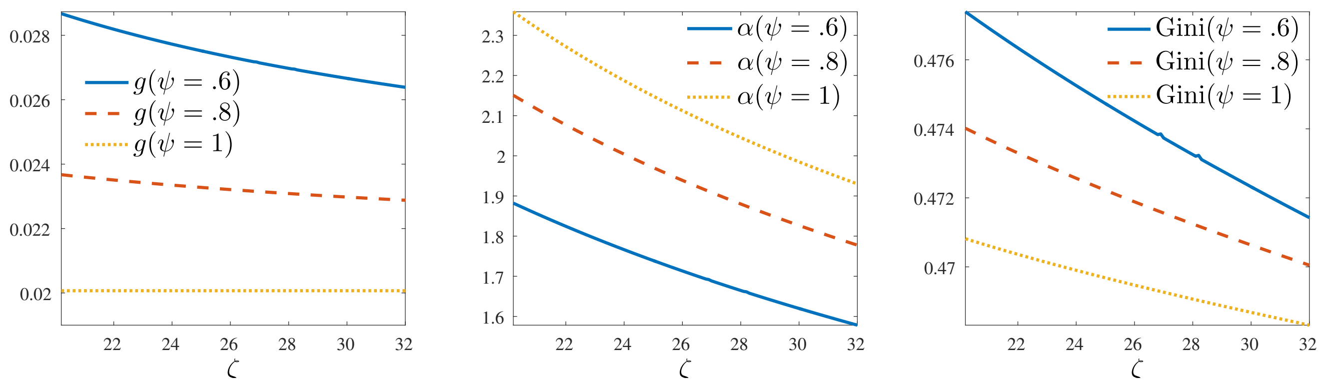

Figure 13: Interaction of Excludability and Adoption Costs

5.2 Directed Adoption: Endogenous κ

In the unbounded BGP, if there is no excludability ($\psi = 1$), then κ has no effect on the aggregate growth rate. Since the option value is essentially zero for innovating firms at the frontier, a change in the probability of adopting better technologies does not impact their innovation incentives, and, thus, does not affect the aggregate growth rate. Changes in the probabilities of adopting good technologies does, however, change the shape of the distribution. This occurs through both altering the incentives to adopt a new technology, as well as distorting the productivities that firms draw. Figure 14 demonstrates how directed adoption affects the shape of the distribution, using the same calibration as in Figure 7, but allowing for an endogenous choice of κ according to varying efficiency in firms' ability to direct their adoption activity, parameterized by ς. The growth rate always equals $g = 0.02$, but the tail parameters and Gini coefficient of the distribution change with ς. The approximate tail parameter and the Gini move in the same direction as the efficiency of directing adoption changes. More efficiency in directed adoption leads to thinner tails and less productivity inequality because changes in κ directly affect the tail parameter by distorting the adoption distribution.

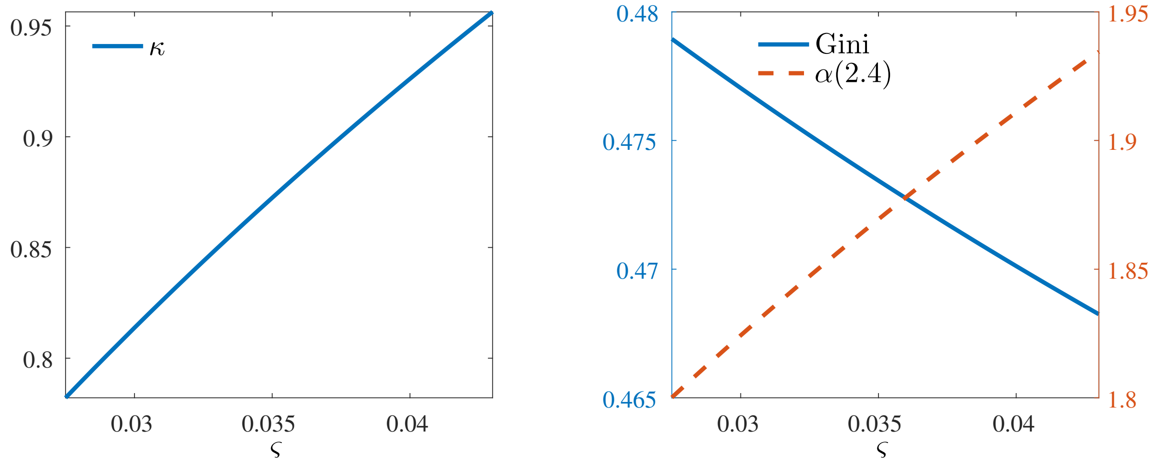

Figure 14: Distribution Shape with Endogenous κ

6 Empirics of the Technology Frontier

The technology frontier is the highest possible productivity and/or quality available to produce a given product. As Mokyr describes "at any moment, there is a large gap between average and best practice technology," and there are many operating firms producing closely related products who have not (or can not) implement the leading-edge technology, resulting in productivity or quality dispersion.

In this section, we provide some preliminary empirical evidence on the size and time-series behavior of the technology frontier. In this paper, we make the distinction between a relative technology frontier that is finite but bounded (in relative terms) and one that is finite but unbounded (in relative terms). Our specification's simple association between a continuous forward drift of the relative technology frontier and an unbounded frontier only occurs with a continuum of firms. Instead, with a finite number of firms the equivalent check on the technology frontier is whether it is a sub-martingale, weakly growing in expectation. Using measures of the frontier for each industry and year, we check the time series properties of the frontier for each industry. The questions guiding this exploration are whether the relative frontier have a positive drift and is it stationary?

Whether $\bar{z}(t)$ diverges or is stationary, and whether it is large or small, implies very different equilibrium behavior for firms and the aggregate economy. While we have analyzed all of the cases in this paper, further work needs to be done on determining if $\bar{z}$ is unbounded or bounded for different industries and economies. Here we provide some exploratory analysis to guide us on the magnitudes of $\bar{z}$ and its time-series trends, and in Appendix D.1 we perform statistical tests for stationarity to get a sense for whether the frontier may be diverging in various industries.

Following the existing literature, with all associated caveats, we use revenue (and/or employment) as a proxy for productivity (and/or quality). Fixing a particular level of aggregation (such as the aggregate economy or at the industry level), we can plot moments of the distribution. In a growing economy, the model consistent measure of the frontier that is the relevant measure of dispersion is not the frontier technology itself, but rather the frontier technology normalized to another moment of the distribution (e.g., the mean). This is in the spirit of Acemoglu, Aghion, Lelarge, Van Reenen, and Zilibotti (2007), which summarizes distance to the frontier as the ratio of the productivity of a firm to the highest productivity within each industry code, or the ratio of the 90th to the 10th percentile. Mapping directly to the model, $\bar{z}$ is defined as the log of the 100th to the 0th percentile, so these $90--10$ proxies provide an approximation, albeit it downwards biased, to the relative frontier.

To give a rough sense of the dynamics of the relative technology frontier, we use data on 416 industries from Compustat with 233,754 observations of firms from 1980-2014, and 176,480 for 1990-2014.[43] In the case of the 1990-2014 sample, when we collapse the firms within each industry to obtain a distribution across industries, this gives us 12,818 observations by SIC and year.[44]

We first present evidence on the aggregate economy before grouping by industry. This has the advantage of protecting against some small sample issues, but has the major disadvantage that many of the size differences are likely compositional across industry and not due to the type of within-product heterogeneity in quality/productivity that motivates our theoretical analysis.

Figure 15 is informative of the broad time-series patterns of the relative frontier using the employment and revenue proxies for productivity for all industries pooled together. The 90th percentile of employment and/or revenue is a proxy for the frontier, and the 10th percentile is used to provide the proxy for the minimum productivity.

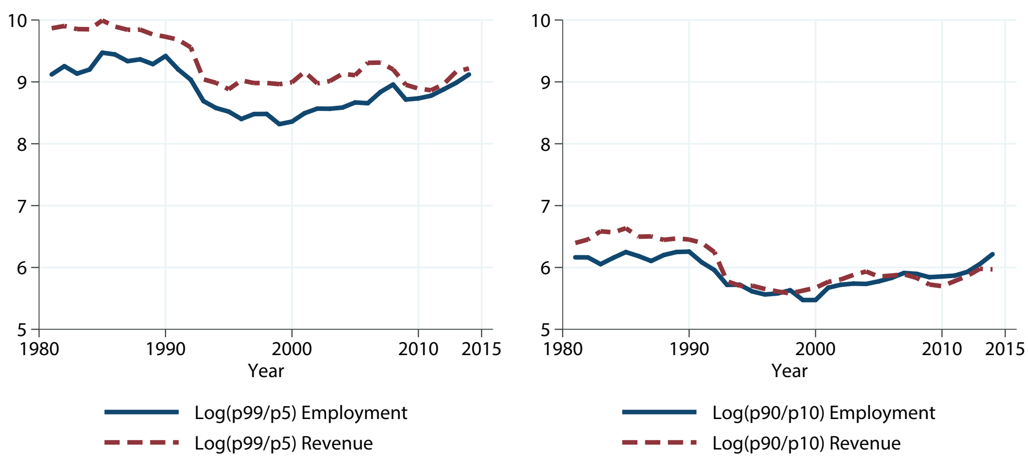

Figure 15: Time Series of the Frontier (Pooled Industries, 1980-2014)

The pooling of industries leads to a dispersed distribution and a distant frontier, with the log of the ratio of the 90th to 10th percentiles around 6. More important than the particular level is a sense of the broad time trends and stationarity. The relative frontier seems to be fairly stable, and does not appear to be diverging in this sample—i.e., the ratio of the most to the least productive seems to remain bounded, albeit large.

Pooling at the SIC level of aggregation, Table 1 and Figure 16 show statistics of the frontier for all of the SIC4 industries across all the years. Presented in Table 1 and Figure 16, the mean frontier is 4.39 when using revenue, and 4.18 when using employment. Not only is there wide dispersion in the productivity distribution represented by large values of the relative frontier, there is wide dispersion across industries in the size of the relative frontier.

While there is substantial dispersion across industries in the size of the frontier, the frontier for a given industry seems to be rather stable over time. Checking the evidence on the trend presented in Table 1, while differing sizably across industries, the median growth rate of the frontier is -0.36 percent using revenue and 0.00 percent using employment. Consequently, we find no strong evidence

[43]While Compustat is imperfect due to its selection of publicly traded firms which tend to be large, it is less of an issue for measuring the absolute frontier since we are primarily interested in the size of the most productive firms within an industry-year. The lower moment used to find the relative frontier is more sensitive to this selection, which motivates robustness checks with the 5th and 10th percentile, or even the median. See Technical Appendix D.1 for additional results and robustness checks. The qualitative results remain the same. Also, a longer panel is given in Technical Appendix Figure 2, but we advise caution in trusting proxies for early years in the sample due to small sample sizes.

[44]If we further require that there are at least 2 firms and 10 observations, then there is a total of 11,097 SIC-year observations. SIC codes for recent data uses NAICS to SIC concordance tables.

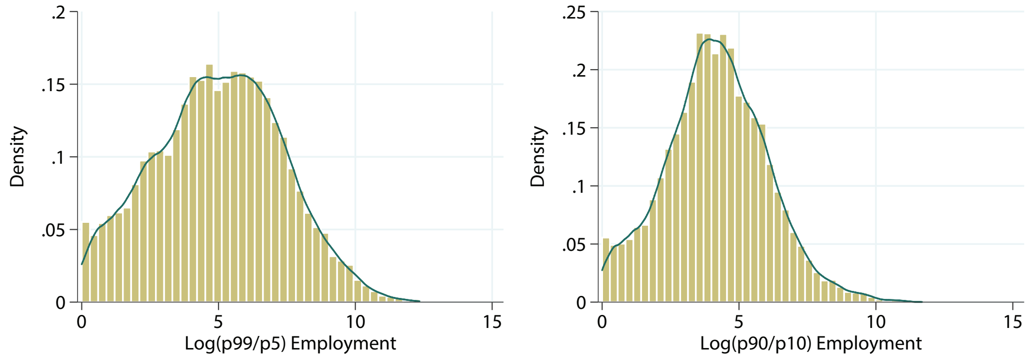

Figure 16: Histogram of Frontier Proxies (by SIC, 1980-2014)

	(1)				
	mean	p25	p50	p75	sd
Log(p90/p10) Revenue	4.39	3.08	4.27	5.55	2.03
Log(p90/p10) Employment	4.18	2.96	4.15	5.39	1.86
% Change Log(p90/p10) Revenue	-1.25	-27.59	-0.36	27.34	72.31
% Change Log(p90/p10) Employment	0.96	-24.42	0.00	25.32	68.59
Observations	12818				

Table 1: Frontier Summary Statistics (by SIC, 1980-2014)

in the panel for a positive drift during this period. However, this measure may be misleading due to high-frequency movements. Smoothing these out by computing the median frontier during 1981-1990, 1991-2000, and 2001-2010, we then analyze how the ratios change—i.e., (1991-2000)/(1980-1990) gives, in log points, the ratio of frontiers over long time periods within industry. Summary statistics for this variable using the $\bar{z} \approx \log(p90/p10)$ proxy is given in Table 2, with histograms given in Figure 17.[45]

The time-series pattern of these proxies suggest that the median frontier by industry remained largely constant when comparing 1981-1990 to 1991-2000 while it declined marginally comparing 1991-2000 to 2001-2010. his modest decline in the long-run is also evident when comparing 1981-1990 with 2001-2010. Hence, we find no evidence of the diverging frontier we would expect if the finite unbounded frontier was the correct specification. That said, since the median frontier in

[45] See Technical Appendix D.1 for robustness tests with the p99/p05 proxy for the frontier.

	(1)			
	mean	p50	p25	p75
Revenue Frontier Ratio (1991-2000)/(1981-1990)	1.06	0.98	0.83	1.18
Employment Frontier Ratio (1991-2000)/(1981-1990)	1.08	0.99	0.83	1.19
Revenue Frontier Ratio (2001-2010)/(1991-2000)	0.94	0.92	0.70	1.11
Employment Frontier Ratio (2001-2010)/(1991-2000)	0.97	0.95	0.72	1.15
Observations	12763			

Table 2: Frontier Ratio Summary Statistics (Using Log(p90/p10) by SIC)

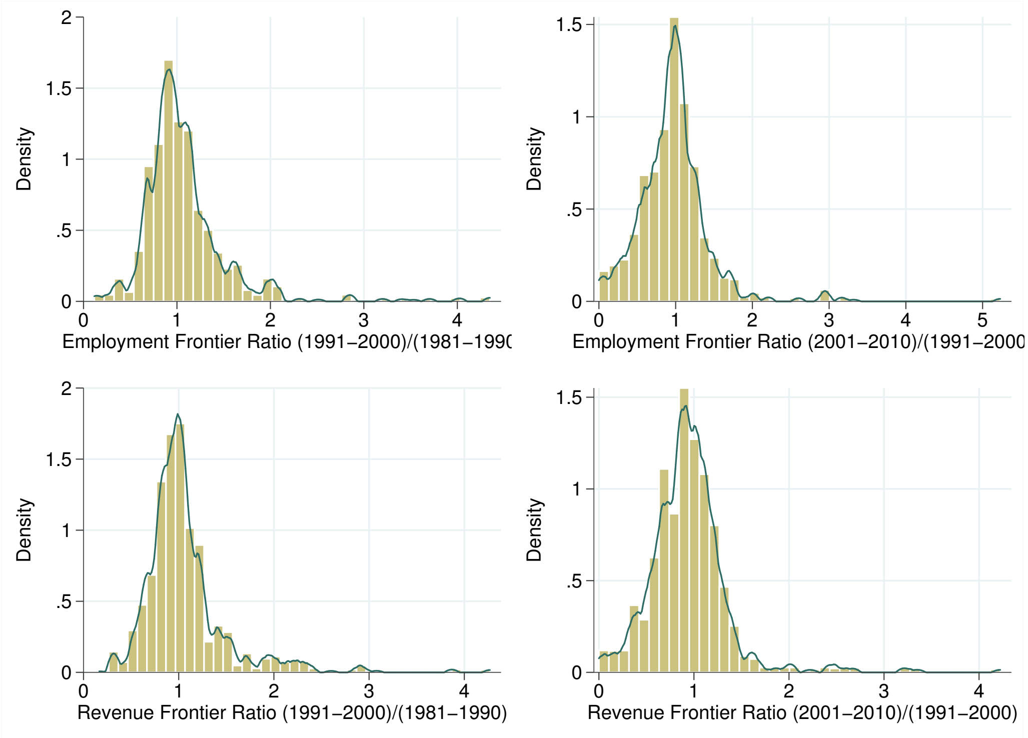

Figure 17: Histogram of Frontier Ratios (Using Log(p90/p10) by SIC)

Table 1 is around 4.2—which corresponds to $e^{4.2}$ in levels—the magnitude may be sufficient to use the unbounded approximation for many industries. We highlight, however, that these estimates of the size of the frontier are larger than those resulting from better microdata, such as in Syverson (2011) and Hsieh and Klenow (2009).

7 Conclusion

Technology adoption, technological innovation and their interaction contribute to economic growth and to the evolution of the productivity distribution. In the various models that we study, growth rates and the distribution of productivities are endogenous, and they depend on the specification of the innovation and adoption processes, as well as on the initial distribution of productivities available for adoption. In particular, whether adoption contributes to long-run growth in addition to innovation can depend on the properties (and tail index) of the initial distribution. The specification of the innovation process (as GBM or a Markov process) can determine whether or not the asymptotic stationary distribution of relative productivities (the ratio of the frontier to the bottom) has finite-support. We show in Propositions 2 and 4 that quality ladder type innovations driven by discrete-state Markov processes, under which a positive fraction of innovators leapfrog to the frontier, guarantee a stationary long-run distribution with a finite-support. We also study the problem of hysteresis: the possible multiplicity of stationary distributions that depend on initial conditions. Multiple stationary distributions occur when (1) the support of the stationary distribution is infinite and adoption contributes to long-run growth or (2) when the intensity of the innovation is endogenously chosen with a positive probability of leapfrogging to the finite frontier of the stationary distribution (see Proposition 6). The various models of innovation and adoption processes that we have studied describe a rich set of long-run productivity distributions and of growth rates that may be useful for empirical work on the evolution of productivities.

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(4), 1759–1799.

Acemoglu, D., P. Aghion, and F. Zilibotti (2006): "Distance to Frontier, Selection and Economic Growth," *Journal of the European Economic Association*, 4(1), 37–74.

Acemoglu, D., U. Akcigit, N. Bloom, and W. R. Kerr (2013): "Innovation, Reallocation and Growth," NBER Working Papers 18993, National Bureau of Economic Research, Inc.

Acemoglu, D., and D. V. Cao (2010): "Innovation by Entrants and Incumbents," NBER Working Papers 16411, National Bureau of Economic Research, Inc.

Acemoglu, D., and M. Dell (2010): "Productivity Differences between and within Countries," *American Economic Journal: Macroeconomics*, 2(1), 169–88.

Achdou, Y., J.-M. Lasry, P.-L. Lions, and B. Moll (2014): "Heterogeneous agent models in continuous time," *Unpublished manuscript, Princeton Univ., Princeton, NJ.*

Aghion, P., U. Akcigit, and P. Howitt (2014): "What Do We Learn From Schumpeterian Growth Theory?," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 2B of *Handbook of Economic Growth*, chap. 1, pp. 515–563. Elsevier.

Aghion, P., and P. Howitt (1992): "A Model of Growth Through Creative Destruction," *Econometrica*, 60(2), 323–51.

Akcigit, U., and W. R. Kerr (2016): "Growth Through Heterogeneous Innovations," NBER Working Papers 16443, National Bureau of Economic Research, Inc.

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

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

Arkolakis, C. (2015): "A Unified Theory of Firm Selection and Growth," *Quarterly Journal of Economics*, Forthcoming.

Arrow, K. J. (1962): "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, 29(3), pp. 155–173.

Atkeson, A., and A. T. Burstein (2010): "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 118(3), 433–484.

Baily, M. N., C. Hulten, D. Campbell, T. Bresnahan, and R. E. Caves (1992): "Productivity Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity. Microeconomics*, 1992, 187–267.

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

Bloom, N., R. Sadun, and J. V. Reenen (2014): "Management as Technology?," *Working Paper.*

Bloom, N., and J. Van Reenen (2010): "Why do management practices differ across firms and countries?," *The Journal of Economic Perspectives*, 24(1), 203–224.

Buera, F. J., and E. Oberfield (2015): "The Global Diffusion of Ideas," *Working Paper*.

Chor, D., and E. L.-C. Lai (2013): "Cumulative Innovation, Growth and Welfare-Improving Patent Policy," CESifo Working Paper Series 4407, CESifo Group Munich.

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

Comin, D., and B. Hobijn (2004): "Cross-country technology adoption: making the theories face the facts," *Journal of Monetary Economics*, 51(1), 39 – 83.

Comin, D., and B. Hobijn (2010): "An Exploration of Technology Diffusion," *American Economic Review*, 100(5), 2031–59.

Comin, D., B. Hobijn, and E. Rovito (2008): "Technology usage lags," *Journal of Economic Growth*, 13(4), 237–256.

Davis, S. J., J. Haltiwanger, and S. Schuh (1996): " Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts," *Small Business Economics*, 8(4), 297–315.

Eeckhout, J., and B. Jovanovic (2002): "Knowledge Spillovers and Inequality," *American Economic Review*, 92(5), 1290–1307.

Gabaix, X. (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1(1), 255–294.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016): "The dynamics of inequality," *Econometrica*, forthcoming.

Garcia-Macia, D., C.-T. Hsieh, and P. J. Klenow (2016): "How Destructive is Innovation?," *Working Paper*.

Grassi, B., and J. Imbs (2016): "Why Risky Sectors Grow Fast," *working paper*.

Grossman, G. M., and E. Helpman (1991): "Quality Ladders in the Theory of Growth," *The Review of Economic Studies*, 58(1), 43–61.

Grossman, G. M., E. Helpman, E. Oberfield, and T. Sampson (2016): "Balanced Growth Despite Uzawa," *American Economic Review*, Forthcoming.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–1448.

Jones, C. I. (1995): "R & D-based models of economic growth," *Journal of Political Economy*, pp. 759–784.

Jovanovic, B., and R. Rob (1989): "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, 56(4), 569–582.

Kemmerer, J. E., and J. Lu (2012): "Profitability and royalty rates across industries: Some preliminary evidence," Discussion paper, KPMG Global Valuation Institute.

Klette, T. J., and S. Kortum (2004): "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 112(5), 986–1018.

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

König, M., and T. Rogers (2016): "Endogenous Technology Cycles in Dynamic R&D Network," *Working Paper*.

Kortum, S. S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65(6), 1389–1420.

Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, 54(1), 159 – 178.

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

Lucas, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3–42.

——— (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(1), pp. 1–51.

Luttmer, E. (2015a): "An Assignment Model of Knowledge Diffusion and Income Inequality," Discussion Paper 509, Federal Reserve Bank of Minneapolis.

——— (2015b): "Four Models of Knowledge Diffusion and Growth," Discussion paper, Federal Reserve Bank of Minneapolis.

Luttmer, E. G. (2007): "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, 122(3), 1103–1144.

——— (2010): "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, 2, 547–576.

——— (2011): "On the Mechanics of Firm Growth," *Review of Economic Studies*, 78(3), 1042–1068.

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

Nelson, R. R., and E. S. Phelps (1966): "Investment in Humans, Technological Diffusion, and Economic Growth," *American Economic Review*, 56(1/2), 69–75.

Perla, J., and C. Tonetti (2014): "Equilibrium Imitation and Growth," *Journal of Political Economy*, 122(1), pp. 52–76.

Perla, J., C. Tonetti, and M. E. Waugh (2015): "Equilibrium Technology Diffusion, Trade, and Growth," Working Paper 20881, National Bureau of Economic Research.

Perron, P., and T. J. Vogelsang (1992): "Nonstationarity and level shifts with an application to purchasing power parity," *Journal of Business & Economic Statistics*, 10(3), 301–320.

Phillips, P. C., and P. Perron (1988): "Testing for a unit root in time series regression," *Biometrika*, 75(2), 335–346.

Raustiala, K., and C. Sprigman (2012): *The knockoff economy: How imitation sparks innovation*. Oxford University Press, USA.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5), S71–102.

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

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

Sutton, J. (1997): "Gibrat's Legacy," *Journal of Economic Literature*, 35(1), 40–59.

Syverson, C. (2011): "What Determines Productivity?," *Journal of Economic Literature*, 49(2), 326–65.

Uzawa, H. (1961): "Neutral inventions and the stability of growth equilibrium," *The Review of Economic Studies*, 28(2), 117–124.

Appendix A Exogenous Markov Innovation

Rather than the log utility in the main paper, the Appendix uses a general firm discount rate r, and the numerical algorithm in Technical Appendix B is implemented for an arbitrary CRRA utility function. Using log utility here simplifies a few of the parameter restrictions and expressions in the algebra compared to linear utility or general CRRA.

A.1 Normalization

This section derives the normalization procedure for the value functions and KFE.

Normalizing the Productivity Distribution Define the normalized distribution of productivity, as the distribution of productivity 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 yields

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

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

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

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

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

$$\boldsymbol{\partial}_t F_\ell(t, z) = -\lambda_\ell F_\ell(t, z) + \lambda_h F_h(t, z) + g(t)\boldsymbol{\partial}_z F_\ell(t, z) + S(t)\hat{F}_\ell(t, z) - S_\ell(t) \tag{A.7}$$

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

Take (9) and (10) and substitute from (A.2),

$$S_\ell(t) = g(t)\boldsymbol{\partial}_z F_\ell(t, 0) \tag{A.9}$$

$$S_h(t) = (g(t) - \gamma)\boldsymbol{\partial}_z F_h(t, 0) \tag{A.10}$$

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

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

Rearrange and differentiate (A.11) with respect to t

$$\boldsymbol{\partial}_t V_i(t,Z) = M'(t) v_i(t, \log(Z/M(t))) - M'(t)\frac{\partial v_i(t, \log(Z/M(t))}{\partial z} + M(t)\frac{\partial v_i(t, \log(Z/M(t))}{\partial t} \tag{A.12}$$

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

$$\frac{1}{M(t)}\boldsymbol{\partial}_t V_i(t,Z) = g(t)v_i(t,z) - g(t)\boldsymbol{\partial}_z v_i(t,z) + \boldsymbol{\partial}_t v_i(t,z) \tag{A.13}$$

Differentiate (A.11) with respect to Z and rearrange

$$\frac{1}{M(t)}\boldsymbol{\partial}_Z V_i(t,Z) = \frac{1}{Z}\boldsymbol{\partial}_z v_i(t,z) \tag{A.14}$$

Divide (2) by $M(t)$ and then substitute from (A.13) and (A.14),

$$r\frac{1}{M(t)}V_h(t,Z) = \frac{Z}{M(t)} + \gamma\frac{M(t)}{M(t)}\frac{Z}{Z}\boldsymbol{\partial}_z v_h(t,z) + g(t)v_h(t,z) - g(t)\boldsymbol{\partial}_z v_h(t,z) + \lambda_h(v_\ell(t,z) - v_h(t,z)) + \boldsymbol{\partial}_t v_h(t,z) \tag{A.15}$$

Use (A.11) and the definition of z and rearrange,

$$(r - g(t))v_h(t,z) = e^z + (\gamma - g(t))\boldsymbol{\partial}_z v_h(t,z) + \boldsymbol{\partial}_t v_h(t,z) \tag{A.16}$$

Similarly, for (1)

$$(r - g(t))v_\ell(t,z) = e^z - g(t)\boldsymbol{\partial}_z v_\ell(t,z) + \lambda_\ell(v_h(t,z) - v_\ell(t,z)) + \boldsymbol{\partial}_t v_\ell(t,z) \tag{A.17}$$

Optimal Stopping Conditions Divide the value-matching condition in (4) by $M(t)$,

$$\frac{V_i(t,M(t))}{M(t)} = \int_{M(t)}^{\bar{Z}(t)} \frac{V_\ell(t,Z)}{M(t)}\boldsymbol{\partial}_Z\hat{\Phi}_\ell(t,Z)\mathrm{d}Z + \int_{M(t)}^{\bar{Z}(t)} \frac{V_h(t,Z)}{M(t)}\boldsymbol{\partial}_Z\hat{\Phi}_h(t,Z)\mathrm{d}Z - \frac{M(t)}{M(t)}\zeta \tag{A.18}$$

Use the substitutions in (A.2) and (A.11), 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 $[\log(M(t)/M(t)), \log(\bar{Z}(t)/M(t))] = [0, \bar{z}(t)]$

$$v_i(t,0) = \int_0^{\bar{z}(t)} v_\ell(t,z)\mathrm{d}\hat{F}_\ell(t,z) + \int_0^{\bar{z}(t)} v_h(t,z)\mathrm{d}\hat{F}_h(t,z) - \zeta \tag{A.19}$$

Evaluate (A.14) at $Z = M(t)$, and substitute this into (6) to give the smooth-pasting condition,[46]

$$\boldsymbol{\partial}_z v_i(t,0) = 0 \tag{A.20}$$

[46]As a model variation, if the cost is proportional to Z, then the only change to the above conditions is that the smooth-pasting condition becomes $\boldsymbol{\partial}_z v(t,0) = -\zeta$. This cost formulation has the potentially unappealing feature that the value is not monotone in Z, as firms close to the adoption threshold would rather have a lower Z to decrease the adoption cost for the same benefit. Furthermore, as the gross value of adoption is independent of the Z, it makes sense that the cost of adoption is independent of the Z.

A.2 Common Adoption Threshold for All Idiosyncratic States

Proof. This section derives sufficient conditions for heterogeneous firms to choose the same adoption threshold. It is kept as general as possible to nest a model variations.

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 controls the infinitesimal generator $\mathbb{Q}_u$ of the Markov process on type i and, potentially, x.[47] 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. $\bar{v}(t) = v_\ell(t, \bar{z}(t)) = v_h(t, \bar{z}(t))$).[48] 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:

$$
\begin{aligned}
(r - g(t))v_i(t, z, x) = \max_{(\hat{\gamma},u)\in U(\cdot)} \Bigg\{ & \pi_i(t, z, x) - c(t, z, \hat{\gamma}, i, x, u) + (\hat{\gamma} - g)\frac{\partial v_i(t, z, x)}{\partial z} + \frac{\partial v_i(t, z, x)}{\partial t} \\
& + \mathbf{e_i} \cdot \mathbb{Q}_u \cdot v(t, z, x) + \eta(\bar{v}(t) - v_i(t, z, x))\Bigg\} \qquad \text{(A.21)}
\end{aligned}
$$

$$v_i(t, \underline{z}(t, i, x), x) = v_s(t) \tag{A.22}$$

$$\frac{\partial v_i(t, \underline{z}(t, i, x), x)}{\partial z} = 0, \tag{A.23}$$

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 $z = 0$,

$$v_i(t, 0, x) = v_s(t) \tag{A.24}$$

$$\frac{\partial v_i(t, 0, x)}{\partial z} = 0. \tag{A.25}$$

Note that equations (A.24) and (A.25) are identical for any i and x. Substitute equations (A.24) and (A.25) into equation (A.21) to obtain

$$(r - g(t))v_s(t) = \max_{(\hat{\gamma},u)\in U(\cdot)} \left\{\pi(t, 0) - c(t, z, \hat{\gamma}, i, x, u) + \mathbf{e_i} \cdot \mathbb{Q}_u \cdot v_s(t) + \eta(\bar{v}(t) - v_s(t)) + v_s'(t)\right\}. \tag{A.26}$$

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,

$$(r - g(t))v_s(t) = \max_{(\hat{\gamma},u)\in U(\cdot)} \left\{\pi(t, 0) - c(t, z, \hat{\gamma}, i, x, u) + v_s'(t) + \eta(\bar{v}(t) - v_s(t))\right\}. \tag{A.27}$$

[47] 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.

[48] 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.

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,

$$(r - g(t))v_s(t) = \pi(t, 0) + v_s'(t) + \eta(\bar{v}(t) - v_s(t)) \quad \text{for all } i. \tag{A.28}$$

Therefore, the necessary conditions for optimal stopping are identical for all i, x, z, confirming our guess. Furthermore, equation (A.28) 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 $\pi(0)$,

$$v(0) \equiv v_s = \frac{\pi(0) + \eta\bar{v}}{r - g + \eta} \tag{A.29}$$

Furthermore, note from equation (97) that $\pi(0) = 1$ for all variations of the model in the body of the paper.

□

A.3 Stationary BGP with a Finite, Unbounded Technology Frontier

Proof of Proposition 2. This proof begins by proving that growth cannot be greater than the innovation rate, and then solves the $g = \gamma$ version. For the $g < \gamma$, see the nested solution in Appendix A.6.

Claim: $g \leq \gamma$ Assume by contradiction there exists an optimally chosen—according to (4)—$g > \gamma$ and $\bar{Z} < \infty$. Note that $M(t) = M(0)e^{gt}$ and $\bar{Z}(t) = \bar{Z}(0)e^{\gamma t}$, so for any initial conditions $M(T) > \bar{Z}(T)$ for all t greater than some T. That is, the distribution compresses until it is a point.

A more subtle question is why an equilibrium cannot exist with $g(t) > \gamma$ for all t, but where $\lim_{t\to\infty} g(t) = \gamma$. That is, why can't $M(t)$ grow faster than the frontier and only converge when the distribution reaches the point $M(T) = \bar{Z}(T)$ at arbitrary large T? The reason is that this would be a contradiction of the optimality of $M(t)$ from (4). Recall that the returns to adoption are related to the current productivity, $M(t)$, relative to the expectation of a draw. As the distribution compresses to a point, the returns to adoption become arbitrarily small. However, the cost of adoption is strictly positive in (4), and equality could not be maintained. Therefore, there would be some point along the transition where they were in balance, the distribution would cease to compress, and $M(t) < \bar{Z}(t)$ would be maintained for all subsequent t.

Case $g = \gamma$: Define the following to simplify notation,

$$\alpha \equiv (1 + \hat{\lambda})\frac{S}{g} \tag{A.30}$$

$$\hat{\lambda} \equiv \frac{\lambda_\ell}{\lambda_h} \tag{A.31}$$

$$\bar{\lambda} \equiv \frac{\lambda_\ell}{r - g + \lambda_h} + 1 \tag{A.32}$$

$$\nu = \frac{(r - g)\bar{\lambda}}{g} \tag{A.33}$$

See Technical Appendix C.4 for a proof that there are no bounded, finite equilibria for any $\kappa > 0$. For the solution to the $\kappa = 1$ case, take (49) and solve for $F_h(z)$

$$F_h(z) = \hat{\lambda}F_\ell(z) \tag{A.34}$$

Substitute into (48)

$$S = gF_\ell'(z) + \left(\hat{\lambda} + 1\right) SF_\ell(z) \tag{A.35}$$

Solve this as an ODE in $F_\ell(z)$, subject to the $F_\ell(0) = 0$ boundary condition in (23)

$$F_\ell(z) = \tfrac{1}{1+\hat{\lambda}} e^{-\alpha z} \tag{A.36}$$

We can check that if $\alpha > 0$ the right boundary conditions hold

$$\lim_{z \to \infty} \left(F_\ell(z) + F_h(z)\right) = 1 \tag{A.37}$$

Differentiate (A.36),

$$F_\ell'(z) = \tfrac{\alpha}{1+\hat{\lambda}} e^{-\alpha z} \tag{A.38}$$

With (A.34), the PDF for the unconditional distribution, $F(z)$,

$$F'(z) = \alpha e^{-\alpha z} \tag{A.39}$$

Solve (46) for $v_h(z)$

$$v_h(z) = \frac{e^z + \lambda_h v_\ell(z)}{r - g + \lambda_h} \tag{A.40}$$

Substitute into (21) to find an ODE in $v_\ell(z)$

$$(r-g)v_\ell(z) = e^z + \lambda_h \hat{\lambda} \left(-v_\ell(z) + \frac{e^z + \lambda_h v_\ell(z)}{r - g + \lambda_h}\right) - g v_\ell'(z) \tag{A.41}$$

Use the constant definitions and simplify

$$(r-g)v_\ell(z) = e^z - \frac{g v_\ell'(z)}{\bar{\lambda}} \tag{A.42}$$

Solve this ODE subject to the smooth-pasting condition in (30) and simplify,

$$v_\ell(z) = \frac{\bar{\lambda}}{g + (r-g)\bar{\lambda}} e^z + \frac{1}{(r-g)(\nu+1)} e^{-z\nu} \tag{A.43}$$

Use the definitions of the constants and (A.43)

$$v_\ell(0) = \tfrac{1}{r-g} \tag{A.44}$$

Substitute (A.39), (A.43) and (A.44) into the value-matching condition in (47) and simplify

$$\frac{1}{r-g} = \int_0^\infty \left[\frac{e^{z\left(\bar{\lambda} - \alpha - \frac{r\bar{\lambda}}{g}\right)} \alpha g}{(g-r)\left(-r\bar{\lambda} + g\left(\bar{\lambda} - 1\right)\right)} + \frac{e^{z - z\alpha} \alpha \bar{\lambda}}{g + r\bar{\lambda} - g\bar{\lambda}}\right] \mathrm{d}z - \zeta \tag{A.45}$$

Evaluate the integral,

$$\zeta = \frac{\alpha\left(-r\bar{\lambda} + g\left(\bar{\lambda} - \alpha + 1\right)\right)}{(g-r)\left(r\bar{\lambda} + g\left(\alpha - \bar{\lambda}\right)\right)(\alpha - 1)} - \frac{1}{(r-g)(\nu+1)} - \frac{\bar{\lambda}}{g + r\bar{\lambda} - g\bar{\lambda}} \tag{A.46}$$

Substitute for α gives an implicit equation in S

$$0 = \zeta + \frac{g\left(\frac{1}{r-g} + \frac{\bar{\lambda}}{S - g + S\hat{\lambda}} - \frac{\bar{\lambda}}{S - g\bar{\lambda} + r\bar{\lambda} + S\hat{\lambda}}\right)}{-r\bar{\lambda} + g\left(\bar{\lambda} - 1\right)} + \frac{1}{(r-g)(\nu+1)} \tag{A.47}$$

As $g = \gamma$ in equilibrium, only S is unknown. This equation is a quadratic in S, and can be analytically in terms of model parameters as,

$$S = \frac{\lambda_h\left(\zeta r(r+\lambda_h+\lambda_\ell) - \sqrt{\zeta((4g+r^2\zeta)(-g+r+\lambda_h)^2 + 2(-2g+(g-r)r\zeta)(g-r-\lambda_h)\lambda_\ell + (g-r)^2\zeta{\lambda_\ell}^2)} + \zeta g^2 2 + \zeta(-g)(3r+2\lambda_h+\lambda_\ell)\right)}{2\zeta(\lambda_h+\lambda_\ell)(g-r-\lambda_h)} \tag{A.48}$$

From this S, α can be calculated through (A.30) and the the rest of the equilibrium follows. $\square$

A.4 Bounded Support

Proof of Proposition 4. Define the following to simplify notation,

$$\alpha \equiv (1+\hat{\lambda})\frac{S-\eta}{g} \tag{A.49}$$

$$\hat{\lambda} \equiv \frac{\lambda_\ell}{\eta+\lambda_h} \tag{A.50}$$

$$\bar{\lambda} \equiv \frac{r-g+\lambda_\ell+\lambda_h}{r-g+\lambda_h} \tag{A.51}$$

$$\nu = \frac{r-g+\eta}{g}\bar{\lambda} \tag{A.52}$$

Solve for $F_h(z)$ in (68),

$$F_h(z) = \hat{\lambda}F_\ell(z) \tag{A.53}$$

Substitute this back into (67) to get an ODE in F_ℓ

$$0 = gF'_\ell(z) + (S-\eta)(1+\hat{\lambda})F_\ell(z) + \eta H\left(z-\bar{z}\right) - S \tag{A.54}$$

Solve this ODE with the boundary condition $F_\ell(0) = 0$

$$F_\ell(z) = \begin{cases} \frac{S}{(S-\eta)(1+\hat{\lambda})}(1-e^{-\alpha z}) & 0 \leq z < \bar{z} \\ \frac{S}{(S-\eta)(1+\hat{\lambda})}(1-e^{-\alpha \bar{z}}) & z = \bar{z} \end{cases} \tag{A.55}$$

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

$$F(z) = (1+\hat{\lambda})F_\ell(\bar{z}) \tag{A.56}$$

$$= \frac{S}{S-\eta}(1-e^{-\alpha z}) \tag{A.57}$$

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

$$\bar{z} = \frac{\log(S/\eta)}{\alpha} \tag{A.58}$$

The PDF of the unconditional distribution is,

$$F'(z) = \frac{\alpha S}{S-\eta}e^{-\alpha z} \tag{A.59}$$

$$\tag{A.60}$$

To solve for the value, solve (66) for $v_h(z)$,

$$v_h(z) = \frac{e^z + (\lambda_h - \eta)v_\ell(z) + \eta v_\ell(\bar{z})}{r-g+\lambda_h} \tag{A.61}$$

Substitute into (65) and simplify

$$(r-g+\eta)v_\ell(z) = e^z + \eta v_\ell(\bar{z}) - \frac{g}{\bar{\lambda}}v'_\ell(z) \tag{A.62}$$

Solve (30) and (A.62) and simplify,

$$v_\ell(z) = \frac{\bar{\lambda}}{g+(r+\eta-g)\bar{\lambda}}e^z + \frac{\eta}{r-g+\eta}v_\ell(\bar{z}) + \frac{1}{(r+\eta-g)(\nu+1)}e^{-z\nu} \tag{A.63}$$

Evaluate (A.63) at $\bar{z}$ and solve for $v_\ell(\bar{z})$,

$$v_\ell(\bar{z}) = \left(-\frac{\eta}{g-r}+1\right)\left(\frac{e^{\bar{z}}\bar{\lambda}}{g+(\eta+r-g)\bar{\lambda}} + \frac{e^{-\nu\bar{z}}}{(\eta+r-g)(\nu+1)}\right) \tag{A.64}$$

Subtitute (A.64) into (A.63) to find an expression for $v_\ell(z)$

$$v_\ell(z) = \frac{\bar{\lambda}}{g(1+\nu)}\left(e^z + \frac{1}{\nu}e^{-\nu z} + \frac{\eta}{r-g}\left(e^{\bar{z}} + \frac{1}{\nu}e^{-\nu\bar{z}}\right)\right) \tag{A.65}$$

Substitute (A.59) and (A.65) into the value-matching condition in (47) and evaluate the integral,

$$\zeta + \frac{1}{r-g} = \frac{S\alpha\bar{\lambda}\left(-\frac{e^{-\nu\bar{z}}\left(-1+e^{-\alpha\bar{z}}\right)\eta}{(-g+r)\alpha\nu} + \frac{e^{\bar{z}}\eta\left(e^{-\alpha\bar{z}}-1\right)}{\alpha(g-r)} + \frac{-e^{-(\alpha+\nu)\bar{z}}+1}{\nu(\alpha+\nu)} + \frac{-e^{\bar{z}-\alpha\bar{z}}+1}{\alpha-1}\right)}{g(S-\eta)(\nu+1)} \tag{A.66}$$

To find an implicit equation for the equilibrium S, take (A.66) and substitute for α and $\bar{z}$ from (A.49) and (A.58)

$$\zeta + \frac{1}{r-g} = \frac{S\bar{\lambda}\left(\hat{\lambda}+1\right)\left(\frac{-\left(\frac{S}{\eta}\right)^{-1-\frac{g\nu}{(S-\eta)\left(1+\hat{\lambda}\right)}}+1}{\nu\left(\nu+\frac{(S-\eta)\left(\hat{\lambda}+1\right)}{g}\right)} + g\left(\frac{1}{-g+(S-\eta)\left(\hat{\lambda}+1\right)} + \frac{\eta\left(-\frac{\left(\frac{S}{\eta}\right)^{-\frac{g\nu}{(S-\eta)\left(1+\hat{\lambda}\right)}}}{\nu} + \frac{\left(\frac{S}{\eta}\right)^{\frac{g}{(S-\eta)\left(\hat{\lambda}+1\right)}}\left(\eta+r-S+\hat{\lambda}(\eta+r-g-S)\right)}{-g+(S-\eta)\left(\hat{\lambda}+1\right)}\right)}{S(g-r)\left(\hat{\lambda}+1\right)}\right)\right)}{g^2(\nu+1)} \tag{A.67}$$

□

A.5 Stationary Stochastic Innovation Equilibrium with Infinite Support

This proof applies to the case of infinite-support and $g > \gamma$.

Proof of Proposition 1. Define $\mathbf{0}, \mathbf{1}, \mathbf{I}$ as a vector of 0, 1, and the identity matrix and the following:

$$A \equiv \begin{bmatrix} \frac{1}{g} \\ \frac{1}{g-\gamma} \end{bmatrix} \qquad B \equiv \begin{bmatrix} \frac{r+\lambda_\ell-g}{g} & -\frac{\lambda_\ell}{g} \\ -\frac{\lambda_h}{g-\gamma} & \frac{r+\lambda_h-g}{g-\gamma} \end{bmatrix} \tag{A.68}$$

$$C \equiv \begin{bmatrix} \frac{gF_\ell'(0)+(g-\gamma)F_h'(0)-\lambda_\ell}{g} & \frac{\lambda_h}{g} \\ \frac{\lambda_\ell}{g-\gamma} & \frac{gF_\ell'(0)+(g-\gamma)F_h'(0)-\lambda_h}{g-\gamma} \end{bmatrix} \qquad D \equiv \begin{bmatrix} F_\ell'(0) \\ F_h'(0) \end{bmatrix} \tag{A.69}$$

$$\vec{F}(z) \equiv \begin{bmatrix} F_\ell(z) \\ F_h(z) \end{bmatrix} \qquad v(z) \equiv \begin{bmatrix} v_\ell(z) \\ v_h(z) \end{bmatrix} \tag{A.70}$$

Then the equilibrium conditions can be written as a linear set of ODEs:

$$v'(z) = Ae^z - Bv(z) \tag{A.71}$$

$$v'(0) = \mathbf{0} \tag{A.72}$$

$$\vec{F}'(z) = -C\vec{F}(z) + D \tag{A.73}$$

$$\vec{F}(0) = \mathbf{0} \tag{A.74}$$

$$\vec{F}(\infty)\cdot\mathbf{1} = 1 \tag{A.75}$$

$$v_\ell(0) = v_h(0) = \int_0^\infty \left(v(z)^T \cdot \vec{F}'(z)\right) \mathrm{d}z - \zeta \tag{A.76}$$

Solve these as a set of matrix ODEs, where e^{Az} is a matrix exponential. Start with (A.71) and (A.72) to get,[49]

$$v(z) = (I + B)^{-1} \left(e^{Iz} + e^{-Bz} B^{-1}\right) A. \tag{A.79}$$

Evaluate at $z = 0$,

$$v(0) = B^{-1}A = \begin{bmatrix} 1/(r-g) & 1/(r-g) \end{bmatrix}. \tag{A.80}$$

Then (A.73) and (A.74) gives

$$\vec{F}(z) = \left(\mathbf{I} - e^{-Cz}\right) C^{-1} D. \tag{A.81}$$

Take the derivative,

$$\vec{F}'(z) = e^{-Cz} D. \tag{A.82}$$

For (A.79) and (A.81) to be well defined as $z \to \infty$, we have to impose parameter restrictions that constrain the growth rate g so that the eigenvalues of B and C are positive or have positive real parts. S_l and S_h are defined in equations (25) and (26) in terms of $F_l'(0)$ and $F_h'(0)$, C and B will have roots with positive real parts iff their determinant and their trace are strictly positive. For C it is straightforward to compute that the conditions for a positive trace and determinant are

$$S_l + S_h > \frac{(g-\gamma)\lambda_l + g\lambda_h}{(g-\gamma) + g} \tag{A.83}$$

$$S_h + S_l > \lambda_h + \lambda_l \tag{A.84}$$

and for B the corresponding conditions are

$$r > g > \gamma \tag{A.85}$$

$$r - g + \lambda_h + \lambda_l > 0. \tag{A.86}$$

With these conditions imposed, we can proceed to characterize the solutions to the value functions and the stationary distribution.

Evaluate (A.82) at $z = 0$,

$$\vec{F}'(0) = D. \tag{A.87}$$

Take the limit of (A.81)

$$\vec{F}(\infty) = C^{-1} D \tag{A.88}$$

[49] The equation $\vec{F}'(z) = AF(z) + b$ subject to $\vec{F}(0) = \mathbf{0}$ has the solution,

$$\vec{F}(z) = \left(e^{Az} - \mathbf{I}\right) A^{-1} b. \tag{A.77}$$

The derivation of these results uses that $\int_0^T e^{tA}\mathrm{d}t = A^{-1}\left(e^{TA} - \mathbf{I}\right)$. With appropriate conditions on eigenvalues, this implies that $\int_0^\infty e^{tA}\mathrm{d}t = -A^{-1}$.
Equations of the form, $v'(z) = Ae^z - B \cdot v(z)$ with the initial condition $v'(0) = \mathbf{0}$ have the solution,

$$v(z) = (I + B)^{-1} \left(e^{Iz} + e^{-Bz} B^{-1}\right) A. \tag{A.78}$$

This derivation exploits commutativity, as both e^{Bz} and $(I + B)^{-1}$ can be expanded as power series of B.

and (A.75) becomes

$$1 = C^{-1} D \cdot \mathbf{1}. \tag{A.89}$$

We can check that, by construction, with the C and D defined by (A.69), (A.89) is fulfilled for any $F'_\ell(0), F'_h(0), \lambda_\ell$, and λ_h.

For $\vec{F}'(z)$ to define a valid PDF it is necessary for $\vec{F}'(z) > 0$ for all z. It can be shown, that for $C > 0$, the only $D > 0$ fulfilling this requirement is one proportional to the eigenvector associated with the dominant eigenvalue of C.[50] The unique constant of proportionality is determined by (A.89). The two eigenvectors of C fulfilling this proportionality are,

$$\nu_i \equiv \begin{bmatrix} -\frac{F'_h(0)\left(\gamma(g-\gamma)F'_h(0) \pm \sqrt{2(g-\gamma)\lambda_l\left(\gamma\left(g\left(F'_h(0)+F'_l(0)\right)-\gamma F'_h(0)\right)+g\lambda_h\right)+\left(\gamma\left(g\left(F'_h(0)+F'_l(0)\right)-\gamma F'_h(0)\right)-g\lambda_h\right)^2+(g-\gamma)^2\lambda_l{}^2}+\gamma g F'_l(0)-g\lambda_h+g\lambda_l-\gamma\lambda_l\right)}{2g\lambda_l} \\ F'_h(0) \end{bmatrix} \tag{A.90}$$

Denote ν as the eigenvector with both positive elements–which is associated with the dominant eigenvalue–then as discussed above, $D \propto \nu$. Use (A.69) and (A.90), and note the eigenvector has already been normalized to match the second parameter.

$$D = \nu. \tag{A.91}$$

The 2nd coordinate already holds with equality by construction, for the first coordinate equating (A.69) and (A.90). Equate the first parameter and choose the positive eigenvector,

$$F'_\ell(0) = -\frac{F'_h(0)\left(\gamma(g-\gamma)F'_h(0) - \sqrt{2(g-\gamma)\lambda_l\left(\gamma\left(g\left(F'_h(0)+F'_l(0)\right)-\gamma F'_h(0)\right)+g\lambda_h\right)+\left(\gamma\left(g\left(F'_h(0)+F'_l(0)\right)-\gamma F'_h(0)\right)-g\lambda_h\right)^2+(g-\gamma)^2\lambda_l{}^2}+\gamma g F'_l(0)-g\lambda_h+g\lambda_l-\gamma\lambda_l\right)}{2g\lambda_l} \tag{A.92}$$

Solve this equation for $F'_\ell(0)$ and choose the positive root

$$F'_\ell(0) = \frac{F'_h(0)\lambda_h}{\gamma F'_h(0) + \lambda_\ell}. \tag{A.93}$$

We can check that with the C and D defined by (A.69), (A.89) is fulfilled by construction. The value-matching in (A.76) becomes,

$$\frac{1}{r-g} + \zeta = \int_0^\infty \left[\left[(I+B)^{-1}\left(e^{Iz} + e^{-Bz}B^{-1}\right)A\right]^T e^{-Cz} D\right] \mathrm{d}z. \tag{A.94}$$

Note that if B has positive eigenvalues, then $\lim_{z\to\infty} v(z) = (1+B)^{-1}(e^z)A$. Therefore, as long as C has a minimal eigenvalue (defined here as α), strictly greater than one, the integral is defined.

The tail index of the unconditional distribution, $F(z) \equiv F_\ell(z) + F_h(z)$ can be calculated from the C matrix in (A.82). As sums of power-law variables inherit the smallest tail index, the endogenous power-law tail is minimum eigenvalue of C. After the substitution for $F'_\ell(0)$ from above, the smallest eigenvalue of C is

$$\alpha \equiv \frac{\left((g-\gamma)F'_h(0) - \lambda_l\right)\left(\gamma(g-\gamma)F'_h(0) + g\left(\lambda_h + \lambda_l\right) - \gamma\lambda_l\right)}{g(g-\gamma)\left(\gamma F'_h(0) + \lambda_l\right)}. \tag{A.95}$$

[50]Since $C > 0$ and irreducible (in this case off diagonals not zero), then by Perron-Frobenius it has a simple dominant real root α and an associated eigenvector $\nu > 0$. Hence, as $\vec{F}(0) = 0, F_\ell(\infty) + F_h(\infty) = 1$, and $\vec{F}'(z) > 0$, we have a valid PDF. This uniqueness of the ν solution only holds if the other eigenvector of C has a positive and negative coordinate, which always holds in our model.

Solve (A.95) for $F_h'(0)$ as a function of α,

$$F_h'(0) = \frac{g\left(\alpha\gamma - \lambda_h + \sqrt{(\lambda_h - \alpha\gamma)^2 + 2\lambda_l\,(\alpha\gamma + \lambda_h) + \lambda_l{}^2} - \lambda_l\right) + 2\gamma\lambda_l}{2\gamma(g-\gamma)}. \tag{A.96}$$

Substitute for $F_i'(0)$ into C and D to get a function in terms of g and α,

$$C = \begin{bmatrix} \frac{-\alpha\gamma+2\alpha g+\lambda_h+\sqrt{(\lambda_h-\alpha\gamma)^2+2\lambda_l(\alpha\gamma+\lambda_h)+\lambda_l{}^2}-\lambda_l}{2g} & \frac{\lambda_h}{g} \\ \frac{\lambda_l}{g-\gamma} & \frac{-\alpha\gamma+2\alpha g-\lambda_h+\sqrt{(\lambda_h-\alpha\gamma)^2+2\lambda_l(\alpha\gamma+\lambda_h)+\lambda_l{}^2}+\lambda_l}{2(g-\gamma)} \end{bmatrix} \tag{A.97}$$

$$D = \begin{bmatrix} \frac{\lambda_h\left(g\left(\alpha\gamma-\lambda_h+\sqrt{(\lambda_h-\alpha\gamma)^2+2\lambda_l(\alpha\gamma+\lambda_h)+\lambda_l{}^2}-\lambda_l\right)+2\gamma\lambda_l\right)}{\gamma g\left(\alpha\gamma-\lambda_h+\sqrt{(\lambda_h-\alpha\gamma)^2+2\lambda_l(\alpha\gamma+\lambda_h)+\lambda_l{}^2}+\lambda_l\right)} \\ \frac{g\left(\alpha\gamma-\lambda_h+\sqrt{(\lambda_h-\alpha\gamma)^2+2\lambda_l(\alpha\gamma+\lambda_h)+\lambda_l{}^2}-\lambda_l\right)+2\gamma\lambda_l}{2\gamma(g-\gamma)} \end{bmatrix} \tag{A.98}$$

As in the example with Geometric Brownian Motion, there are multiple stationary equilibria. While both $F_i'(0)$ could conceivably parameterize a set of solutions for each g, they are constrained by the eigenvector proportionality condition, which ensures that the manifold of solutions is 1 dimensional.

□

(41) shows that the positivity of the tail index α is now equivalent to C having positive eigenvalues. For the decomposition of the option value, (43) shows that positive eigenvalues of B ensure the option values in the vector $v(z)$ converges to 0 as z increases.

In Technical Appendix Propositions 1 and 2 we characterized the stationary distributions in terms of the tail index of the initial distribution of productivities, given in Technical Appendix (E.39) explicitly by $\alpha = \kappa F'(0)$, a scalar. In this section, the stationary distribution is a vector $\vec{F}(z)$ solving (33) and (34) , a system of linear ODEs. If we define the unconditional distribution $F(z) \equiv F_\ell(z)+F_h(z)$, and if both $F_\ell(z)$ and $F_h(z)$ are power-laws, any mixture of these distributions inherits the smallest (i.e. thickest) tail parameter (as discussed in Gabaix (2009)). Since there are now two dimensions of heterogeneity, the tail index, α, is defined as that of the unconditional distribution, $F(z)$. The ODE solution for the vector $\vec{F}(z)$ given in Proposition 1 by (41) will depend on the roots of C (both positive, see Appendix A.5). The smallest root of C, representing the slower rate of decay for both elements of $F(z)$, is the tail index α by the construction of (38).

Note that in Technical Appendix Propositions 1 to 3, the tail index is determined by the single initial condition $F'(0) > 0$, a scalar. In Proposition 1 the initial condition $F'(0)$ is a vector, so in principle this raises the possibility that the continuum of stationary equilibria could be two dimensional, parametrized by $F_\ell'(0) > 0$ and by $F_h'(0) > 0$. However as shown in Appendix A.5 this is not possible since the only initial condition that ensures that $F_\ell(z)$ and $F_h(z)$ remain positive and satisfy (21) and (22) is exactly the eigenvector of C corresponding to its dominant (Frobenius) eigenvalue. Since the eigenvector is determined only up to a multiplicative constant, the continuum of stationary distributions is therefore one dimensional. We use the smallest eigenvalue of C, defined as the tail index α, to solve for $F_h'(0)$, which then determines $F_\ell'(0)$ from the eigenvector restriction. This then allows us to obtain the expressions (38) and (39) in terms of parameters, α and g. Then value-matching, (A.76) and (A.94), gives us expression (40) to define g in terms α, so we end up with a continuum of stationary equilibria parametrized by α.

A.6 Stationary Stochastic Innovation Equilibrium with $g < \gamma$

This derivation applies to the case with $g < \gamma$ and either unbounded or infinite-support.

Proof of Proposition 3. Unlike the case with a infinite-support and $g > \gamma$ in Appendix A.5, we can use the simple case of draws are from the unconditional distribution, i.e. $\hat{\Phi}(t, Z) = \Phi_\ell(t, Z) + \Phi_h(t, Z)$.

Given our $g < \gamma$ assumption, $S_h = 0$ as no agents in the high state will cross the barrier. For the support of the distribution, the minimum of support has been normalized to $\log(M(t)/M(t)) = 0$. The maximum of support starts at $\log(\bar{Z}(0)/M(0)) \equiv \bar{z}(0)$ and grows as $\bar{z}(t) = \log\left(\bar{z}(0)\frac{e^{\gamma t}}{e^{gt}}\right)$ or

$$\bar{z}(t) = \bar{z}(0) + (\gamma - g)t \tag{A.99}$$

Hence, asymptotically, $\lim_{t\to\infty} \bar{z}(t) = \infty$ given our $\gamma > g$ assumption. Alternatively, if $\bar{Z}(0) = \infty$, then this solution fully nests the $g < \gamma$ case as well. Following the notation of Appendix A.5 as much as possible, define

$$A \equiv \begin{bmatrix} \frac{1}{g} \\ \frac{1}{\gamma - g} \end{bmatrix} \qquad B \equiv \begin{bmatrix} \frac{r+\lambda_\ell - g}{g} & -\frac{\lambda_\ell}{g} \\ -\frac{\lambda_h}{\gamma - g} & \frac{r+\lambda_h - g}{\gamma - g} \end{bmatrix} \tag{A.100}$$

$$C \equiv \begin{bmatrix} f_\ell(0) - \frac{\lambda_\ell}{g} & f_\ell(0) + \frac{\lambda_h}{g} \\ \frac{-\lambda_\ell}{\gamma - g} & \frac{\lambda_h}{\gamma - g} \end{bmatrix} \qquad D \equiv \begin{bmatrix} F'_\ell(0) \\ 0 \end{bmatrix} \tag{A.101}$$

Otherwise, the $\vec{F}(z)$, and $v(z)$ are identical to those in (A.68) and (A.69), and the ODEs for the KFE and the value function fulfill the same (A.71) to (A.75).[51] The only change in A and B from the previous definition was to swap the order to $\gamma - g$. Under these definitions, the matrix A and the eigenvalues of B are positive.

As we have unconditional draws ending up with the ℓ type, the value-matching condition is replaced with the standard one, as in (47)

$$v_\ell(0) = v_h(0) = \int_0^\infty v_\ell(z) F'(z) \mathrm{d}z - \zeta \tag{A.102}$$

The other matrices and generic solutions to the ODEs for the $v(z)$ and $F(z)$ are identical to those in (A.79), (A.81) and (A.82).

In order to better parameterize the solutions, reorganize the algebra to be in terms of the asymptoptic tail parameter of the distribution, α, instead of $F'_\ell(0)$. Using this, there may be a continuum of $\{g, \alpha\}$ which generate stationary distributions (and have an accompanying $F'_\ell(0)$). Define the following,

$$\bar{\alpha} \equiv \frac{\lambda_h}{\gamma - g} - \frac{\lambda_\ell}{g} - \frac{\alpha\gamma\lambda_\ell}{g\left(\lambda_\ell + \lambda_h - \alpha(\gamma - g)\right)} \tag{A.103}$$

$$F'_\ell(0) \equiv \alpha\left(1 - \frac{\gamma\lambda_\ell}{g\left(-(\gamma - g)\alpha + \lambda_h + \lambda_\ell\right)}\right) \tag{A.104}$$

$$= \alpha + \bar{\alpha} - \frac{\lambda_h}{\gamma - g} + \frac{\lambda_\ell}{g} \tag{A.105}$$

Here, $\{\alpha, \bar{\alpha}\}$ are the eigenvalues of the matrix C, ordered so that $\bar{\alpha} > \alpha$

Take the solution (A.82) and expand using the constant definitions above,

$$\vec{F}'(z) = \frac{\alpha\left(\lambda_h - \left(\frac{\gamma}{g} - 1\right)\lambda_\ell - (\gamma - g)\alpha\right)}{(\gamma - g)\left(\bar{\alpha} - \alpha\right)} \frac{1}{\lambda_h + \lambda_\ell - \alpha(\gamma - g)} \begin{bmatrix} (\lambda_h - \alpha(\gamma - g))\, e^{-\alpha z} - (\lambda_h - \bar{\alpha}(\gamma - g))\, e^{-\bar{\alpha} z} \\ \lambda_\ell \left(e^{-\alpha z} - e^{-\bar{\alpha} z}\right) \end{bmatrix} \tag{A.106}$$

[51] Unlike the infinite horizon case with $g > \gamma$, we no longer require an argument based on Perron-Froebenius in the proof. The reason is that $F'_h(0) = 0$ trivially, so the manifold of the solution in $\{g, F'_\ell(0)\}$ is already of the correct dimension.

The solution for the unconditional PDF, $F'(z) \equiv F'_\ell(z) + F'_h(z)$, is,

$$F'(z) = b_1\left(e^{-\alpha z} - b_2 e^{-\bar{\alpha} z}\right) \tag{A.107}$$

Where,

$$b_1 \equiv \frac{\alpha\left(-\alpha(\gamma-g)+\left(1-\frac{\gamma}{g}\right)\lambda_l+\lambda_h\right)}{(\bar{\alpha}-\alpha)(\gamma-g)} \tag{A.108}$$

$$b_2 \equiv \frac{\bar{\alpha}(g-\gamma)+\lambda_h+\lambda_l}{\alpha(g-\gamma)+\lambda_h+\lambda_l} \tag{A.109}$$

To solve the value function in (A.81), first define the value β_1 and β_2 as,

$$\beta_{1/2} \equiv -\frac{-g\left(\gamma-\lambda_h+\lambda_l\right) \pm \sqrt{\left(g\ \lambda_h+(\gamma-g)\lambda_l+\gamma(r-g)\right)^2+4g\ (g-\gamma)(g-r)\left(g-\lambda_h-\lambda_l-r\right)}+\gamma\ \left(\lambda_l+r\right)}{2g(g-\gamma)} \tag{A.110}$$

Using the eigenvectors and eigenvalues, the matrix exponential in (A.79) can be written through a standard eigendecomposition with the eigenvalues β_1, β_2 and associated eigenvectors. Use this technique and rearrange (A.79) to find,

$$v_\ell(z) = a_1 e^z + a_2\left(e^{-\beta_1 z} - a_3 e^{-\beta_2 z}\right) \tag{A.111}$$

where,

$$a_1 \equiv \frac{\gamma-2g+\lambda_h+\lambda_l+r}{-2g\left(\lambda_l+r\right)+r\left(\gamma+\lambda_h+r\right)+\lambda_l(\gamma+r)} \tag{A.112}$$

$$a_2 \equiv \frac{\left(-\beta_2\gamma+\left(\beta_2-1\right)g+r\right)\left((g-\gamma)\lambda_h\lambda_l+g\left(-\gamma+2g-\lambda_h-r\right)\left(g\left(\beta_1-1\right)+\lambda_h-\gamma\beta_1+r\right)\right)}{(g-\gamma)(r-g)\lambda_h\left(\beta_1-\beta_2\right)\left(-2g\left(\lambda_l+r\right)+r\left(\gamma+\lambda_h+r\right)+\lambda_l(\gamma+r)\right)} \tag{A.113}$$

$$a_3 \equiv \frac{\left(g\left(\beta_1-1\right)-\gamma\beta_1+r\right)\left((g-\gamma)\lambda_h\lambda_l+g\left(-\gamma+2g-\lambda_h-r\right)\left(-\beta_2\gamma+\left(\beta_2-1\right)g+\lambda_h+r\right)\right)}{\left(-\beta_2\gamma+\left(\beta_2-1\right)g+r\right)\left((g-\gamma)\lambda_h\lambda_l+g\left(-\gamma+2g-\lambda_h-r\right)\left(g\left(\beta_1-1\right)+\lambda_h-\gamma\beta_1+r\right)\right)} \tag{A.114}$$

To finalize the solution, for any given α within the set of admissible parameters, substitute (A.107) and (A.111) into (A.102) and solve as an implicit function of g. Substitute and simplify to find,

$$\frac{1}{r-g} = a_2 b_1\left(\frac{a_3}{\alpha+\beta_2} - \frac{a_3 b_2}{\bar{\alpha}+\beta_2} - \frac{1}{\alpha+\beta_1} + \frac{b_2}{\bar{\alpha}+\beta_1}\right) + a_1 b_1\left(\frac{1}{1-\alpha} + \frac{b_2}{\bar{\alpha}-1}\right) - \zeta \tag{A.115}$$

For a given α (or, equivalently, a g), this implicit equation provides a solution for the corresponding g given all of the g and α dependent constants in (A.103), (A.108) to (A.110) and (A.112) to (A.114).

Parameter Restrictions A set of parameter restrictions are required to ensure that $F'(z) > 0$ and that the eigenvalues of both B and C are strictly positive to ensure a non-explosive root.

To ensure that the PDF is positive at every point in (A.107), note that since $\bar{\alpha} > \alpha$, the term $0 < \frac{\lambda_\ell+\lambda_h-(\gamma-g)\bar{\alpha}}{\lambda_\ell+\lambda_h-(\gamma-g)\alpha} < 1$ is positive as long as $\lambda_\ell+\lambda_h-(\gamma-g)\bar{\alpha} > 0$. Similarly, for $F'_\ell(z) > 0$ in (A.106), a requirement is that $0 < \frac{\lambda_h-(\gamma-g)\bar{\alpha}}{\lambda_h-(\gamma-g)\alpha} < 1$ and $\lambda_h-(\gamma-g)\bar{\alpha} > 0$, which reduce to

$$g > \gamma - \frac{\lambda_\ell+\lambda_h}{\alpha} \tag{A.116}$$

Note that the thicker the tail (i.e. smaller α), the greater the range of possible g to match the α.

For (A.81) to be well defined as $z \to \infty$, we have to impose parameter restrictions that constrain the growth rate g so that the eigenvalues of C are positive or have positive real parts. Note that,

$$\det\{C\} = \frac{\lambda_h + \lambda_\ell}{\gamma - g} f_\ell(0) > 0 \tag{A.117}$$

And,

$$\text{Trace}\{C\} = f_\ell(0) + \frac{\lambda_h}{\gamma - g} - \frac{\lambda_\ell}{g} \tag{A.118}$$

Hence, since the determinant is positive, a necessary condition for C to have two positive eigenvalues is for

$$f_\ell(0) > \frac{\lambda_\ell}{g} - \frac{\lambda_h}{\gamma - g} \tag{A.119}$$

From (A.82) we see that $\vec{F}'(0) = D$ and $f_h(0) = 0$. If C does have two positive eigenvalues, then from (A.81), $\vec{F}(\infty) = \frac{1}{\lambda_\ell + \lambda_h}\begin{bmatrix}\lambda_h & \lambda_\ell\end{bmatrix}$, which fulfills $F(\infty) = 1$.

With the new constants, the condition in (A.119) to ensure positive eigenvalues becomes,

$$\alpha < \sqrt{\frac{(\lambda_h + \lambda_l)\left(g\left(\lambda_h + \lambda_l\right) - \gamma\lambda_l\right)}{g(g-\gamma)^2}} \tag{A.120}$$

Furthermore, as α has been constructed to be the smallest eigenvalue, and hence is the tail parameter, a necessary condition to have a finite mean is that $\alpha > 1$. From this, the g in equilibrium must fulfill,

$$2g\left(\lambda_h + \lambda_l\right) > \sqrt{4g^2(g-\gamma)^2 + \gamma^2\lambda_l^2} + \gamma\lambda_l \tag{A.121}$$

To ensure that the eigenvalues of B are positive, and hence $v_i(z)$ is well-defined, note that the determinant of B is always positive, and the trace of B requires $g < \frac{\gamma(r+\lambda_\ell)}{\gamma - \lambda_h + \lambda_\ell}$. However, in practice this upper-bound is greater than γ.

To summarize: the key conditions parameter restrictions are, (A.116), (A.120) and (A.121)

□

As an example of this equilibrium, with the same parameters as Section 3.3, Figure 18 plots the equilibrium growth rate as a function of α. For smaller α values, the growth rate approaches $\gamma = 0.02$, and matches it exactly at the calibrated $g = \gamma$ case.

A.7 No Equilibrium Exist with $g < \gamma$ and Jumps

Proof of Uniqueness of Proposition 4. In order to show that there are no stationary equilibrium with $g < \gamma$ and $\eta > 0$, we will: (1) assume there is some constant $g < \gamma$ which is the consumer's optimal choice (i.e. balancing adoption costs and benefits); (2) use the constant g to calculate the expectation of the stationary productivity distribution; (3) show that the mean cannot be stationary 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), add in the jumps as in (67) and (68). We maintain the assumption of a constant g and S with

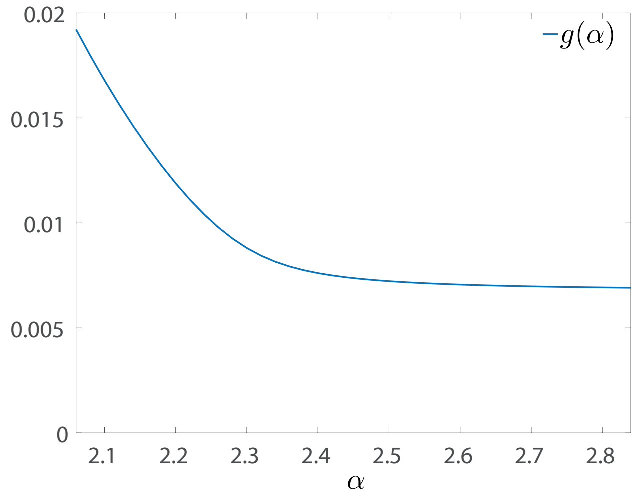

Figure 18: Exogenous $g < \gamma$ examples, for various $\alpha > 1$

draws from the unconditional $F(t, \cdot)$. Use that the derivative of the Heaviside is the dirac-delta to find,

$$\boldsymbol{\partial}_t f_\ell(t,z) = g\boldsymbol{\partial}_z f_\ell(t,z) + (S - \lambda_\ell - \eta) f_\ell(t,z) + (S + \lambda_h) f_h(t,z) - S\boldsymbol{\delta}(z) + \eta\boldsymbol{\delta}(z - \bar{z}) \tag{A.122}$$

$$\boldsymbol{\partial}_t f_h(t,z) = (g-\gamma)\boldsymbol{\partial}_z f_h(t,z) + \lambda_\ell f_\ell(t,z) - (\lambda_h + \eta) f_h(t,z) \tag{A.123}$$

Here, after differentiating, we have the dirac-delta, $\boldsymbol{\delta}(z)$, in (A.122) to remove those adopting agents at the threshold (which was normalized to $z = 0$), and the insertion of η arrival rate of agents at $\bar{z}$.

Taking inspiration from Gabaix, Lasry, Lions, and Moll (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.122) and (A.123) gives,[52]

$$\boldsymbol{\partial}_t \mathcal{F}_\ell(t,\xi) = g\xi\mathcal{F}_\ell(t,\xi) + (S - \lambda_\ell - \eta)\mathcal{F}_\ell(t,\xi) + (S + \lambda_h)\mathcal{F}_h(t,\xi) - S + \eta e^{-\bar{z}\xi} \tag{A.124}$$

$$\boldsymbol{\partial}_t \mathcal{F}_h(t,\xi) = (g-\gamma)\xi\mathcal{F}_h(t,\xi) + \lambda_\ell\mathcal{F}_\ell(t,\xi) - (\eta + \lambda_h)\mathcal{F}_h(t,\xi) \tag{A.125}$$

From Gabaix, Lasry, Lions, and Moll (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.124) and (A.125) to find,

$$0 = -g\mathcal{F}_\ell(t,-1) + (S - \lambda_\ell - \eta)\mathcal{F}_\ell(t,-1) + (S + \lambda_h)\mathcal{F}_h(t,-1) - S + \eta e^{\bar{z}} \tag{A.126}$$

$$0 = -(g-\gamma)\mathcal{F}_h(t,-1) + \lambda_\ell\mathcal{F}_\ell(t,-1) - (\eta + \lambda_h)\mathcal{F}_h(t,-1) \tag{A.127}$$

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 $\mathcal{F}(t,-1) = \mathcal{F}_\ell(t,-1) + \mathcal{F}_h(t,-1)$,

$$\mathbb{E}_t\left[Z/M(t)\right] = \mathcal{F}(t,-1) = \frac{g - \gamma + \eta + \lambda_h + \lambda_\ell}{(g - S + \eta)(g - \gamma + \eta + \lambda_h) - \lambda_l(S - g + \gamma - \eta)}\left(\eta e^{\bar{z}} - S\right) \tag{A.128}$$

Since $g < \gamma$, from (A.99), $\lim_{t\to\infty} \bar{z}(t) = \infty$. Therefore, (A.128) 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, $\zeta M(t)$. Furthermore, since $v_\ell(z) > z$, so (47) cannot hold with equality, and the $M(t)$ leading to the g cannot have been optimal. □

[52] 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: $\mathcal{L}\{f'(z)\} = \xi\mathcal{F}(\xi)$. The other important formula is that $\mathcal{L}\{\boldsymbol{\delta}(z-c)\} = e^{-c\xi}$, and $\mathcal{L}\{\boldsymbol{\delta}(z)\} = 1$

Appendix B Endogenous Markov Innovation

Proof of Propositions 5 and 6 . Note that Section 4.2 nests Section 4.1 when $\eta = 0$.

Nested Derivation of Stationary HBJE To create a stationary solution for the value function define a change of variables,[53]

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

From (30) and (31),

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

Differentiate (B.1) and reorganize ,

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

Assuming an interior solution, take the first order necessary condition of the Hamilton-Jacobi-Bellman equation in (87), and reorganize

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

Substitute this back into (87) to get a non-linear ODE,

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

Alternatively, (87) and (B.4) could be kept separate to form a differential-algebraic equation (DAE), which can be more numerically stable. Differentiate (86),

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

As before, for simplicity, assume that if $\psi < 1$, then $\kappa = 1$. Multiply (B.6) by e^{-z} and use (100), (B.1) and (B.3).[54]

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

Note that using (B.3),

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

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

Differentiate (B.5), multiply by e^{-z}, and use (100), (B.1), (B.3) and (B.9)

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

From (B.4),

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

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

[53] 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. An alternative approach to solving the HJBE numerically might be to use up/downwind methods as in Achdou, Lasry, Lions, and Moll (2014).

Define the integrated marginal utility from 0 to z as,

$$\hat{w}_i(z) \equiv \int_0^z v_i'(\hat{z})\mathrm{d}\hat{z} = \int_0^z e^{\hat{z}} w_i(\hat{z})\mathrm{d}\hat{z} \tag{B.13}$$

Integrate (B.1) with the initial value from (88) and use (B.13) to get,

$$v_i(z) = v(0) + \hat{w}_i(z) \tag{B.14}$$

Substitute (99) and (B.14) into (A.29) and rearrange to get an expression for $v(0)$ in terms of $\hat{w}_\ell$ and intrinsics,

$$v(0) = \frac{1 + \eta v_\ell(\bar{z})}{r - g + \eta} = \frac{1 + \eta \hat{w}_\ell(\bar{z})}{r - g} \tag{B.15}$$

Value Matching with Endogenous choice of θ and κ: A change to $w_i(z)$ space will also be useful for simplifying integrals. Note that,[55]

$$\int_0^{\bar{z}} v_\ell(z)\mathrm{d}F(z)^\kappa = v(0) + \int_0^{\bar{z}} e^z w_\ell(z)\left(1 - F(z)^\kappa\right)\mathrm{d}z \tag{B.17}$$

And expanding when $\bar{z} < \infty$,

$$\int_0^{\bar{z}} v_\ell(z)\mathrm{d}F(z)^\kappa = v_\ell(\bar{z}) - \int_0^{\bar{z}} e^z w_\ell(z) F(z)^\kappa \mathrm{d}z \tag{B.18}$$

Take the value-matching condition for the choice of the idiosyncratic $\hat{\theta}$ and $\hat{\kappa}$ given equilibrium θ and κ choices of the other firms.

$$v(0) = \max_{\hat{\theta}\geq 0,\, \hat{\kappa}>0} \left\{(1-\hat{\theta})\int_0^{\bar{z}} v_\ell(z)\mathrm{d}F(z)^{\hat{\kappa}} + \hat{\theta} v_\ell(\bar{z}) - \frac{1}{\psi}\left(\zeta + \tfrac{1}{\vartheta}\hat{\theta}^2 + \tfrac{1}{\varsigma}\hat{\kappa}^2\right)\right\} \tag{B.19}$$

Use (B.14) and (B.17)

$$v(0) = \max_{\hat{\theta}\geq 0,\, \hat{\kappa}>0} \left\{(1-\hat{\theta})\left(v(0) + \int_0^{\bar{z}} e^z w_\ell(z)(1 - F(z)^{\hat{\kappa}})\mathrm{d}z\right) + \hat{\theta}\left(v(0) + \hat{w}_\ell(\bar{z})\right) - \frac{1}{\psi}\left(\zeta + \tfrac{1}{\vartheta}\hat{\theta}^2 + \tfrac{1}{\varsigma}\hat{\kappa}^2\right)\right\} \tag{B.20}$$

Simplify,

$$0 = \max_{\hat{\theta}\geq 0,\, \hat{\kappa}>0} \left\{(1-\hat{\theta})\int_0^{\bar{z}} e^z w_\ell(z)(1 - F(z)^{\hat{\kappa}})\mathrm{d}z + \hat{\theta}\hat{w}_\ell(\bar{z}) - \frac{1}{\psi}\left(\zeta + \tfrac{1}{\vartheta}\hat{\theta}^2 + \tfrac{1}{\varsigma}\hat{\kappa}^2\right)\right\} \tag{B.21}$$

In the case of $\theta = 0$ this simplifies to,

$$0 = \max_{\hat{\kappa}>0} \left\{\int_0^{\bar{z}} e^z w_\ell(z)(1 - F(z)^{\hat{\kappa}})\mathrm{d}z - \frac{1}{\psi}\left(\zeta + \tfrac{1}{\varsigma}\hat{\kappa}^2\right)\right\} \tag{B.22}$$

[54] The general form of $\kappa \neq 1$ and $\psi < 1$ is covered in (98). We are avoiding this due to numerical difficulties rather than anything intrinsic in the model.

[55] These come out of using integration by parts on the calculation of the expectation. For example, if $F(z)$ is the CDF for a random variable Z with minimum and maximum support $\underline{z}$ and $\bar{z}$, then the following holds for any reasonable $h(z)$,

$$\mathbb{E}\left[h(Z)\right] = \int_{\underline{z}}^{\bar{z}} h'(z)(1 - F(z))\mathrm{d}z + h(\underline{z}) \tag{B.16}$$

In the case of $\theta > 0$, (B.21) simplifies to,

$$0 = \max_{\hat{\theta}\geq 0,\, \hat{\kappa}>0} \left\{ \hat{w}_\ell(\bar{z}) - (1-\hat{\theta}) \int_0^{\bar{z}} e^z w_\ell(z) F(z)^{\hat{\kappa}} \mathrm{d}z - \frac{1}{\psi}\left(\zeta + \tfrac{1}{\vartheta}\hat{\theta}^2 + \tfrac{1}{\varsigma}\hat{\kappa}^2\right) \right\} \tag{B.23}$$

Crucially, if the firm chooses a $\hat{\theta} \neq \theta$, they are infinitesimal and have no influence on the value or equilibrium distributions. Take the first order condition of (B.23) with respect to $\hat{\theta}$ and then let $\hat{\theta} = \theta$ in equilibrium,

$$\theta = \frac{\psi\vartheta}{2} \int_0^{\bar{z}} e^z w_\ell(z) F(z)^\kappa \mathrm{d}z \tag{B.24}$$

Take the first order condition of (B.21) and equate $\hat{\kappa} = \kappa$ in the economy. Note that $F(z)^\kappa = \exp(\kappa \log F(z))$, so that $\boldsymbol{\partial}_\kappa F(z)^\kappa = \log(F(z))F(z)^\kappa$ and assume conditions to differentiate under the integral

$$\kappa = \frac{-\varsigma\psi(1-\theta)}{2} \int_0^{\bar{z}} e^z w_\ell(z) \log(F(z)) F(z)^\kappa \mathrm{d}z \tag{B.25}$$

This is an implicit equation in κ. Note that as $0 < F(z) < 1$ and $\log(F(z)) < 0$, the sign of this term is correct to ensure a positive κ.

KFE and Value Matching From (89), for $z < \bar{z}$ the KFE is,

$$0 = gF_\ell'(z) + \lambda_h F_h(z) - \lambda_\ell F_\ell(z) - \eta F_\ell(z) + (1-\theta)(S_\ell + S_h)F(z)^\kappa - S_\ell,\ z < \bar{z} \tag{B.26}$$

In the limit as $z \to \bar{z}$, we know both $\gamma(z)$ and $F(z)$ are continuous. Assume that $\lim_{z\to\bar{z}} F_h'(z) < \infty$ and use $g - \gamma(\bar{z}) = 0$ in (24) to (26) and (90) to get the system of equations,[56]

$$0 = (\lambda_h + \eta)F_h(\bar{z}) - \lambda_\ell F_\ell(\bar{z}) + gF_h'(0) \tag{B.27}$$
$$1 = F_\ell(\bar{z}) + F_h(\bar{z}) \tag{B.28}$$

Solve to find a boundary condition for $F(\bar{z})$ for a given $F'(0)$

$$F_\ell(\bar{z}) = \frac{1}{\lambda_\ell + \lambda_h + \eta}\left(gF_h'(0) + \eta + \lambda_h\right) \tag{B.29}$$
$$F_h(\bar{z}) = \frac{1}{\lambda_\ell + \lambda_h + \eta}\left(-gF_h'(0) + \lambda_\ell\right) \tag{B.30}$$

From (B.30) , for $F_h(\bar{z}) < 1$, it must be that $F_h'(0) < \lambda_\ell/g$, which provides a bound for possible guesses.[57]

[56] The KFE for h in (89) is co-linear with (B.26) at $\bar{z}$, and hence wouldn't provide an additional equation.

[57] Using (B.29) gives the same equation due to collinearity. An alternative approach is to rearrange (B.30) and get the $F_h'(0)$ given the particular $F_h(\bar{z})$ guess,

$$F_h'(0) = \frac{\lambda_\ell - (\eta + \lambda_h + \lambda_\ell)F_h(\bar{z})}{g} \tag{B.31}$$

Upper bound on g: For the unbounded case where $\eta = \theta = 0$, and $\bar{z} \to \infty$, we can check the asymptotic value comes from (B.1)

$$\lim_{z\to\infty} w_i(z) = c_i \tag{B.32}$$

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, $\lim_{z\to\infty} w_i'(z) = 0$ as $w_i(z)$ have been constructed to be stationary. Furthermore, note that the maximum g from (B.12) is,

$$g = \lim_{\bar{z}\to\infty} \tfrac{\chi}{2} w_h(\bar{z}) = \tfrac{\chi}{2} c_h \tag{B.33}$$

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

$$(r + \lambda_\ell + \eta - (1-\psi)\tfrac{\chi}{2} c_h F'(0))c_\ell = 1 + \lambda_\ell c_h \tag{B.34}$$

$$(r + \lambda_h + \eta)c_h = 1 + \left(\lambda_h + (1-\psi)\tfrac{\chi}{2} c_h F'(0)\right) c_\ell + \tfrac{\chi}{4} c_h^2 \tag{B.35}$$

Given a $F'(0)$, (B.34) and (B.35) provide a quadratic system of equations c_l and c_h—and ultimately g through (B.33). While analytically tractable given an $F'(0)$, this quadratic has a complicated solution—except if $\psi = 0$. For that case, define

$$\bar{\lambda} \equiv \frac{r + \eta + \lambda_\ell + \lambda_h}{r + \eta + \lambda_\ell} \tag{B.36}$$

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

$$g < \bar{\lambda}(r+\eta)\left[1 - \sqrt{1 - \frac{\chi}{\bar{\lambda}(r+\eta)^2}}\right] \tag{B.37}$$

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

$$g = \bar{\lambda} r \left[1 - \sqrt{1 - \frac{\chi}{\bar{\lambda} r^2}}\right] \tag{B.38}$$

where a necessary condition for an interior equilibrium is

$$r > \sqrt{\frac{\chi}{\bar{\lambda}}} \tag{B.39}$$

□

Summarizing the full set of equations to solve for $F_i(z)$ and $w_i(z)$ from (23) to (26), (90), (B.2), (B.7), (B.10) to (B.12), (B.14), (B.15), (B.22) to (B.26) and (B.36).

Appendix C Further Discussion and Analysis

This appendix collects further analysis of the model, including the role of hysteresis and the empirical plausibility of endogenous innovation.

Support	Exogenous $\gamma \implies g$ Hysteresis?	Endogenous $\gamma(\bar{z}) \implies g$ and $\gamma(\cdot)$ Hysteresis?
Infinite	Continuum of $g \geq \gamma$ and $g < \gamma$	Unique $\gamma(\infty)$ with $g \geq \gamma(\infty)$ and $g < \gamma(\infty)$
Finite Unbounded	Continuum of $g \leq \gamma$	Unique $\gamma(\bar{z})$ with $g \leq \gamma(\infty)$
Finite Bounded	Unique $g = \gamma$ and corresponding $\bar{z}$	Continuum of $\gamma(\bar{z})$ and corresponding $\bar{z}$, with unique $g = \gamma(\bar{z})$

Table 3: Summary of Hysteresis and Uniqueness

C.1 Summary of Hysteresis and Multiplicity

The results of uniqueness of stationary equilibria are summarized in Table 3. The table compares models with an infinite-support coming from initial conditions, to those with unbounded vs. bounded finite-support (i.e., models in Sections 3.1 to 3.3 for the exogenous γ, and in Sections 4.1 and 4.2 for the endogenous γ).

There are several possible sources of multiplicity in the stationary equilibrium—all of which are due to initial condition dependence (i.e., hysteresis). The first is that, when the innovation rate $\gamma(z)$ is chosen endogenously, that the growth rate of the frontier $\gamma(\bar{z})$ could be unique or could have a continuum of solutions depending on the initial conditions. The other source of multiplicity is that, given a γ (either exogenously or endogenously), there may be a continuum of aggregate growth rates, g, each with a corresponding stationary distribution $F(z)$ and relative frontier $\bar{z}$.

These sources of multiplicity interact, but may come different economic forces. The endogenous choice of innovation, γ, can give rise to a feedback where the particular productivity distribution induces different innovation incentives at the frontier. Alternatively, the endogenous choice of technology adoption can give rise to hysteresis in g, in the sense that for a particular γ, the productivity growth rates of the distribution as a whole can depend on the initial distribution of productivities.

Aggregate Growth Rate Multiplicity Where $g > \gamma$: The fatter the tail of the initial distribution, the richer will be the opportunities to adopt superior technologies, and therefore the overall economy-wide growth rate will be higher. In the limit, therefore, the stationary distribution may depend on the initial productivity distribution. With an initial condition with an infinite-support, if the adoption opportunities remain profitable, the limiting growth rate of the economy may forever exceed its growth rate from innovation alone. This is the case for all models considered in Technical Appendix E. It also can be the case in Section 2 if the initial distribution has infinite-support and bounded growth rates as in Proposition 1. By contrast, if the initial productivity distribution has finite-support, the growth rate of limiting distribution must be weakly less than the exogenous innovation rate, $g \leq \gamma$. As the growth above γ comes from outside of the model of innovation, we label it "latent growth".

Aggregate Growth Rate Multiplicity Where $g < \gamma$: Looking at the other direction, there will be initial conditions (with both infinite and finite unbounded support) where the incentives for technology adoption are relatively low, and the aggregate growth rate is unable to keep up with the innovation rate. In those cases, as analyzed in Section 3.2.1, the growth rate can be $g < \gamma$ if $\bar{z} = \infty$ or $\lim_{t\to\infty} \bar{z}(t) = \infty$. In terms of the adoption incentives: the mass of firms close to the technology frontier growing at rate γ is strictly positive, but asymptotically it disappears, so the growth rate of the distribution as a whole doesn't need to keep pace with the growth rate of the frontier. In that case, since the incentives for technology adoption are related to the mean (rather than the maximum) productivity, the growth rate of firms exactly at the frontier is not crucial, and the far tail can diverge in relative terms.

However, in the case of a finite bounded solution (i.e., the $\lim_{t\to\infty} \bar{z}(t) < \infty$ of Proposition 4), the upper tail of the distribution is kept from becoming asymptotically infinitesimal, and the growth rate, g, converges to the innovation rate γ. Economically, the presence of a small (but strictly positive) mass of firms arbitrarily close to the technology frontier induces the technology adoption rate to keep pace. Otherwise, the benefits of technology adoption would diverge relative to the cost. See the proof in Appendix A.7 for an explanation. We also label this sort of multiplicity as latent growth, but recognize that the sign is negative—i.e., initial conditions are dragging the aggregate growth rate to below the innovation rate.

Multiplicity of Both the Aggregate Growth Rate and the Innovation Rate: When innovation is chosen by firms in Sections 4.1 and 4.2, a new source of multiplicity can emerge. In the unbounded case, the innovation rate is unique, since the incentives for innovation are independent of the incentives for technology adoption as $z \to \infty$. Ultimately, the the option value $\to 0$, and hence there is no feedback from the distribution to the innovation rate of firms with large z, which leads to uniqueness. The aggregate growth rate, g, in that case may still have multiplicity for the same reasons it would in an exogenous innovation setup.

But when jumps are added to the distribution, and it becomes bounded, a new sort of hysteresis enters: multiplicity of the aggregate innovation rate. Unlike the case of $\bar{z} \to \infty$, when the frontier is bounded, the option value of technology diffusion at the frontier as strictly positive. This creates a feedback between the shape of the distribution (and the accompanying $\bar{z} < \infty$) and the choice of $\gamma(\bar{z})$. The stationary distributions can then be parametrized by $\bar{z}$, the ratio of frontier productivity to the mean of the distribution (Proposition 6).[58] In this case the support and the position of the stationary distribution, as well as the growth rate, are parametrically determined by $\bar{z}$.

If the distribution started off very compact, with a relatively small $\bar{z}$, then the option value of technology diffusion at the frontier is relatively large, leading to less investment in innovation. On the other hand, if the distribution started off with a very large $\bar{z}$, then the growth rate would converge to one with a higher innovation rate that maintains the more dispersed distribution. In all cases, the bounded distribution leads to $g = \gamma$ on a BGP.

C.2 Discussion of Firm Dynamics: Growth Rates Conditional on Size

As the intensity of innovation, $\gamma(z)$, is increasing in z, the larger and more productive firms do the most innovation. While the economics and model are very different, this is related to Acemoglu, Aghion, and Zilibotti (2006) and Benhabib, Perla, and Tonetti (2014), who feature innovation rates weakly increasing in a firm's relative productivity. In our model, this is driven by the adoption option value, in that agents closer to the endogenous adoption threshold have less incentive to invest in incremental productivity enhancement and, accordingly, decrease their endogenous investment in $\gamma(z)$.

Because $\gamma(z)$ is increasing, the growth rate conditional on being a high type is also increasing in z. While this may appear to contradict Gibrat's law and some modern evidence on non-Gibrat's growth, as surveyed in Sutton (1997) and modeled in Luttmer (2007) and Arkolakis (2015), consider that: technology adoption is a key component of growth for small firms but is not measured by γ; this model does not have endogenous exit, which is important for reconciling growth rates of small firms; and the growth process is not an iid random walk, but has auto-correlation due to the Markov chain.

The first consideration is that smaller firms at the adoption threshold are growing rapidly. Therefore, the model does have small firms tending to grow faster than larger firms. Here, we

[58] Note that unlike Proposition 4 $\bar{z}$ can be jointly solved with $F_\ell'(0)$, using (73) and (74). In Proposition 6, $\bar{z}$ can be chosen parametrically for various stationary distributions.

have simplified the model to ensure that only a single adoption barrier exists and that firms make immediate productivity jumps. With more frictions and heterogeneity leading to a continuum of adoption barriers, the average growth rates might be more empirically plausible, while the same economic forces present in this stark model would remain.

Second, many models investigating the empirics of Gibrat's law have emphasized that the higher growth rates for small firms are only conditional on survival. As small firms are more likely to exit, this implies that the average growth rate for smaller firms in the sample is higher. As we have purposely shut off exit in our model to focus on the interaction between adoption and innovation, this effect is not present. Davis, Haltiwanger, and Schuh (1996) find that when selection into exit and mean reversion in stochastic processes are taken into account, the inverse relationship between size and growth can disappear. However, Arkolakis (2015) discusses how the inverse relationship tends to still exist even after selection, and describes how the Davis, Haltiwanger, and Schuh (1996) adjustment does not apply to random walks. Finally, due to the Markov chain process for growth, there is auto-correlation of growth rates for firms. This is in contrast to growth being simply a random walk, and, hence, the Davis, Haltiwanger, and Schuh (1996) results may still apply.

The model presented is stripped down for expositional purposes to highlight the importance of finite frontiers and how adoption and innovation interact to generate productivity dispersion and aggregate growth. Extensions such as directed adoption (Section 5), which determines the conditional growth rate of adopters, may be necessary to better match firm dynamics in the panel data.

Appendix D Data and Calibration

This section documents our calibration strategy using firm-level data, and further empirics on the technology distribtuion

D.1 More Empirics of the Technology Frontier

The slight downward movement in the aggregated Figure 15 might have been caused by compositional changes. We should be cautious, however, due to the high volatility of these growth rates.

The question of stationarity is also important, as a non-stationary process would eventually diverge (even with a mean 0 growth rate since the log of the ratio can never go below 0). To answer that question, we will consider tests of stationarity, treating each SIC code as having its own time-series. The first test to use is Kwiatkowski, Phillips, Schmidt, and Shin (1992) (i.e., KPPS), which tests the null hypothesis of level-stationarity of the stochastic process. From the other side, we use Phillips and Perron (1988) (i.e. PPerron), which tests the null hypothesis of a unit root.[59] Using the separate tests with opposite null hypothesis directions helps us better distinguish the degree of certainty in our results.

See Figure 19 for a histogram of these test statistics across industries, and a further summary in Technical Appendix Table 2. The results are that that KPSS accepts stationarity about 75% of the time, while the Phillips-Perron rejects a unit root about 20% of the time.

The results, then, on stationarity are inconclusive. Analyzing both a finite-bounded, and a finite-unbounded frontier are important and may apply under different conditions, but leave for future research with better data a tighter sense of the conditions for divergence or stationarity of

[59] In both cases, the bandwidth (i.e., number of lags) is set to 4, which was calculated as the optimal bandwidth selected using a Newey and West-style procedure. The PPerron test was chosen rather than Dickey Fuller-style tests to be consistent with automatic-bandwidth selection of the KPSS test, and are more sensitive to the choice of lags).

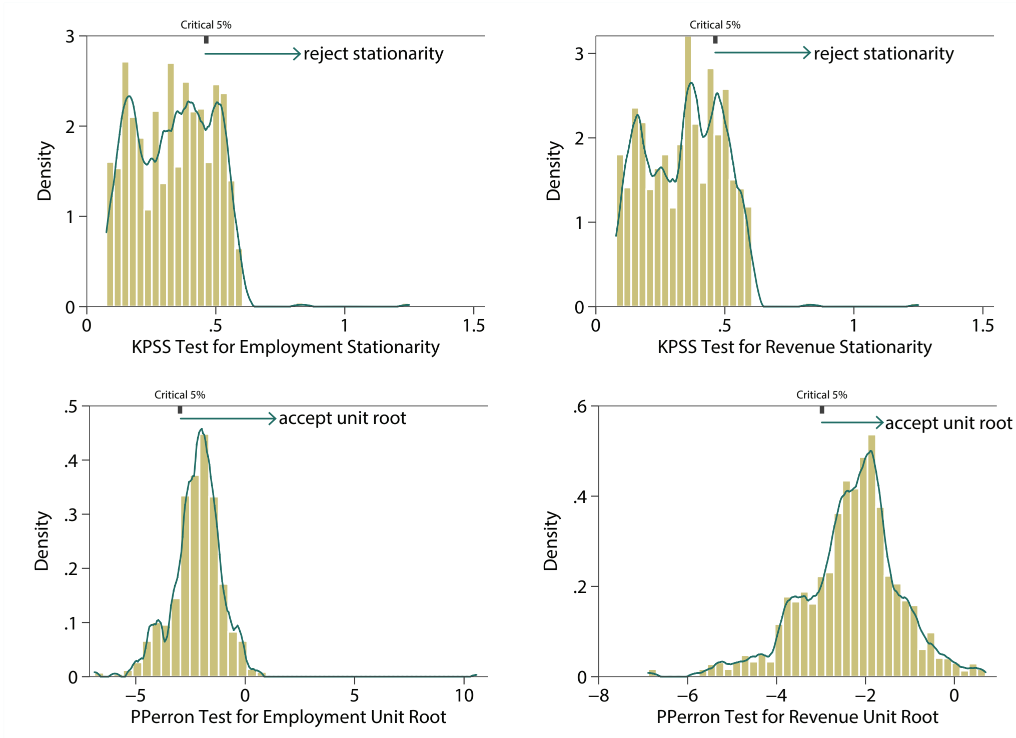

Figure 19: Histogram of Frontier Test Statistics (by SIC, 1990-2014)

the relative frontier.[60]

D.2 Transition Probabilities

The parameters for the transition rates are roughly calibrated from growth rates of firms in Compustat. While this data source has significant selection issues, this is less of an issue for calibrating growth rates and transitions (since our theory says that R&D based growth for incumbents is driven by the upper tail of the distribution).

The data comes from Compustat from 1971-2014, for all firms with primary SIC 2000-3999. We can use a longer-panel than the proxies for the frontier, since individual firm growth rates are not sensitive to the bias on the calculation of percentiles with small samples. The demeaned yearly growth rate in real revenue is calculated for each firm. Any firms with a growth rate $> .05$ are assigned the high type, and others are assigned the low type.

Note that the transition matrix P for t years is e^{tP}. From the data, at yearly frequency, the transition matrix of the changes in types is calculated and equated

$$\exp\left(1.0 \times \begin{bmatrix} -\lambda_\ell & \lambda_\ell \\ \lambda_h & -\lambda_h \end{bmatrix}\right) = \begin{bmatrix} 0.74 & 0.26 \\ 0.55 & 0.45 \end{bmatrix}$$

The solution to this system of equations is $\lambda_\ell = 0.533074$ and $\lambda_h = 1.12766$. While the transition probabilities themselves are sensitive to the thresholds for choosing ℓ vs. h types, this has little impact on the equilibrium itself.

[60]Another consideration is the role of regime-changes. Nearly all tests for stationarity are senstive to regime switches, and will bias towards rejection of stationarity in those cases. The aggregated time series in Figure 15 shows a likely regime shift around 1990, which led choosing the statistics in Figure 19 for data in 1990-2014. But it is possible that industries have structural breaks in the frontier at different times within the 1990-2014 sample, as would be uncovered by tests such as Perron and Vogelsang (1992).

D.3 Tail Parameters

When looking at a more realistic model with product differentiation, Technical Appendix (G.24) shows how higher markups lead to changes in the tail parameter of the size distribution, $\hat{\alpha}$, compared to the underlying productivity distribution. With the α generated by the model, Technical Appendix (G.30) shows a rough adjustment of $\alpha = (\varpi - 1)\hat{\alpha}$ is necessary to compare to the tail parameter in the data, where ϖ is the elasticity of substitution. If $\varpi = 3$, as in Perla, Tonetti, and Waugh (2015), then markups are 50%, and an $\alpha = 2.12$ corresponds to the $\hat{\alpha} = 1.06$ tail parameter in the size or profits distribution—as used in Luttmer (2007)

D.4 Bargaining Power

From Kemmerer and Lu (2012), the reported royalty rates (i.e., the proportion of profits a licensor must pay for the intellectual property) from a data source called RoyaltySource are heterogeneous between industries, ranging between roughly 4% and 13%, with a median royalty rate of 5%.

In our model, from (95), the surplus going to the licensor is $(1-\psi)(v_\ell(z)-v(0))$. If $v_\ell(z) \gg v(0)$, such as the adoption of a frontier technology, then this translates to a $\psi = .95$ to match the royalty rate. The effective royalty rates for technologies with lower $\bar{z}$ diminish towards 0 due to the role of the outside option.

As we suspect that royalty rates that are actually reported are skewed towards those with enforceable IP near the frontier, we use the $\psi = .95$ parameter and do a robustness check to test the sensitivity.

D.5 Frontier Productivity

The frontier productivity depends on the interpretation and selection by industry. Two approaches to consider:

TFP from Industry Studies Syverson (2011) surveys the evidence on productivity dispersion. Within the US, the ratio of the top to the bottom decile is approximately 1.92:1 . This provides a minimum bound on productivity differences, implying a $\bar{z}$ of at least 0.651. Within places such as China and India, Hsieh and Klenow (2009) finds the ratio is closer to 5:1, or a minimum $\bar{z}$ of 1.61.

One consideration is that this data tends to rely on fairly homogeneous manufacturing industries, where TFP can be easily compared. Hence, it is likely to significantly underestimate productivity dispersion for more differentiated industries such as the services sector. Another consideration is that adoption in our model isn't necessarily about remaining in narrowly defined industries, and perhaps revenue TFP should be pooled across industries when considering the dispersion.

Firm Size Distribution An alternative approach is to use the values calculated in Section 6. From Table 1, the proxy for $\bar{z} \approx \log(\text{90th percentile/10th percentile})$ is around 4.2 across 416 industries from 1980-2014. If all industries are pooled, as in Figure 15, then $\bar{z}$ is closer to 6.

D.6 Calibration Summary

The calibration and targets for our baseline model are documented in Table 4.

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
ψ	0.95	Matches median 5% royalties of large firms reported from RoyaltySource in Kemmerer and Lu (2012)
Λ	1	Log utility baseline
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 $\eta = 0$, 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.) See Syverson (2011) and Appendix D for discussion of bounds.
$\bar{\kappa}$	1	Simple baseline
$\bar{\theta}$	0	Simple baseline

Table 4: Summary of Calibration