Knowledge is (Market) Power

Job Market Paper

Jinglun Yao *
London Business School

February 25, 2024
Version 2.2.4 Click here for the latest version.

Abstract

US corporate concentration has been persistently rising over the past century and productivity growth has been concurrently declining. This paper builds a continuous-time Schumpeterian growth model in which a *uniform* decline in research efficiency increases the *relative* growth of leading firms compared to laggards and endogenously thickens the Pareto tail of firms' productivity distribution. With a demand system featuring realistic variable demand elasticities, the model explains a large part of the dynamics of firms' productivity, corporate concentration, markup, labor share, R&D cost, entry and exit rates, as well as job creation and destruction rates in the US since the 1980s. The model can also accommodate increasing concentration with a stable markup and labor share in the pre-1980 period by accounting for the role of economic integration.

Keywords: Schumpeterian Growth, Corporate Concentration, Market Power, Labor Share, Business Dynamism, Heterogeneous Firms, Mean Field Game, Pareto Tail

*Department of Economics, London Business School, Regent's Park, London NW1 4SA, United Kingdom. Email: jyao@london.edu. This paper could not have taken its current form without the continuous and committed support of Hélène Rey. I am indebted to Philippe Aghion, Nicolas Coeurdacier, Pierre-Louis Lions, Joseba Martinez and John Van Reenen for their insights and assistance during repeated discussions, and grateful to Ufuk Akcigit, Fernando Alvarez, David Argente, Martin Beraja, Florin Bilbiie, Johannes Boehm, Santiago Caicedo, Julieta Caunedo, Maarten de Ridder, Gustavo de Souza, Miguel Faria-e-Castro, Simon Gilchrist, Basile Grassi, Chenlin Gu, Stéphane Guibaud, Wei Hua, Anders Humlum, Loukas Karabarbounis, Peter Klenow, Francis Kramarz, Howard Kung, Francesco Lippi, Isabelle Méjean, Virginia Minni, Johan Moen-Vorum, Elias Papaioannou, Christina Patterson, Stefan Pollinger, Marta Prato, Lucrezia Reichlin, Pascual Restrepo, Natalie Rickard, Andrew J Scott, Petr Sedláček, Yongseok Shin, Paolo Surico, Maria Micaela Sviatschi, Stephen J Terry, Nizar Touzi, Conor Walsh and seminar participants at Sciences Po, Collège de France/INSEAD, LBS, TADC 2023, EEA-ESEM Congress 2023, LUISS Workshop on Mean Field Games in Economics, EWMES 2023, Peking (NSD), CUHK (Shenzhen), PHBS, HKU, NUS, HEC Paris, IIES and Nottingham for sharing their insights and valuable suggestions. I thank the UniCredit Foundation and the European Economic Association for the Econ Job Market Best Paper Award. The paper was previously presented under the title "A Tale of Sunset Industries". All errors are my own.

1 Introduction

US corporate concentration has been rising since at least the 1930s (Figure 1a, Kwon et al. (2022)). While many explanations have been proposed to account for this phenomenon since the 1980s, its strikingly long-run nature remains a puzzle. Moreover, rising concentration takes a specific form of fattening Pareto tail as documented in Kwon et al. (2022) and Chen (2022): the share of top 0.1% firms within top 1% firms, as well as top 1% firms within top 10% firms, has increased over time. That is, mega firms have pulled ahead relative to large firms. Understanding the rise of concentration is not only important for its own sake, but may also shed light on other major issues, such as the rise of market power and the decline of labor share since the 1980s.[1] Such an endeavor is, however, challenging: while we observe an increase in corporate concentration since the 1930s, markups were roughly stable until the 1980s and have increased since then in US data.[2] Corroborating the stable market power before the 1980s is the stable labor share known as one of Kaldor's facts (Kaldor (1961)). How can we reconcile the fattening tail of firm size distribution with the stable market power until the 1980s?

This paper argues that declining research productivity, which has been identified as a widespread feature across scientific fields or industrial classifications,[3] can explain all these facts. According to Gordon (2016) and Nordhaus (2021), US Total Factor Productivity (TFP) growth peaked in the 1930s and 1940s with an annual rate of around 2.5%, then subsequently declined gradually to near 0 today (Figure 1b). Against the backdrop of declining TFP growth is not a decline in research input but rather a tremendous *increase*: since the 1930s, research effort has risen by a factor of 23, an average annual growth rate of 4.3%. Following the semi-endogenous growth literature such as Jones (1995), Bloom et al. (2020) defines research productivity as the ratio of TFP growth to R&D input, allowing for possible decreasing returns to scale in the production function for ideas. Research productivity has fallen by a factor of 19 since the 1930s with an annual growth rate of -3.67% (Figure 1b).[4] As Bloom et al. (2020) puts it, "ideas are getting harder to find".

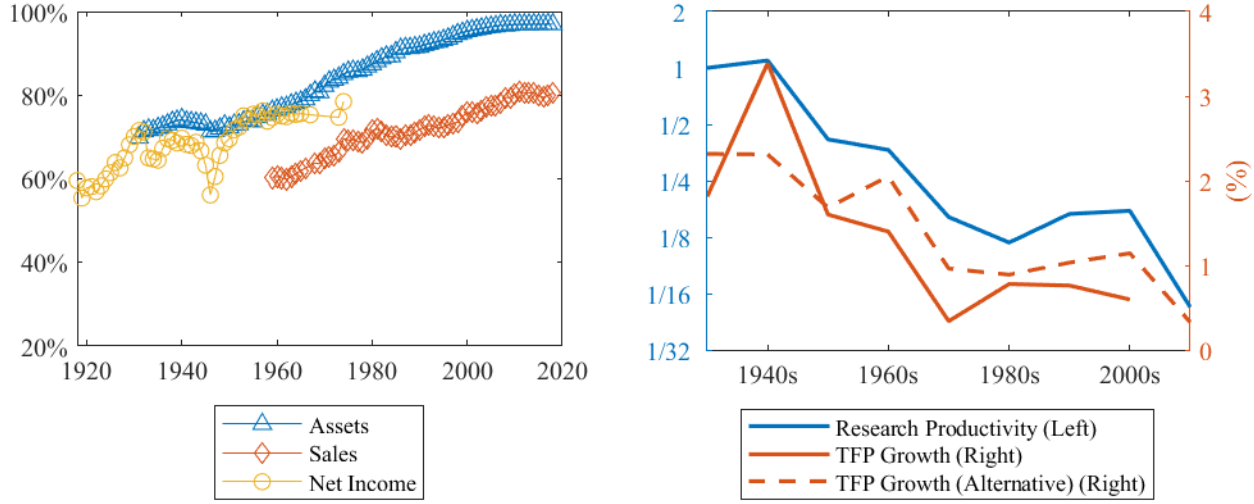
To establish the link between increasing research difficulty and rising concentration, I build a continuous-time Schumpeterian growth model which endogenizes the growth decisions of individual firms and generates a productivity distribution with a Pareto tail. In the model, each firm can

[1]See Karabarbounis and Neiman (2014), Autor et al. (2020) and Kehrig and Vincent (2021) for the labor share, and De Loecker et al. (2020) and Edmond et al. (2023) for markups. If demand elasticity decreases with firm size, as assumed in standard macroeconomic models with heterogeneous markups (e.g. Atkeson and Burstein (2008) in an oligopolistic competition and Melitz and Ottaviano (2008) in a monopolistic competition), then understanding higher concentration paves the way for understanding higher market power.

[2]See Edmond et al. (2023) for cost-weighted markup and De Loecker et al. (2020) for sales-weighted markup between 1950 and 1980.

[3]See Gordon (2016), Bloom et al. (2020) and Park et al. (2023).

[4]To be consistent with my model, I assume decreasing returns to scale in the idea production function with 0.5 elasticity of growth with respect to research input. In the baseline setting of Bloom et al. (2020) with constant returns to scale of the idea production function, research productivity decreases by a factor of 41 since the 1930s with an annual growth rate of -5.1%.



(a) Corporate Concentration

(b) Decadal TFP Growth and Research Productivity

Figure 1: US secular trends

Notes: Corporate concentration is the share of top 1% firms in terms of assets, receipts (sales) and net income from Kwon et al. (2022). TFP growth comes from Gordon (2016) and the alternative measure from Nordhaus (2021). Research productivity is measured using the methodology of Bloom et al. (2020) with decreasing returns to the idea production function.

conduct step-by-step innovation on a stand-alone basis to improve its productivity.[5] According to firm-level evidence in Bloom et al. (2020), firms find it more difficult to improve upon more advanced technology. Thus in a cross section of firms, technological laggards grow more quickly than leaders. Over time, research becomes more difficult *uniformly* across all firms: knowledge becomes more difficult with economic growth, as "standing on the shoulders of giants" is hard (Jones (2009)). This first of all decreases the growth of all firms so that the aggregate growth rate declines. Second, the higher growth of laggards and their reliance on growth to catch up with leaders implies that they are particularly hurt by harder research. This increases the *relative growth* of leaders versus laggards and allows the former to stretch further into the right tail, resulting in a fatter tail of productivity distribution and higher concentration. Under the assumptions of the model, I demonstrate that there is a one-to-one correspondence between TFP growth rate and the Pareto tail index of firm size distribution, so that the secular rise of corporate concentration manifests the secular decline of TFP growth in the data. Moreover, even though growth is *negatively* correlated with the *level* of concentration, it is *positively* correlated with the *speed of concentration increase*. The latter is another empirical regularity documented by Kwon et al. (2022). In the model, higher TFP growth translates into a faster increase in research difficulty during a fixed period of time following the semi-endogenous growth logic. Consequently, concentration increases faster at higher levels of growth.

[5]In an extended model, a firm can also learn from more productive firms. See Section 8.

It should be emphasized that the negative correlation between growth and concentration level should *not* be understood as higher concentration causing lower growth, which is the emphasis of the existing Schumpeterian growth literature (e.g. Aghion et al. (2005), Akcigit and Ates (2021)), but rather as lower growth generating higher concentration. Both directions are present in my model, as it describes jointly firms' growth decisions based on profit incentives and hence concentration, with the endogenous determination of the market structure based on growth.[6] I use the newly developed Mean Field Game methodology to solve the model and derive the productivity distributions of firms. In the model, higher research difficulty reduces growth and hurts the dynamic advantage of laggards, consequently increasing concentration. Instead of reducing growth, higher concentration partly mitigates the impact of harder research and encourages growth by increasing the Schumpeterian profits of innovation. Compared to the existing Schumpeterian literature, the emphasis is now shifted from how market structure determines growth to how growth determines market structure.

In addition to the productivity improvements discussed above, the model introduces idiosyncratic productivity shocks (i.e. random growth) in order to generate Pareto-tailed distributions, like many existing papers (e.g. Gabaix (1999), Gabaix et al. (2016)).[7] There is entry and exit of firms. Potential entrants enter the market by learning imperfectly from incumbents: when an entrant is matched with an incumbent and learns from the latter, the entrant can jump to the productivity level of the incumbent *but also* faces a positive probability of moving to some productivity below it. Learning is imperfect as the probability of jumping to the incumbent is less than 1. The specification can be seen as the continuous counterpart of the discrete case in König et al. (2016). I take advantage of tools in the continuous setting to prove that the tail of the equilibrium distribution is uniquely determined as long as learning is imperfect. Thus the model links growth to equilibrium productivity distributions and concentration, *without resorting to any additional assumption on the initial productivity distribution*.[8]

To study the implications of higher concentration on markups and labor share, I design a demand system featuring variable demand elasticities and assume that firms compete monopolistically with each other. The demand system is a special specification of the Kimball (1995) preferences and can be seen as a counterpart in monopolistic competition of nested CES in oligopolistic competition (Atkeson and Burstein (2008)). It retains key features of nested CES regarding demand elasticities, pass-throughs and firm-size distributions, while eliminating the strategic interactions of oligopolistic

[6] Technically, the model couples a Hamilton-Jacobi-Bellman (HJB) equation with a Kolomogorov Forward (KF) equation.

[7] Another commonly-used way of generating the Pareto tail is assuming an initial distribution with Pareto tail and perfect learning to perpetuate that tail, e.g. Lucas and Moll (2014).

[8] Luttmer (2012) and Perla et al. (2021) assume the initial productivity distribution to be sufficiently light-tailed in order to uniquely pin down the equilibrium Pareto tail. See Section 4.2.3 on why the assumption is needed in their models and why imperfect learning makes such an assumption unnecessary.

competition. This is necessary for the Mean Field Game methodology, as each firm is infinitesimal in the case of monopolistic competition and does not exert a noticeable impact on the aggregate economy on its own. Like a typical macroeconomic model in a static setting, more productive firms are also larger and enjoy a higher market power. The fatter tail of the productivity distribution thus reallocates market share towards firms with higher markups. Both the increase in markups and the decrease in labor share mainly come from the between-firm reallocation component rather than the within-firm component in an Olley-Pakes or Melitz-Palanec decomposition, as is consistent with the empirical evidence (De Loecker et al. (2020), Autor et al. (2020)). The between-firm component, however, can be largely compensated by the within-firm component if additional forces other than harder research are at play. Before the 1980s, basic infrastructure projects such as highways, airline facilities, etc. connecting different regions of the US (Gordon (2016)), as well as deregulations such as the 1978 Airline Deregulation Act which removed market access restriction across regions, motivates us to consider economic integration as another first-order economic force during the pre-1980 period.[9] Economic integration reduces markups and increases labor share through the within-firm component as it brings more firms into competition with each other.[10] The reallocation component from harder research is thus largely compensated by the within-firm component of economic integration, and aggregate markup and labor share can remain roughly stable while concentration increases with a fatter tail.[11] The model can thus make sense of the pre-1980 evidence.

Decker et al. (2020) has shown that declining job reallocation rates since the 1980s are due to firms' reduced sensitivity to idiosyncratic shocks and not the reduced volatility of these shocks. The model is consistent with this evidence, as higher market power is a manifestation of lower demand elasticity, and with lower elasticity the same idiosyncratic productivity shock translates into lower job creation or job destruction. The exit rate declines for similar reasons due to reduced reaction to productivity shocks, while the entry rate declines as research becomes more difficult. The model is thus also consistent with diminished business dynamics.

Using sectoral-level data from the US since 1987, I find empirical evidence for the mechanism of the model which links research productivity, market structure and business dynamism. Research productivity, measured as TFP growth over R&D cost, is significantly negatively correlated with the

[9] Other examples of rising concentration but stable markup and labor share include European countries in the last few decades. See Bighelli et al. (2023), Gutiérrez et al. (2022) and Bauer and Boussard (2020). Perhaps not coincidentally, this is the era during which European countries are being integrated into the European Union, which is a typical example of economic integration.

[10] The productivity distribution is Pareto at the right tail but log-concave in log productivity overall. According to Autor et al. (2020), within-firm reduction in markup dominates between-firm reallocation towards firms with higher markups when the number of firms increases. The net effect coming from static integration is thus a decrease in markup due to the within-firm component. I thank Pascual Restrepo for pointing this out.

[11] When the productivity distribution is exactly Pareto, the sole force of economic integration without any change in the productivity distribution can explain higher concentration with stable aggregate markup and labor share as in Melitz and Ottaviano (2008), but cannot explain why higher concentration takes the form of a fattening Pareto tail.

market concentration and profitability of the largest firms in the sector, and positively correlated with labor share and major indicators of business dynamism in the Business Dynamics Statistics (BDS): job creation rate by entrants, total job creation rate, job destruction rate by death, total job destruction rate, job reallocation rate, and the number of entrants/deaths over the total number of establishments. To quantify the model, I calibrate the demand system to match the heterogeneous labor shares across the establishment size distribution in the Census of Manufacturing. Other parameters are structurally estimated to match key moments around 1980. I calibrate a 6.84-fold decrease of research productivity for the 1980-2020 period using the methodology of Bloom et al. (2020) with aggregate US data, and a decrease in interest rate from 0.0469 to 0.0109 during the same period following Liu et al. (2022). Since the interest rate affects how a firm discounts future profits in making their growth decision, a lower discount rate encourages growth and counteracts some of the increase in research difficulty. This effect must be taken into account in order to evaluate the model against historical evidence. Moreover, in a partial equilibrium, harder innovation leads to lower R&D expenditure.[12] The increase in R&D expenditure *in conjunction with* harder research in the data needs to be understood in a setting where compensating forces, such as a lower interest rate, mitigate the growth effects of harder research.[13,14] Despite the pro-growth effect of a lower interest rate and the Schumpeterian incentives provided by higher concentration, the large increase in research difficulty dominates the landscape. The model can explain the major part of the changes in US TFP growth, concentration, markup, labor share, entry rate, exit rate, job creation rate, job destruction rate and job reallocation rate since the 1980s.

The above discussion leaves unanswered the crucial question of why research has become more difficult. To be sure, this has not always been the case: TFP growth went up by 1.5% from the 1900s to 1930s and peaked in the 1930s and 1940s according to Gordon (2016) and Nordhaus (2021). In Philippon (2022), productivity is best described as “additive” *per historical era*, meaning that the growth rate declines during each era but can temporarily increase at some key moments. In the Schumpeterian literature, this pattern of the rise and fall of growth is called the “Schumpeterian Wave”.[15] The high moments of waves could be due to the arrival and *widespread* adoption of *disruptive* general-purpose technologies, such as electricity and internal combustion engine which characterized the Second Industrial Revolution. For the era under study, the lack of disruptive technologies has been extensively documented by the recent literature. Gordon (2016) provides an

[12] To see this, consider a simple but general case in which the cost of R&D is $\frac{1}{2}\alpha\lambda^s$, where $s > 1$ and λ is the rate of success to be optimally chosen. Denote $V > 0$ to be the benefit if λ is successfully realized. Then the optimal λ is $\lambda^* = (\frac{2V}{\alpha s})^{1/(s-1)}$ and the R&D expenditure is $(\frac{1}{2}\alpha)^{-1/(s-1)}(\frac{V}{s})^{s/(s-1)}$. The latter is decreasing in α, i.e. decreasing in research difficulty in a partial equilibrium. If, however, compensating forces such as a lower interest rate make the change of λ less than what is implied by the sole force of harder research, then R&D expenditure can increase in a general equilibrium.

[13] In a standard consumption-based asset pricing model, the decrease in interest rate can be endogenized by a decrease in growth rate with appropriate relative risk aversion.

[14] These compensating forces should also be expected to exist given the large increase in research difficulty, otherwise the economic system seems too fragile.

[15] See Aghion and Howitt (1992) for a brief discussion.

extraordinary historical account of the rise and fall of US economic growth since 1870s, emphasizing the disruptiveness and broadness of the Second Industrial Revolution. Garcia-Macia et al. (2019) demonstrates that most growth since the 1980s has resulted from incumbents' innovations on existing products. Park et al. (2023) shows a decline in the disruptiveness of patents and academic papers over time, at least from the end of World War II. Nonetheless, the optimistic implication is that faster growth is possible in the future, as long as fundamental advances in science and technology can start another wave. Policies can be designed to stimulate R&D, particularly *fundamental* and *disruptive* ones that have a *wide scope* of applicability. The contribution of my paper is to analyze the effect of research productivity on the capabilities to innovate and the market structure by studying the endogenous dynamics of firms' productivity and size distributions.

Related Literature This paper first of all belongs to the Schumpeterian growth literature and owes its inspiration to the seminal works of Aghion and Howitt (1992) and Aghion et al. (2001). Compared to Aghion et al. (2001), this paper innovates by generating a well-defined productivity distribution with Pareto tail.[16] Such a distribution allows us to understand why market concentration has increased via the specific form of a fattening Pareto tail. It also allows us to evaluate economic integration in the spirit of the international trade literature, and thus to study dynamic growth effects and static market structure changes in a unified framework. The paper shifts from the traditional Schumpeterian emphasis of how market structure determines growth to how growth determines market structure. Akcigit and Ates (2021), in the spirit of the traditional Schumpeterian literature, interprets lower growth since the 1980s through a change in market structure, i.e. markets become more concentrated when the exogenous technological diffusion rate decreases.[17] In the present paper, lower growth is not due to higher concentration but to harder research, while higher concentration actually has a pro-growth effect. Harder research hurts the dynamic growth advantage of laggards and allows leaders to stretch out in relative terms. In this sense, the emphasis is on how growth determines the market structure, with potential dramatic implications for how anti-trust policies should be conducted.

Declining research productivity, the driving force in this paper, draws on fundamental insights of the semi-endogenous growth literature, e.g. Jones (1995), Kortum (1997) and Segerstrom (1998).[18] Bloom et al. (2020) provides recent empirical evidence based on micro data. Existing theoretical frameworks typically assume a positive exogenous population growth which cancels out the effect of harder research by the increasing number of researchers, and consequently maintains constant

[16] The baseline framework does generate a productivity *gap* distribution, but the productivity distribution dissipates over time as innovation and diffusion intensities do not depend on sectoral productivity *level*. Acemoglu et al. (2018) implicitly assumes harder innovation for more productive sectors and generates a stationary productivity distribution, but the distribution is thin-tailed.

[17] However, a decrease in the technological diffusion rate also has a pro-growth effect for leaders, which is the standard argument of intellectual property right protection. When Akcigit and Ates (2023) calibrates the model to the data, the two forces compensate each other so that growth decreases little.

[18] See Jones (2022) for a review.

productivity growth. When Jones (2002) brings the model to the data, however, less than 20% of growth is explained by population growth. Thus I assume a fixed population except when I analyze economic integration, and study the effect of declining research productivity on declining growth.

The long-run rise of US corporate concentration is based on the estimates of Kwon et al. (2022). Chen (2022) supports the empirical evidence and shows in addition that concentration is positively correlated with GDP per capita in a cross section of countries. Kwon et al. (2022) shows that rising concentration in an industry aligns closely with investment intensity in R&D and information technology, and interprets these investments as fixed costs so that increasing concentration is a result of higher fixed costs. However, total R&D expenditure over GDP has never exceeded 3.5% in the past century in the US.[19] Such a small change in fixed costs is unlikely to match the large increase in corporate concentration through the traditional sense of economies of scale, and cannot speak to the flattening Pareto tail.[20] The present paper concurs with Kwon et al. (2022) in that R&D is the key for understanding increasing concentration, but interprets its effect through endogenous growth which increases concentration via a change in the endogenous productivity distribution. Put differently, instead of considering increasing returns to scale due to fixed costs, this paper emphasizes increasing returns to scale of the production function when productivity can be changed in addition to capital and labor. Such change has been one of the central pillars of endogenous growth models (see Romer (1990) and Jones (2005)).[21] Chen (2022) also takes a growth perspective on the secular rise in concentration. While his paper focuses on the transitional dynamics when the initial distribution is assumed to be lighter-tailed than the equilibrium distribution,[22] my paper focuses on the equilibrium distribution for each historical period and the transition is modelled as moving from one equilibrium to another due to changes in research productivity.[23] While concentration increases and growth is constant in Chen (2022), in my paper increasing concentration is associated with declining growth due to harder research.

The paper is also closely related to the vibrant literature on US growth, market power and business

[19]Total private investment in intellectual property products, which includes investment in IT, has never exceeded 5% of GDP during the same period.

[20]In Melitz (2003), a higher fixed cost implies higher concentration but not a heavier tail of employment. In a demand system with variable demand elasticities, a higher fixed cost may imply a *lighter* tail of employment, as the productivity tail remains the same but demand elasticity is lower.

[21]Romer (1990) discusses how the non-rivalry of ideas implies increasing returns to scale of the production function when the idea and rival inputs (such as production labor, human capital, etc.) are jointly considered. Conversely, increasing returns to scale also implies non-rivalry of ideas. Suppose by contradiction that ideas are rival, then the standard replication argument implies that the production function should exhibit constant returns to scale with respect to idea and other rival inputs, which contradicts increasing returns to scale.

[22]Tail transition is fast in Chen (2022) as learning is perfect and targets only more productive firms. Consequently, when the log-productivity distribution has a tail of order Ce^{-kx}, successful learners with tail Cxe^{-kx} are injected, pushing the distribution towards a heavier tail.

[23]Thus the present paper compares different historical periods based on comparative statics, each period with a balanced growth path solution. The transitional dynamics in this baseline model is too slow to match the data, as explained by Luttmer (2011), Gabaix et al. (2016) and Jones and Kim (2018). Like these papers, different types of firms can potentially be introduced in an extended model to speed up the transition to a reasonable scale.

dynamism since the 1980s. Notable works on the theory side include Akcigit and Ates (2021, 2023) on technological diffusion, Aghion et al. (2023) and De Ridder (2019) on intangibles, and Liu et al. (2022) on low interest rates. While each theory sheds light on a specific mechanism, they face the common challenge of interpreting the long-run trend of declining growth and rising concentration since the 1930s.[24] Lower technological diffusion in Akcigit and Ates (2021, 2023) is interpreted as a result of more stringent intellectual property right (IPR) enforcement. While this could be the case since the 1980s, there is little evidence of the secular increase of IPR enforcement since the 1930s. Even if the latter is true, its pro-growth effect on leaders largely compensates its anti-growth effect from higher concentration in this type of model, making it difficult to speak to the secular decline in growth.[25] But my paper does concur with Akcigit and Ates (2021, 2023) in that the technological diffusion rate, i.e. the incumbent learning rate in the extended model of this paper, declines. The declining diffusion rate is now understood to be *not* due to institutional reasons, but rather to increasing technological difficulty. Aghion et al. (2023) and De Ridder (2019) assume ex ante differences in firms' abilities to utilize intangibles such as IT. More IT-capable firms can span their control with the arrival of IT, increasing concentration and depressing growth. IT was not widespread before the 1980s, making its historical relevance limited. Despite this, my paper's mechanism is consistent with higher intangibles like these papers, as measured intangibles predominantly consist of capitalized R&D expenditures (Hall et al. (2005), Bloom et al. (2013), Peters and Taylor (2017) and Haskel and Westlake (2017)) and the R&D share of GDP increases in my model. Liu et al. (2022) investigates the anti-growth effect of a lower interest rate when the latter is close to 0 and if there is no quick catch-up of laggards. In the present paper, a lower interest rate has a pro-growth effect, regardless of whether there is a quick catch-up for laggards or not, and acts as a compensating force to mitigate the effects of harder research.[26] Closest in spirit to the present paper are Olmstead-Rumsey (2019) and Cavenaile et al. (2019). Olmstead-Rumsey (2019) studies the decrease in innovativeness of all firms which corresponds to declining research productivity in this paper, and Cavenaile et al. (2019) compares different mechanisms and argues that harder research is likely to be the driver behind lower growth. Neither of these papers can account for the increasing concentration of very top firms or accommodate different patterns of concentration and market power across historical periods.

This paper is able to account for the key moments in the data including TFP growth (Gordon (2016), Nordhaus (2021)), corporate concentration (Kwon et al. (2022)), markup (De Loecker et al. (2020), Edmond et al. (2023)), labor share (Karabarbounis and Neiman (2014), Autor et al. (2020), Kehrig and Vincent (2021)) and various business dynamism indicators (Decker et al. (2016),

[24] One long-run driver other than declining research productivity is declining population growth. Peters and Walsh (2020) formalizes the idea that declining population growth can reduce business dynamism and growth, but its effect on market power and labor share is limited.

[25] See also Aghion et al. (2001) for the non-monotonic relationship between aggregate growth and diffusion rate.

[26] Moreover, if equity cost is considered in addition to debt cost to form the Required Rate of Return on Capital (RRRC) for discounting future profits, RRRC has always been above 10% since the 1980s according to Barkai (2020).

Decker et al. (2020)). It explains why market power and labor share may or may not change with a flattening Pareto tail depending on economic integration and market structure. The model is able to encompass and discuss some main drivers which have been put forward in the literature: technological diffusion rate, intangibles, interest rate and innovativeness of R&D.[27] In this sense, my work attempts to resolve the double accounting issue in the literature and hopes to avoid "explaining the labor share decline many times over" as advocated by Grossman and Oberfield (2022).

Importantly, the framework takes advantage of the newly developed methodology of Mean Field Game in the mathematical literature (e.g. Lasry and Lions (2007) and Lions (2006)). The methodology has been adopted in economics by Achdou et al. (2022), Alvarez et al. (2022) and Benhabib et al. (2021). Closest in spirit to my model is Benhabib et al. (2021), which generates a stationary productivity distribution via the innovation and learning decisions of individual firms. This paper differs by generating Pareto-tailed distributions and linking its flattening tail to lower growth when ideas get harder to find. It also studies the implications for market power and labor share with a demand system featuring variable demand elasticities.

Roadmap Section 2 presents empirical evidence on the relationships between research productivity and various indicators of market structure and business dynamism. Section 3 develops a simple model which illustrates the mechanism behind the long-run rise in concentration and decline in growth. Section 4 constructs a full model with endogenous innovation, generates a travelling wave solution of productivity distribution, and describes business dynamics with entry and exit. The distribution reflects the dynamics of firm heterogeneity and market power in a monopolistic competition with variable demand elasticities. Section 5 calibrates the model parameters based on US data. When ideas get harder to find, Section 6 explores the implications for growth, market structure and business dynamism through a comparative statics analysis. Section 7 reconciles the earlier US experience of increasing concentration with a stable markup and labor share by introducing economic integration alongside harder innovation. Section 8 extends the model by introducing incumbent learning in addition to innovation, and shows the robustness of the results under the alternative model. Section 9 concludes with a discussion of potential policy implications.

2 Empirical Evidence

I use US sectoral level data between 1987 and 2018 to show the correlations between research productivity on the one hand, and TFP growth, indicators of market structure and business dynamics on the other. There are 33 sectors in total, including 19 manufacturing sectors at 3-digit NAICS level and 14 non-manufacturing sectors at 2-digit NAICS level. I follow Bloom

[27] The relationship with other drivers is less clear, such as automation in Martinez (2021). Nevertheless, Martinez (2021) emphasizes the key role of automation distribution in generating aggregate effects in the long run, which resonates with the role of productivity distribution in the present paper.

et al. (2020) and define research productivity as the ratio between TFP growth and the effective number of researchers at the sectoral level, taking into account decreasing returns to scale of the idea production function.[28] The effective number of researchers is R&D expenditure deflated by scientist wage which is proxied by average annual earnings for men with four or more years of college education. R&D expenditure comes from the aggregation of firm-level R&D in Compustat at the sectoral level.[29] TFP growth comes from the KLEMS data of the US Bureau of Labor Statistics (BLS) and Bureau of Economic Analysis (BEA).[30] To remove the cyclical variations of TFP growth in a business cycle, I divide the 1987-2018 history into four equal periods, each with eight years. All variables are calculated as per period average. Such a division is also conceptually reasonable as a change in "research productivity" only seems meaningful when a longer horizon is involved.[31]. See Appendix A.1 for more details on the definition of research productivity.

Dependent variables are constructed to reflect the market structure and business dynamics. (1) I construct an indicator of market concentration from Business Dynamics Statistics (BDS) which reports employment by firm size category and sector. For each sector*period, market concentration is defined as the employment share of firms with more than 5000 employees. (2) EBIT/Sales of the largest four or eight firms by sector in Compustat is used to indicate the profitability of the largest firms in the sector. (3) Sectoral labor shares come from KLEMS. (4) For business dynamics, I use entry rate in terms of the number of establishments, entry rate in terms of job creation from birth, total job creation rate, exit rate in terms of the number of establishments, exit rate in terms of job destruction from death, total job destruction rate, and job reallocation rate, all from BDS.

For each dependent variable $y_{s,t}$, where s stands for sector and t for period, I run the following regression:

$$y_{s,t} = \beta_0 + \beta_1 \log(\text{IdeaProd}_{s,t}) + \gamma_t + \epsilon_{s,t},$$

where $\text{IdeaProd}_{s,t}$ is the research productivity of the idea production function and γ_t is period fixed effect. Standard errors are clustered at sectoral level. Table 1 shows the results for growth and market structure indicators, and Table 2 for business dynamism indicators. Within-R2 reports the R^2 without fixed effects. Research productivity is significantly negatively correlated with market concentration and profitability of the largest firms in a sector, and positively correlated with labor share and all indicators of business dynamics. These empirical patterns, together with the stylized facts in the introduction, are consistent with a model in which declining research

[28] The results remain robust with constant returns to scale of the idea production function (see Appendix A.2).

[29] Compustat covers public firms whose share among all firms varies across sectors and years. To alleviate the concern about the R&D data coverage in Compustat, Appendix A.2 adjusts sectoral R&D by the employment share of Compustat firms within each sector. The correlations remain robust after the adjustment.

[30] BEA treats R&D as intermediate input when calculating TFP and reports separately the contribution of R&D to productivity. Following Bloom et al. (2020), this contribution is added back to TFP to be consistent with the conceptual framework of endogenous growth.

[31] Using divisions other than eight years per period, as long as each period is not too short so that cyclical variations can be removed, and not too long so that there are sufficient data points, generates similar results (See Appendix A.2)

productivity decreases growth and increases market concentration. Higher market concentration manifests itself via higher market power and a lower labor share in a demand system with variable demand elasticities. Job reallocation decreases as firms react less to idiosyncratic shocks due to lower demand elasticities. The next section presents a simple model focusing on growth and concentration, and Section 4 presents a full model consistent with all the empirical findings.

	(1) Growth	(2) Concentration	(3) Ebit/Sales 4	(4) Ebit/Sales 8	(5) Labor Share
log(IdeaProd)	0.472***	-4.594***	-0.723*	-0.817*	2.664***
	(0.0853)	(1.526)	(0.414)	(0.432)	(0.927)
N	132	132	132	132	132
Within R2	0.187	0.147	0.0547	0.0729	0.0877
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes

Notes: Concentration is the employment share of firms with more than 5000 employees from BDS. Top 4(8) profit is the EBIT/Sales ratio of the largest 4(8) firms in the sector based on Compustat. Labor share comes from KLEMS. Intercepts are omitted. Standard errors are clustered at sectoral level and shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1: Regressions of TFP growth and market structure indicators on log research productivity.

	(1) Entry (num)	(2) Exit (num)	(3) Entry (job)	(4) Exit (job)	(5) Job Creation	(6) Job Destruction	(7) Job Reallocation
log(IdeaProd)	0.610**	0.593***	0.459***	0.393***	0.967***	0.848***	1.867***
	(0.232)	(0.157)	(0.150)	(0.0992)	(0.279)	(0.228)	(0.493)
N	132	132	132	132	132	132	132
Within R2	0.0740	0.103	0.0991	0.110	0.0909	0.0969	0.117
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Entry (num) is the number of new establishments over the number of existing establishments. Entry (job) is the number of jobs created by new establishments over total jobs. Exit (num) and Exit (job) are defined analogously for exits. Job creation is the number of new jobs created by entrants and incumbents over total jobs. Job destruction is the number of jobs destroyed by exits and continuers over total jobs. Job reallocation is the sum of job creation rate and job destruction rate. Intercepts are omitted. Standard errors are clustered at sectoral level and shown in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Regressions of business dynamism indicators on log research productivity.

3 A Simple Model

This section presents an illustrative model in a continuous-time setting, illustrating the key mechanism that links lower research productivity with increased concentration of the productivity distribution of firms. It also paves the way for introducing additional elements in the full model. For simplicity, I skip the consumption side of the economy and model only the production side.

For notational conciseness, time index t is dropped whenever possible.

3.1 Simple Model and Evidence on the Mechanism

There is a continuum of firms indexed by $\theta \in [0, 1]$, each of which produces one differentiated good. Denote A_θ as firm-level productivity, $a_\theta = \log(A_\theta)$ as its log level, and A as aggregate productivity. Assume firm θ's per-period profit as $\Pi_\theta = C\tilde{A}_\theta^{\sigma-1}A$, where $\tilde{A}_\theta = A_\theta/A$ is firm θ's productivity relative to the aggregate level, $C > 0$ is a constant, and $\sigma > 1$ is a constant. The profit function can be generated by a static model in which the demand system has Constant Elasticity of Substitution (CES) σ and firms compete monopolistically with each other. Strictly speaking, the constant C depends on the productivity distribution of firms, which will be case in the full model. In this illustrative model, however, we consider it to be a constant regardless of the distribution. In this way, we allow the distribution to be determined by firms' growth decisions, but shut down the feedback from the distribution to profits and hence to growth, which greatly simplifies the model. Under the CES interpretation, σ is also the demand elasticity of each variety of goods, and aggregate productivity is a homothetic aggregator of firms' productivities with the form $A = [\int_0^1 A_\theta^{\sigma-1}\,\mathrm{d}\theta]^{\frac{1}{\sigma-1}}$.

Each firm can invest in research to consciously improve its productivity step-by-step. By hiring $\frac{1}{2}\alpha\tilde{A}_\theta^{\beta}\lambda^2$ number of researchers per period, firm θ's innovation has Poisson rate of arrival λ, the success upon which allows its log productivity to increase from a_θ to $a_\theta + q$.[32] Figure 2 illustrates a successful innovation.[33] The factor $\tilde{A}_\theta^{\beta}$ with $\beta > 0$ captures harder research upon more advanced technology in a cross section of firms, while α captures the economy-wide research difficulty at a specific time. $\beta > 0$ is supported by firm-level evidence in Bloom et al. (2020) which shows declining research productivity as a firm grows. Since all firms are synonymous in my model except for their productivities, lower research productivity associated with higher productivity of a specific firm translates into the same relationship in a cross section of firms. A larger α is used for a later historical period to capture uniformly harder research over time. Adjusting researchers' wage by aggregate productivity A, the research cost associated with Poisson rate λ is $R = \frac{1}{2}\alpha\tilde{A}_\theta^{\beta}\lambda^2 A$. For simplicity, each firm is myopic and considers only marginal profit when taking the innovation decision, i.e. it does not solve an infinite-horizon optimization problem with discounted future profits.

[32] The quadratic cost function is a typical assumption in Schumpeterian growth models which matches the elasticity of R&D with respect to the user costs of around -1. See Bloom et al. (2002), Acemoglu et al. (2018) and Akcigit and Kerr (2018).

[33] The equilibrium distribution is virtually the same under an alternative model in which innovation is modelled as a continuous process (i.e. a $\mathrm{d}t$ term) rather than as a Poisson process, which is testified by the quasi-equivalence between equation A1 and equation A2. However, the interpretations are different when looking at the data. In the current specification only λ share of firms realize innovation and hence fast growth within each size category, and other firms are dictated by random growth. We only need to look at fast growing firms to check the validity of the mechanism.

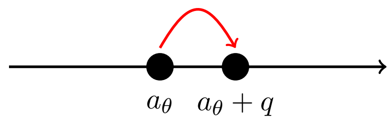


Figure 2: Improvement of log productivity upon a successful innovation.

The innovation decision of each firm is thus:

$$\max_{\lambda} \left\{ \lambda q \frac{\mathrm{d}\Pi_\theta}{\mathrm{d}a_\theta} - R \right\}$$

whose solution is

$$\lambda^* = \frac{Cq(\sigma-1)}{\alpha} \tilde{A}_\theta^{\sigma-1-\beta} \tag{1}$$

Note that λ^* decreases exponentially with $\tilde{a}_\theta = \log(\tilde{A}_\theta)$ if $\beta > \sigma - 1$, which is assumed in this section. The validity of the last assumption will be discussed when introducing the more general model. On the profit side of the innovation decision, a one percent increase in productivity increases profit by $\sigma - 1$ *percent* because of the constant demand elasticity. The *amount* of profit increase is however larger if the firm already reaps a higher profit. Larger firms benefit from economies of scale when innovating: an improved technology can be used for all units of production. As firm size is intimately tied to productivity level in the model, marginal profit increases with productivity level with an elasticity of $\sigma - 1$. On the other hand, the marginal cost of innovation increases with productivity level with an elasticity of β, as technological improvement becomes more difficult with more advanced technology. Innovation intensity thus decreases with productivity if harder innovation dominates economies of scale. Laggards enjoy a dynamic advantage due to the relative ease of growth.

In addition to innovation, each firm is subjected to idiosyncratic log-productivity shocks which is modelled as a Brownian motion with a standard deviation of ν.[34] Keeping α constant, we focus on the travelling wave equilibrium in which the log-productivity Probability Density Function (PDF) ϕ maintains its shape and travels at a constant speed. As A is a homothetic aggregator of $\{A_\theta\}_{\theta\in[0,1]}$, the travelling speed is also the aggregate growth rate g. After normalizing all firms' productivities by aggregate productivity, i.e. normalizing by aggregate growth rate after each period, the distribution $\tilde{\phi}$ is stationary, where I have used a tilde to emphasize the normalization. Taken together, each firm's log productivity after normalization follows the following stochastic process:

$$\mathrm{d}\tilde{a}_{\theta,t} = (-\frac{1}{2}\nu^2 - g)\,\mathrm{d}t + \nu\,\mathrm{d}B_{\theta,t} + \mathrm{d}J_{\theta,t} \tag{2}$$

where $B_{\theta,t}$ is Brownian motion independent across firms, $-\frac{1}{2}\nu^2\,\mathrm{d}t$ is a normalization term for the

[34] As is typical in continuous-time models, the probability of Brownian motion and innovation happening together is of order of magnitude $o(\mathrm{d}t)$ so that it can be ignored. Thus when making the innovation decision, each firm ignores the Brownian motion.

Brownian motion so that the expected productivity change due to Brownian motion is null, and $J_{\theta,t}$ is the jump process due to innovation. A Kolmogorov Forward (KF) equation describes the dynamics of distribution $\tilde{\phi}$ whose equilibrium solution can be approximated by:

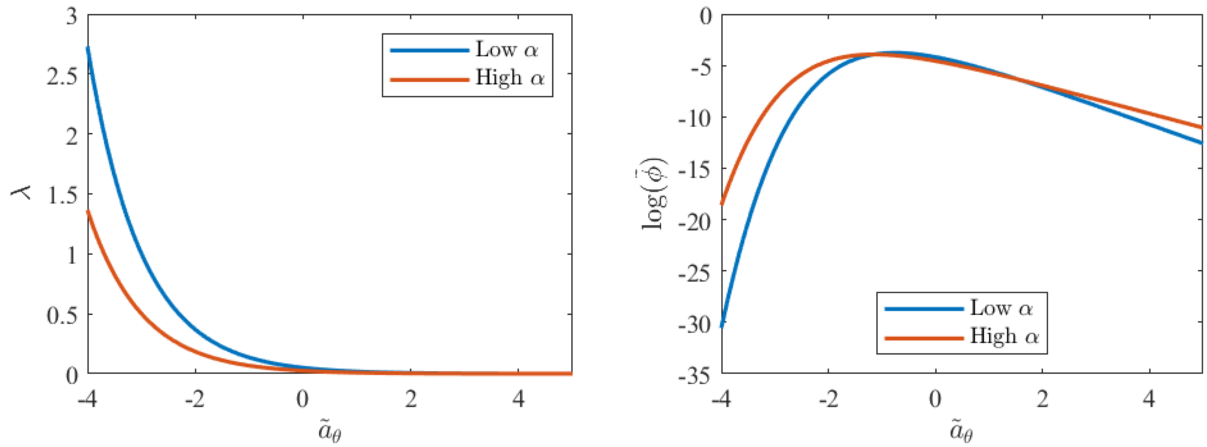
$$\tilde{\phi}(\tilde{a}) = C_1 \exp(-C_2\tilde{a} - C_3 e^{-(\beta-\sigma+1)\tilde{a}}), \quad \tilde{a} \in \mathbb{R} \tag{3}$$

where $C_1 > 0$ is a normalizing constant which ensures that $\tilde{\phi}$ integrates into 1, $C_2 = 1 + 2g/\nu^2 > 0$ and $C_3 = \frac{2Cq^2(\sigma-1)}{\alpha\nu^2(\beta-\sigma+1)} > 0$. See Appendix B.1 for the derivations. Asymptotically, $\tilde{\phi}(\tilde{a}) \sim C_1 \exp(-C_2\tilde{a})$ as $\tilde{a} \to +\infty$:[35] $\tilde{\phi}(\tilde{a})$ has an exponential right tail with tail index C_2, or equivalently, the productivity distribution has a Pareto right tail with tail index C_2. The index decreases with aggregate growth, meaning that a fatter tail is associated with lower growth.[36] Like in random growth models, I have assumed idiosyncratic shocks to generate the Pareto tail. The heaviness of the tail depends on the joint force of random growth and endogenous growth. Aggregate growth g is the one which normalizes aggregate productivity to 1.[37]

We solve one travelling wave equilibrium for one constant α, and then compare different travelling wave equilibria. Figure 3 shows two solutions, one with low α and another with high α: innovation becomes *uniformly* more difficult for all firms in the high α case. Thus marginal research cost increases by the same percentage for all firms, while marginal benefit remains unchanged in this simple model. This implies that growth decreases by the same proportion for all firms. But what matters for a laggard's ability to catch up with a leader is not the ratio between their growth rates but rather the difference between them. A higher growth rate of laggards before the increase in research difficulty means that the reduction in laggards' growth is more pronounced in absolute terms. Laggards' advantage in growth thus implies that they are more negatively affected when research becomes more difficult. Consequently, leaders' *relative* growth compared to laggards increases, allowing the leaders to stretch out further to the right and fattening the Pareto tail of the productivity distribution (Figure 3b). At the core of the mechanism is the semi-endogenous-growth assumption that it is more difficult to improve upon more advanced technology, both in a cross section and over time. With a CES demand, the fatter tail of the productivity distribution translates into fatter tails of sales and employment, implying a higher corporate concentration of top firms. In this illustrative model, we have shut down the feedback loop from concentration to growth decisions. Unlike most of the papers on Schumpeterian growth, the relationship between higher concentration and lower growth is not the former leading to the latter. Rather, higher concentration is understood *as a result of* lower growth when research becomes more difficult.

Innovation intensity decreases with firm size, which is a departure from Gibrat's law and consistent with the data as explained below. The departure is, however, slight as only a small proportion of

[35] As with standard mathematical notations, $f(x) \sim g(x)$ as $x \to +\infty$ means $\lim_{x\to+\infty} \frac{f(x)}{g(x)} = 1$.
[36] As a reminder, a smaller Pareto tail index means a fatter tail.
[37] In other words, it is implicitly determined by: $[\int_{-\infty}^{+\infty} e^{(\sigma-1)\tilde{a}}\tilde{\phi}(\tilde{a})\,\mathrm{d}\tilde{a}]^{\frac{1}{\sigma-1}} = 1$.



(a) Innovation intensity

(b) Log productivity distribution in log scale

Figure 3: Illustrative model solutions for illustrative purposes. $\sigma = 2$, $\beta = 2$, $C = 1$, $q = 0.2$, $\nu = 0.15$, $\alpha = 4$ for low α case and $\alpha = 8$ for high α case.

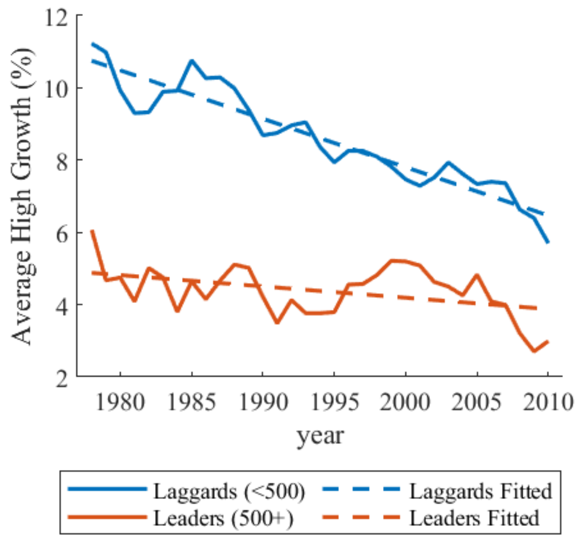
firms successfully realize innovation at a specific time. I interpret these firms as high-growth firms when investigating the data. The great majority of firms are subjected to idiosyncratic shocks at any specific time so that their average growth is insensitive to size.[38] To determine whether the model's mechanism is consistent with the data, I categorize firms into leaders and laggards based on employment. Following Decker et al. (2016), high growth firms are defined as those with more than 30% annual growth in employment. I calculate the weighted average growth of high growth firms within leaders and laggards respectively, which takes into account both the employment share of high growth firms and the growth rates of these firms. Figure 4a shows that the high growth firms among the laggards grow more quickly than those of leaders in any given year. Both leaders and laggards witness a decline in growth, but the decline is more pronounced for laggards. All of these empirical patterns are consistent with the model in Figure 3a. Moreover, Figure 4b shows the histograms of log TFP dispersion between the 99th quantile and the 90th quantile, where each observation is a 4-digit NAICS sector.[39] The productivity dispersion at the tail has in general increased,[40] which is consistent with the model's prediction that the Pareto tail of productivity distribution fattens.

To sum up, this simple model conveys the mechanism according to which an aggregate increase in research difficulty translates into an increase in the concentration of firms. The key mechanism (differential growth between laggards and leaders, and the fatter tail of the productivity distribution) is borne out in the data. However, many ingredients are missing form this simple model.

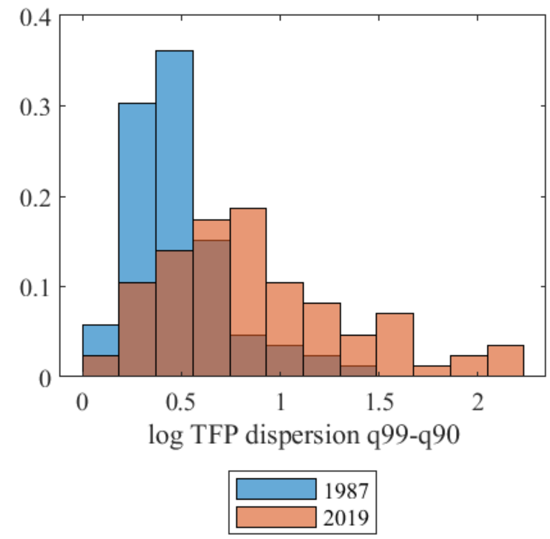
[38]Averaging across both endogenous-growth firms and random-growth firms in the model still implies higher average growth for laggards. Such a slight departure from Gibrat's law is consistent with the empirical evidence in Akcigit and Kerr (2018).

[39]Due to data limitation, the sectors comprise manufacturing only.

[40]If I track each 4-digit sector and calculate its change in log TFP dispersion between 1987 and 2019, 84% of sectors see an increase in the dispersion.



(a) High growth among leaders and laggards



(b) Histogram of TFP dispersion at tail

Figure 4: Checking the model's mechanism with data.

Notes: Panel (a): High growth firms are those with more than 30% annual employment growth. Laggards are firms with less than 500 employees and leaders more than 500. For each year, the plot shows the employment-weighted average growth of high growth firms among leaders and laggards respectively. Data source: Decker et al. (2016). Panel (b): The histogram shows the log TFP dispersion between the 99th quantile and the 90th quantile, where each observation is one 4-digit NAICS sector of manufacturing. Data source: US Census Bureau, Dispersion Statistics on Productivity (DiSP).

3.2 Necessity of a Richer Model

Apart from illustrating the key mechanism at work, the simple model also clarifies the necessity of a richer model:

1. Heterogeneous demand elasticities

   The assumption of $\beta > \sigma - 1$ is convenient in this section but tenuous in the data if the demand system is CES. Fortunately, we do not need the assumption to be true for all firms to make the mechanism work. As the paper focuses on the right tail, the assumption only needs to be valid for large firms so that the largest firm can stretch out relative to moderately large firms. Figure A2 re-plots Figure 4a with finer size bin, and shows this is the case in the data. Following the macroeconomic literature on market power, I assume demand elasticity decreases with firm size in the full model. Intuitively, the markets in which larger firms are involved are more saturated, so that these firms' ability to attract customers by cutting prices is more limited. The full model only requires $\beta > \underline{\sigma} - 1$ where $\underline{\sigma}$ is the lowest possible demand elasticity, which is easily satisfied by the data. Introducing heterogeneous demand elasticities also has implications for market power naturally linked to increases in concentration.

2. Entry and Exit

   If demand elasticity σ declines with firm size such that $\sigma - 1 - \beta$ is first positive and then negative, the innovation intensity in equation 1 then features an inverted U that is often found in Schumpeterian growth models. The low growth of the least productive firm implies that a dynamic equilibrium may not exist, and additional stabilizing forces are needed. Existing Schumpeterian growth models typically introduce technological diffusion to allow laggards grow more quickly than leaders.[41] Another way to stabilize the distribution is by introducing entry and exit. The least productive firms that cannot catch up with the rest of the distribution die. Entrants join the pool of incumbents so that the total number of firms remain stable. I shall proceed with entry and exit in the baseline model due to their analytical simplicity, and introduce incumbent learning as an extension. Adding entry and exit also gives a more realistic description of job dynamics.

3. Feedback from market structure to growth decisions

   The simple model keeps the functional relationship between profit and firm-specific log productivity unchanged. The relationship should also depend on the productivity distribution which has changed with research difficulty, i.e. C should depend on $\tilde{\phi}$. A full model thus requires two directions: how growth determines market structure, which is present in the illustrative model; and how market structure shapes the Schumpeterian incentives of individual firms and hence determines growth. Moreover, the latter involves the forward-looking behavior of individual firms so that each firm optimizes discounted future profits instead of instantaneous profit, which is adequately described by a Hamilton-Jacobi-Bellman (HJB) equation. Joining the HJB equation with the KF equation gives rise to the Mean Field Game (MFG) system that the full model is based on.

4 Model

Time t is continuous. The economy has a representative consumer who consumes differentiated goods per period. Each good is produced by one firm and the firms compete monopolistically with each other. Firms' productivities are heterogeneous, and each firm can improve its productivity by innovating on a stand-alone basis.[42] Incumbent firms can die due to large idiosyncratic shocks to their productivities or due to being too unproductive, and entrants keep joining the pool of incumbent firms by learning from the latter. Figure 5 illustrates the structure of the baseline full model. For notational simplicity, I shall drop the time index t whenever possible.

[41] This is introduced as incumbent learning in Section 8 of this paper, which explains higher growth of laggards in the data.

[42] An extended model with incumbent firms learning from others is presented in Section 8.

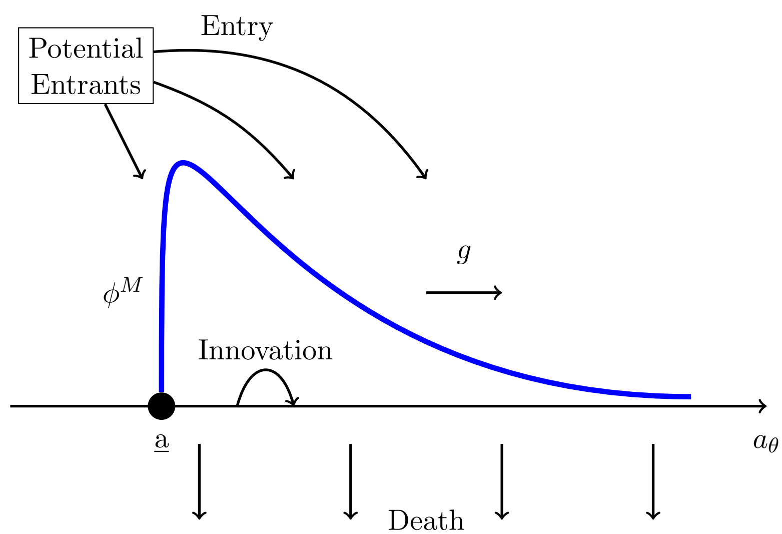


Figure 5: Illustration of the baseline full model. a_θ is an individual firm's log-productivity, and the bold blue line represents its distribution.

4.1 Market Structure

4.1.1 Preference

The representative consumer's utility Y from goods consumption is defined implicitly by a homothetic Kimball aggregator:

$$M \int_0^1 \gamma(\frac{Y_\theta}{Y}) \mathrm{d}\theta = 1 \tag{4}$$

where M denotes the total measure of varieties, Y_θ is the consumption of variety θ and γ is an increasing and concave function with $\gamma(0) = 0$ and $\gamma(1) = 1$.[43]

Each variety is produced by one firm indexed by $\theta \in [0, 1]$ and firms differ in their productivities. Note that θ is *not* necessarily ordered in a way such that higher productivity corresponds to a higher θ: θ tracks a firm and firms' productivities evolve in a dynamic setting introduced below. In any static moment, nevertheless, one can re-index firms such that $\hat{\theta} \in [0, 1]$ indexes log-productivities in an orderly way. With an abuse of notation, I shall use θ as $\hat{\theta}$ for denoting the share of the bottom θ firms in the size distribution. Then θ is the cumulative probability of the log-productivity distribution with $\theta = \Phi(a_\theta)$, where a_θ is the log productivity of firm θ and Φ is the Cumulative

[43]To see why $\gamma(1) = 1$ as a normalization condition, consider the homogenous case in which $Y_\theta \equiv Y_0$. Assume in addition that $M = 1$. Then it is reasonable to have $Y = Y_0$ as a normalization, which implies $\gamma(1) = 1$.

Distribution Function (CDF) of log productivities.[44]

It should be emphasized that equation 4 implicitly incorporates the distribution of productivities. To see this more clearly, change variable from θ to a_θ to obtain:

$$M \int_{\mathbb{R}} \gamma(\frac{Y(a_\theta)}{Y}) \phi(a_\theta) \mathrm{d}a_\theta = 1$$

where ϕ is the Probability Distribution Function (PDF) of log productivities. The variable to be integrated in the above equation can of course be written as any variable such as x instead of a_θ, but I shall often use a_θ to emphasize the nature of individual firms. Denote $\phi^M = M\phi$ to be the generalized PDF which has a total measure of M. The distribution and its change will be the key concern in the paper.

The representative consumer maximizes his/her utility Y subject to the budget constraint:

$$M \int_0^1 P_\theta Y_\theta \,\mathrm{d}\theta = W \tag{5}$$

where P_θ is the price of variety θ and W the income of the consumer. Solving the consumer's maximization problem gives the inverse-demand curve for each variety θ:

$$\frac{P_\theta}{P} = \gamma'(\frac{Y_\theta}{Y}) \tag{6}$$

where the price aggregator P satisfies:

$$\frac{P}{\zeta} Y = W \tag{7}$$

and ζ is defined as

$$\zeta = \big[M \int_0^1 \gamma'(\frac{Y_\theta}{Y}) \frac{Y_\theta}{Y} \,\mathrm{d}\theta\big]^{-1} \tag{8}$$

Note that there are two price indices when the preference is Kimball: P and ζ.[45] The ideal price index, which is used to deflate nominal income to evaluate welfare, is P/ζ and *not* P.

For notational simplicity, I shall denote $Z_\theta = \frac{Y_\theta}{Y}$ as the relative output and $z_\theta = \log(Z_\theta)$. The price elasticity of demand for variety θ is:

$$\sigma_\theta = \frac{\gamma'(Z_\theta)}{-Z_\theta \gamma''(Z_\theta)} \tag{9}$$

The constant Elasticity of Substitutions (CES) preference is a special case of the Kimball preference

[44] To see why cumulative probability is used for indexing firms, consider the discrete counterpart of equation 4: $M \frac{1}{I} \sum_{i=1}^{I} \gamma(\frac{Y_{i/I}}{Y}) = 1$. In this case, $\theta(a_{i/I}) = \frac{i}{I}$ for the i-th firm.

[45] This is a general feature of Homothetic Direct Implicit Additivity (HDIA) preferences of which Kimball is a special example, see Matsuyama (2023).

with $\gamma(Z) = Z^{\frac{\sigma-1}{\sigma}}$, where $\sigma > 1$ is the elasticity of substitution between varieties of goods. With CES, $\gamma'(Z_\theta) = \frac{\sigma-1}{\sigma} Z_\theta^{-\frac{1}{\sigma}}$ and $\zeta = \frac{\sigma}{\sigma-1}$. The elegance of the Kimball aggregator can be seen from equations 6 and 9: by introducing a different functional form of γ, we can generate heterogeneous demand elasticities and hence heterogeneous markups for heterogeneous firms. The current literature has often adopted the specification of Klenow and Willis (2016) for the Kimball aggregator. However, as Baqaee et al. (2023) remarks, the Klenow and Willis (2016) specification implies that the demand elasticity converges to 0 too quickly as the consumption of this variety increases, which implies excessively high markups and excessively low sales for these varieties. I adopt a new parametric specification in the spirit of Atkeson and Burstein (2008) which is standard in the literature on market power.[46] In particular, define γ' for the inverse demand curve 6:

$$\gamma'(Z_\theta) = C\Big[\frac{1}{\underline{\sigma}} Z_\theta^{\frac{k\underline{\sigma}}{\overline{\sigma}-\underline{\sigma}}} + \frac{1}{\overline{\sigma}} Z_\theta^{\frac{k\overline{\sigma}}{\overline{\sigma}-\underline{\sigma}}}\Big]^{-\frac{\overline{\sigma}-\underline{\sigma}}{k\overline{\sigma}\underline{\sigma}}} \tag{10}$$

where $1 < \underline{\sigma} \leq \overline{\sigma}$, $k > 0$ and $C > 0$. $\overline{\sigma}$ is the upper bound of demand elasticities and $\underline{\sigma}$ is the lower bound.[47] k governs the transition from the highest demand elasticity to the lowest when output increases. $C > 0$ is a constant pinned down by normalizing conditions. The special case of $\underline{\sigma} = \overline{\sigma}$ corresponds to CES. When $\underline{\sigma} < \overline{\sigma}$, larger firms face lower demand elasticities and charge higher markups. The specification resembles the nested CES of Atkeson and Burstein (2008) as it combines two power functions, each corresponding to one limiting demand elasticity. Section 4.1.3 discusses the properties of the demand system and shows that the resemblance is not only in appearance but also in essence.

To see the transition of demand elasticity more clearly, use equation 9 for calculating the demand elasticity:

$$\sigma(z_\theta) = \overline{\sigma} + \frac{\underline{\sigma} - \overline{\sigma}}{1 + \exp(-kz_\theta)} \tag{11}$$

i.e. σ is a logistic function of z_θ, which decreases monotonically with z_θ and has limits:

$$\lim_{z_\theta \to +\infty} \sigma(z_\theta) = \underline{\sigma}, \qquad \lim_{z_\theta \to -\infty} \sigma(z_\theta) = \overline{\sigma} \tag{12}$$

4.1.2 Firms' Production Decision

Each variety of good is produced by one firm with labor as the only factor of production:

$$Y_\theta = A_\theta L_\theta \tag{13}$$

where A_θ is the productivity of the firm and L_θ the amount of labor used in the production. Firms engage in monopolistic competition with each other and take into account the inverse demand curve

[46] In principle, I can also use a non-parametric form of Kimball aggregator following Baqaee et al. (2023). Nonetheless, the parametric form makes numerical calculations easier and more accurate. It also gives a portable demand system for further studies.

[47] See Appendix H.1 for the numerical calculation of γ.

6 for pricing decisions. Optimal pricing implies the markup:

$$\mu_\theta = \frac{P_\theta/P}{MC_\theta/P} = \frac{\sigma(z_\theta)}{\sigma(z_\theta)-1} \tag{14}$$

where $MC_\theta = \frac{w}{A_\theta}$ is the marginal cost of production and w is the wage rate. Equation 14 is similar to the markup formula under CES, with the sole exception that σ now depends on z_θ. With given aggregate variables w and P, the intersection of the inverse demand curve 6 and the pricing equation 14 solves a firm's static production problem at each point in time. Both price and quantity are functions of the firm's productivity.

Production incurs a fixed cost of fA per period, where A is the aggregate TFP to reflect higher wages as the economy grows. The least productive firms with operational profits less than fA per period die immediately from the market. Denote the cutoff log productivity to be $\underline{a}$ above which operational profit is higher than fA. $\underline{a}$ is determined by the break-even condition:

$$\Pi(\underline{a};\phi^M) = fA \tag{15}$$

4.1.3 Properties of the Demand System

Aggregate productivity is defined by:

$$A = \frac{Y}{L} \tag{16}$$

where $L = M\int_0^1 L_\theta \,\mathrm{d}\theta$ is the total amount of labor inelastically supplied by the representative consumer. It is easy to check that

$$A = \left[M\int_0^1 Z_\theta A_\theta^{-1}\,\mathrm{d}\theta\right]^{-1} \tag{17}$$

i.e. A is a harmonic mean of A_θ weighted by relative production. Given the measure of firms M, all variables are determined by the productivity distribution so that the above equation can be represented as $A = \mathcal{A}(\{A_\theta\}_{0\le\theta\le1})$. The following property simplifies the model and numerical calculations:

Proposition 1 (Homotheticity of the Productivity Aggregator)**.** *$\mathcal{A}$ is a homothetic aggregator of firms' productivities $\{A_\theta\}_{0\le\theta\le1}$, i.e. $\mathcal{A}(\{sA_\theta\}_{0\le\theta\le1}) = s\mathcal{A}(\{A_\theta\}_{0\le\theta\le1})$, $\forall s>0$*

Proof. Consider the case in which each A_θ is multiplied by s. To show the homotheticity of A with respect to $\{A_\theta\}_{0\le\theta\le1}$, by equation 17 we only need to show that $Z_\theta = \frac{Y_\theta}{Y}\quad(\forall\theta)$ does not change.

Combining equations 6 and 14 gives:

$$\frac{\gamma'(Z_\theta)}{MC_\theta/P} = \frac{\sigma(z_\theta)}{\sigma(z_\theta)-1} \tag{18}$$

which pins down Z_θ as an implicit function of A_θ and P/w:

$$Z_\theta = Z(a_\theta, P/w) \tag{19}$$

Plugging this function into the Kimball aggregator 4 pins down P/w as a function of $\{A_\theta\}_{0\leq\theta\leq 1}$ and M. If each A_θ is multiplied by s, it is easy to check that $\{Z_\theta\}_{0\leq\theta\leq 1}$ and $\frac{P}{sw}$ is the new solution to this system of equations. Price-wage ratio decreases by s but the relative quantity does not change. This concludes the proof. $\square$

In particular, with a constant adjustment of all productivities, wage-price ratio increases by s but relative quantities $\{Z_\theta\}_{0\leq\theta\leq 1}$ do not change. In the special case of CES, $A = [M \int_0^1 A_\theta^{\sigma-1}\,\mathrm{d}\theta]^{\frac{1}{\sigma-1}}$. Homotheticity means that we can normalize the aggregate productivity A to 1 by dividing all A_θ by A.

Proposition 2 (Variable Demand Elasticities and Passthroughs, Pareto Tail)**.**
If $1 < \underline{\sigma} < \overline{\sigma}$*, then*

1. $\frac{\partial \sigma}{\partial a_\theta} < 0$*, i.e. Demand elasticity decreases with a firm's productivity, so that more productive firms charge higher markups.*

2. *There exists a threshold* $\hat{a}$*, such that passthrough* $\frac{\partial \log(P_\theta)}{\partial \log(MC_\theta)}$ *as a function of* a_θ *is decreasing on* $(-\infty, \hat{a}]$ *and increasing on* $[\hat{a}, +\infty)$*, i.e. the passthrough of cost shocks to price decreases with a firm's productivity when the latter does not exceed a certain threshold* $\hat{a}$*.*

3. $P_\theta Y_\theta \sim C_1 A_\theta^{\underline{\sigma}-1}$ *and* $L_\theta \sim C_2 A_\theta^{\underline{\sigma}-1}$ *as* $a_\theta \to +\infty$*, where* C_1 *and* C_2 *are positive constants. Moreover, the distributions of these variables are Pareto-tailed if the productivity distribution is Pareto-tailed.*

See Appendix C.2 for the proof. We have used the notation ∂ to emphasize that we only consider the direct impact of a_θ on the respective variables, without taking into account the indirect effect through aggregate variables, as each firm is infinitesimal by assumption.

The lower demand elasticities and passthroughs of more productive firms are consistent with the empirical evidence, as surveyed by Arkolakis and Morlacco (2017), and are primordial for Schumpeterian incentives of innovation. An improvement in productivity reduces the marginal cost one-to-one at the log level, which then translates into a reduction of price at the rate of passthrough. This reduction in price is more than compensated by an increase in quantity as $\sigma > 1$ so that sales and profits increase. If the passthrough is artificially high for productive firms due to a mis-specification of the demand system, the decrease in price and the increase in profit would be too great, giving high productivity firms an excessively stronger incentive to innovate. Similarly, if demand elasticity σ is mistakenly high for productive firms, their growth rate would also be excessive. In fact, the demand system in the Schumpeterian growth models of Aghion et al. (2001)

and Aghion et al. (2005) is a simplified version of the one in Atkeson and Burstein (2008), which the current paper mimics.

I design this new specification of demand because commonly used demand systems under monopolistic competition do not satisfy these properties very well. The benchmark CES does not feature heterogeneous demand elasticities. Translog preference (Feenstra (2003)) and Generalized CES (Arkolakis et al. (2019)) imply counterfactually *higher* passthroughs for more productive firms. Even though the quadratic preference (Melitz and Ottaviano (2008)) and the Klenow and Willis (2016) specification of Kimball satisfy the first two properties, they imply a bounded firm size when $a_\theta \to +\infty$ and hence are unable to match Pareto tails. Moreover, they imply excessively high markups for productive firms from an empirical point of view.

As my specification of Kimball is in the same spirit as Atkeson and Burstein (2008), which relates a firm' market power to its market share and features bounded demand elasticities, it is of no surprise that it satisfies the properties of markup, passthrough and large firm asymptotics like Atkeson and Burstein (2008).[48] The oligopolistic competition setting of Atkeson and Burstein (2008), however, implies strategic interactions between firms. When growth decisions are involved, as in my paper, such interactions make even numerical solutions intractable. Thus my demand system can be seen as an approximation of Atkeson and Burstein (2008) under monopolistic competition. The basic assumption of monopolistic competition, which allows individual firms to interact with the whole distribution but not with specific firms, matches exactly the setting of Mean Field Games which simplifies the model by assuming infinitesimal agents/firms.

The specification is also reasonable from an empirical point view. Figure 6 shows labor shares across US manufacturing establishments, where each point represents a bin of employment size range. Recall that I have abused the notation and rearranged firms in the static setting so that θ is the share of bottom firms. $\omega = -\log(1-\theta)$ is used as index instead of θ to take into account the Pareto-tail nature of the distributions,[49] with a larger value corresponding to a larger firm. Labor share decreases with establishment size in the Census of Manufacturing with possible bounds on the two extremities (see Appendix I for the data table). A logistic function is a natural choice to capture this feature.

[48] The only concern is increasing passthrough when productivity is excessively large. This feature also exists in Atkeson and Burstein (2008): in this case, the firm is effectively the monopoly in its sector and competes with other sectors as if in a monopolistic competition between sectors. Since the sectoral aggregator is a CES, the passthrough approaches 1 when the firm's productivity tends to $+\infty$.

[49] See Appendix I.1 for detailed discussions.

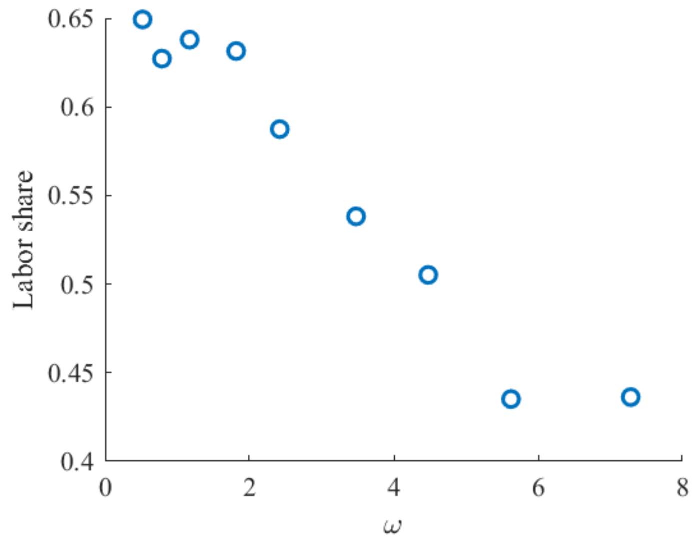


Figure 6: Labor share in a cross section of establishments, US Census of Manufacturing 2002 corrected by KLEMS and the Annual Survey of Manufacturing. $\omega = -\log(1-\theta)$ to take into account the Pareto tail. Each dot is the average labor share of firms within the employment distribution bin $\omega \in [\omega_i, \omega_{i+1}]$.

4.2 Exit and Entry

4.2.1 Exit

Incumbents face idiosyncratic shocks of log productivity modelled as Brownian motion with a standard deviation of ν, regardless of their current level of productivity.[50] The manifestation of this productivity shock in labor, output, sales or profits depends on the demand elasticity faced by the firm. Sales are more volatile for smaller firms with higher demand elasticities, and vice versa for larger firms. I assume that there is a cutoff log-point change of sales $-\kappa$, where $\kappa > 0$, such that if the decrease in sales per period exceeds κ, the firm dies. The death rate $\lambda_{d,\theta}$ is thus monotonically decreasing in firm size, which is consistent with the empirical evidence shown in Figure 7. A demand system with variable demand elasticities, in addition to the benefits discussed in Section 4.1.3, also lends itself to heterogeneous death rates.

Formally,[51]

$$\lambda_{d,\theta} = \lambda_d(a_\theta; \phi^M) = \mathbb{P}\Big(\frac{\partial \log(P_\theta Y_\theta)}{\partial a_\theta} \cdot \nu B_1 < -\kappa\Big) \tag{20}$$

where $B_1 \sim \mathcal{N}(0,1)$.

A second source of exit concerns only firms close to the left boundary $\underline{a}$. A negative shock of

[50] Put in discrete terms, during each period a productivity shock of $\mathcal{N}(0, \nu^2)$ is realized at the log level.

[51] To incentivize the definition, I have stayed slightly outside the continuous-time setting and considered its discrete counterpart. Once definition 20 is put down, however, it only depends on the variables at time t and is equally suitable for the continuous-time setting.

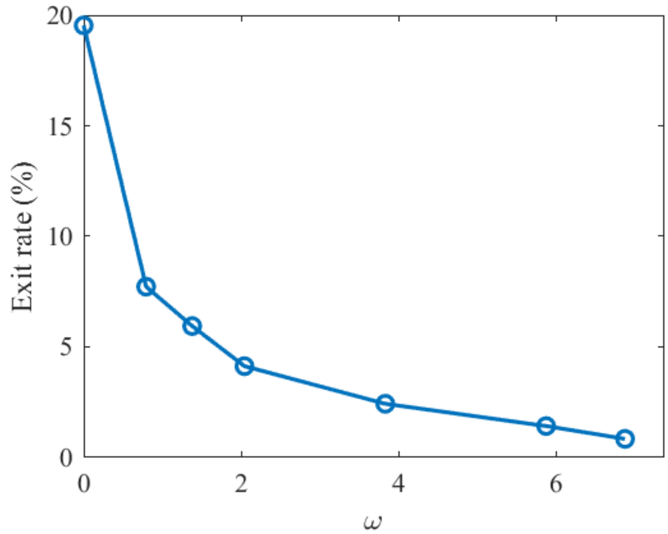


Figure 7: Exit rate in a cross section of establishments, average of US Business Dynamic Statistics 1978-1982, all sectors. $\omega = -\log(1-\theta)$ to take into account the Pareto tail.

productivity can bring these firms to the left of $\underline{a}$ and hence force them to exit. The total number of firms dying from this second channel is $M\frac{\nu^2}{2}\frac{\mathrm{d}\phi}{\mathrm{d}a}(\underline{a})$ per period. See Appendix F.2 for the proof.

4.2.2 Entry

At any point in time, there are M_e potential entrants or outsiders. Each of them invests in learning and enters the market if the learning is successful. If it is not successful, the opportunity to enter lapses and a new set of M_e outsiders prepares for entry. Outsiders are assumed to be equipped with an initial productivity of $\underline{a}$: since any log productivity below $\underline{a}$ is not profitable to exploit, I assume them to be freely available for any firm including outsiders. The learning cost is assumed to be quadratic with respect to the Poisson success rate:

$$R_e = \frac{1}{2}\alpha_e \lambda_e^2 A \tag{21}$$

where λ_e is the success rate. If the entry is successful, the outsider becomes an incumbent according to a matching procedure. A firm with log productivity a_j is randomly drawn from the log productivity distribution of existing incumbents ϕ. The entrant enters at position $a_\psi \in (\underline{a}, a_j]$ with probability

$$\psi(a_\psi; \underline{a}, a_j) = k_\psi e^{-k_\psi(a_\psi - \underline{a})} + e^{-k_\psi(a_j - \underline{a})}\delta_{a_j}(a_\psi) \tag{22}$$

where δ_{a_j} is the Dirac mass function at point a_j and $k_\psi > 0$ suggests imperfect learning. Figure 8 illustrates two matchings, each with a different incumbent. Section 4.2.3 explains the intuition behind the specification by its discrete counterpart. I have included $\underline{a}$ in the notation $\psi(a_\psi; \underline{a}, a_j)$ to emphasize that entrants come from the initial position $\underline{a}$.

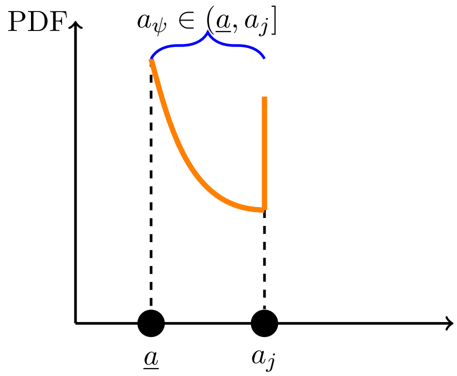


Figure 8: Illustration of an entrant being matched with an incumbent a_j or a_j'. The vertical line at a_j or a_j' illustrates the Dirac mass.

Taking into account all possible matchings, entry occurs at position a_ψ with probability distribution:

$$\psi_e(a_\psi;\phi) = \int_{\underline{a}}^{+\infty} \psi(a_\psi;\underline{a},a_j)\phi(a_j)\,\mathrm{d}a_j \tag{23}$$

$$= [k_\psi\overline{\Phi}(a_\psi) + \phi(a_\psi)]e^{-k_\psi(a_\psi-\underline{a})} \tag{24}$$

where $\overline{\Phi}(a_\psi) = 1-\Phi(a_\psi) = \mathbb{P}(a_\theta > a_\psi)$ is the survival function of the log productivity distribution. Moreover, $\psi_e(a_\psi;\phi) = \frac{\mathrm{d}}{\mathrm{d}a_\psi}\left[1-\overline{\Phi}(a_\psi)e^{-k_\psi(a_\psi-\underline{a})}\right]$ so that $\overline{\Phi}_e(a_\psi) = \overline{\Phi}(a_\psi)e^{-k_\psi(a_\psi-\underline{a})}$ is the survival function of entrants' log-productivity distribution. It is easy to see that $\overline{\Phi}_e(a_\psi) \leq \overline{\Phi}(a_\psi)$, $\forall a_\psi \geq \underline{a}$, i.e. Φ first-order stochastically dominates Φ_e. We will shortly compare the right tail of the two distributions.

4.2.3 Properties of the Entry Specification

The functional form of equation 22 can be understood intuitively by resorting to a discrete setting. It is simply a succinct description of step-by-step learning *with commitment* under the continuous setting. Consider its discrete counterpart illustrated in Figure 9 to appreciate this point.[52] Discretize the segment between $\underline{a}$ and a_j into equally-spaced N steps. Define $p = e^{-\frac{1}{N}k_\psi(a_j-\underline{a})}$ and consider the learning process as following a geometric distribution with probability p. In other words, a potential entrant starts from position $\underline{a}$ and goes up step by step, each step with success probability p. If the potential entrant has successfully gone through all the N steps, it has learned all that can be learned from firm a_j and enters at position a_j, i.e. the incumbent a_j sets the upper bound for learning. The probability of entering at position n for $n < N$ is $p^n(1-p)$. Using the fact that at position n, $a_\psi = \underline{a} + \frac{n}{N}(a_j - \underline{a})$, we have

$$p^n(1-p) = e^{-k_\psi(a_\psi-\underline{a})}(1-e^{-\frac{1}{N}k_\psi(a_j-\underline{a})}) \tag{25}$$

$$\approx k_\psi e^{-k_\psi(a_\psi-\underline{a})}\cdot\frac{1}{N}(a_j-\underline{a}) \tag{26}$$

[52]See also König et al. (2016) for the discrete specification.

which is just the continuous part of equation 22 times the step size. The Dirac mass of equation 22 simply says that if the potential entrant has gone through all N steps, it enters at position a_j. In fact, in the discrete case such a probability is p^N, which is exactly the coefficient before the delta function.

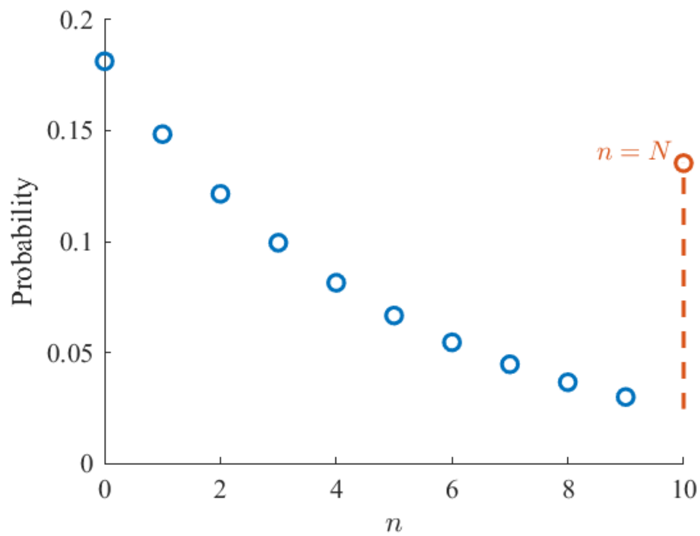


Figure 9: Discrete counterpart of an entrant learning from a_j. $n = 0$ corresponds to $a_\psi = \underline{a}$, and $n = N$ to $a_\psi = a_j$. For illustration purposes with $\underline{a} = -1$, $k_\psi = 1$, $a_j = 1$ and $N = 10$.

In most papers with learning such as Lucas and Moll (2014), a learning firm jumps *with certainty* to a_j if it has been successfully matched with a_j, which corresponds to $k_\psi = 0$ in equation 22. Why have I specified imperfect learning with $k_\psi > 0$, instead of the seemingly easier assumption of perfect learning with $k_\psi = 0$?

The reason is that $k_\psi > 0$ not only is empirically plausible, but also eliminates multiple tails of the solution to the KF equation. Consider a specific example in which the productivity is Pareto distributed, or equivalently its log is exponentially distributed: $\phi(a) = k_\phi e^{-k_\phi(a-\underline{a})}$, $a \geq \underline{a}$. Then $\psi_e(a_\psi; \phi) = (k_\phi + k_\psi)e^{-(k_\phi+k_\psi)(a_\psi-\underline{a})}$, which has the same tail as ϕ if $k_\psi = 0$ and a thinner tail if $k_\psi > 0$. Thinking about the evolution of productivity distribution over time, during any time period $\mathrm{d}t$, a total mass of $M_e\lambda_e\mathrm{d}t$ entrants with PDF ψ_e are injected into the market, so that the incumbent distribution ϕ is linearly combined with the entrant distribution to form a new distribution. When two exponential distributions are linearly combined, one with a heavier tail than the other, the heavier tail always dominates in the combined distribution. Hence, if $k_\psi = 0$ and if the model starts from an initial condition with a sufficiently heavy tail, the injected distribution keeps interfering with the incumbent distribution, preventing the latter from evolving to a lighter tail. The model then features multiple equilibria with the equilibrium distribution depending on the initial condition of ϕ like in Lucas and Moll (2014). Adding the Brownian motion as in Luttmer

(2012) and Perla et al. (2021) does not essentially make the equilibrium distribution unique, as both papers require the initial condition to be thin-tailed enough to avoid the interference.[53] While the assumption is mild for the objective of these papers, it amounts to assuming what the present paper seeks to explain. On the other hand, if $k_\psi > 0$, the entrant distribution has a lighter tail than the incumbent distribution, making sure that it does not interfere with the tail of the latter and allows the incumbent tail to be shaped by other forces. While multiple equilibria in the case of $k_\psi = 0$ is not a problem per se, the above argument shows that it is the result of a cutting-edge assumption unlikely to hold empirically.

In fact, the non-interference condition under $k_\psi > 0$ is valid for virtually any distribution ϕ one can think of with an infinite right endpoint, proved in the following proposition.[54] The condition will be the basis for proving the unique tail index of the KF equation and is a precondition for the numerical algorithm to start from *any* initial condition of ϕ.[55]

Proposition 3 (Non-interference Condition of Entry)**.** *If ϕ is ultimately monotone (i.e. ϕ is monotone on $[\hat{a}, +\infty)$ for some $\hat{a}$), then $\forall k_s \in (-\infty, k_\psi)$, $\lim_{a_\psi \to +\infty} \frac{\psi_e(a_\psi;\phi)}{\phi(a_\psi)} e^{k_s a_\psi} = 0$.*
In particular, if $k_\psi > 0$, then $\lim_{a_\psi \to +\infty} \frac{\psi_e(a_\psi;\phi)}{\phi(a_\psi)} = 0$ by setting $k_s = 0$.

Sketch of proof. By definition,

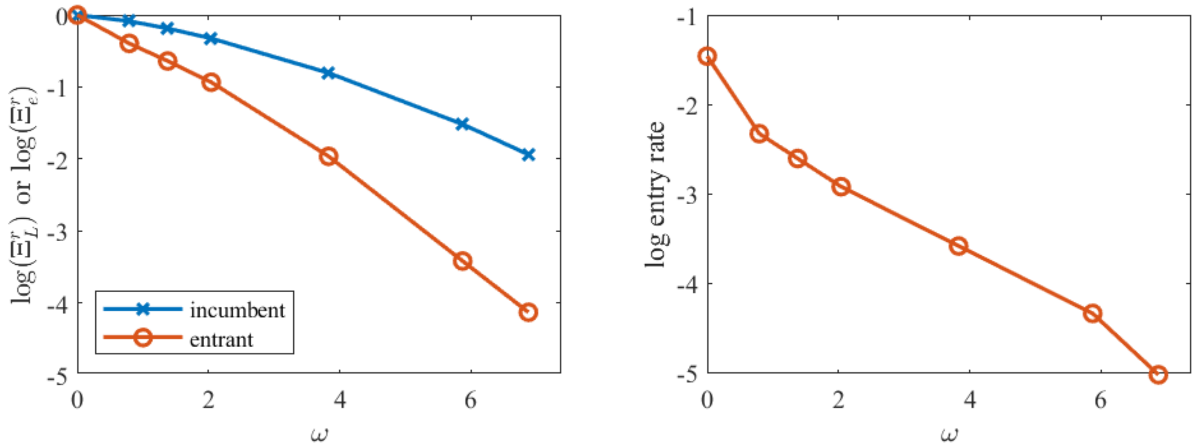
$$\frac{\psi_e(a_\psi;\phi)}{\phi(a_\psi)} e^{k_s a_\psi} = \left[k_\psi \frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)} + 1\right] e^{-(k_\psi - k_s)a_\psi} e^{k_\psi \underline{a}} \tag{27}$$

Hence it suffices to prove $\lim_{a_\psi \to +\infty} \frac{\phi(a_\psi)}{\overline{\Phi}(a_\psi)} e^{(k_\psi - k_s)a_\psi} = +\infty$ if $k_\psi - k_s > 0$, i.e. the hazard rate $\frac{\phi(\cdot)}{\overline{\Phi}(\cdot)}$ cannot decrease more quickly than any exponential function. See Appendix E for the proof. □

The ultimate-monotonicity condition is a very weak condition whose violation virtually never appears in practice. Proposition 4 below shows that any equilibrium distribution ϕ is ultimately monotone. The assumption of imperfect learning ($k_\psi > 0$) is not only theoretically convenient, but also empirically plausible: it simply says that learning has a positive probability of failure, no matter how small it is. US BDS data reports the number of establishments, incumbent employment and job creation from entrants by the size bin of establishments. $k_\psi > 0$ can be validated by a lighter tail of entrants than incumbents. To this end, I define $\omega = -\log(1-\theta) \in [0, +\infty)$ and calculate the complementary cumulated share of employment respectively for incumbents and entrants, $\log(\xi_L^r)$ and $\log(\xi_e^r)$, where r stands for "right" to emphasize the accumulation from the right tail. If incumbent (entrant) employment is Pareto-tailed, then $\log(\xi_L^r)$ ($\log(\xi_e^r)$) should be linear in large

[53]The property of the Fisher-KPP type of KF equation in Luttmer (2012) is well understood; see McKean (1975). In particular, if the initial condition is sufficiently heavy-tailed, multiple equilibria emerge.
[54]As Brownian motion has an unbounded support, we only consider ϕ with an infinite right endpoint.
[55]Luttmer (2020) discusses how bounded learning can ensure the uniqueness of the solution to the KF equation. In the present model, this means $\psi_e(a_\psi;\phi) = 0$ if a_ψ exceeds a certain threshold and that the non-interference condition is trivially satisfied. In Gabaix et al. (2016), the non-interference condition is directly assumed when death and rebirth are introduced as "stabilizing forces".

values of ω. A flatter line suggests a heavier tail.



(a)

(b)

Figure 10: Check $k_\psi > 0$.

Left panel: complementary cumulated share of employment for incumbents and entrants, as a function of $\omega = -\log(1-\theta)$, where the same ω is used for incumbents and entrants of the same bin. Right panel: log entry rate as a function of ω. Entry rate per bin is the number of new jobs created by entrants divided by the number of jobs employed by incumbents for the bin. Data source: Average of US Business Dynamics Statistics 1978-1982, all sectors.

Figure 10a shows that entrants indeed exhibit a lighter tail than incumbents. Another way to see this is to look at job creation rate from birth by the size bin of establishments. If the exponential specification in equation 23 is true and $k_\psi > 0$, then the log job creation rate from birth should be a linear and decreasing function of ω at the right tail. Figure 10b shows this is indeed the case.

4.3 Evolution of Incumbents

Apart from idiosyncratic shocks, incumbents can consciously improve their productivities via innovation on a stand-alone basis. Like in the illustrative model, a firm with log productivity a_θ can improve its log productivity by $q > 0$ by incurring an innovation cost $R_{i,\theta}$.[56] Denote $\lambda_{i,\theta}$ as the Poisson rate of innovation success and assume the innovation cost to be quadratic in it:

$$R_{i,\theta} = R(\lambda_{i,\theta}; \tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta \tilde{a}_\theta} \lambda_{i,\theta}^2 A \tag{28}$$

where $\tilde{a}_\theta = a_\theta - a$ is the relative log TFP of firm θ compared to the aggregate economy. Unless otherwise stated, a tilde will always denote normalization by the aggregate TFP.[57] $\beta > 0$ captures

[56]In a basic Poisson process with success rate λ, the number of arrivals within one unit of time follows a Poisson distribution with parameter λ. Thus a firm can advance multiple steps per unit of time. Moreover, the expected per-period log-productivity improvement is λq: choosing the success rate λ or the step size q are equivalent.

[57]At the original level, such a normalization takes the form of division: $\tilde{A}_\theta = A_\theta / A$.

the fact that ideas are harder to find at higher levels of productivity (Jones (1995), Bloom et al. (2020)). It governs heterogeneous research difficulties in a cross section of firms at a specific time, while α governs the general difficulty for all firms over time. During each period, α is given a constant value for solving a balanced growth path. Comparative statics (i.e. 1980 vs 2020) are conducted between two balanced growth paths, with a higher value of α in the more recent period to reflect harder research.

Taken together, firm θ's log productivity a_θ evolves according to the following stochastic process:

$$\mathrm{d}a_{\theta,t} = -\frac{1}{2}\nu^2 \mathrm{d}t + \nu \mathrm{d}B_{\theta,t} + \mathrm{d}J_{\theta,t} \tag{29}$$

where $-\frac{1}{2}\nu^2$ is a normalization for the Brownian motion to ensure that $A_{\theta,t}$ does not change in expectation due to idiosyncratic noises, and $B_{\theta,t}$ is Brownian motion independent across θ. $J_{\theta,t}$ is a jump term and consists of two parts: death and innovation. If death is realized with Poisson intensity $\lambda_{d,\theta,t}$, then $a_{\theta,t}$ jumps directly to $-\infty$ and firm θ exits. If innovation is realized with intensity $\lambda_{i,\theta,t}$, then the log productivity jumps up by q.

4.4 Innovation Decisions and Evolution of Productivity Distribution

Given the building blocks in previous sections, each incumbent firm solves its optimal innovation decision by a Hamilton-Jacobi-Bellman (HJB) equation, and each potential entrant solves its optimal entry. The productivity distribution evolves based on these optimal decisions via a Kolmogorov forward (KF) equation.

Each incumbent's infinite-horizon optimisation problem can be formulated as an HJB equation:

$$\begin{aligned} r_t V(a_\theta, t) = \max_{\lambda_i} \Big\{ &\Pi(a_\theta; \phi_t^M) - f A_t - R_t(\lambda_i; \tilde{a}_\theta) - \frac{1}{2}\nu^2 \frac{\partial V}{\partial a}(a_\theta, t) + \frac{1}{2}\nu^2 \frac{\partial^2 V}{\partial a^2}(a_\theta, t) \\ &+ \lambda_i \big[V(a_\theta + q, t) - V(a_\theta, t)\big] - \lambda_d(a_\theta; \phi_t^M) V(a_\theta, t) + \frac{\partial V}{\partial t}(a_\theta, t) \Big\} \end{aligned} \tag{30}$$

where

$$R_t(\lambda_i; \tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta \tilde{a}_\theta} \lambda_i^2 A_t, \tag{31}$$

r_t is the interest rate, $\phi_t^M = M_t \phi_t$ is the generalized PDF of log productivity at time t, Π is the per-period production profit which depends on the firm's log productivity a_θ and the market structure ϕ_t^M, $\lambda_d(a_\theta; \phi_t^M)$ is the death rate defined in equation 20. See Appendix F.3 for derivations of the HJB equation.

Equation 30 has an intuitive interpretation: the per period flow value should be equal to the sum of current-period profit net of fixed cost and innovation cost, plus a change of value due to the normalizing drift $-\frac{1}{2}\nu^2$, plus a second order derivative due to idiosyncratic Brownian motion, plus an increase in value due to innovation, plus a decrease in value due to death and plus a change of value function over time.

The first order condition of innovation, which describes its optimal intensity, is:

$$\lambda_i^*(a_\theta, t) = \frac{1}{\alpha A_t} e^{-\beta \tilde{a}_\theta} \big[V(a_\theta + q, t) - V(a_\theta, t) \big] \tag{32}$$

I have introduced in Section 4.1.2 the left boundary $\underline{a}_t$ below which a firm dies immediately, i.e. an "absorbing barrier" on the left:[58]

$$V(\underline{a}_t, t) = 0 \tag{33}$$

For entry, each potential entrant solves the maximisation problem:

$$\max_{\lambda_e} \Big\{ -R_{e,t}(\lambda_e) + \lambda_e \int_{\mathbb{R}} V(a_\psi, t) \psi_e(a_\psi; \phi_t) \, \mathrm{d}a_\psi \Big\} \tag{34}$$

where

$$R_{e,t}(\lambda_e) = \frac{1}{2} \alpha_e \lambda_e^2 A_t \tag{35}$$

The maximisation problem involves only the instant period as the opportunity to enter lapses if not taken advantage of.

The first order condition of entry, which gives its optimal intensity, is:

$$\lambda_e^*(t) = \frac{1}{\alpha_e A_t} \int_{\mathbb{R}} V(a_\psi, t) \psi_e(a_\psi; \phi_t) \, \mathrm{d}a_\psi \tag{36}$$

Given the choice of λ_i and λ_e, the Kolmogorov Forward (KF) equation describes the evolution of the log-productivity distribution $\phi^M(a_\theta, t)$:

$$\begin{aligned} \frac{\partial \phi^M}{\partial t}(a_\theta, t) = -\frac{\partial[-\frac{1}{2}\nu^2 \phi^M]}{\partial a}(a_\theta, t) + \frac{1}{2}\nu^2 \frac{\partial^2 \phi^M}{\partial a^2}(a_\theta, t) + \lambda_e(t) M_e \psi_e(a_\theta; \phi_t) - \lambda_d(a_\theta; \phi_t^M) \phi^M(a_\theta, t) \\ + \lambda_i(a_\theta - q, t) \phi^M(a_\theta - q, t) - \lambda_i(a_\theta, t) \phi^M(a_\theta, t) \end{aligned} \tag{37}$$

See Appendix F.4 for a proof of the KF equation based on the law of motion of firms. The terms have intuitive interpretations: apart from the usual terms for drift and Brownian motion, $\lambda_e(t) M_e \psi_e(a_\theta; \phi_t)$ describes the change of distribution due to entry, $-\lambda_d(a_\theta; \phi_t^M)\phi^M(a_\theta, t)$ describes the first type of firm death due to sales shocks,[59] and the last two terms describe the change of

[58] At the algorithm level, solving HJB 30 numerically requires another boundary condition on the right as a computer can only handle finite segments. A natural choice for the right boundary $\overline{a}_t$ is a "reflecting barrier", i.e. $a_{\theta,t}$ is brought back to $\overline{a}_t$ if it goes above $\overline{a}_t$. Formally, $\frac{\partial V}{\partial a}(\overline{a}_t, t) = 0$. Note that the reflecting barrier is *not* used for stabilizing the productivity distribution. As shall be seen, the most productive firms have growth rates lower than average, which already stabilizes the distribution.

[59] The second type of firm death near the left boundary is incorporated into the boundary condition.

distribution due to innovation. Like in the illustrative model, $\lambda_i(a_\theta - q, t)\phi^M(a_\theta - q, t)$ captures the inflow into position a_θ from position $a_\theta - q$ due to successful innovation from the latter, while $\lambda_i(a_\theta, t)\phi^M(a_\theta, t)$ captures the outflow from position a_θ. The boundary condition for the left absorbing barrier is:[60]

$$\phi^M(\underline{a}_t, t) = 0 \tag{38}$$

For notational elegance and computational simplicity, it is useful to define a functional operator $\mathcal{F}_t$ and write both the HJB and the KF in terms of it:

$$\mathcal{F}_t V(a_\theta, t) = -\frac{1}{2}\nu^2 \frac{\partial V}{\partial a}(a_\theta, t) + \frac{1}{2}\nu^2 \frac{\partial^2 V}{\partial a^2}(a_\theta, t) + \lambda_i \big[V(a_\theta + q, t) - V(a_\theta, t)\big] - \lambda_d(a_\theta; \phi_t^M) V(a_\theta, t) \tag{39}$$

HJB 30 and KF 37 can now be written as:

$$r_t V(a_\theta, t) = \max_{c,\, \lambda_c} \Big\{ \Pi(a_\theta; \phi_t^M) - f A_t - R_t(\lambda_c; \tilde{a}_\theta) + \mathcal{F}_t V(a_\theta, t) + \frac{\partial V}{\partial t}(a_\theta, t) \Big\}$$

$$\frac{\partial \phi^M}{\partial t}(a_\theta, t) = \mathcal{F}_t^\intercal \phi^M(a_\theta, t) + \lambda_e(t) M_e \psi_e(a_\theta; \phi_t)$$

where $\mathcal{F}_t^\intercal$ is the adjoint operator of $\mathcal{F}_t$. See Appendix F.5 for the proof of adjointness. In numerical calculations where $\mathcal{F}_t$ is discretized into a matrix, the discrete counterpart of $\mathcal{F}_t^\intercal$ is then the transpose of this matrix. Hence the notation $\intercal$ for representing the adjointness.

4.5 Travelling Wave Equilibrium

We will focus on the balanced growth path in which the log-productivity distribution keeps its shape and travels at a constant speed, i.e. a travelling-wave solution. As aggregate productivity is a homothetic aggregator of firms' productivities (Proposition 1), aggregate growth is the same as the travelling wave speed. After normalization by the aggregate growth, the log-productivity distribution becomes stationary and all the macroeconomic variables are constant. For the balanced growth path to exist, interest rate r, innovation cost parameter α, and entry cost parameter α_e are kept constant. The comparative statics of different historical periods will be based on comparing different balanced growth paths with different sets of parameters.

The following definition summarizes the equations constituting a travelling wave equilibrium. Variables are normalized by aggregate growth, where a tilde denotes normalization.

Definition 1 (Travelling Wave Equilibrium)**.**
Denote $\tilde{\mathcal{F}}$ to be the functional operator:

$$\tilde{\mathcal{F}} \tilde{V} = \Big[-\frac{1}{2}\nu^2 - g\Big] \tilde{V}'(\tilde{a}_\theta) + \frac{1}{2}\nu^2 \tilde{V}''(\tilde{a}_\theta) + \lambda_i(\tilde{a}_\theta)\big[\tilde{V}(\tilde{a}_\theta + q) - \tilde{V}(\tilde{a}_\theta)\big] - \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M)\tilde{V} \tag{40}$$

[60] At the algorithm level, the boundary condition for the right reflecting barrier is $\frac{1}{2}\nu^2\phi^M(\overline{a}_t, t) + \frac{1}{2}\nu^2 \frac{\partial \phi^M}{\partial a}(\overline{a}_t, t) = 0$, where I have not removed $\frac{1}{2}\nu^2$ to make it consistent with the general form in the case of reflecting barriers.

Then incumbent HJB in the case of a travelling wave is:

$$(r-g)\tilde{V} = \max_{\lambda_i}\Big\{\Pi(\tilde{a}_\theta;\tilde{\phi}^M) - f - \tilde{R}(\lambda_i;\tilde{a}_\theta) + \tilde{\mathcal{F}}\tilde{V}\Big\} \tag{41}$$

with boundary condition

$$\tilde{V}(\underline{\tilde{a}}) = 0 \tag{42}$$

where

$$\tilde{R}(\lambda_i;\tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta\tilde{a}_\theta}\lambda_i^2 \tag{43}$$

The entry decision is:

$$\max_{\lambda_e}\Big\{-\tilde{C}_e(\lambda_e) + \lambda_e\int_{\mathbb{R}}\tilde{V}(\tilde{a}_\psi)\psi_e(\tilde{a}_\psi;\tilde{\phi})\,\mathrm{d}\tilde{a}_\psi\Big\} \tag{44}$$

where

$$\tilde{C}_e(\lambda_e) = \frac{1}{2}\alpha_e\lambda_e^2 \tag{45}$$

The KF equation is equivalent to two equations, one for $\tilde{\phi}$

$$0 = \tilde{\mathcal{F}}^{\intercal}\tilde{\phi}(\tilde{a}_\theta) + \frac{\lambda_e M_e}{M}\psi_e(\tilde{a}_\theta;\tilde{\phi}) \tag{46}$$

with boundary condition

$$\tilde{\phi}(\underline{\tilde{a}}) = 0 \tag{47}$$

and another one for M which represents balanced entry and exit:

$$\lambda_e M_e = M\Big[\int_{\mathbb{R}}\lambda_d(\tilde{a}_\theta;\tilde{\phi}^M)\tilde{\phi}(\tilde{a}_\theta)\,\mathrm{d}\tilde{a}_\theta + \frac{\nu^2}{2}\tilde{\phi}'(\underline{\tilde{a}})\Big] \tag{48}$$

The left boundary $\underline{\tilde{a}}$ is pinned down by the zero-profit condition:

$$\Pi(\underline{\tilde{a}};\tilde{\phi}^M) = f \tag{49}$$

The travelling wave equilibrium (i.e. balanced growth path) is defined by equations 40-49, and aggregate growth g which, after solving the KF equation, normalizes aggregate productivity A to 1.[61]

See Appendix F.6 and F.7 for the proof of the normalizations and Appendix H for the computational algorithm. Compared to the general setting, all t indices and partial derivatives with respect to t are dropped, as the problem is time-invariant after normalization by the aggregate growth. Two $-g$ terms appear in HJB 41, one on the left hand side and another one in $\tilde{\mathcal{F}}\tilde{V}$. The first one is due to profits growing at the growth rate g. When discounted at the interest rate r, it is as if the profits do not grow while the interest rate decreases by g. The second $-g$ term comes from the fact that

[61] At the algorithm level, the right reflecting barrier implies the boundary conditions $\tilde{V}'(\bar{\tilde{a}}) = 0$ for the HJB and $\frac{1}{2}\nu^2\tilde{\phi}(\bar{\tilde{a}}) + \frac{1}{2}\nu^2\tilde{\phi}'(\bar{\tilde{a}}) = 0$ for the KF.

we are now reasoning in term of the normalized log productivity, whose stochastic process features an additional $-g\mathrm{d}t$ term compared to equation 29.

The following proposition is one of the most important results of this paper: the equilibrium log-productivity distribution $\tilde{\phi}$ features a unique exponential tail, regardless of the initial condition for solving the KF equation. Consequently, productivity, sales and employment are all Pareto-tailed with a unique Pareto index. Moreover, there is a one-to-one correspondence between lower aggregate growth and a fatter Pareto tail, regardless of market power. Compared to the existing literature, we can link aggregate growth to rising tail concentration in Figure 1, without resorting to additional assumptions about the initial distribution when solving the KF equation.

Proposition 4 (Unique Tail of Equilibrium Distribution)**.**
If $\beta > \underline{\sigma} - 1$ and $k_\psi > 0$, then

1. *Any solution $\tilde{\phi}$ to KF equation 46 must be ultimately monotone.*

2. *$\tilde{\phi}$ has an exponential tail with unique tail index $-\xi$, where ξ is the unique negative root of*

$$-\frac{1}{2}\nu^2\xi^2 - (\frac{1}{2}\nu^2 + g)\xi + \lambda_d(\infty) = 0 \tag{50}$$

   and where $\lambda_d(\infty) = \Phi_{\mathcal{N}}(-\frac{\kappa}{\nu(\underline{\sigma}-1)})$ and $\Phi_{\mathcal{N}}$ is the CDF of standard normal distribution $\mathcal{N}(0,1)$. In other words, the equilibrium productivity distribution has a unique Pareto tail index of $-\xi$. Moreover, equilibrium sales and employment distributions have a unique Pareto tail index of $-\xi/(\underline{\sigma}-1)$.

3. *$-\xi$ is strictly increasing in g, i.e. lower growth is associated with fatter Pareto tails.*

See Appendix G for the proof. The non-interference condition under the imperfect learning assumption $k_\psi > 0$ (Propositions 3) is key for ensuring the uniqueness, as the proof reveals.[62] In the calibrated model, a uniform increase in research difficulty α across all firms decreases aggregate growth and Pareto-tail indices, i.e. tails become fatter. The mechanism is similar to the one in the illustrative model and will be discussed with the numerical results in Section 6.

For the period before 1980, we will consider economic integration as an additional economic force on top of harder research. Economic integration has two effects: an increase in market size L, and an increase in the number of firms M. The latter is pinned down by the balanced entry/exit condition with an increase in the potential number of entrants M_e. The following proposition analyzes the first effect of market size L expansion, keeping M_e unchanged. It shows that expanding the market size can annihilate the harder research if the magnitudes of both changes happen to be the same:

[62]Note that Proposition 4 only ensures the uniqueness of the solution to the KF equation, but not the uniqueness of the whole mean field game system. The latter remains an open question, even though numerical solutions have been reassuringly unique. Moreover, the existence of a solution to the KF equation has not been analytically proved, though the numerical algorithm has consistently found solutions.

Proposition 5 (Indistinguishable Economies). *An economy with labor L, fixed cost f, innovation cost parameter α and entry cost parameter α_e is indistinguishable from another economy with parameters NL, Nf, $N\alpha$ and $N\alpha_e$ ($N > 0$), keeping all the other parameters the same, except that all the firms and their value functions in the second economy are N times large.*

The indistinguishability is in both the static and the dynamic senses; see Appendix F.8 for a proof by verification.[63] Entry cost parameter α_e will be calibrated to $\alpha e^{\beta \underline{\tilde{a}}}$ to reflect the difficulty of knowledge, so that its change is tied to α. The proposition says that more difficult research can be compensated by a larger market, as firms can earn higher profits with a larger market size L to incentivize innovation. The pro-growth effect of a larger market size should be distinguished from semi-endogenous growth models in which population growth cancels the harder research and maintains constant growth. In the models summarized by Jones (2005), for instance, population growth allows for more people to be allocated as researchers through the binding resource constraint. In the present paper, by contrast, it is by higher profits incentivising Schumpeterian growth. The pro-growth effect of a larger market size is reminiscent of the argument in Chandler (1990) that firms expand their geographical reach as a growth strategy.

It should not be surprising that the proposition yields a clean result when the static fixed cost f is assumed to change by the same proportion. As research features economies of scale, research costs can be seen as a dynamic fixed cost. Both the dynamic and the static fixed costs are diluted by a larger market size. For the numerical exercise on the pre-1980 period, we shall keep the static fixed cost constant like in the international trade literature so that firms benefit from (static) economies of scale when the market size increases. We increase L and M_e by the same proportion to reflect economic integration. The increase in M_e endogenously increases the number of incumbent firms M and introduces a competition effect. The implications of economic integration in addition to harder research will be discussed in Section 7.

5 Calibration

Comparative statics in Section 6 will be between 1980 and 2020, and in Section 7 between 1960 and 1980. The choice of 1980 as the watershed is motivated by most papers in the literature which investigate declining growth, increasing concentration and market power, declining labor share, and diminishing business dynamism for the post-1980 period. Such a focus in the literature is partly driven by data availability, as US Business Dynamics Statistics (BDS) has only been available since 1978, and partly by data patterns, as the declining labor share and increasing markup have been most apparent since the 1980s.

Most parameters remain unchanged from 1980 to 2020, except the interest rate r, research cost parameter α, and entry cost parameter α_e which is intimately tied to α. Table 3 summarizes all

[63]In mathematical terms, we are looking for a "characteristic line" of the model.

the parameters for 1980 and 2020.

Parameter	Description	Value	Source
$\overline{\sigma}$	Upper bound of demand elasticity	6.5	External
$\underline{\sigma}$	Lower bound of demand elasticity	1.93	Estimated from US Census
k	Transition speed of demand elasticity	0.32	Estimated from US Census
$\alpha_{t=2020}/\alpha_{t=1980}$	Change in research cost	6.84	Estimated from US aggregate
β	Harder research for higher TFP	1.6	Estimated from Compustat
$r_{t=1980}$	Interest rate	0.0469	External
$r_{t=2020}$	Interest rate	0.0109	External
L	Labor force	1	Normalization
$\alpha_{e,t}$	Research cost of entrants	$\alpha_t \underline{\tilde{A}_t}^{\beta}$	Model assumption
$\alpha_{t=1980}$	Research cost parameter in 1980	557	Structural estimation
q	Innovation step	0.201	Structural estimation
ν	Std of idiosyncratic shocks	0.118	Structural estimation
κ/ν	Relative death threshold	2.99	Structural estimation
M_e	Mass of potential entrants	0.71	Structural estimation
f	Fixed cost	0.104	Structural estimation
k_ψ	Learning parameter	1.37	Structural estimation

Table 3: Summary of parameters for 1980 and 2020. Parameters that can change between 1980 and 2020 are indexed by t.

5.1 Demand Parameters

Demand parameters $\underline{\sigma}$, $\overline{\sigma}$ and k can be estimated by labor shares in a cross section of firms. The approach is similar to the Kimball demand estimation of Baqaee et al. (2023), except that they use pass-through instead of labor share, and they estimate the demand non-parametrically while I take advantage of the parametric form. For a few vintages, the US Census Bureau publishes data on payrolls, material costs, sales, value added and number of establishments for each employment size range of manufacturing establishments. I use the 2002 vintage for estimating the demand parameters.[64] A few adjustments of the labor shares are needed before the estimation: (1) the Census Bureau does not collect information on service input and thus does not deduce it from sales when calculating value added, overstating the value added and thus understating the labor share; (2) payroll is narrowly defined in the data and excludes fringe benefits; (3) the model does not feature heterogeneous wages, so the average payroll per employee of each size bin is adjusted to the sectoral average; (4) the model does not feature capital, so the labor share in the data is adjusted to be consistent with the model. See Appendix I.2 and I.3 for detailed discussions.[65]

[64]Using the 2012 vintage, for which the Census Bureau also publishes the information, yields similar results.

[65]See also the appendix in Autor et al. (2020) for the first two points. They show that the labor share in the Census of Manufacturing after these two adjustments is close to that in the National Income and Product Accounts (NIPA). Figure 6 presents the labor share after the first two adjustments.

The adjusted labor share is close to 1 for the smallest firms. As $\overline{\sigma} = \frac{1}{1-\overline{\chi}}$, where $\overline{\chi}$ is the upper bound of the labor share corresponding to the smallest firms, $\overline{\sigma}$ is sensitive to a small change in $\overline{\chi}$ when the latter is close to 1. Thus I have calibrated $\overline{\sigma}$ as 6.5, which is in the middle of the range used in other papers with nested-CES preferences under oligopolistic competition (see Atkeson and Burstein (2008), De Loecker et al. (2021), Gaubert and Itskhoki (2021) and Burstein et al. (2020)). $\underline{\sigma}$ is estimated as 1.93 which corresponds to the lowest labor share of the largest establishments in the data, which is consistent with the estimates of these papers. It remains to estimate k, which governs the transition of demand elasticity as firm size increases. See Appendix I.4 for details.

5.2 Interest Rates and Research Parameters

I follow Liu et al. (2022) in using the US AA corporate bond rate net of current inflation for calibrating interest rates in 1980 and 2020, as it is more relevant than the 10-year treasury rate as firms' discount rate. They are 0.0469 for 1980 and 0.0109 for 2020.

The research cost function 43 implies that $\lambda_c = (\frac{2}{\alpha})^{\frac{1}{2}} e^{-\frac{\beta}{2}\tilde{a}_\theta} \tilde{R}^{\frac{1}{2}}$, i.e. growth increases with research input $\tilde{R}$ but features decreasing returns to scale with the degree of $\frac{1}{2}$. I follow the methodology in Bloom et al. (2020) with decreasing returns to estimate research productivity, i.e.

$$\text{Research productivity} = \text{Growth}/\text{Effective number of researchers}^{\frac{1}{2}} \tag{51}$$

where the effective number of researchers captures the research cost and is measured as R&D expenditures deflated by the average male wage with four years or more of college education. With aggregate data, the decrease in research productivity over time reflects a general increase in α across all firms. α is estimated to increase by 6.84-fold from 1980 to 2020. For β, which governs the harder research in a cross section of firms, I use micro-level evidence and estimate it with Compustat. Following the methodology of Bloom et al. (2020), β is estimated as 1.6.

5.3 Structurally Estimated Parameters

According to Proposition 5, we can normalize the labor force in 1980 as 1 so that the other parameters can be identified. I set α_e to be $\alpha e^{\beta\underline{\tilde{a}}}$ so that entrants' learning cost function also reflects the industry-average knowledge difficulty α and cross-sectional differences $e^{\beta\underline{\tilde{a}}}$. The adjustment factor $e^{\beta\underline{\tilde{a}}}$ reflects the more difficult knowledge at higher levels of technology, independently from the distribution of productivities to be learned from. The content of knowledge is harder if entrants start from a higher level of productivity, irrespective of whether they can be more or less effectively matched with more productive firms.[66] The specification is also consistent with the incumbents'

[66]Despite the same cost function of entry and incumbent innovation, k_ψ and q can adjust the relative effectiveness of one with respect to the other. See Appendix D for a more detailed discussion of the cross-sectional adjustment factor $e^{\beta\underline{\tilde{a}}}$.

learning cost function 52 in the extended model.

The remaining parameters are structurally estimated by targeting data moments in 1980. While a change in one parameter moves virtually all the moments, some moments are especially useful for identifying specific parameters. The structurally estimated parameters and their main identifying moments are summarized in Table 4. $\alpha_{t=1980}$ targets the average TFP growth of 0.95% in the 1970s and 1980s according to BLS-KLEMS data, and q targets the 2.4% share of R&D expenditure over GDP in the National Income and Product Accounts (NIPA) around 1980. Idiosyncratic shock is the predominant reason for the incumbent's job creation and job destruction: ν is thus identified by the job reallocation rate of 29% around 1980 according to the BDS.[67] The travelling wave equilibrium features balanced entry/exit, so I take the average of the entry rate and exit rate around 1980 to be the targets for both moments. The entry rate is measured as the job creation rate from birth and the exit rate as the job destruction rate from death. The average of these around 1980 is 5.6% according to the BDS. By the definition of the first kind of death in equation 20, which is the predominant form of death, the threshold-volatility ratio κ/ν rather than κ is what ultimately matters for the death rate. Thus I use κ/ν to target the death rate of 5.6%. The number of potential entrants M_e targets the entry rate of 5.6%. Per-period fixed cost f determines the lower-bound productivity through the zero-profit condition 49, and thus shapes the demand elasticities and exit rates of the least productive firms. I use it to match the 20% exit rate of the smallest size bin in the BDS, corresponding to the bottom 55% firms in the employment size distribution. Finally, this paper takes a productivity view of firm size distribution. As entrants' size distribution closely match the dying firms' size distribution in the BDS, the contribution of net entry to growth should be tiny from this point of view and for consistency with the model. Without a better measure of the contribution, I estimate k_ψ to match a 0 contribution, as k_ψ shapes the entry distribution.[68]

The parameters are estimated using the Generalized Method of Moments (GMM). Weights are the inverse of the square moments to represent errors in percentage terms, except for the growth contribution of net entry whose weight is the same as that of TFP growth.

5.4 Travelling Wave Equilibrium of 1980

Figure 11 shows the travelling wave equilibrium of 1980 based on calibrated parameters. Figure 11a displays the optimal innovation intensity of incumbent firms featuring an inverted-U relationship with log productivity. As demand elasticity decreases with productivity, a similar inverted-U relationship exists between demand elasticity and innovation intensity, as discussed in Section 3.2. My paper follows Aghion et al. (2001) in understanding competition as the demand elasticity a firm is facing. Even though the inverted-U relationship between competition and innovation is typical

[67]See Appendix H.7 for the computation of the job creation rate and job destruction rate.
[68]See Appendix H.8 for the computation of the net entry contribution to growth.

Parameter	Main Identifying Moment	Moment Data	Moment Model
$\alpha_{t=1980}$	TFP growth rate	0.95%	0.98%
q	R&D share of GDP	2.4%	2.3%
ν	Job reallocation rate	29%	28%
κ/ν	Exit rate	5.6%	5.6%
$M_{e,t=1980}$	Entry rate	5.6%	5.6%
f	Bottom 55% firms' exit rate	20%	20%
k_ψ	Growth contribution of net entry	0%	-0.08%

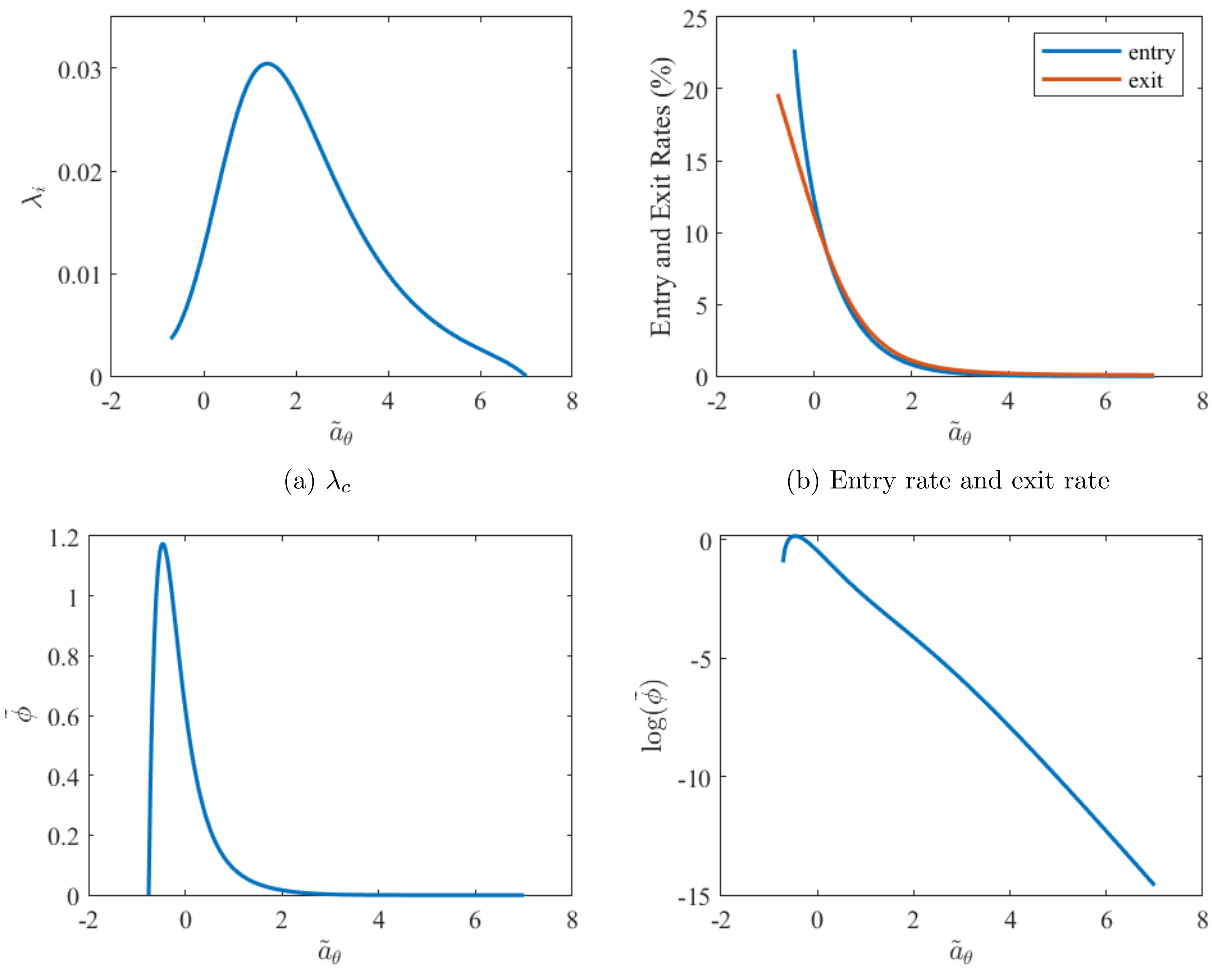
Table 4: Estimated parameters for 1980 by moment matching.

in Schumpeterian growth models, the mechanism is different from that in Aghion et al. (2005). In an economic environment with more elastic demand and hence higher competition, given the same percentage difference in productivity, the firm size difference is larger. The monotonicity of innovation intensity depends on how effectively a more productive firm can benefit from the economies of scale of innovation. Thus more productive firms in a more competitive environment are better able to take advantage of the economies of scale associated with size, so that it dominates the higher marginal innovation cost. Innovation intensity is thus increasing in productivity when demand elasticity is high, and vice versa when demand elasticity is low.[69]

While the higher innovation cost associated with higher productivity is a trait of knowledge, the Schumpeterian incentive works through the lens of market structure and thus depends on demand elasticity. The lower demand elasticity of a larger output can be interpreted as market saturation: this is evidently the case with the linear demand curve in Melitz and Ottaviano (2008) where each variety of goods has an upper bound of demand. The same intuition applies for other demand systems in which demand elasticity decreases with quantity: with a higher consumption in a specific variety of goods, the same reduction of price in percentage terms translates into a lower percentage increase in demand. Market saturation implies that at the right tail, increasing the firm size is not sufficiently effective for benefiting from economies of scale. The harder innovation effect thus dominates and innovation intensity decreases with firm size.

Because the model assumes imperfect learning, the entry rate declines exponentially with log productivity along the distribution in Figure 11b. The exit rate declines in a similar way, as demand elasticity declines with firm size: to knock out a larger firm by sales shock, a larger productivity shock is needed, whose probability of occurrence is lower. Both entry rate and exit rate closely resemble the pattern in the data.

[69]Such an inverted-U between a firm's innovation intensity and relative step size vis-à-vis its competing peer can also be found in models following Aghion et al. (2001), if the innovations of both leaders and laggards are assumed to be step-by-step. This is because the value function has a lower bound and an upper bound, hence close to 0 marginal change at the two extremities, despite their assumption that marginal innovation cost does not increase with technological level.



(a) λ_c

(b) Entry rate and exit rate

(c) $\tilde{\phi}$

(d) $\log(\tilde{\phi})$

Figure 11: Travelling wave equilibrium of 1980.

Notes: $\tilde{a}_\theta$ is firm's log productivity. Aggregate log productivity is normalized to 0. Panel (a): incumbent innovation intensity. Panel (b): entry rate $\frac{\lambda_e M_e \psi_e}{M\tilde{\phi}}$ and exit rate λ_d along the distribution of firms. The entry rate close to $\underline{\tilde{a}}$ is not shown as $\tilde{\phi}$ in the denominator is close to 0. Panel (c): log-productivity distribution. Panel (d): log-productivity distribution in log scale to emphasize the tail.

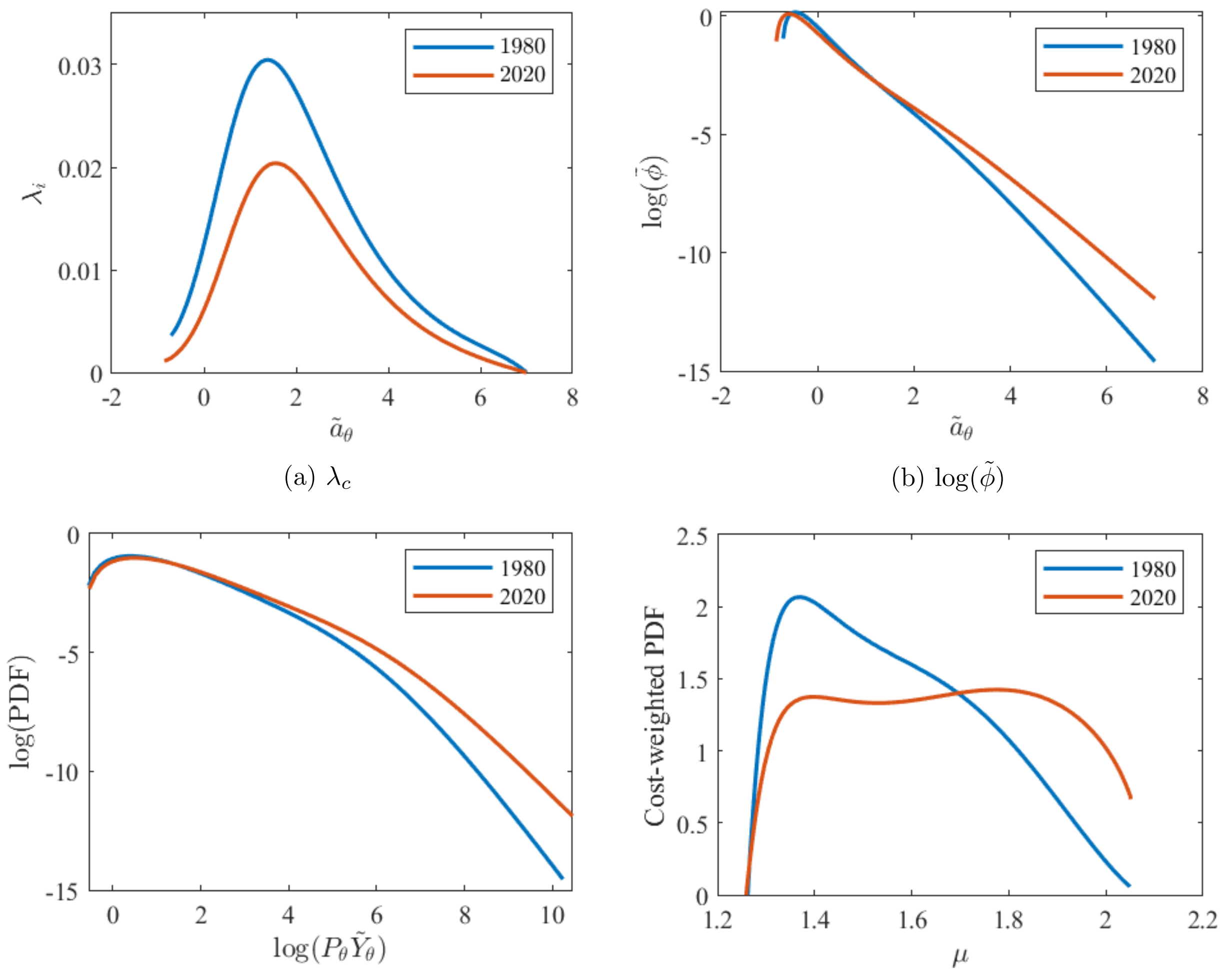
Figure 11c shows the equilibrium log-productivity distribution $\tilde{\phi}$ which has an exponential tail. By construction $\tilde{\phi}(\underline{\tilde{a}}) = 0$ at the left boundary because of absorbing barrier: firms immediately die if touching the boundary. To emphasize the exponential tail, Figure 11d plots the distribution again in log scale, with the right tail being linear. The exponential tail of log-productivity distribution is the same thing as the Pareto tail of productivity distribution, which then translates into Pareto tails of sales and employment distributions through the demand structure (Proposition 2). When the productivity distribution has a fatter tail due to dynamic effects of growth, markup increases because of reallocation towards high markup firms, as will be discussed in Section 6. The log-productivity distribution is log linear at right tail but log concave overall: this feature is shared by many other commonly-used distributions with Pareto tail, such as Fréchet distribution. In a static model with fixed productivity distribution, such log-concavity implies that market size expansion decreases within-firm markup more than it reallocates market share towards high markup firms.[70] If both forces are present, i.e. fatter tail due to dynamic growth effects and market size expansion due to static economic integration, the latter force may compensate higher markup of fatter tail, as will be discussed in Section 7.

6 Comparative Statics When Ideas Are Getting Harder to Find

What happens to growth, concentration, market power, labor share and business dynamics when ideas get harder to find? As explained in Section 5, I calibrate a 6.84-fold increase of α from 1980 to 2020 using the methodology of Bloom et al. (2020). Entrants' learning cost parameter α_e increases accordingly due to its tight relationship with α. The interest rate decreases and promotes growth like in standard macroeconomic models, which partially mitigates the effects of harder research. Since the model is sensitive to interest rate, its effect should be jointly evaluated to speak to the data. Moreover, given the large magnitude of the increase in research difficulty, compensating forces should be expected to exist in a well-functioning economic system. In standard consumption-based asset pricing models, such a decrease in interest rate can be endogenized by a decrease in the growth rate. As the interest rate can decline for reasons other than the growth rate, for instance the global saving glut (Bernanke (2005)), I remain agnostic about the source of interest rate decline and calibrate an exogenous change.

Figure 12a compares the innovation intensities in both periods. Given the *uniform* increase in research difficulty across all firms, the intensity curve shifts down so that aggregate growth declines. The decrease in growth is however larger for firms that previously enjoyed a higher growth: these

[70] If the log-productivity distribution is log linear, i.e. the productivity distribution is exactly Pareto, then the two forces exactly cancel each other so that the aggregate markup does not change. If it is log convex, then between-firm reallocation dominates so that markup increases in a market size expansion. The results hold for a wide range of demand systems with variable demand elasticities; see the appendix in Autor et al. (2020). To the best of my knowledge, I have not seen any log convex assumption in the macroeconomics and trade literature. As the exact Pareto is a cutting-edge case, the endogenously-generated distribution with overall log-concavity in this paper is consistent with the typically assumed distributions in static models.



(a) λ_c (b) $\log(\tilde{\phi})$

(c) log(PDF) of $\log(P_\theta \tilde{Y}_\theta)$ (d) Markup μ

Figure 12: Comparison between 1980 and 2020. Panel (a): innovation intensity; Panel (b): log PDF of log productivity; Panel (c): log PDF of log sales; Panel (d): cost-weighted markup distribution.

firms have a dynamic advantage of growth and are particularly hurt when innovation becomes more difficult. Thus, even though the growth rate of the most productive firms declines in absolute terms, it increases *in relative terms* compared to that of medium-sized firms. This allows leaders to stretch out further to the right tail *relative to* the latter, implying a fatter tail of the productivity distribution and firm size distributions. To emphasize the heaviness of the tail, Figures 12b and 12c plot the productivity distribution and sales distribution in log-log terms. Both distributions show a flatter linear tail in 2020 than in 1980, meaning a fatter Pareto tail in 2020. The higher concentration due to fatter tails translates into a higher markup, as shown in the cost-weighted markup distributions in Figure 12d.

Table 5 summarizes the changes in key macroeconomic moments from 1980 to 2020 in the model and in the data. The model explains a majority of the data changes in all cases despite all these moments being non-targeted. (1) The sales share of the top 1% firms increases as the Pareto tail of the sales distribution fattens. (2) TFP growth declines as harder research dominates all other mitigating forces including the lower interest rate. (3) Because we have assumed a demand system with variable demand elasticities, aggregate markup increases with the tail fatness of the productivity distribution *other things being equal*. (4) The aggregate labor share decreases, as labor is the only factor of production in the model so that labor share is simply the inverse of the markup. (5) Table 6 conducts a Melitz and Polanec (2015) decomposition of markup change in the model into a within-firm component, a reallocation component between firms, and a net entry component. A similar analysis is done for labor share. Both the increase in markup and the decrease in labor share predominantly come from reallocation from low markup (i.e. high labor share) firms to high markup (i.e. low labor share) firms, which is consistent with the micro-level empirical evidence (see De Loecker et al. (2020) for markup, Autor et al. (2020) and Kehrig and Vincent (2021) for labor share). (6) The lower demand elasticity that accompanies higher markup also means that the same idiosyncratic productivity shock translates into less employment and sales fluctuations. This is consistent with the evidence in Decker et al. (2020) that declining job fluctuations stem not from a decreasing volatility of idiosyncratic shocks but from a weaker responsiveness of firms to these shocks. Incumbent job creation and job destruction rates decline, as does their death rate. (7) The entry rate declines as learning becomes more difficult for potential entrants. (8) Combining (6) and (7) together, the job reallocation rate decreases.

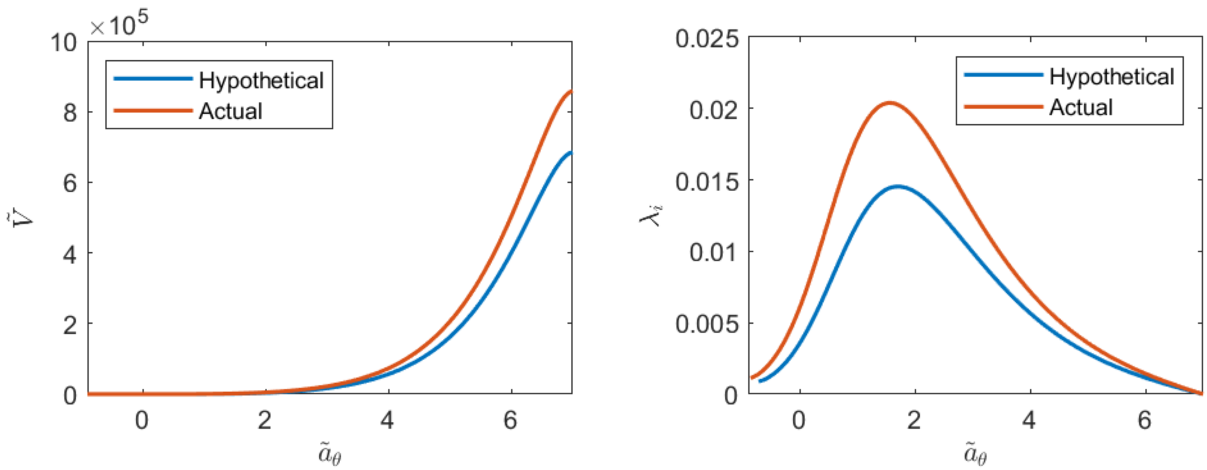
Moment	Data Δ	Model Δ
Sales share of top 1% firms	10.07%	10.31%
TFP growth	-0.41%	-0.44%
Cost-weighted markup	11.40%	8.86%
Labor share	-5.39%	-3.41%
Job creation by birth (entry)	-2.70%	-1.93%
Job destruction by death (exit)	-1.67%	-1.67%
Job creation rate	-5.13%	-2.82%
Job destruction rate	-4.06%	-2.88%
Job reallocation rate	-8.12%	-5.69%
R&D over value added	64.50%	75.83%

Table 5: Moment changes between 1980 and 2020, Model and Data. All the changes are non-targeted.

	Within	Between	Net entry	Total
ΔLabor share	-0.22%	-3.15%	-0.03%	-3.41%
ΔMarkup	0.59%	8.31%	-0.04%	8.86%

Table 6: Melitz-Polanec decomposition of markup and labor share changes between 1980 and 2020.

In the existing Schumpeterian growth literature à la Aghion et al. (2001), growth often declines as a result of higher concentration. That framework features a continuum of sectors each with a fixed market size and with two firms competing oligopolistically with each other. Market concentration is interpreted as the share of sectors with a specific technological gap between the two firms. A higher concentration *per se* does not affect individual firms' growth decisions,[71] as the market size of each sector is fixed, but has a composition effect as polarized sectors grow less quickly. In the present paper, however, growth does *not* decline due to a more concentrated market structure. Higher concentration actually promotes growth in the model and acts as a mitigating force against harder research, as higher rents better incentivize innovation. To grasp this point, consider a hypothetical case in which market structure $\tilde{\phi}^M$ retains its 1980 value in 2020, while all the other parameters have changed to their 2020 value. Figure 13 compares the solution to the HJB equation in the hypothetical case with that in the actual solution of 2020. In the hypothetical case, the value function increases less quickly with productivity than in the actual case because of lower profits. Consequently, growth is lower in the hypothetical case (Figure 13b). Higher concentration is a result of lower growth, not the reason behind it.



(a) Value function

(b) Innovation intensity λ_i

Figure 13: Comparison between the hypothetical case and actual case in 2020. In the hypothetical case, the market structure $\tilde{\phi}^M$ takes the value of its 1980 solution, while other parameters take the value in 2020. The red line in Panel (b) is the same as the red line in Figure 12a.

7 Introducing Economic Integration

While the previous section well accounts for the key macroeconomic moments since 1980s, a major puzzle remains: before the 1980s, how could concentration increase with a fattening Pareto tail without a trend in markup or labor share? Before 1980, labor share according to the US Bureau of Economic Analysis (BEA) had no apparent trend, and both sales-weighted markup (De Loecker

[71]If higher concentration is an endogenous outcome, e.g. due to a lower technological diffusion rate, then firms' growth decisions can be different.

et al. (2020)) and cost-weighted markup (Edmond et al. (2023)) had similar values in 1980 as in 1950, even though TFP growth declined (Gordon (2016), Nordhaus (2021)) and corporate concentration was increasing rapidly (Kwon et al. (2022)). I choose 1960 as the starting point for the earlier period, as sales concentration as measured by Kwon et al. (2022) is only available from that date.[72]

This section argues that introducing economic integration in addition to harder research can account for the pattern, acting as another first-order economic force during the earlier periods. Historical evidence motivates such an investigation. According to Gordon (2016), the 1920s-1970s saw the heyday of US government spending on highways, interrupted only by World War II. The inflation-adjusted domestic airfare per mile decreased by -4.55% per year between 1940 and 1980, and only by -0.1% per year after 1980. The construction of highway and airline facilities, as well as other basic infrastructure, could have significantly reduced non-tariff trade costs across US regions in earlier periods. Deregulations could also have played a role in facilitating integration. One of the most notable examples is the 1978 Airline Deregulation Act which removed restrictions on the routes and market access of airline companies. The effect of economic integration is potentially large as US regions are close to each other, and each of them can be construed as a small economy. As is well-known from the trade literature, smaller economies are more open to trade, and geographical distance is key.[73] The above historical examples thus prompt us to consider economic integration as another important force shaping the economic landscape of earlier periods. Perhaps not coincidentally, some European countries witnessed rising concentration with stable market power in the last few decades (Bighelli et al. (2023), Gutiérrez et al. (2022) and Bauer and Boussard (2020)), during the archetypal integration into the European Union. Such a similar experience also motivates us to consider economic integration as a confounding economic force.

I study the 1960-1980 period in the US to show the effect of economic integration *in conjunction with* harder research. Consider a thought experiment in which there are N identical economies in autarky in 1960. All the economies are on the same travelling wave equilibrium as described by the model. As in the previous section, ideas are harder to find in 1980 than in 1960, and the interest rate declines which partially mitigates the harder research. Based on the same methodology as in Section 5, I calibrate a 7.93-fold increase in α from 1960 to 1980. For the interest rate, the paper takes a secular perspective and views the decline in the natural interest rate from 1960 to 1980 as the change in firms' discount rate during the period.[74] Such a change is added to the 1980 real corporate rate to calibrate firms' discount rate in 1960.[75]

[72] Asset concentration measured by Kwon et al. (2022) has a longer coverage, but the model cannot speak to it.

[73] Globalization during the post-1980 period should in principle also have an effect through the logic of this section. However, international trade openness measured as the trade-to-GDP ratio increased at least as rapidly in the 1940-1980 period as in 1980-2020. Moreover, the US economy as a whole is large and thus more closed to trade. Thus the current section emphasizes the role of regional integration within the US rather than the international integration of the US with other countries.

[74] According to Holston et al. (2017), the US natural interest rate declined persistently from 1960 onwards.

[75] I am not using the real corporate rate for the pre-1980 period, as the US experienced high inflation before Volcker's disinflation policies around 1980. High inflation implies that the measured real rate was lower before 1980 than

The key difference in this section compared to the previous sections is the integration of N economies into one: labor force L and the potential number of entrants M_e are multiplied by N.[76] Based on the pro-growth effect of integration by Proposition 5, I estimate the degree of economic integration N by targeting the TFP growth rate of 1.6% around 1960. Concentration, markup and labor share remain non-targeted. Table 7 summarizes the related parameters for 1960.

Parameter	Description	Value	Source
$\alpha_{t=1980}/\alpha_{t=1960}$	Change of research cost	7.93	Estimated from US aggregate
$r_{t=1960}$	Interest rate	0.0669	External
N	Magnitude of economic integration	3.96	Structural estimation

Table 7: Summary of parameters for 1960.

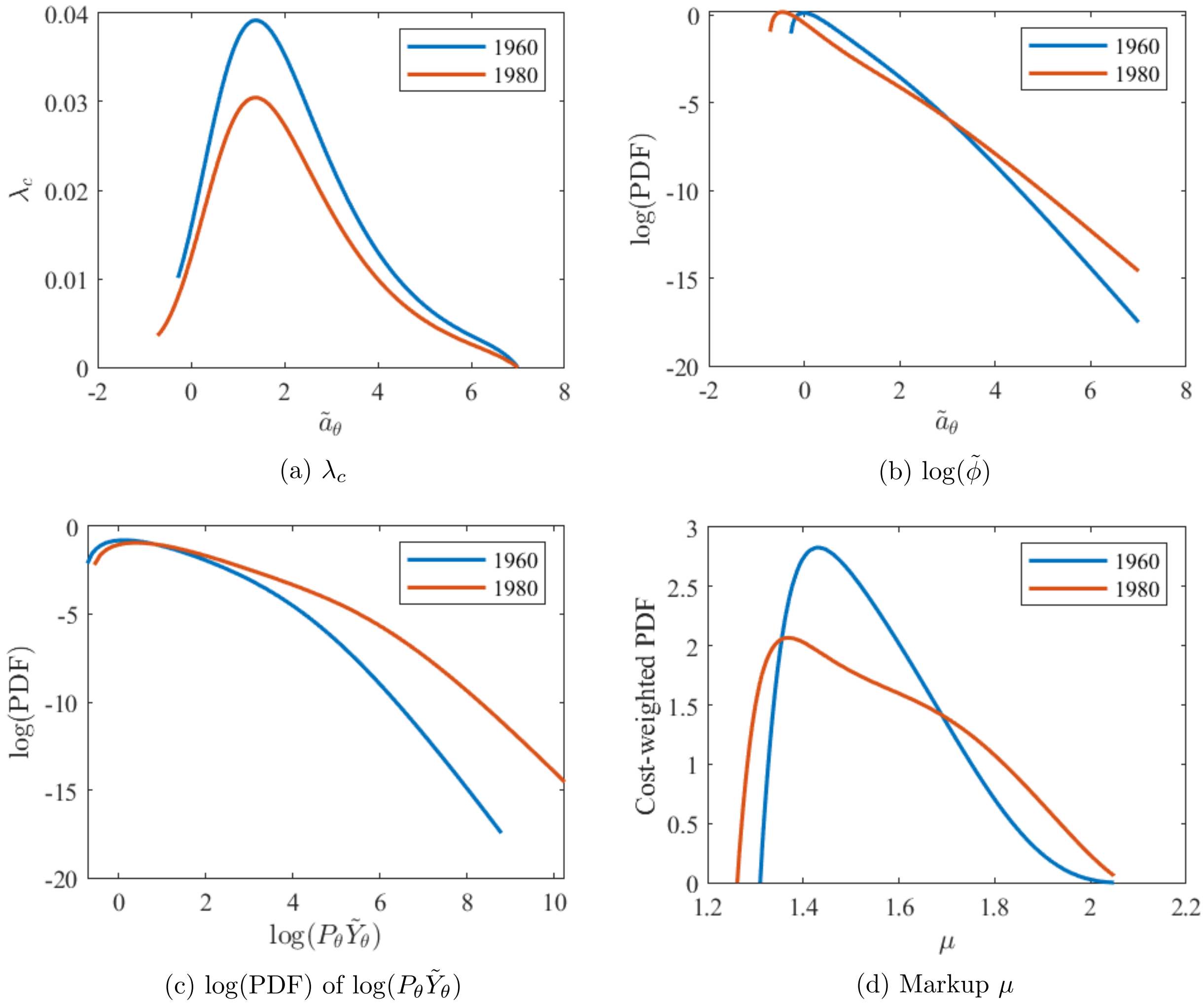
I compare the new travelling wave equilibrium of the integrated economy in 1980 with the old one of segregated economies in 1960. As the N economies in 1960 are identical, the macro moments of growth, concentration, markup and labor share are the same regardless of whether the N economies are seen separately or together.

Figure 14 compares the equilibrium in 1960 with that in 1980. Similar to 12, growth declines for all firms but especially for laggards, and the productivity distribution and sales distribution stretch out to a heavier tail. Declining growth, like in the previous section, is due to the dominating role of harder research despite the mitigating forces of a lower interest rate and larger market size. The top 1% firm sales concentration increases by 14.43%, which is in the ballpark of the estimate in Kwon et al. (2022) and slightly greater than the change from 1980 to 2020. As growth rates in 1960 (this section) and in 1980 (Section 5.3) have been targeted, the increase in concentration resonates with Proposition 4, which links fatter tail with lower growth regardless of market power. Changes in markup and labor share, however, are largely attenuated. Table 8 shows that while market share is reallocated towards high markup and low labor share firms as in the 1980-2020 period, the reallocation component is largely compensated by the within-firm component due to economic integration. As integration increases the number of firms competing with each other, in a purely static setting it reduces within-firm markups more than it reallocates markups towards higher-markup firms because the equilibrium log-productivity distribution is overall log-concave.[77] The net effect of integration is thus a reduction of markup and can act as a compensating force against higher markup due to the endogenous fatter tail. The combined effect of integration and harder research means strongly attenuated markup and labor share changes, despite a slightly higher increase in concentration. To summarize, the dynamic growth effect of harder research and

afterwards, even though the natural interest rate was higher before 1980.

[76] Aggregating N representative consumers into one is feasible, as the Kimball preference is homothetic and hence belongs to the Gorman form.

[77] See further discussions in Section 5.4.



(a) λ_c

(b) $\log(\tilde{\phi})$

(c) log(PDF) of $\log(P_\theta \tilde{Y}_\theta)$

(d) Markup μ

Figure 14: Comparison between 1960 and 1980. Panel (a): innovation intensity; Panel (b): log PDF of log productivity; Panel (c): log PDF of log sales; Panel (d): cost-weighted markup distribution.

the static effect of economic integration jointly reconcile fatter Pareto tails with the stable markup and labor share. The stable labor share in this context is not understood as an economic regularity as in Kaldor (1961), but rather as a historical coincidence. Such an analysis is only possible with a framework in which a well-defined Pareto-tailed productivity distribution is endogenously generated by firm-specific growth. The model provides a way to study the question in a unified setting.

8 Extended Model with Incumbent Learning

I now generalize the model to allow for incumbent learning. Incumbent firms can not only conduct innovation on a stand-alone basis but also learn from other firms. Learning improves a firm's productivity in a quicker fashion. It captures learning-by-doing (Arrow (1962)), technology diffusion

	Within	Between	Net entry	Total
ΔMarkup	-6.18%	9.00%	-0.41%	2.41%
ΔLabor share	3.24%	-4.39%	0.16%	-0.99%

Table 8: Melitz-Polanec decomposition of markup and labor share changes between 1960 and 1980.

(Aghion et al. (2001)), technology adoption (Comin and Gertler (2006)) or imitation (König et al. (2022)). This section shows that all the results hold true in the extended model if both innovation and learning become more difficult over time to reflect more difficult knowledge.

Denote $\lambda_{l,\theta}$ as the Poisson rate of learning success and assume the learning cost $R_{l,\theta}$ to be the same quadratic function as 28:

$$R_{l,\theta} = R(\lambda_{l,\theta}; \tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta\tilde{a}_\theta}\lambda_{l,\theta}^2 A \tag{52}$$

Similar to the entrant's learning cost function, the factor $e^{\beta\tilde{a}_\theta}$ reflects the more difficult knowledge to be learned at higher levels of technology, *irrespective of* how effective an incumbent can be matched with a more productive firm. See Appendix D for a more detailed discussion. If learning is successfully realized with Poisson rate $\lambda_{l,\theta}$, the firm's log productivity jumps to a higher level according to a PDF $\psi_{l,\theta}$. The process takes a similar form of matching as that of entry. If Poisson rate $\lambda_{l,\theta}$ is realized, firm a_θ is randomly matched with an incumbent firm with log productivity a_j.[78] If $a_j \leq a_\theta$, no learning occurs (Figure 15a). If $a_j > a_\theta$, then firm θ jumps to position $a_\psi \in [a_\theta, a_j]$ according to the PDF $\psi(a_\psi; a_\theta, a_j)$:

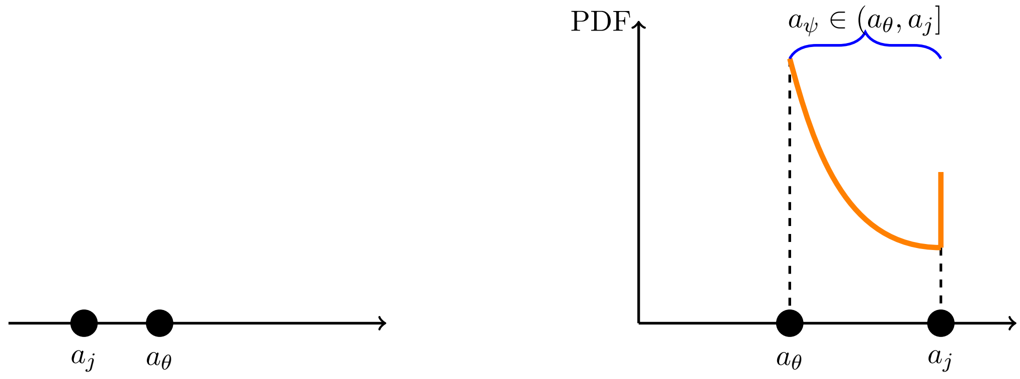
$$\psi(a_\psi; a_\theta, a_j) = k_\psi e^{-k_\psi(a_\psi - a_\theta)} + e^{-k_\psi(a_j - a_\theta)}\delta_{a_j}(a_\psi) \tag{53}$$

Figure 15b illustrates such matching with a more productive firm. The only difference from entry is that the firm now starts from position a_θ, instead of from $\underline{a}$. Integrating over all possible draws of a_j to get the learning outcome distribution ψ_l:

$$\begin{aligned}\psi_l(a_\psi; a_\theta, \phi) &= \int_{a_\theta}^{+\infty} \psi(a_\psi; a_\theta, a_j)\phi(a_j)\,\mathrm{d}a_j \\ &= [k_\psi\overline{\Phi}(a_\psi) + \phi(a_\psi)]e^{-k_\psi(a_\psi - a_\theta)}, \quad \forall a_\psi > a_\theta \end{aligned}\tag{54}$$

I have included a_θ and ϕ in the parentheses of ψ_l to emphasize that the latter depends on the origin of the jump and the existing distribution. By construction, $\psi_l(a_\psi; a_\theta, \phi) = 0$ for $a_\psi < a_\theta$ and $\psi_l(a_\psi; a_\theta, \phi) = \Phi(a_\theta)\delta_{a_\theta}(a_\psi)$ for $a_\psi = a_\theta$, where δ_{a_θ} is the Dirac mass function at point a_θ. The latter comes from the fact that all draws from the left of a_θ result in firm θ staying at a_θ. Taken

[78] The draw is *not* conditional on the drawn firm being more productive than the learning firm, i.e. the learning firm does not target more productive firms for the draw. While this is a typical assumption in the literature (see Buera and Lucas Jr (2018) for a review), one can consider an alternative model in which learning targets only more productive firms like in Chen (2022). As long as learning is imperfect, the results remain robust except that firms may never choose innovation: as learning is very effective even for leaders, they also choose learning over innovation.



(a) Matching with a less productive firm

(b) Matching with a more productive firm

Figure 15: Illustration of incumbent learning.

together,

$$\psi_l(a_\psi; a_\theta, \phi) = [k_\psi \overline{\Phi}(a_\psi) + \phi(a_\psi)] e^{-k_\psi(a_\psi - a_\theta)} \mathbb{1}_{(a_\theta, +\infty)}(a_\psi) + \Phi(a_\theta)\delta_{a_\theta}(a_\psi), \quad \forall a_\psi \geq \underline{a} \tag{55}$$

where $\mathbb{1}$ is the indicator function. It is easy to check that 55 integrates into 1 so that it is a well-defined PDF.

To make the notation succinct, denote the respective destination PDF of innovation as:

$$\psi_i(a_\psi; a_\theta) = \delta_{a_\theta + q}(a_\psi) \tag{56}$$

where $\delta_{a_\theta+q}$ is the Dirac mass function at $a_\theta + q$.

If the log-productivity distribution ϕ has an exponential tail, which is the relevant case for my model, then $\psi_{l,\theta}$ is also exponentially tailed. Learning thus allows a firm to jump far ahead and improve its productivity more quickly than innovation. Equation 55 also makes clear that learning is not an easy process however: a productive firm finds it difficult to be matched with another one which is even more productive. The probability of such a match, $1 - \Phi(a_\theta)$, decreases and approaches 0 when $a_\theta \to +\infty$.

At a specific time, an incumbent chooses only innovation or learning as its growth strategy, as well as the Poisson intensity of that strategy. We shall once again consider a travelling wave equilibrium and use a tilde to denote normalization. The HJB equation of each incumbent in such an equilibrium now becomes:

$$(r - g)\tilde{V} = \max_{c,\, \lambda_c} \left\{ \Pi(\tilde{a}_\theta; \tilde{\phi}^M) - f - \tilde{R}(\lambda_c; \tilde{a}_\theta) + \tilde{\mathcal{F}}\tilde{V} \right\} \tag{57}$$

with the same boundary condition as in 42, where

$$\tilde{\mathcal{F}}\tilde{V} = \Big[-\frac{1}{2}\nu^2 - g\Big]\tilde{V}'(\tilde{a}_\theta) + \frac{1}{2}\nu^2\tilde{V}''(\tilde{a}_\theta) + \lambda_c(\tilde{a}_\theta)\int_{\mathbb{R}}\big[\tilde{V}(\tilde{a}_\psi) - \tilde{V}\big]\psi_c(\tilde{a}_\psi;\tilde{a}_\theta,\tilde{\phi})\,\mathrm{d}\tilde{a}_\psi - \lambda_d(\tilde{a}_\theta;\tilde{\phi}^M)\tilde{V}, \tag{58}$$

$$\tilde{R}(\lambda_c;\tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta\tilde{a}_\theta}\lambda_c^2, \tag{59}$$

$c(\tilde{a}_\theta) \in \{i, l\}$ is the choice between innovation i and learning l and I have used the shorthand c to simplify the notation, ψ_c is the destination PDF of innovation ($c = i$) or learning ($c = l$). For succinctness, I have slightly abused the notation and used $\psi_i(\tilde{a}_\psi;\tilde{a}_\theta,\tilde{\phi})$ to represent $\psi_i(\tilde{a}_\psi;\tilde{a}_\theta)$, even though ψ_i does not depend on $\tilde{\phi}$.

Given the same cost function of innovation and learning, choice c is based on the expected value gain of each strategy:

$$c^*(\tilde{a}_\theta) = \begin{cases} i, & \text{if } \int_{\mathbb{R}}\tilde{V}(\tilde{a}_\psi)\psi_i(\tilde{a}_\psi;\tilde{a}_\theta)\,\mathrm{d}\tilde{a}_\psi \geq \int_{\mathbb{R}}\tilde{V}(\tilde{a}_\psi)\psi_l(\tilde{a}_\psi;\tilde{a}_\theta,\phi)\,\mathrm{d}\tilde{a}_\psi \\ l, & \text{otherwise.} \end{cases} \tag{60}$$

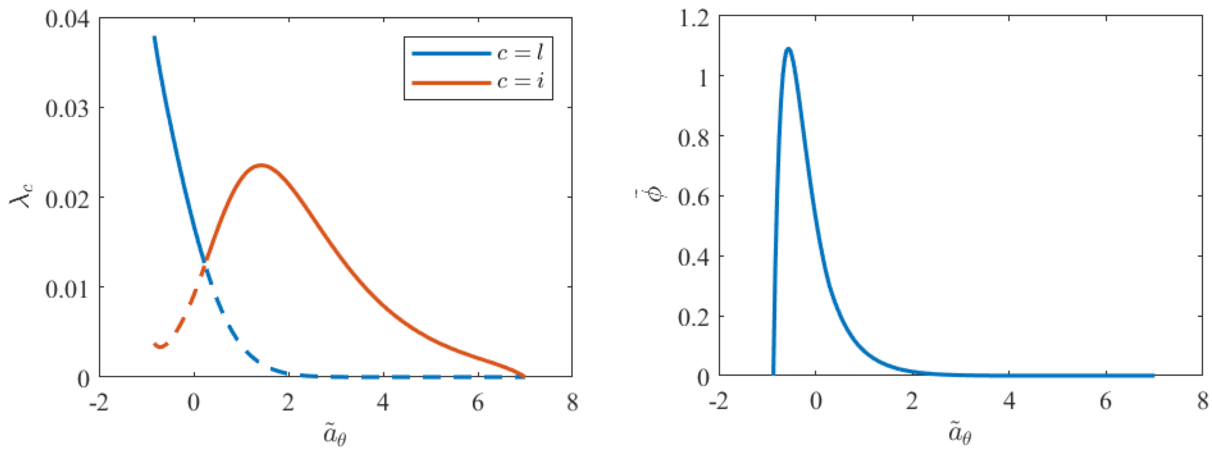
The entrants' problem remains the same as in equation 44. Given the choice of c, λ_c and λ_e, the KF equation now becomes:

$$\begin{aligned} 0 = {} & (\frac{1}{2}\nu^2 + g)\tilde{\phi}' + \frac{1}{2}\nu^2\tilde{\phi}'' - \lambda_d(\tilde{a}_\theta;\tilde{\phi}^M)\tilde{\phi}(\tilde{a}_\theta) + \frac{\lambda_e M_e}{M}\psi_e(\tilde{a}_\theta;\tilde{\phi}) \\ & + \int_{(\underline{\tilde{a}},\tilde{a}_\theta)}\lambda_c(\tilde{a}_\psi)\tilde{\phi}(\tilde{a}_\psi)\psi_c(\tilde{a}_\theta;\tilde{a}_\psi,\tilde{\phi})\,\mathrm{d}\tilde{a}_\psi - \tilde{\phi}(\tilde{a}_\theta)\lambda_c(\tilde{a}_\theta)\int_{(\tilde{a}_\theta,+\infty)}\psi_c(\tilde{a}_\psi;\tilde{a}_\theta,\tilde{\phi})\,\mathrm{d}\tilde{a}_\psi \end{aligned} \tag{61}$$

with the same boundary condition as in 47. The last two terms are different from the baseline model and correspond respectively to inflows into $\tilde{a}_\theta$ and outflows from $\tilde{a}_\theta$ due to innovation/learning. Appendix F.9 shows that the KF equation 61 can be equally written in the form of equation 46, where $\tilde{\mathcal{F}}^\intercal$ is now the adjoint operator of the one defined in 58. The balanced entry/exit condition, zero profit net of fixed cost at the left boundary, and aggregate growth g have the same form as in Definition 1.

Given the similar specification of incumbent learning as that of entry, a non-interference condition under imperfect learning also holds for incumbent learning. Proposition A8 in Appendix E.2 states and proves the result. Like in the baseline model, these non-interference conditions are key for proving the unique tail index of the equilibrium productivity distribution. Appendix G.2 proves the counterpart of Proposition 4 under the extended model. Proposition 5 on market size expansion remains valid under the extended model. All the externally calibrated parameters remain the same as in the baseline model, and the structural estimation follows the same procedure as in Section 5.

Figure 16 shows some results of the travelling wave equilibrium in 1980 under the extended model.



(a) λ_c (b) $\tilde{\phi}$

Figure 16: Travelling wave equilibrium of 1980 in the extended model.

Notes: $\tilde{a}_\theta$ is a firm's log productivity. Aggregate log productivity is normalized to 0. Panel (a): incumbent learning ($c = l$) or innovation ($c = i$) intensity. Solid lines represents the choice between learning and innovation, as well as their intensity; dotted lines indicate the growth strategy (l or i) is not chosen but show the intensity had the firm been forced to use the strategy. Panel (b): log-productivity distribution.

Figure 16a displays the optimal learning ($c = l$) and innovation ($c = i$) intensities of incumbent firms where solid lines indicate the actual choices of firms. There is a cutoff log productivity above which firms choose innovation over learning and vice versa. Dotted lines represent the hypothetical optimal intensities if firms are forced to use a strategy (i.e. innovation or learning) which is not optimally chosen. The overall innovation intensity, i.e. hypothetical for laggards and actual for leaders, shows a similar inverted-U relationship as that in the baseline model. The overall learning intensity is decreasing in firm's productivity, as productive firms find it difficult to be matched with more productive firms, and learning also becomes more difficult with more advanced knowledge. Since I assume a higher cost for both learning and innovation for more productive firms, such firms' optimal choice of innovation over learning reflects the difficulty of being matched with even more productive firms. As innovation is conducted on a stand-alone basis, it allows a leader to stretch out the technological frontier without being restrained by existing knowledge, and is thus more effective than learning for a leader. Figure 16b shows the equilibrium log-productivity distribution with an exponential tail similar to that in the baseline model.

As knowledge becomes more difficult to acquire over time, I calibrate the same increase in α from 1980 to 2020 as in the baseline model, which uniformly increases learning, innovation and entry costs. Appendix J shows comparative statics across the two travelling wave equilibria and calculates moment changes. All the results remain robust under the extended model.

9 Conclusion

The present study constructs a continuous-time Schumpeterian growth model which complements the existing literature by emphasizing how growth shapes market structure. From a methodological point of view, the model benefits from the recently developed Mean Field Game to generate an equilibrium productivity distribution with Pareto tail. As ideas are getting harder to find, the distribution shifts to a heavier tail and the market becomes more concentrated. The model explains a majority of the changes in growth, concentration, markup, labor share, R&D cost, entry and exit rates, and job creation and destruction rates in the US in the last four decades. The explanation is compatible with the growth decline and increasing concentration in a longer historical horizon. The framework can accommodate fatter a Pareto tail with stable markup and labor share in the pre-1980 period by introducing economic integration in addition to harder research. It strives to provide a unified explanation linking key macroeconomic moments, and suggests more difficult knowledge as the "fundamental cause" in the words of Grossman and Oberfield (2022).

Changing the perspective from how market structure determines growth to how growth determines market structure implies dramatically different policy recommendations. In particular, anti-trust policies are generally *not* key for promoting growth, contrary to what most papers in the literature suggest. In the model, higher market power actually *encourages* growth and mitigates harder research, as firms reap higher profits. Undoubtedly, anti-trust policies can still be pro-growth in specific cases where the traditional Schumpeterian logic of higher concentration prohibiting growth is present, or should be applied for ethical and social considerations when they are justified. But growth policies should center around promoting technological opportunities for every firm, large or small. This approach is reminiscent of the East Asian Miracles which involved strong development policies designed by the government.[79] Even though the model cannot fully speak to that history, as the catch-up growth of these countries involves capital accumulation and learning from foreign countries which are outside the model's scope, it shares the same focus of promoting technology to encourage growth. In the case of the US, Antolin-Diaz and Surico (2022) shows that military spending increases productivity in the long run through public R&D. Gruber and Johnson (2019) interprets the post-war experience of the US in a similar fashion, suggesting policy implications centering around public R&D. This framework can hopefully open the door for further fruitful discussions.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, reallocation, and growth. *American Economic Review*, 108(11):3450–3491. 6, 12, 61

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022). Income and wealth distribution

[79]See for instance Evans (1995), Lane (2022), Choi and Levchenko (2021), Liu (2019) and Lan (2021).

in macroeconomics: A continuous-time approach. *The Review of Economic Studies*, 89(1):45–86. 9, 106

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2023). A theory of falling growth and rising rents. *The Review of Economic Studies*, page rdad016. 8

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728. 3, 23, 39

Aghion, P., Harris, C., Howitt, P., and Vickers, J. (2001). Competition, imitation and growth with step-by-step innovation. *The Review of Economic Studies*, 68(3):467–492. 6, 8, 22, 38, 39, 43, 48

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–351. 5, 6

Akcigit, U. and Ates, S. T. (2021). Ten facts on declining business dynamism and lessons from endogenous growth theory. *American Economic Journal: Macroeconomics*, 13(1):257–98. 3, 6, 8

Akcigit, U. and Ates, S. T. (2023). What happened to US business dynamism? *Journal of Political Economy*, 131(8):2059–2124. 6, 8

Akcigit, U. and Kerr, W. R. (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374–1443. 12, 15, 61

Alvarez, F., Argente, D., Lippi, F., Méndez, E., and Van Patten, D. (2022). Strategic complementarities in a dynamic model of fintech adoption. Working paper. 9

Antolin-Diaz, J. and Surico, P. (2022). The long-run effects of government spending. Working paper. 52

Arkolakis, C., Costinot, A., Donaldson, D., and Rodríguez-Clare, A. (2019). The elusive pro-competitive effects of trade. *The Review of Economic Studies*, 86(1):46–80. 23

Arkolakis, C. and Morlacco, M. (2017). Variable demand elasticity, markups, and pass-through. mimeo. 22

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, 29(3):155–173. 47

Atkeson, A. and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031. 1, 3, 20, 23, 37, 117

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709. 1, 4, 8, 36, 41, 43

Baqaee, D. R., Farhi, E., and Sangani, K. (2023). The Darwinian returns to scale. *The Review of Economic Studies*, page rdad061. 20, 36

Barkai, S. (2020). Declining labor and capital shares. *The Journal of Finance*, 75(5):2421–2463. 8, 117

Barles, G. and Souganidis, P. E. (1991). Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4(3):271–283. 106

Bauer, A. and Boussard, J. (2020). Market power and labor share. *Economie et Statistique/Economics and Statistics*, pages 520–521. 4, 45

Benhabib, J., Perla, J., and Tonetti, C. (2021). Reconciling models of diffusion and innovation: A theory of the productivity distribution and technology frontier. *Econometrica*, 89(5):2261–2301. 9

Bernanke, B. (2005). The global saving glut and the US current account deficit. Technical report, Board of Governors of the Federal Reserve System (US). 41

Bighelli, T., Di Mauro, F., Melitz, M. J., and Mertens, M. (2023). European firm concentration and aggregate productivity. *Journal of the European Economic Association*, 21(2):455–483. 4, 45

Bloom, N., Griffith, R., and Van Reenen, J. (2002). Do R&D tax credits work? Evidence from a panel of countries 1979–1997. *Journal of Public Economics*, 85(1):1–31. 12, 61

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–44. 1, 2, 5, 6, 9, 10, 12, 30, 37, 41, 61, 62

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393. 8

Buera, F. J. and Lucas Jr, R. E. (2018). Idea flows and economic growth. *Annual Review of Economics*, 10:315–345. 48

Burstein, A., Carvalho, V. M., and Grassi, B. (2020). Bottom-up markup fluctuations. Working paper, National Bureau of Economic Research. 37, 117

Cavenaile, L., Celik, M. A., and Tian, X. (2019). Are markups too high? Competition, strategic innovation, and industry dynamics. *Competition, Strategic Innovation, and Industry Dynamics (August 1, 2019)*. 8

Chandler, A. D. (1990). *Scale and scope: The dynamics of industrial capitalism*. Harvard University Press. 35

Chen, Z. (2022). Economic growth and the rise of large firms. Working paper. 1, 7, 48

Choi, J. and Levchenko, A. A. (2021). The long-term effects of industrial policy. Working paper, National Bureau of Economic Research. 52

Comin, D. and Gertler, M. (2006). Medium-term business cycles. *American Economic Review*, 96(3):523–551. 48

Cont, R. and Tankov, P. (2003). *Financial modelling with jump processes*. Chapman and Hall/CRC. 106

Cont, R. and Voltchkova, E. (2005). A finite difference scheme for option pricing in jump diffusion and exponential lévy models. *SIAM Journal on Numerical Analysis*, 43(4):1596–1626. 106

De Loecker, J., Eeckhout, J., and Mongey, S. (2021). Quantifying market power and business dynamism in the macroeconomy. Working paper, National Bureau of Economic Research. 37, 117

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644. 1, 4, 8, 43, 44

De Ridder, M. (2019). Market power and innovation in the intangible economy. Working paper. 8

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2016). Where has all the skewness gone? The decline in high-growth (young) firms in the US. *European Economic Review*, 86:4–23. 8, 15, 16

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2020). Changing business dynamism and productivity: Shocks versus responsiveness. *American Economic Review*, 110(12):3952–3990. 4, 9, 43

Dixit, A. (2013). *The art of smooth pasting*. Routledge. 80

Edmond, C., Midrigan, V., and Xu, D. Y. (2023). How costly are markups? *Journal of Political Economy*, 131(7):1619–1675. 1, 8, 45

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: For insurance and finance*, volume 33. Springer Science & Business Media. 74, 75, 76

Evans, P. B. (1995). Embedded autonomy. In *Embedded autonomy*. Princeton University Press. 52

Feenstra, R. C. (2003). A homothetic utility function for monopolistic competition models, without constant price elasticity. *Economics Letters*, 78(1):79–86. 23

Gabaix, X. (1999). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, 114(3):739–767. 3

Gabaix, X., Lasry, J.-M., Lions, P.-L., and Moll, B. (2016). The dynamics of inequality. *Econometrica*, 84(6):2071–2111. 3, 7, 28, 91

Garcia-Macia, D., Hsieh, C.-T., and Klenow, P. J. (2019). How destructive is innovation? *Econometrica*, 87(5):1507–1541. 6

Gaubert, C. and Itskhoki, O. (2021). Granular comparative advantage. *Journal of Political Economy*, 129(3):871–939. 37, 117

Gordon, R. J. (2016). *The rise and fall of American growth: The US standard of living since the civil war*, volume 70. Princeton University Press. 1, 2, 4, 5, 8, 45

Grossman, G. M. and Oberfield, E. (2022). The elusive explanation for the declining labor share. *Annual Review of Economics*, 14:93–124. 9, 52

Gruber, J. and Johnson, S. (2019). *Jump-starting America: How breakthrough science can revive economic growth and the American dream.* Hachette UK. 52

Gutiérrez, G., Martinez, J., Philippon, T., and Piton, S. (2022). The evolution of firm heterogeneity in Europe: Facts and explanations. Working paper. 4, 45

Hall, B. H., Jaffe, A., and Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, pages 16–38. 8

Haskel, J. and Westlake, S. (2017). *Capitalism without capital.* Princeton University Press. 8

Holston, K., Laubach, T., and Williams, J. C. (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, 108:S59–S75. 45

Jones, B. F. (2009). The burden of knowledge and the "death of the Renaissance man": Is innovation getting harder? *The Review of Economic Studies*, 76(1):283–317. 2

Jones, C. I. (1995). R&d-based models of economic growth. *Journal of Political Economy*, 103(4):759–784. 1, 6, 30

Jones, C. I. (2002). Sources of US economic growth in a world of ideas. *American Economic Review*, 92(1):220–239. 7

Jones, C. I. (2005). Growth and ideas. In *Handbook of economic growth*, volume 1, pages 1063–1111. Elsevier. 7, 35

Jones, C. I. (2022). The past and future of economic growth: A semi-endogenous perspective. *Annual Review of Economics*, 14:125–152. 6

Jones, C. I. and Kim, J. (2018). A Schumpeterian model of top income inequality. *Journal of Political Economy*, 126(5):1785–1826. 7

Kaldor, N. (1961). Capital accumulation and economic growth. In *The Theory of capital: Proceedings of a conference held by the International Economic Association*, pages 177–222. Springer. 1, 47

Karabarbounis, L. and Neiman, B. (2014). The global decline of the labor share. *The Quarterly Journal of Economics*, 129(1):61–103. 1, 8

Kehrig, M. and Vincent, N. (2021). The micro-level anatomy of the labor share decline. *The Quarterly Journal of Economics*, 136(2):1031–1087. 1, 8, 43

Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking*, 27(4):1241–1277. 3

Klenow, P. J. and Willis, J. L. (2016). Real rigidities and nominal price changes. *Economica*, 83(331):443–472. 20, 23

König, M., Storesletten, K., Song, Z., and Zilibotti, F. (2022). From imitation to innovation: Where is all that Chinese R&D going? *Econometrica*, 90(4):1615–1654. 48

König, M. D., Lorenz, J., and Zilibotti, F. (2016). Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics*, 11(3):1053–1102. 3, 26

Kortum, S. S. (1997). Research, patenting, and technological change. *Econometrica*, 65(6):1389–1419. 6

Kwon, S. Y., Ma, Y., and Zimmermann, K. (2022). 100 years of rising corporate concentration. Working paper. 1, 2, 7, 8, 45, 46

Lan, X. (2021). *Being involved: The Chinese government and economic development (Zhi shen shi nei: zhong guo zheng fu yu jing ji fa zhan)*. Shanghai People's Press. 52

Lane, N. (2022). Manufacturing revolutions: Industrial policy and industrialization in South Korea. Working paper. 52

Lasry, J.-M. and Lions, P.-L. (2007). Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260. 9

Lions, P.-L. (2006). Cours au collège de france, 2006-2012. 9

Liu, E. (2019). Industrial policies in production networks. *The Quarterly Journal of Economics*, 134(4):1883–1948. 52

Liu, E., Mian, A., and Sufi, A. (2022). Low interest rates, market power, and productivity growth. *Econometrica*, 90(1):193–221. 5, 8, 37

Lucas, R. E. and Moll, B. (2014). Knowledge growth and the allocation of time. *Journal of Political Economy*, 122(1):1–51. 3, 27

Luttmer, E. G. (2011). On the mechanics of firm growth. *The Review of Economic Studies*, 78(3):1042–1068. 7

Luttmer, E. G. J. (2012). *Eventually, noise and imitation implies balanced growth.* Federal Reserve Bank of Minneapolis, Research Department. 3, 27, 28, 90

Luttmer, E. G. J. (2020). Bounded learning from incumbent firms. Working paper, Federal Reserve Bank of Minneapolis. 28

Martinez, J. (2021). Putty-clay automation. Working paper. 9

Matsuyama, K. (2023). Non-CES aggregators: A guided tour. *Annual Review of Economics*, 15:235–265. 19

McKean, H. P. (1975). Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Communications on Pure and Applied Mathematics*, 28(3):323–331. 28

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725. 7

Melitz, M. J. and Ottaviano, G. I. (2008). Market size, trade, and productivity. *The Review of Economic Studies*, 75(1):295–316. 1, 4, 23, 39

Melitz, M. J. and Polanec, S. (2015). Dynamic Olley-Pakes productivity decomposition with entry and exit. *The Rand Journal of Economics*, 46(2):362–375. 43

Mimica, A. (2016). Exponential decay of measures and Tauberian theorems. *Journal of Mathematical Analysis and Applications*, 440(1):266–285. 91, 92

Nordhaus, W. D. (2021). Are we approaching an economic singularity? Information technology and the future of economic growth. *American Economic Journal: Macroeconomics*, 13(1):299–332. 1, 2, 5, 8, 45

Olmstead-Rumsey, J. (2019). Market concentration and the productivity slowdown. Working paper. 8

Park, M., Leahey, E., and Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, 613(7942):138–144. 1, 6

Perla, J., Tonetti, C., and Waugh, M. E. (2021). Equilibrium technology diffusion, trade, and growth. *American Economic Review*, 111(1):73–128. 3, 28, 90

Peters, M. and Walsh, C. (2020). Population growth and firm dynamics. Working paper. 8

Peters, R. H. and Taylor, L. A. (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics*, 123(2):251–272. 8

Philippon, T. (2022). Additive growth. Working paper, National Bureau of Economic Research. 5

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5, Part 2):S71–S102. 7

Segerstrom, P. S. (1998). Endogenous growth without scale effects. *American Economic Review*, pages 1290–1310. 6

Stein, E. M. and Shakarchi, R. (2010). *Complex analysis*, volume 2. Princeton University Press. 90

Widder, D. V. (1946). *The Laplace transformation.* Princeton University Press. 91, 92

Appendix

Format Conventions

Sections with Arial font demarcated by an upper and a lower line explain the mathematical intuitions behind the propositions. While the propositions are more general, these sections give specific examples to illustrate the propositions. They also give a sense of direction for the proofs. Readers not interested in the proofs can gain intuitions through these examples.

```
Algorithms are in boxes with typewriter font.
```

A Empirical evidence

A.1 Methodology

While TFP growth cannot be negative in the model, it can be negative in the data, especially for more recent periods at the sectoral level. This causes a considerable problem for the definition of research productivity as the economic intuition does not coincide with the mathematical definition: given a negative growth rate, a higher research input suggests lower research productivity according to economic reasoning, but the ratio of growth/research input goes *up* because of the negative sign. Moreover, the ratio is negative so we cannot take the log of it, while it is the log level that matters in the model. To make the mathematical definition consistent with economic intuition, I define the "generalized log", or glog, which extends the definition and monotonicity to the domain of $\mathbb{R}$. The function considers a growth rate very close to 0 as being indifferent from 0, and is similar to winsorization in the applied microeconomics literature. Formally, choose a small $g_0 > 0$ and let:

$$\text{glog}(g) = \begin{cases} \log(g) - \log(g_0), & \text{if } g > g_0 \\ 0, & \text{if } -g_0 < g \le g_0 \\ -\log(-g) + \log(g_0), & \text{if } g \le -g_0. \end{cases}$$

Figure A1 shows an illustration. The idea is simply that a growth rate sufficiently close to 0 in the data should be regarded as indifferent from 0. In mathematics, however, $\lim_{g\to 0+} \log(g) = -\infty$ which is irrelevant for the economic intuition. Thus for $g \in (-g_0, g_0)$, where g_0 is a small positive number, $\text{glog}(g)$ should be given a constant value which I normalize to 0. When $g > g_0$, the usual log function applies, with only a parallel shift of $-\log(g_0)$ to ensure monotonicity between $(-g_0, g_0)$ and $(g_0, +\infty)$. Finally, I can apply a similar log transformation for $g < -g_0$ while ensuring monotonicity on $g \in \mathbb{R}$.

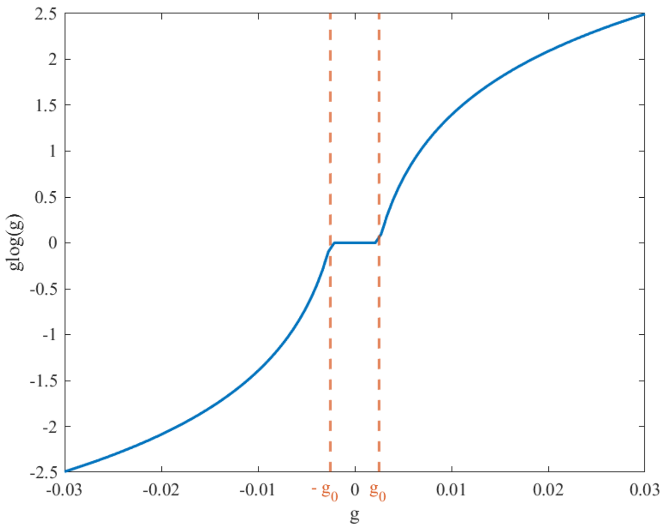


Figure A1: glog(g) with $g_0 = e^{-6}$

In the model, R&D cost is a quadratic function of growth, i.e. $R = Cg^2$ for some $C > 0$. The quadratic cost function is a typical assumption in Schumpeterian growth models which matches the elasticity of R&D with respect to the user costs of around -1.[80] Rearranging the terms, $g = C^{-\frac{1}{2}} R^{\frac{1}{2}}$, which is the form of the idea production function in Bloom et al. (2020) with decreasing returns to research input. $C^{-\frac{1}{2}}$ is then the research productivity of the idea production function which I shall denote IdeaProd, i.e. $\log(\text{IdeaProd}) = \log(g) - \frac{1}{2}\log(R)$.

In the data, research input R is always positive at the sector*period level. The definition of $\log(g/R^{1/2}) = \log(g) - \frac{1}{2}\log(R)$ for $g > 0$ can now be extended to $g \in \mathbb{R}$ by using $\widetilde{\log}(g/R^{1/2}) = \text{glog}(g) - \frac{1}{2}\log(R)$, where I have used a tilde to emphasize the difference. $\widetilde{\log}(g/R)$ is monotone in both $g \in \mathbb{R}$ and $R > 0$, making the mathematical monotonicity consistent with economic intuition. I shall use $g_0 = e^{-6}$ for the following results. Correlations remain robust under alternative specifications of g_0.

A.2 Robustness check

This section shows that the correlations are robust under an alternative assumption of constant returns to scale in the idea production function, or with alternative ways of dividing historical periods, or with alternative values of threshold g_0.

Divide the history since 1987 into three periods instead: 1987-1996, 1997-2006, 2010-2019. We have excluded the Great Recession in this alternative division and each period comprises 10 years.

[80] See Bloom et al. (2002), Acemoglu et al. (2018) and Akcigit and Kerr (2018).

	(1) Growth	(2) Concentration	(3) Ebit/Sales 4	(4) Ebit/Sales 8	(5) Labor Share
log(IdeaProd)	0.403***	-5.147***	-0.980**	-1.123**	2.738**
	(0.0824)	(1.675)	(0.413)	(0.427)	(1.029)
N	99	99	99	99	99
Within R2	0.137	0.166	0.0954	0.135	0.0840
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A1: Regressions of TFP growth and market structure indicators on log research productivity, with alternative time period division (1987-1996, 1997-2006, 2010-2019).

	(1) Entry (num)	(2) Exit (num)	(3) Entry (job)	(4) Exit (job)	(5) Job Creation	(6) Job Destruction	(7) Job Reallocation
log(IdeaProd)	0.559**	0.656***	0.461***	0.451***	0.955***	0.971***	1.908***
	(0.254)	(0.183)	(0.160)	(0.0965)	(0.295)	(0.210)	(0.486)
N	99	99	99	99	99	99	99
Within R2	0.0559	0.119	0.0909	0.137	0.0805	0.120	0.112
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Regressions of business dynamism indicators on log research productivity, with alternative time period division (1987-1996, 1997-2006, 2010-2019).

Tables A1 and A2 show that the correlations are robust.

The second robustness check modifies the definition of research productivity: the idea production function is assumed to feature constant returns to scale as in the baseline setting of Bloom et al. (2020). Tables A3 and A4 show that the correlations are robust.

Thirdly, I use a different threshold $g_0 = e^{-5}$ for defining the generalized log. Tables A5 and A6 show that the correlations are robust.

Finally, I adjust the sectoral R&D expenditure aggregated from Compustat firm-level data by the total employment share of Compustat firms per sector.[81] Tables A7 and A8 show that the correlations are robust.

[81] The share can be larger than 1, as Compustat is reported at the group level.

	(1) Growth	(2) Concentration	(3) Ebit/Sales 4	(4) Ebit/Sales 8	(5) Labor Share
log(IdeaProd)	0.0509	-3.976***	-0.475	-0.562*	1.922***
	(0.0763)	(0.944)	(0.304)	(0.312)	(0.626)
N	132	132	132	132	132
Within R2	0.00607	0.306	0.0657	0.0959	0.127
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Regressions of TFP growth and market structure indicators on log research productivity, with constant returns to scale of the idea production function.

	(1) Entry (num)	(2) Exit (num)	(3) Entry (job)	(4) Exit (job)	(5) Job Creation	(6) Job Destruction	(7) Job Reallocation
log(IdeaProd)	0.439**	0.390***	0.304**	0.238***	0.673***	0.520***	1.195***
	(0.180)	(0.131)	(0.120)	(0.0831)	(0.221)	(0.188)	(0.395)
N	132	132	132	132	132	132	132
Within R2	0.107	0.124	0.121	0.112	0.123	0.102	0.133
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Regressions of business dynamism indicators on log research productivity, with constant returns to scale of the idea production function.

	(1) Growth	(2) Concentration	(3) Ebit/Sales 4	(4) Ebit/Sales 8	(5) Labor Share
log(IdeaProd)	0.156	-8.111***	-0.853	-1.020	3.853**
	(0.108)	(1.782)	(0.612)	(0.621)	(1.408)
N	132	132	132	132	132
Within R2	0.0142	0.318	0.0529	0.0789	0.128
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Regressions of TFP growth and market structure indicators on log research productivity, with $g_0 = e^{-5}$.

	(1) Entry (num)	(2) Exit (num)	(3) Entry (job)	(4) Exit (job)	(5) Job Creation	(6) Job Destruction	(7) Job Reallocation
log(IdeaProd)	0.894**	0.804***	0.583**	0.484**	1.345***	1.083**	2.400***
	(0.393)	(0.294)	(0.254)	(0.183)	(0.486)	(0.413)	(0.853)
N	132	132	132	132	132	132	132
Within R2	0.111	0.132	0.111	0.116	0.122	0.110	0.134
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Regressions of business dynamism indicators on log research productivity, with $g_0 = e^{-5}$.

	(1) Growth	(2) Concentration	(3) Ebit/Sales 4	(4) Ebit/Sales 8	(5) Labor Share
log(IdeaProd)	0.706***	-1.218	-0.651*	-0.706*	1.152
	(0.140)	(1.700)	(0.364)	(0.389)	(1.202)
N	132	132	132	132	132
Within R2	0.361	0.00887	0.0381	0.0468	0.0141
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Regressions of TFP growth and market structure indicators on log research productivity, with adjusted Compustat R&D data.

	(1) Entry (num)	(2) Exit (num)	(3) Entry (job)	(4) Exit (job)	(5) Job Creation	(6) Job Destruction	(7) Job Reallocation
log(IdeaProd)	0.404	0.572***	0.355*	0.410***	0.563	0.776**	1.355**
	(0.284)	(0.197)	(0.175)	(0.111)	(0.385)	(0.289)	(0.656)
N	132	132	132	132	132	132	132
Within R2	0.0280	0.0827	0.0512	0.103	0.0266	0.0700	0.0530
Period Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Regressions of business dynamism indicators on log research productivity, with adjusted Compustat R&D data.

B Additional materials for the illustrative model

B.1 Derivations for the equilibrium distribution

The Kolmogorov Forward (KF) equation associated with stochastic process 2 determines the equilibrium distribution $\tilde{\phi}$:

$$0 = (\frac{1}{2}\nu^2 + g)\tilde{\phi}'(\tilde{a}) + \frac{1}{2}\nu^2\tilde{\phi}''(\tilde{a}) + \lambda(\tilde{a} - q)\tilde{\phi}(\tilde{a} - q) - \lambda(\tilde{a})\tilde{\phi}(\tilde{a}) \tag{A1}$$

On the right hand side of the equation, the first two terms correpond to the dt term and the d$B_{\theta,t}$ term of the stochastic process. $\lambda(\tilde{a}-q)\tilde{\phi}(\tilde{a}-q)$ captures the inflow into position $\tilde{a}$ from position $\tilde{a}-q$ due to innovation, while $\lambda(\tilde{a})\tilde{\phi}(\tilde{a})$ captures outflow from position $\tilde{a}$. The net effect of all the terms on the right hand side is 0 as we are considering a stationary distribution after normalization. No boundary condition is needed for the KF equation at the model level, as the optimal Poisson rate of innovation $\lambda(\tilde{a})$ satisfies $\lim_{\tilde{a}\to-\infty}\lambda(\tilde{a}) = +\infty$ and $\lim_{\tilde{a}\to+\infty}\lambda(\tilde{a}) = 0$ if $\beta > \sigma - 1$.[82] Thus the least productive firms on average grow more quickly than aggregate growth, while the most productive ones grow less quickly: this drives the mean reversion when log productivities are normalized by g per period, so that a stationary distribution $\tilde{\phi}$ emerges.

As q is small, equation A1 can be approximated by

$$0 = (\frac{1}{2}\nu^2 + g)\tilde{\phi}'(\tilde{a}) + \frac{1}{2}\nu^2\tilde{\phi}''(\tilde{a}) - q\frac{\mathrm{d}[\lambda(\tilde{a})\tilde{\phi}(\tilde{a})]}{\mathrm{d}\tilde{a}} \tag{A2}$$

One can use the guess-and-verify method to see that equation 3 is the solution.

B.2 Further empirical evidence

Figure A2 is a counterpart of Figure 4a and disaggregates laggards into small (< 5 employees) and medium ($5 - 500$ employees) firms. The growth decline of medium firms is more pronounced than that of large firms, allowing large firms to stretch out further to the right tail.

C Derivations for the demand system

C.1 Utility maximisation

The representative consumer with Kimball preference solves the following optimisation problem:

$$\max_{Y,\{Y_\theta\}_{0\leq\theta\leq1}} \log(Y)$$

[82] At the algorithm level, two boundary conditions (one on the left end, the other one on the right end of the distribution) are still needed, as a computer can only handle finite segments. A natural choice is a reflecting barrier: if a firm's log productivity goes beyond the barrier, it is brought back to the barrier.

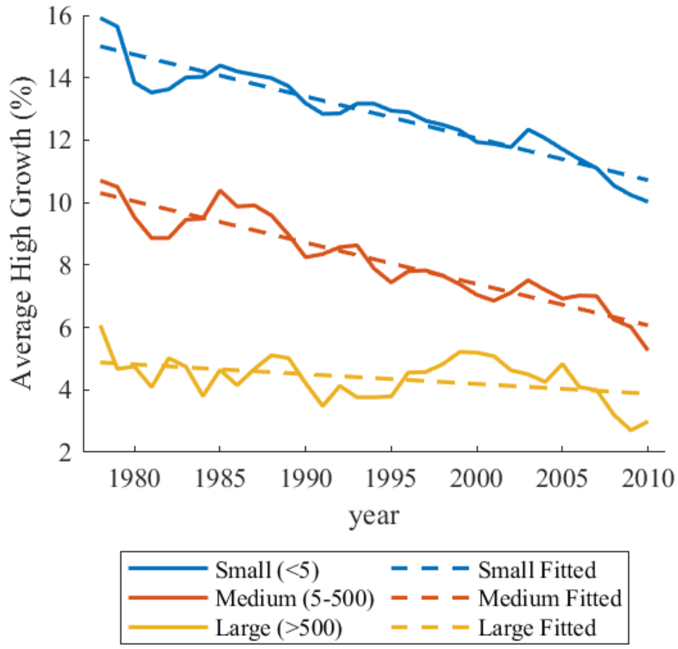


Figure A2: High growth among small, medium and large firms

s.t.

$$M \int_0^1 \gamma(\frac{Y_\theta}{Y})\,\mathrm{d}\theta = 1$$

$$M \int_0^1 P_\theta Y_\theta\,\mathrm{d}\theta \leq wL + \Pi$$

where Π is total profits from firms. Define the Lagrangian multiplier:

$$\mathcal{L} = \log(Y) + \lambda_1(M \int_0^1 \gamma(\frac{Y_\theta}{Y})\,\mathrm{d}\theta - 1) + \lambda_2(wL + \Pi - M \int_0^1 P_\theta Y_\theta\,\mathrm{d}\theta)$$

First order conditions:

$$\frac{\partial \mathcal{L}}{\partial Y_\theta} = \lambda_1 M \gamma'(\frac{Y_\theta}{Y})\frac{1}{Y} - \lambda_2 M P_\theta = 0$$

$$\frac{\partial \mathcal{L}}{\partial Y} = \frac{1}{Y} - \lambda_1 M \int_0^1 \gamma'(\frac{Y_\theta}{Y})\frac{Y_\theta}{Y^2}\,\mathrm{d}\theta = 0$$

Define two price indices $P = \frac{\lambda_1}{\lambda_2 Y}$ and $\zeta = \lambda_1 = [M \int_0^1 \gamma'(\frac{Y_\theta}{Y})\frac{Y_\theta}{Y}\,\mathrm{d}\theta]^{-1}$. Rewrite the FOCs in terms of P and ζ. Then

$$\gamma'(\frac{Y_\theta}{Y}) = \frac{P_\theta}{P}$$
$$M\int_0^1 P_\theta Y_\theta \,\mathrm{d}\theta = \frac{P}{\zeta}Y$$

C.2 Proof of proposition 2

Proof. 1. By 11, it is clear that $\frac{\partial\sigma}{\partial z_\theta} < 0$, it suffices to show $\frac{\partial z_\theta}{\partial a_\theta} > 0$.

In fact, combining equations 6 and 14 gives:

$$\frac{\gamma'(Z_\theta)}{MC_\theta/P} = \frac{\sigma(z_\theta)}{\sigma(z_\theta)-1}$$

which is an implicit function of Z_θ with respect to a_θ. Take the first derivative w.r.t. a_θ and rearrange the terms to give:

$$\frac{\partial Z_\theta}{\partial a_\theta}\big[\gamma''(Z_\theta)(1-\sigma^{-1}) + \gamma'(Z_\theta)\sigma^{-2}\frac{\partial\sigma}{\partial Z_\theta}\big] = -\frac{MC_\theta}{P}$$

Using equations 9 and 14 to recollect terms, this simplifies to:

$$\frac{\partial Z_\theta}{\partial a_\theta} = \frac{\sigma(\sigma-1)}{\frac{\sigma-1}{Z_\theta} - \frac{\partial\sigma}{\partial Z_\theta}}$$

or

$$\frac{\partial z_\theta}{\partial a_\theta} = \frac{\sigma(\sigma-1)}{\sigma - 1 - \frac{\partial\sigma}{\partial z_\theta}} \tag{A3}$$

It suffices to note $\sigma > 1$ and $\frac{\partial\sigma}{\partial z_\theta} < 0$ to see $\frac{\partial z_\theta}{\partial a_\theta} > 0$

2. By definition, the passthrough is:

$$\begin{aligned}\frac{\partial\log(P_\theta)}{\partial\log(MC_\theta)} &= -\frac{\partial\log(P_\theta/P)}{\partial a_\theta} \\ &= -\frac{\partial\log(P_\theta/P)}{\partial z_\theta}\frac{\partial z_\theta}{\partial a_\theta} \\ &= \frac{1}{1+\frac{-\frac{\partial\sigma}{\partial z_\theta}}{\sigma-1}} \end{aligned} \tag{A4}$$

where we have used equation A3.

To show $\frac{\partial\log(P_\theta)}{\partial\log(MC_\theta)}$ is a U-shaped function of a_θ, we only need to show it is a U-shaped function of z_θ, or equivalently, $\frac{-\frac{\partial\sigma}{\partial z_\theta}}{\sigma-1}$ is a hump-shaped function of z_θ. It is easy to check that $\frac{-\frac{\partial\sigma}{\partial z_\theta}}{\sigma-1}$ is the markup elasticity to price $-\frac{\partial \log \mu_\theta}{\partial \log \frac{P_\theta}{P}}$. Taking the derivative with respect to z_θ gives:

$$\frac{\partial}{\partial z_\theta}\left(-\frac{\partial \log \mu_\theta}{\partial \log \frac{P_\theta}{P}}\right) = -\frac{\partial^2 \sigma}{\partial z_\theta^2}\frac{1}{\sigma-1} + \left(\frac{\partial \sigma}{\partial z_\theta}\right)^2 \frac{1}{(\sigma-1)^2}$$

Under the logistic specification of equation 11,

$$\begin{aligned}
\frac{\partial \sigma}{\partial z_\theta} &= -k\frac{(\overline{\sigma}-\sigma)(\sigma-\underline{\sigma})}{\overline{\sigma}-\underline{\sigma}} \\
\frac{\partial^2 \sigma}{\partial z_\theta^2} &= k\frac{2\sigma-(\overline{\sigma}+\underline{\sigma})}{\overline{\sigma}-\underline{\sigma}}\frac{\partial \sigma}{\partial z_\theta}
\end{aligned}$$

Thus

$$\frac{\partial}{\partial z_\theta}\left(-\frac{\partial \log \mu_\theta}{\partial \log \frac{P_\theta}{P}}\right) = -\frac{1}{(\sigma-1)^2}\frac{\partial \sigma}{\partial z_\theta}\frac{k}{\overline{\sigma}-\underline{\sigma}}\left[\sigma^2 - 2\sigma + \overline{\sigma} + \underline{\sigma} - \overline{\sigma}\underline{\sigma}\right]$$

The quadratic term in the last bracket has two roots $1 \pm \sqrt{(\overline{\sigma}-1)(\underline{\sigma}-1)}$, only one of which is within the range of $(1, +\infty)$. Thus $\frac{\partial}{\partial z_\theta}\left(-\frac{\partial \log \mu_\theta}{\partial \log \frac{P_\theta}{P}}\right) < 0$ for $\sigma \in (\underline{\sigma}, 1 + \sqrt{(\overline{\sigma}-1)(\underline{\sigma}-1)})$ and > 0 for $\sigma \in (1 + \sqrt{(\overline{\sigma}-1)(\underline{\sigma}-1)}, \overline{\sigma})$. Thus $-\frac{\partial \log \mu_\theta}{\partial \log \frac{P_\theta}{P}}$ is a U-shaped function of σ. Because σ is monotonically decreasing in z_θ, $-\frac{\partial \log \mu_\theta}{\partial \log \frac{P_\theta}{P}}$ is hump-shaped in z_θ.

3. First note that

$$\frac{\partial \sigma}{\partial z_\theta} = -\frac{(\overline{\sigma}-\sigma)(\sigma-\underline{\sigma})}{\overline{\sigma}-\underline{\sigma}}k$$

Inserting it into A3, we get:

$$\begin{aligned}
\frac{\partial z_\theta}{\partial a_\theta} &= \frac{\sigma}{1 + \frac{(\overline{\sigma}-\sigma)(\sigma-\underline{\sigma})}{(\overline{\sigma}-\underline{\sigma})(\sigma-1)}k} \\
&\geq \frac{\underline{\sigma}}{1 + \frac{(\overline{\sigma}-\underline{\sigma})^2}{(\overline{\sigma}-\underline{\sigma})(\underline{\sigma}-1)}k} \\
&= \frac{\underline{\sigma}}{1 + \frac{\overline{\sigma}-\underline{\sigma}}{\underline{\sigma}-1}k}
\end{aligned} \tag{A5}$$

The last is a positive constant. Hence $\lim_{a_\theta \uparrow +\infty} z_\theta = +\infty$. This result, together with the fact that z_θ is strictly increasing in a_θ, means that we do not need to distinguish between $a_\theta \to +\infty$ and $z_\theta \to +\infty$.

By A3,

$$\begin{aligned}
\frac{\partial a_\theta}{\partial z_\theta} &= \frac{1}{\sigma} - \frac{1}{\sigma(\sigma-1)}\frac{\partial \sigma}{\partial z_\theta} \\
&= \frac{1}{\underline{\sigma}} + (\sigma-\underline{\sigma})\Big[-\frac{1}{\sigma\underline{\sigma}} + \frac{\overline{\sigma}-\sigma}{\sigma(\sigma-1)(\overline{\sigma}-\underline{\sigma})}k\Big]
\end{aligned}$$

Denote the term in squared brackets to be $f(z_\theta)$, i.e. $f(z_\theta) = -\frac{1}{\sigma\underline{\sigma}} + \frac{\overline{\sigma}-\sigma}{\sigma(\sigma-1)(\overline{\sigma}-\underline{\sigma})}k$. Then

$$\lim_{z_\theta \to +\infty} f(z_\theta) = \frac{(k-1)\underline{\sigma}+1}{\underline{\sigma}^2(\underline{\sigma}-1)}$$

Moreover,

$$\sigma - \underline{\sigma} = (\overline{\sigma} - \underline{\sigma})\frac{\exp(-kz_\theta)}{1+\exp(-kz_\theta)} \leq C\exp(-kz_\theta)$$

for some constant $C > 0$ as $z_\theta \to +\infty$.

Thus

$$|(\sigma - \underline{\sigma})f(z_\theta)| \leq C\exp(-kz_\theta)$$

for another positive constant $C > 0$ and when z_θ is sufficiently large. Denote z^* to be one point above which the inequality holds. Denote a^* to be the corresponding log productivity. Then $\forall z_\theta > z^*$,

$$\frac{1}{\underline{\sigma}} - C\exp(-kz_\theta) \leq \frac{\partial a_\theta}{\partial z_\theta} \leq \frac{1}{\underline{\sigma}} + C\exp(-kz_\theta)$$

Together with $a_\theta = a^* + \int_{z^*}^{z_\theta} \frac{\partial a_\theta}{\partial z_\theta}(z)\,\mathrm{d}z$, we get:

$$\begin{aligned}
a_\theta &\geq a^* + \frac{1}{\underline{\sigma}}(z_\theta - z^*) + \frac{C}{k}[\exp(-kz_\theta) - \exp(-kz^*)] \geq a^* + \frac{1}{\underline{\sigma}}(z_\theta - z^*) - \frac{C}{k}\exp(-kz^*),\\
a_\theta &\leq a^* + \frac{1}{\underline{\sigma}}(z_\theta - z^*) - \frac{C}{k}[\exp(-kz_\theta) - \exp(-kz^*)] \leq a^* + \frac{1}{\underline{\sigma}}(z_\theta - z^*) + \frac{C}{k}\exp(-kz^*)
\end{aligned}$$

Thus there is some $C_1 \in \mathbb{R}$ and $C_2 \in \mathbb{R}$ such that $a_\theta \in [\frac{1}{\underline{\sigma}}z_\theta + C_1, \frac{1}{\underline{\sigma}}z_\theta + C_2]$ for sufficiently large z_θ, which implies $A_\theta \sim \mathcal{O}(Z_\theta^{\frac{1}{\underline{\sigma}}})$ or equivalently $Z_\theta \sim \mathcal{O}(A_\theta^{\underline{\sigma}})$ as $a_\theta \to +\infty$.

To show the stronger result of $A_\theta \sim CZ_\theta^{\frac{1}{\underline{\sigma}}}$ as $z_\theta \to +\infty$, first note that $\frac{A_\theta}{Z_\theta^{\frac{1}{\underline{\sigma}}}}$ is bounded as $A_\theta \sim \mathcal{O}(Z_\theta^{\frac{1}{\underline{\sigma}}})$. This means that we only need to show $\frac{A_\theta}{Z_\theta^{\frac{1}{\underline{\sigma}}}}$ is ultimately monotone. In fact,

$$\begin{aligned}
\frac{A_\theta}{Z_\theta^{\frac{1}{\underline{\sigma}}}} &= \exp\left(a_\theta - \frac{1}{\underline{\sigma}}z_\theta\right)\\
&= \exp\left(a^* + \int_{z^*}^{z_\theta} \frac{\partial a_\theta}{\partial z_\theta}(z)\,\mathrm{d}z - \frac{1}{\underline{\sigma}}z_\theta\right)\\
&= \exp\left(a^* + \int_{z^*}^{z_\theta} (\sigma - \underline{\sigma})f(z)\,\mathrm{d}z - \frac{1}{\underline{\sigma}}z^*\right)
\end{aligned}$$

If the limit of $f(z_\theta)$, $\frac{(k-1)\underline{\sigma}+1}{\underline{\sigma}^2(\underline{\sigma}-1)}$, is positive, then the integrand is always positive for sufficiently large z. This means that we can modify the choice of z^* such that $f(z) > 0$ for all $z > z^*$, in addition to the requirement above. Then it is clear that $\frac{A_\theta}{Z_\theta^{\frac{1}{\underline{\sigma}}}}$ is increasing for all $z_\theta > z^*$.

Similarly, if the limit of $f(z_\theta)$ is negative, then $\frac{A_\theta}{Z_\theta^{\frac{1}{\underline{\sigma}}}}$ is ultimately decreasing.

In the case where $\lim_{z_\theta \to +\infty} f(z_\theta) = 0$, $(k-1)\underline{\sigma} + 1 = 0$. Plug this relationship into the definition of f:

$$\begin{aligned} f(z_\theta) &= -\frac{1}{\sigma\underline{\sigma}} + \frac{\overline{\sigma} - \sigma}{\sigma(\sigma - 1)(\overline{\sigma} - \underline{\sigma})} \frac{\underline{\sigma} - 1}{\underline{\sigma}} \\ &= \frac{\underline{\sigma} - \sigma}{\sigma\underline{\sigma}(\overline{\sigma} - \underline{\sigma})} < 0 \end{aligned}$$

Thus $\frac{A_\theta}{Z_\theta^{\frac{1}{\underline{\sigma}}}}$ is ultimately decreasing.

In summary, $\frac{A_\theta}{Z_\theta^{\frac{1}{\underline{\sigma}}}}$ is ultimately monotone and $A_\theta \sim C Z_\theta^{\frac{1}{\underline{\sigma}}}$ for some positive constant C.

In order to show $P_\theta Y_\theta \sim C A_\theta^{\underline{\sigma}-1}$ as $a_\theta \to +\infty$, it suffices to show $\gamma'(Z_\theta) Z_\theta \sim C A_\theta^{\underline{\sigma}-1}$. This can be easily seen as by definition $\gamma'(Z_\theta) \sim C Z_\theta^{-\frac{1}{\underline{\sigma}}}$ when $Z_\theta \to +\infty$. Hence

$$\gamma'(Z_\theta) Z_\theta \sim C Z_\theta^{1-\frac{1}{\underline{\sigma}}} \sim C A_\theta^{\underline{\sigma}-1} \quad \text{as} \quad a_\theta \to +\infty$$

Finally, it is easy to see $L_\theta = \frac{Y_\theta}{A_\theta} \sim C \frac{Z_\theta}{A_\theta} \sim C A_\theta^{\underline{\sigma}-1}$ as $a_\theta \to +\infty$.

This concludes our proof.

□

D Micro-foundation of the learning cost function

This section gives more texture to the entrants' learning cost function 21 with $\alpha_e = \alpha e^{\beta \tilde{\underline{a}}}$ and the incumbents' learning cost function 52. Recall that in the case of innovation, I have used $e^{\beta \tilde{a}_\theta}$ for adjusting the innovation cost in a cross section of firms. The adjustment factor depends on $\tilde{a}_\theta$, as innovation is a marginal improvement when q is small and thus is local in nature. Why do I use the same factor $e^{\beta \tilde{a}_\theta}$ for adjusting the difficulty of learning, despite the fact that learning is non-local?

We have derived the learning destination distribution $\psi_l(a_\psi; a_\theta, \phi)$ in 55 by matching with incumbent firms and imperfect learning from the latter. Once the specification is laid down, it can be equally regarded as jumping according to the distribution ψ_l, without reference to a specific a_j to be learned from.[83] Viewed in this way, if the learning firm a_θ happens to jump to a_ψ, it jumps with certainty. We start by considering learning costs along the learning path to a

[83] An analogy with fields in physics may better clarify the point. An electric field can be generated by electrically charged particles. Once the field is established, its force exerted on other charged particles in the field can be analyzed by taking the field as the object of study, without reference to the particles that generate the field. Similarly, ψ_l is the "learning field" generated by incumbent firms.

specific a_ψ and investigating related properties, then consider the more general case with a non-degenerate distribution of learning destination. Like in Section 4.2.3, I interpret the learning jump as consecutive marginal improvements: if firm with log-TFP a_θ jumps to a_ψ due to learning, then it moves up to a_ψ by incremental steps. A local learning cost of factor $e^{\beta\tilde{a}_k}\mathrm{d}a_k$ is incurred at each position $a_k \in [a_\theta, a_\psi]$, where constants have been omitted to focus on the harder learning effect in the cross section. The total learning cost from a_θ to a_ψ is adjusted by:

$$\int_{a_\theta}^{a_\psi} e^{\beta\tilde{a}_k}\,\mathrm{d}a_k = C(e^{\beta\tilde{a}_\psi} - e^{\beta\tilde{a}_\theta}) = Ce^{\beta\tilde{a}_\theta}(e^{\beta(\tilde{a}_\psi-\tilde{a}_\theta)} - 1)$$

where $e^{\beta\tilde{a}_\theta}$ reflects the local adjustment due to harder learning, and $e^{\beta(\tilde{a}_\psi-\tilde{a}_\theta)} - 1$ reflects an additional adjustment due to the distance of jump.

Based on the above argument, suppose for the moment a learning function with the form $R(a_\theta, a_\psi) = \frac{1}{2}\alpha(e^{\beta\tilde{a}_\psi} - e^{\beta\tilde{a}_\theta})\lambda^2$, where α is some positive constant. If λ is realized, then a_θ moves with certainty to a_ψ. Denote $\mathbb{E}(A_{\theta,t+\Delta t}) = \lambda\Delta t A_\psi + (1-\lambda\Delta t)A_\theta$ as the expected position after the realization of λ. The following proposition says that such a cost function satisfies properties consistent with our considered convictions:

Proposition A1.

1. *If $0 < \beta < 2$, then with the same cost, $\mathbb{E}(A_{\theta,t+\Delta t})$ is monotonically increasing with a_ψ. In other words, with the same spending on learning, learning from a more productive firm is more effective.*

2. *With the same cost and the same a_ψ, $\mathbb{E}(A_{\theta,t+\Delta t})$ is monotonically increasing with a_θ. In other words, if two firms learn from the same firm and incur the same cost, the more productive firm among the two learning firms is expected to end up at a more productive position than the other one.*

Proof.

1. First note that

$$\lambda = \Big[\frac{2R}{\alpha(\tilde{A}_\psi^\beta - \tilde{A}_\theta^\beta)}\Big]^{\frac{1}{2}} \tag{A6}$$

   To prove the monotonicity, it suffices to prove the monotonicity of $\lambda(\tilde{A}_\psi - \tilde{A}_\theta)$, i.e. the monotonicity of

$$f(\tilde{A}_\psi) = \frac{(\tilde{A}_\psi - \tilde{A}_\theta)^2}{\tilde{A}_\psi^\beta - \tilde{A}_\theta^\beta}$$

   Its derivative is:

$$f'(\tilde{A}_\psi) = \frac{(\tilde{A}_\psi - \tilde{A}_\theta)\tilde{A}_\psi^\beta[2 - \beta - 2\exp(-\beta(\tilde{a}_\psi - \tilde{a}_\theta)) + \beta\exp(-(\tilde{a}_\psi - \tilde{a}_\theta))]}{(\tilde{A}_\psi^\beta - \tilde{A}_\theta^\beta)^2}$$

To show $f'(\tilde{A}_\psi) > 0$ for any $\tilde{A}_\psi > \tilde{A}_\theta$, it suffices to show that

$$g(x) = 2 - \beta - 2\exp(-\beta x) + \beta \exp(-x)$$

is positive for any $x > 0$.

Note that $g(0) = 0$, $\lim_{x \to +\infty} = 2 - \beta > 0$. Moreover,

$$g'(x) = \beta(2\exp(-\beta x) - \exp(-x))$$

is at first positive then negative on $(0, +\infty)$, meaning that $g(x)$ is at first increasing then decreasing on $(0, +\infty)$. It then follows that $g(x) > 0$ for any positive x and that we have proved the monotonicity.

2. Let $a_j < a_k < a_\psi$. We want to prove that $\mathbb{E}(A_{j,t+\Delta t}) < \mathbb{E}(A_{k,t+\Delta t})$.

   By equation A6, $\lambda_j < \lambda_k$, hence

$$\begin{aligned}
\mathbb{E}(A_{j,t+\Delta t}) &= A_j + \lambda_j \Delta t (A_\psi - A_j) \\
&< A_j + \lambda_k \Delta t (A_\psi - A_j) \\
&= \lambda_k \Delta t A_\psi + (1 - \lambda_k \Delta t) A_j \\
&< \lambda_k \Delta t A_\psi + (1 - \lambda_k \Delta t) A_k \\
&= \mathbb{E}(A_{k,t+\Delta t})
\end{aligned}$$

□

The discussions above consider a specific destination of jump a_ψ. In the model, a_ψ comes from some distribution. We need to integrate the cost factor $e^{\beta \tilde{a}_\psi} - e^{\beta \tilde{a}_\theta}$ across a_ψ to obtain the relevant adjustment factor of the learning cost function. Importantly, the factor should capture the difficulty of learning *irrespective of* the *actual* learning destination ψ_l: if $a_{\theta_1} < a_{\theta_2}$ and thus a higher proportion of incumbents are situated above a_{θ_1} than above a_{θ_2}, such an effect should not be captured by the cost factor. This ensures that the adjustment factor captures the difficulty of knowledge and not the difficulty of being matched. In the case of $a_{\theta_1} < a_{\theta_2}$, to have the same condition of learning, the learning destination distribution shifts in a parallel way to the right in the case of θ_2 compared to that of θ_1.[84] I shall denote the *hypothetical* destination distribution as f in order to emphasize its difference from the *actual* learning distribution ψ_l.

[84] Using the field analogy in the previous footnote, the hypothetical learning field relative to the learning firm is the same in the two cases. This ensures that the cost factor captures the difficulty of knowledge and not a change of the field.

Given the above discussion, f should be a function of $a_\psi - a_\theta$, $\forall a_\psi \geq a_\theta$. Then the cost factor is given by:

$$
\begin{aligned}
\int_{a_\theta}^{+\infty} (e^{\beta\tilde{a}_\psi} - e^{\beta\tilde{a}_\theta}) f(a_\psi - a_\theta)\,\mathrm{d}a_\psi &= \int_{a_\theta}^{+\infty} (e^{\beta\tilde{a}_\psi} - e^{\beta\tilde{a}_\theta}) f(\tilde{a}_\psi - \tilde{a}_\theta)\,\mathrm{d}a_\psi \\
&= \int_{\tilde{a}_\theta}^{+\infty} (e^{\beta\tilde{a}_\psi} - e^{\beta\tilde{a}_\theta}) f(\tilde{a}_\psi - \tilde{a}_\theta)\,\mathrm{d}\tilde{a}_\psi \\
&= e^{\beta\tilde{a}_\theta} \int_{\tilde{a}_\theta}^{+\infty} (e^{\beta(\tilde{a}_\psi - \tilde{a}_\theta)} - 1) f(\tilde{a}_\psi - \tilde{a}_\theta)\,\mathrm{d}\tilde{a}_\psi \\
&= e^{\beta\tilde{a}_\theta} \int_{0}^{+\infty} (e^{\beta x} - 1) f(x)\,\mathrm{d}x \\
&= C e^{\beta\tilde{a}_\theta}
\end{aligned}
$$

Hence the adjustment factor of $e^{\beta\tilde{a}_\theta}$ in equation 52.

E Proof of non-interference conditions

As seen in the main text, the key is to prove

$$\lim_{x \to +\infty} \frac{\phi(x)}{\bar{\Phi}(x)} e^{(k_\psi - k_s)x} = +\infty$$

for any $k_s < k_\psi$ and any ultimately monotone PDF ϕ with an infinite right endpoint. The ultimate-monotonicity condition rules out pathological cases in which $\phi(x)$ touches 0 infinitely many times as $x \to +\infty$, for example when an ordinary PDF is multiplied by $1 + \sin(x)$. These strange cases do not appear in practice and are assumed away.

As ϕ in the model has an infinite right endpoint, the propositions below show that it can only be of two types: the Fréchet type and the Gumbel type. Take specific examples to understand the intuition:

- For the Fréchet type, take a Pareto-tailed distribution with $\bar{\Phi}(x) \sim Cx^{-\alpha}$ and $\phi(x) \sim C\alpha x^{-\alpha-1}$ as an example. Then $\frac{\phi(x)}{\bar{\Phi}(x)} e^{(k_\psi - k_s)x} = \alpha \frac{e^{(k_\psi - k_s)x}}{x} \to +\infty$ as $x \to +\infty$.
- For the Gumbel type, take the exponential-tailed distribution with $\bar{\Phi}(x) \sim Ce^{-\alpha x}$ and $\phi(x) \sim C\alpha e^{-\alpha x}$ as an example. This turns out to be the case for the log-productivity distribution in the model. Then $\frac{\phi(x)}{\bar{\Phi}(x)} e^{(k_\psi - k_s)x} = \alpha e^{(k_\psi - k_s)x} \to +\infty$ as $x \to +\infty$.

Propositions 3 and A8 establish the general result for *any* ultimately monotone distribution with an infinite right endpoint. The result paves the way for proving Proposition 4 in the baseline model or its counterpart in the extended model, which establishes the unique exponential tail of the equilibrium log-productivity distribution. At the present stage, however, the equilibrium tail is not known a priori. The non-interference conditions are thus needed for *any* distribution, as long as some weak conditions are satisfied.

The propositions used for proving Propositions 3 and A8 in this section can be found in Embrechts et al. (2013). Their proof can be found in the book and references therein.

Proposition A2 (Fisher Tippett Theorem for Maximum Domain of Attraction, Embrechts et al. (2013) Theorem 3.2.3)**.** *Let $\{X_n\}_{n\leq 1}$ be a sequence of i.i.d. random variables. If there exist norming constants $c_n > 0$, $d_n \in \mathbb{R}$ and some non-degenerate CDF H such that*

$$c_n^{-1}(M_n - d_n) \xrightarrow{\mathcal{L}} H \tag{A7}$$

where $M_n = \max(\{X_i\}_{1\leq i\leq n})$, then H belongs to one of the following three types of distribution:

- *Fréchet:*

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases} \quad \alpha > 0$$

- *Weibull:*

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^{\alpha}), & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0$$

- *Gumbel:*

$$\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}$$

The theorem says that any distribution must belong to one to one of three types in the sense of A7, if the norming constants exist. For practical purposes and following virtually all economics papers using the theorem, I assume that the constants exist. Distributions that belong to the Fréchet type constitute a family of distributions which is called the "Maximum Domain of Attraction" of Φ_α and shall be denoted $MDA(\Phi_\alpha)$. Similar notations are adopted for the other two types. The Fisher-Tippett theorem can be seen as a counterpart to the Central Limit Theorem (CLT): the maximum M_n is replaced by the sum $S_n = \sum_{i=1}^n X_i$ in the latter. Unlike the CLT, there is no unique Maximum Domain of Attraction in the Fisher Tippett Theorem. Nonetheless, the categorisation of *any* (i.e. both $\mathcal{L}^2$ and non-$\mathcal{L}^2$) distribution into only three types is very powerful and becomes *the* basis of extreme value theory.

The notion of a slowly varying function is useful in the extreme value theory:

Definition A1 (Slow Variation and Regular Variation)**.**

1. *A function $L : \mathbb{R} \to \mathbb{R}^+$ is slowly varying at ∞, denoted as $L \in \mathcal{R}_0$, if $\forall t > 0$*

$$\lim_{x\to+\infty} \frac{L(tx)}{L(x)} = 1$$

2. A function $L : \mathbb{R} \to \mathbb{R}^+$ is regularly varying at ∞, denoted as $L \in \mathcal{R}_\alpha$ where $\alpha \in \mathbb{R}$, if $\forall t > 0$

$$\lim_{x \to +\infty} \frac{L(tx)}{L(x)} = t^\alpha$$

Roughly speaking, a regularly varying function behaves like x^α as $x \to +\infty$. The intuitive idea behind a slowly varying function is that it changes more slowly than any power function:

Proposition A3. *If $L \in \mathcal{R}_0$, then*

$$\lim_{x \to +\infty} x^\alpha L(x) = \begin{cases} +\infty, & \text{if } \alpha > 0 \\ 0, & \text{if } \alpha < 0. \end{cases}$$

Proof. By definition, $x^\alpha L(x)$ is a regularly varying function at ∞. It suffices to apply Corollary A3.4 of Embrechts et al. (2013) to conclude. □

Examples of slowly varying functions include functions that converge to a positive constant, $\log(x)$ and $1/\log(x)$.

The following propositions fully characterize the Maximum Domain of Attraction of each type of distribution. We shall denote $\bar{F}(x) = 1 - F(x)$ as the survival function of a distribution F.

Proposition A4 (Characterization of $MDA(\Phi_\alpha)$, Embrechts et al. (2013) Theorem 3.3.7)**.** *A distribution F belongs to $MDA(\Phi_\alpha)$, $\alpha > 0$, if and only if its survival function can be represented as $\bar{F} = x^{-\alpha} L(x)$ for some slowly varying function L.*

Proposition A5 (Characterization of $MDA(\Psi_\alpha)$, Embrechts et al. (2013) Theorem 3.3.12)**.** *A distribution F belongs to $MDA(\Psi_\alpha)$, $\alpha > 0$, if and only if its right endpoint $x_F < +\infty$ and $\bar{F}(x_F - x^{-1}) = x^{-\alpha} L(x)$ for some slowly varying function L.*

Proposition A6 (Characterization of $MDA(\Lambda)$, Embrechts et al. (2013) Theorems 3.3.26 and 3.3.27)**.** *A distribution F with right endpoint $x_F \leq +\infty$ belongs to $MDA(\Lambda)$ if and only if there exists some positive function a such that*

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x + t a(x))}{\bar{F}(x)} = e^{-t}, \quad t \in \mathbb{R}$$

In particular, one can choose

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} \, \mathrm{d}t, \quad x < x_F \tag{A8}$$

which is absolutely continuous with respect to the Lebesgue measure and has density $a'(x)$ with $\lim_{x \uparrow x_F} a'(x) = 0$.

The following proposition will be useful for characterizing monotone density functions whose distributions belong to $MDA(\Phi_\alpha)$:

Proposition A7 (Monotone Density Theorem, Embrechts et al. (2013) Theorem A3.7)**.** *Let* $\bar{F}(x) = \int_x^{+\infty} f(y)\,\mathrm{d}y$ *where* f *is ultimately monotone (i.e.* f *is monotone on* $(z, +\infty)$ *for some* $z > 0$*). If*

$$\bar{F}(x) \sim cx^{-\alpha}L(x), \quad x \to +\infty$$

with $c > 0$*,* $\alpha \in \mathbb{R}$ *and* $L \in \mathcal{R}_0$*, then*

$$f(x) \sim c\alpha x^{-\alpha-1}L(x), \quad x \to +\infty$$

The following lemma is a variant of L'Hôpital's rule. It can be easily proved using Cauchy's Mean Value Theorem, similar to the proof of L'Hôpital's rule.

Lemma A1. *If* $f(x)$ *and* $g(x)$ *are differentiable on* $(a, +\infty)$*,* $g'(x) \neq 0\ \forall x \in (a, +\infty)$*,* $\lim_{x\to+\infty} f(x) = \lim_{x\to+\infty} g(x) = 0$*, and* $\lim_{x\to+\infty} \frac{f(x)}{g(x)}$ *exists. Then* $\lim_{x\to+\infty} \frac{f'(x)}{g'(x)}$ *exists and*

$$\lim_{x\to+\infty} \frac{f'(x)}{g'(x)} = \lim_{x\to+\infty} \frac{f(x)}{g(x)}$$

Proof. By a change of variable $y = \frac{1}{x}$ and define $\tilde{f}(y) = f(\frac{1}{y})$ and $\tilde{g}(y) = g(\frac{1}{y})$, the conditions become: $\lim_{y\to0+} \tilde{f}(y) = \lim_{y\to0+} \tilde{g}(y) = 0$ and $\lim_{y\to0+} \frac{\tilde{f}(y)}{\tilde{g}(y)}$ exists. The result to be proven becomes $\lim_{y\to0+} \frac{\tilde{f}'(y)}{\tilde{g}'(y)} = \lim_{y\to0+} \frac{\tilde{f}(y)}{\tilde{g}(y)}$. These two formulations are equivalent. We can redefine $\tilde{f}(0) = \tilde{g}(0) = 0$ to ensure that $\tilde{f}$ and $\tilde{g}$ are continuously at $y = 0$.

Then according to Cauchy's Mean Value Theorem, there exists a $\xi \in (0, y)$ such that

$$\frac{\tilde{f}(y)}{\tilde{g}(y)} = \frac{\tilde{f}(y) - \tilde{f}(0)}{\tilde{g}(y) - \tilde{g}(0)} = \frac{\tilde{f}'(\xi)}{\tilde{g}'(\xi)}$$

for any $y \in (0, \frac{1}{a})$. Taking the limit of $y \downarrow 0+$, we get:

$$\lim_{y\to0+} \frac{\tilde{f}(y)}{\tilde{g}(y)} = \lim_{y\to0+} \frac{\tilde{f}'(\xi)}{\tilde{g}'(\xi)} = \lim_{\xi\to0+} \frac{\tilde{f}'(\xi)}{\tilde{g}'(\xi)} = \lim_{y\to0+} \frac{\tilde{f}'(y)}{\tilde{g}'(y)}$$

This completes the proof. □

E.1 Non-interference condition of entry

We are now ready to prove **Proposition 3**.

Proof. As argued in the main text, it suffices to show

$$\lim_{x\to+\infty} \frac{\phi_t(x)}{\bar{\Phi}_t(x)} e^{(k_\psi - k_s)x} = +\infty$$

if $k_\psi - k_s > 0$.

I shall drop the time subscript for notational simplicity. Since ϕ has an infinite endpoint, the distribution cannot belong to $MDA(\Psi_\alpha)$ and must belong to either $MDA(\Phi_\alpha)$ or $MDA(\Lambda)$ according to the Fisher Tippett Theorem.

- If $\Phi \in MDA(\Phi_\alpha)$,
  then according to Proposition A4, $\bar{\Phi} = x^{-\alpha}L(x)$ for some slowly varying function L. Since ϕ is ultimately monotone, we can use the monotone density theorem to get:

  $$\phi \sim \alpha x^{-\alpha-1}L(x), x \to +\infty$$

  Hence

  $$\frac{\phi(x)}{\bar{\Phi}(x)}e^{(k_\psi - k_s)x} \sim \frac{\alpha x^{-\alpha-1}L(x)}{x^{-\alpha}L(x)}e^{(k_\psi - k_s)x} = \frac{\alpha e^{(k_\psi - k_s)x}}{x} \to +\infty, \text{ as } x \to +\infty$$

- If $\Phi \in MDA(\Lambda)$,
  then according to Proposition A6, there exists a positive function $a(x)$ with density $a'(x)$ having $\lim_{x\uparrow+\infty} a'(x) = 0$ such that

  $$\lim_{x\uparrow+\infty} \frac{\bar{\Phi}(x + ta(x))}{\bar{\Phi}(x)} = e^{-t}, \quad t \in \mathbb{R} \tag{A9}$$

  Fix a $t < 0$, then

  $$\frac{\bar{\Phi}(x + ta(x))}{\bar{\Phi}(x)} = \frac{\bar{\Phi}(x) - ta(x)\phi(\xi(x))}{\bar{\Phi}(x)} \leq 1 - \frac{ta(x)\phi(x + ta(x))}{\bar{\Phi}(x)}$$

  for some $\xi(x) \in [x + ta(x), x]$. The inequality uses the ultimate monotonicity of ϕ.

  Since $\lim_{x\uparrow+\infty} a'(x) = 0$, we know that $\forall \epsilon_1 > 0$, there exists a cutoff x_0 such that $|a'(x)| < \epsilon_1$ $\forall x \geq x_0$. Hence $|a(x) - a(x_0)| \leq \epsilon_1(x - x_0)$. Thus

  $$\frac{\bar{\Phi}(x + ta(x))}{\bar{\Phi}(x)} \leq 1 - \frac{t[a(x_0) + \epsilon_1(x - x_0)]\phi(x + ta(x))}{\bar{\Phi}(x)} \tag{A10}$$

  By A9 and A10, there exists a sufficiently small $\epsilon_2 > 0$, such that for sufficiently large values of x,

  $$-\frac{t[a(x_0) + \epsilon_1(x - x_0)]\phi(x + ta(x))}{\bar{\Phi}(x)} \geq e^{-t} - 1 - \epsilon_2 > 0$$

Using Lemma A1, we get:

$$\lim_{x\to+\infty}\frac{\phi(x)}{\phi(x+ta(x))}\frac{1}{1+ta'(x)}=\lim_{x\to+\infty}\frac{\bar{\Phi}(x)}{\bar{\Phi}(x+ta(x))}$$
$$\Rightarrow\lim_{x\to+\infty}\frac{\phi(x)}{\phi(x+ta(x))}=e^t$$

Thus

$$\frac{\phi(x)}{\bar{\Phi}(x)}e^{(k_\psi-k_s)x}=\frac{e^{(k_\psi-k_s)x}}{-t[a(x_0)+\epsilon_1(x-x_0)]}\frac{-t[a(x_0)+\epsilon_1(x-x_0)]\phi(x+ta(x))}{\bar{\Phi}(x)}\frac{\phi(x)}{\phi(x+ta(x))}$$
$$\to+\infty,\text{ as }x\to+\infty$$

□

E.2 Non-interference condition of incumbent learning in the extended model

Given the similar specification of incumbent learning as that of entry, a non-interference condition also holds for incumbent learning. The case is slightly more complex than entry, as we need to consider jumps from all $a_\theta\in(\underline{a},a_\psi)$ when thinking about the inserted mass at a_ψ due to learning:

$$\psi_l^{\text{in}}(a_\psi;\phi)\triangleq\int_{(\underline{a},a_\psi)}\psi_l(a_\psi;a_\theta,\phi)\phi(a_\theta)\lambda_l(a_\theta)\,\mathrm{d}a_\theta \tag{A11}$$

Proposition A8 (Non-interference Condition of Incumbent Learning)**.**
If ϕ is ultimately monotone, and if there exists $k'\in\mathbb{R}$ such that $\lambda_l(x)\sim o(e^{-k'x})$ as $x\to+\infty$, then $\lim_{a_\psi\to+\infty}\frac{\psi_l^{in}(a_\psi;\phi)}{\phi(a_\psi)}e^{k_s a_\psi}=0$, $\forall k_s<\min\{k',k_\psi\}$.
In particular, if $k_\psi>0$ and if $\lambda_l(x)$ decreases at least exponentially with x, then $\lim_{a_\psi\to+\infty}\frac{\psi_l^{in}(a_\psi;\phi)}{\phi(a_\psi)}=0$.

Remark:

1. In cases where leaders choose innovation over learning, there is a cutoff $\hat{a}$ such that $\lambda_l(a_\theta)=0$, $\forall a_\theta>\hat{a}$. Then k' can take any real value and thus k_s can be set to any number smaller than k_ψ.

2. Cases where $\lambda_l(x)\sim O(e^{-k_l x})$ as $x\to+\infty$ with $k_l>0$, i.e. $\lambda_l(x)$ decreases exponentially, are a special case of $\lambda_l(x)\sim o(e^{-k'x})$ by considering $k'=k_l-\epsilon$ for any $\epsilon>0$. k_s can be set to any number smaller than $\min\{k',k_\psi\}$.

Proof.

$$\begin{aligned}
\frac{\psi_l^{\text{in}}(a_\psi)}{\phi(a_\psi)}e^{k_s a_\psi} &= \frac{e^{k_s a_\psi}}{\phi(a_\psi)}\int_{\underline{a}}^{a_\psi}[k_\psi\overline{\Phi}(a_\psi)+\phi(a_\psi)]e^{-k_\psi(a_\psi-a_\theta)}\phi(a_\theta)\lambda_l(a_\theta)\,\mathrm{d}a_\theta \\
&= [k_\psi\frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)}+1]e^{(k_s-k_\psi)a_\psi}\int_{\underline{a}}^{a_\psi}e^{k_\psi a_\theta}\phi(a_\theta)\lambda_l(a_\theta)\,\mathrm{d}a_\theta \\
&\leq [k_\psi\frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)}+1]e^{(k_s-k_\psi)a_\psi}C_1\int_{\underline{a}}^{a_\psi}e^{(k_\psi-k')a_\theta}\,\mathrm{d}a_\theta
\end{aligned}$$

where the inequality arises from the fact that $\lambda_l(a_\theta)\sim o(e^{-k'a_\theta})$ as $a_\theta\to+\infty$ and $\phi(a_\theta)$ is bounded.

Depending on whether k_ψ or k' is larger, we consider three cases:

If $k_\psi<k'$, then the integral converges to a finite number as $a_\psi\to+\infty$. Hence by setting $k_s<k_\psi$, we prove the result by using the proof of Proposition 3.

If $k_\psi=k'$, then

$$\begin{aligned}
\frac{\psi_l^{\text{in}}(a_\psi)}{\phi(a_\psi)}e^{k_s a_\psi} &\leq [k_\psi\frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)}+1]e^{(k_s-k_\psi)a_\psi}C_1(a_\psi-\underline{a}) \\
&= [k_\psi\frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)}+1]e^{(k_s-k_\psi+\epsilon)a_\psi}C_1e^{-\epsilon a_\psi}(a_\psi-\underline{a}) \\
&\to 0
\end{aligned}$$

by fixing an $\epsilon\in(0,k_\psi-k_s)$ and by using the proof of Proposition 3.

If $k_\psi>k'$, then

$$\begin{aligned}
\frac{\psi_l^{\text{in}}(a_\psi)}{\phi(a_\psi)}e^{k_s a_\psi} &\leq [k_\psi\frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)}+1]e^{(k_s-k_\psi)a_\psi}C_2[e^{(k_\psi-k')a_\psi}-C_3] \\
&= [k_\psi\frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)}+1]e^{(k_s-k')a_\psi}C_2[1-C_3e^{(k'-k_\psi)a_\psi}]
\end{aligned}$$

When setting $k_s<k'$, we can see that $[1-C_3e^{(k'-k_\psi)a_\psi}]\to 1$ and $[k_\psi\frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)}+1]e^{(k_s-k')a_\psi}\to 0$ as $a_\psi\to+\infty$. This concludes our proof. □

F Proofs for the HJB-KF system

F.1 Random walk representation of the Brownian motion

The random walk representation of Brownian motion is intuitive and shall be used for proving the HJB-KF system by taking the limit of $\Delta t \to 0$.[85]

Consider a Brownian motion with drift in the continuous-time setting:

$$\mathrm{d}x_t = \mu \mathrm{d}t + \nu \mathrm{d}B_t \tag{A12}$$

Divide time into discrete periods of length Δt and define

$$\Delta h = \nu\sqrt{\Delta t} \tag{A13}$$

and

$$p = \frac{1}{2}\left[1 + \frac{\mu}{\nu}\sqrt{\Delta t}\right], \quad q = \frac{1}{2}\left[1 - \frac{\mu}{\nu}\sqrt{\Delta t}\right] \tag{A14}$$

Note that $p + q = 1$. Consider a two-dimensional space of time t and variable x. At each point on the grid, x has a probability p of going up Δh and a probability q of going down Δh. The discrete process then converges to the continuous process A12 as $\Delta t \to 0$.[86]

F.2 Prove the death rate from the left boundary

Using the left boundary condition of $\phi(\underline{a}) = 0$, we have:

$$\phi(\underline{a} + \Delta h) = \phi'(\underline{a})\Delta h + o(\Delta h)$$

Thus the number of firms dying from the left boundary per period is

$$\frac{M\phi(\underline{a} + \Delta h)\Delta h q}{\Delta t} = M\phi'(\underline{a})\nu^2 q + o(1) \to \frac{1}{2}\nu^2 M\phi'(\underline{a}) \text{ as } \Delta t \to 0$$

Note that it depends only on the Brownian motion and not on the drift term μ.

F.3 Prove the HJB equation

Proof. Each incumbent's optimization problem is given by:

[85] See for instance Dixit (2013) for an introduction to the representation.
[86] The convergence is based on the functional central limit theorem, i.e. Donsker's invariance principle.

$$V(a_{\theta,t}, t) = \max_{\lambda_{i,\theta,s}\}_{s\geq t}} \mathbb{E}_t \int_t^{+\infty} e^{-\int_t^s r_u \, \mathrm{d}u} \Big[\Pi(a_{\theta,s}; \phi_s^M) - fA_s - R_s(\lambda_{i,\theta,s}; \tilde{a}_{\theta,s})\Big] \, \mathrm{d}s \tag{A15}$$

$$\text{s.t. equation 29} \tag{A16}$$

Denote $\Omega_t(a_{\theta,t}; \lambda_{i,\theta,t}) = \Pi(a_{\theta,s}; \phi_s^M) - fA_s - R_s(\lambda_{i,\theta,s}; \tilde{a}_{\theta,s})$ to be the period t payoff function. Then the optimization problem A15 can be written recursively as:

$$\begin{aligned}
&V(a_{\theta,t}, t) \\
=& \max_{\lambda_{i,\theta,t}} \mathbb{E}_t \Big\{ \Omega_t(a_{\theta,t}; \lambda_{i,\theta,t})\Delta t + \max_{\{\lambda_{i,\theta,s}\}_{s\geq t+\Delta t}} \mathbb{E}_{t+\Delta t} \int_{t+\Delta t}^{+\infty} e^{-\int_t^s r_u \, \mathrm{d}u} \Omega_s(a_{\theta,s}; \lambda_{i,\theta,s}) \, \mathrm{d}s \Big\} + o(\Delta t) \\
=& \max_{\lambda_{i,\theta,t}} \mathbb{E}_t \Big\{ \Omega_t(a_{\theta,t}; \lambda_{i,\theta,t})\Delta t + e^{-\int_t^{t+\Delta t} r_u \, \mathrm{d}u} V(a_{\theta,t+\Delta t}, t + \Delta t) \Big\} + o(\Delta t) \\
=& \max_{\lambda_{i,\theta,t}} \mathbb{E}_t \Big\{ \Omega_t(a_{\theta,t}; \lambda_{i,\theta,t})\Delta t + (1 - r_t\Delta t) V(a_{\theta,t+\Delta t}, t + \Delta t) \Big\} + o(\Delta t) \\
=& \max_{\lambda_{i,\theta,t}} \Big\{ \Omega_t(a_{\theta,t}; \lambda_{i,\theta,t})\Delta t + (1 - r_t\Delta t) \mathbb{E}_t V(a_{\theta,t+\Delta t}, t + \Delta t) \Big\} + o(\Delta t)
\end{aligned} \tag{A17}$$

Use the random walk representation of the Brownian motion with drift for the continuous part of the stochastic process 29. Then

$$\begin{aligned}
\mathbb{E}_t V(a_{\theta,t+\Delta t}, t + \Delta t) &= pV(a_{\theta,t} + \Delta h, t + \Delta t) + qV(a_{\theta,t} - \Delta h, t + \Delta t) \\
&+ \lambda_{i,\theta,t}\Delta t V(a_{\theta,t} + q, t + \Delta t) + (1 - \lambda_{i,\theta,t}\Delta t - p - q) V(a_{\theta,t}, t + \Delta t) + o(\Delta t)
\end{aligned} \tag{A18}$$

where Δh, p and q are defined in A13 and A14. Note that we have proceeded as if only one of the two parts of the stochastic process can happen during the time period $[t, t + \Delta t)$: the continuous part of the Brownian motion with drift, the discontinuous part of jumps. The reason is that the term corresponding to the two happening together is of order $o(\Delta t)$ and thus absorbed in $o(\Delta t)$ in the equation. Because the continuous part is of the order of magnitude $O(\sqrt{\Delta t})$ or $o(\sqrt{\Delta t})$ as $\Delta t \to 0+$ (the drift term is of order Δt and the Brownian motion without drift is of order $\sqrt{\Delta t}$), and the discontinuous part is of order Δt, their joint term is of order $o(\Delta t)$.

Take the Taylor expansion of V with respect to a and t respectively to the second and first order in order to have residuals of order $o(\Delta t)$. For the first two terms of equation A18:

$$
\begin{aligned}
& pV(a_{\theta,t}+\Delta h, t+\Delta t) + qV(a_{\theta,t}-\Delta h, t+\Delta t) \\
=& p\Big[V(a_{\theta,t},t) + \frac{\partial V}{\partial a}(a_{\theta,t},t)\Delta h + \frac{1}{2}\frac{\partial^2 V}{\partial a^2}(a_{\theta,t},t)(\Delta h)^2 + \frac{\partial V}{\partial t}(a_{\theta,t},t)\Delta t + o(\Delta t)\Big] \\
& + q\Big[V(a_{\theta,t},t) - \frac{\partial V}{\partial a}(a_{\theta,t},t)\Delta h + \frac{1}{2}\frac{\partial^2 V}{\partial a^2}(a_{\theta,t},t)(\Delta h)^2 + \frac{\partial V}{\partial t}(a_{\theta,t},t)\Delta t + o(\Delta t)\Big] \\
=& V(a_{\theta,t},t) + \mu\frac{\partial V}{\partial a}(a_{\theta,t},t)\Delta t + \frac{1}{2}\nu^2\frac{\partial^2 V}{\partial a^2}(a_{\theta,t},t)\Delta t + \frac{\partial V}{\partial t}(a_{\theta,t},t)\Delta t + o(\Delta t)
\end{aligned} \tag{A19}
$$

For the second line of A18:

$$
\begin{aligned}
& \lambda_{i,\theta,t}\Delta t V(a_{\theta,t}+q, t+\Delta t) + (1-\lambda_{i,\theta,t}\Delta t - p - q)V(a_{\theta,t}, t+\Delta t) \\
=& \lambda_{i,\theta,t}\Delta t\Big[V(a_{\theta,t}+q, t+\Delta t) - V(a_{\theta,t}, t+\Delta t)\Big] \\
=& \lambda_{i,\theta,t}\Delta t\Big[V(a_{\theta,t}+q, t) - V(a_{\theta,t}, t)\Big] + o(\Delta t)
\end{aligned} \tag{A20}
$$

Combining equations A17, A18, A19 and A20, we get:

$$
\begin{aligned}
& V(a_{\theta,t},t) \\
=& \max_{\lambda_{i,\theta,t}}\Big\{\Omega_t(a_{\theta,t};\lambda_{i,\theta,t})\Delta t + (1-r_t\Delta t)\Big[V(a_{\theta,t},t) + \mu\frac{\partial V}{\partial a}(a_{\theta,t},t)\Delta t + \frac{1}{2}\nu^2\frac{\partial^2 V}{\partial a^2}(a_{\theta,t},t)\Delta t \\
& + \frac{\partial V}{\partial t}(a_{\theta,t},t)\Delta t + \lambda_{i,\theta,t}\Delta t[V(a_{\theta,t}+q,t) - V(a_{\theta,t},t)]\Big]\Big\} + o(\Delta t) \\
=& \max_{\lambda_{i,\theta,t}}\Big\{\Omega_t(a_{\theta,t};\lambda_{i,\theta,t})\Delta t + V(a_{\theta,t},t) + \mu\frac{\partial V}{\partial a}(a_{\theta,t},t)\Delta t + \frac{1}{2}\nu^2\frac{\partial^2 V}{\partial a^2}(a_{\theta,t},t)\Delta t \\
& + \frac{\partial V}{\partial t}(a_{\theta,t},t)\Delta t + \lambda_{i,\theta,t}\Delta t\Big[V(a_{\theta,t}+q,t) - V(a_{\theta,t},t)\Big] - r_t\Delta t V(a_{\theta,t},t)\Big\} + o(\Delta t)
\end{aligned}
$$

Cancelling $V(a_{\theta,t},t)$ on both sides of the equation, dividing both sides by Δt, and taking the limit of $\Delta t \to 0+$, we get the HJB equation 30.

□

F.4 Prove the KF equation

Proof. Thinking about the evolution of mass at position a_θ,

$$
\begin{aligned}
\phi^M(a_\theta, t+\Delta t) - \phi^M(a_\theta, t) =& p\phi^M(a_\theta-\Delta h, t) + q\phi^M(a_\theta-\Delta h, t) - (p+q)\phi^M(a_\theta, t) \\
& + \lambda_i(a_\theta-q,t)\Delta t\phi^M(a_\theta-q,t) - \lambda_i(a_\theta,t)\Delta t\phi^M(a_\theta,t) \\
& + \lambda_{e,t}\Delta t M_e\psi_e(a_\theta;\phi_t) - \lambda_d(a_\theta;\phi_t^M)\Delta t\phi^M(a_\theta,t) + o(\Delta t)
\end{aligned} \tag{A21}
$$

The first line of A21 on the right hand side of the equality corresponds to inflow into a_θ and outflow from a_θ due to Brownian motion with drift. The second line corresponds to inflow and outflow due to jumps and the third line corresponds to entry and exit. Like in the proof of HJB, we have ignored the interactions between these movements as they are of order $o(\Delta t)$.

For the first line,

$$
\begin{aligned}
&p\phi^M(a_\theta - \Delta h, t) + q\phi^M(a_\theta - \Delta h, t) - (p+q)\phi^M(a_\theta, t) \\
=&p\Big[\phi^M(a_\theta, t) - \Delta h\frac{\partial \phi^M}{\partial a}(a_\theta, t) + \frac{1}{2}(\Delta h)^2\frac{\partial^2\phi^M}{\partial a^2}(a_\theta, t) + o(\Delta t)\Big] \\
&+ q\Big[\phi^M(a_\theta, t) + \Delta h\frac{\partial \phi^M}{\partial a}(a_\theta, t) + \frac{1}{2}(\Delta h)^2\frac{\partial^2\phi^M}{\partial a^2}(a_\theta, t) + o(\Delta t)\Big] - \phi^M(a_\theta, t) \\
=&-\mu\frac{\partial \phi^M}{\partial a}(a_\theta, t)\Delta t + \frac{1}{2}\nu^2\frac{\partial^2\phi^M}{\partial a^2}(a_\theta, t)\Delta t + o(\Delta t)
\end{aligned}
\tag{A22}
$$

Plugging A22 into A21, dividing both sides of A21 by Δt and taking the limit of $\Delta t \to 0+$, we get the KF equation 37. $\square$

F.5 Prove the adjointness of the HJB operator and KF operator

We want to prove that for any $f : [\underline{a}, +\infty) \to \mathbb{R}$ with $f(\underline{a}) = 0$ and any test function $g : [\underline{a}, +\infty) \to \mathbb{R}$ with compact support and $g(\underline{a}) = 0$,

$$
< \mathcal{F}_t f, g > = < f, \mathcal{K}_t g > \tag{A23}
$$

where

$$
\mathcal{K}_t g = -\mu g' + \frac{1}{2}\nu^2 g'' - \lambda_d(a_\theta; \phi_t^M)g + \lambda_i(a_\theta - q, t)g(a_\theta - q) - \lambda_i(a_\theta, t)g(a_\theta) \tag{A24}
$$

i.e. the operator in the Kolmogorov equation 37. Equation A23 means that $\mathcal{K}_t = \mathcal{F}_t^\intercal$ and hence simplifies numerical calculations: when discretized, the operator matrix for the KF equation is simply the transpose of the operator matrix for the HJB equation.

In fact,

$$
< \mathcal{F}_t f, g > = \int_{\underline{a}}^{+\infty} g(a_\theta)\Big\{\mu f'(a_\theta) + \frac{1}{2}\nu^2 f''(a_\theta) + \lambda_i(a_\theta, t)\big[f(a_\theta + q) - f(a_\theta)\big] - \lambda_d(a_\theta; \phi_t^M)f(a_\theta)\Big\}\,\mathrm{d}a_\theta \tag{A25}
$$

where

$$
\begin{aligned}
&\int_{\underline{a}}^{+\infty} g(a_\theta)\Big\{\mu f'(a_\theta) + \frac{1}{2}\nu^2 f''(a_\theta)\Big\}\,\mathrm{d}a_\theta \\
=&\mu g(a_\theta)f(a_\theta)\Big|_{a_\theta=\underline{a}}^{+\infty} - \int_{\underline{a}}^{+\infty} \mu f(a_\theta)g'(a_\theta)\,\mathrm{d}a_\theta + \frac{1}{2}\nu^2 f'(a_\theta)g(a_\theta)\Big|_{a_\theta=\underline{a}}^{+\infty} - \int_{\underline{a}}^{+\infty} \frac{1}{2}\nu^2 f'(a_\theta)g'(a_\theta)\,\mathrm{d}a_\theta \\
=&-\int_{\underline{a}}^{+\infty} \mu f(a_\theta)g'(a_\theta)\,\mathrm{d}a_\theta - \frac{1}{2}\nu^2 f(a_\theta)g'(a_\theta)\Big|_{a_\theta=\underline{a}}^{+\infty} + \int_{\underline{a}}^{+\infty} \frac{1}{2}\nu^2 f(a_\theta)g''(a_\theta)\,\mathrm{d}a_\theta \\
=&-\int_{\underline{a}}^{+\infty} \mu f(a_\theta)g'(a_\theta)\,\mathrm{d}a_\theta + \int_{\underline{a}}^{+\infty} \frac{1}{2}\nu^2 f(a_\theta)g''(a_\theta)\,\mathrm{d}a_\theta
\end{aligned}
\tag{A26}
$$

and

$$
\begin{aligned}
&\int_{\underline{a}}^{+\infty} g(a_\theta)\lambda_i(a_\theta,t)\big[f(a_\theta+q) - f(a_\theta)\big]\,\mathrm{d}a_\theta \\
=&\int_{\underline{a}}^{+\infty} g(a_\theta)\lambda_i(a_\theta,t)f(a_\theta+q)\,\mathrm{d}a_\theta - \int_{\underline{a}}^{+\infty} g(a_\theta)\lambda_i(a_\theta,t)f(a_\theta)\,\mathrm{d}a_\theta \\
=&\int_{\underline{a}}^{+\infty} g(a_\theta-q)\lambda_i(a_\theta-q,t)f(a_\theta)\,\mathrm{d}a_\theta - \int_{\underline{a}}^{+\infty} g(a_\theta)\lambda_i(a_\theta,t)f(a_\theta)\,\mathrm{d}a_\theta \\
=&\int_{\underline{a}}^{+\infty} f(a_\theta)\Big[g(a_\theta-q)\lambda_i(a_\theta-q,t) - g(a_\theta)\lambda_i(a_\theta,t)\Big]\,\mathrm{d}a_\theta
\end{aligned}
\tag{A27}
$$

.
Plugging A27 and A26 into A25,

$$
\begin{aligned}
&< \mathcal{F}_t f, g > \\
=&\int_{\underline{a}}^{+\infty} f(a_\theta)\Big[-\mu g'(a_\theta) + \frac{1}{2}\nu^2 g''(a_\theta) + g(a_\theta-q)\lambda_i(a_\theta-q,t) - g(a_\theta)\lambda_i(a_\theta,t)\Big]\,\mathrm{d}a_\theta \\
=&< f, \mathcal{K}_t g >
\end{aligned}
$$

meaning that the KF operator $\mathcal{K}_t$ is the adjoint operator of the HJB operator $\mathcal{F}_t$.

F.6 Normalization

Since we are in a growth model, it is natural to normalize the distribution ϕ by the aggregate TFP a. Denote $\tilde{a}_\theta = a_\theta - a$ and define $\tilde{\phi}(\tilde{a}_\theta) \triangleq \phi(a_\theta)$ as the normalized log-productivity distribution. Moreover, define $\tilde{\phi}^M(\tilde{a}_\theta) \triangleq M\tilde{\phi}(\tilde{a}_\theta)$. It is easy to verify the following results based on the definitions:

Lemma A2.

1. $\psi(a_\psi; a_\theta, a_j) = \psi(\tilde{a}_\psi; \tilde{a}_\theta, \tilde{a}_j)$
2. $\psi_e(a_\psi; \phi) = \psi_e(\tilde{a}_\psi; \tilde{\phi})$

Proposition A9.

1. $\Pi(a_\theta; \phi^M) = A \cdot \Pi(\tilde{a}_\theta; \tilde{\phi}^M)$

2. $\lambda_d(a_\theta; \phi^M) = \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M)$

Proof.

1. According to the proof in Proposition 1, if each A_θ is divided by A, then the price-wage ratio P/w increases by A but the relative quantity $Z_\theta = \frac{Y_\theta}{Y}$ does not change. The increase of P/w reflects a decrease of w by A, as wage tracks productivity in such a growth model. Hence

$$
\begin{aligned}
P(\tilde{a}_\theta; \tilde{\phi}^M) &= P(a_\theta; \phi^M) \\
Z(\tilde{a}_\theta; \tilde{\phi}^M) &= Z(a_\theta; \phi^M)
\end{aligned}
$$

Since labor is inelastically supplied, $Y = AL$ decreases by A after normalization. Thus $Y_\theta = YZ_\theta$ decreases by A and $L_\theta = Y_\theta / A_\theta$ does not change. Moreover, $P_\theta = P\frac{P_\theta}{P} = P\gamma'(Z_\theta)$ does not change. Put together,

$$
\begin{aligned}
\Pi(a_\theta; \phi^M) &= P_\theta Y_\theta - wL_\theta \\
&= P(a_\theta; \phi^M)Y(a_\theta; \phi^M) - wL(a_\theta; \phi^M) \\
&= P(\tilde{a}_\theta; \tilde{\phi}^M)AY(\tilde{a}_\theta; \tilde{\phi}^M) - A\tilde{w}L(\tilde{a}_\theta; \tilde{\phi}^M) \\
&= A\Pi(\tilde{a}_\theta; \tilde{\phi}^M)
\end{aligned}
$$

2. Using the arguments above and the definition of λ_d, the result is obvious.

□

We can now proceed with the normalization of the HJB and KF equations.

Denote $\tilde{V}(\tilde{a}_\theta, t) = \frac{1}{A_t} V(a_\theta, t)$, then based on equation 30:

$$
\begin{aligned}
r_t A_t \tilde{V}(\tilde{a}_\theta, t) = \max_{\lambda_i} \Big\{ & A_t \Pi(\tilde{a}_\theta; \tilde{\phi}_t^M) - fA_t - R_t(\lambda_i; \tilde{a}_\theta) - \frac{1}{2}\nu^2 \frac{\partial \tilde{V}}{\partial \tilde{a}}(\tilde{a}_\theta, t)A_t + \frac{1}{2}\nu^2 \frac{\partial^2 \tilde{V}}{\partial \tilde{a}^2}(\tilde{a}_\theta, t)A_t \\
&+ \lambda_i A_t [\tilde{V}(\tilde{a}_\theta + q, t) - \tilde{V}(\tilde{a}_\theta, t)] - \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M) A_t \tilde{V}(\tilde{a}_\theta, t) + \frac{\partial A_t}{\partial t} \tilde{V}(\tilde{a}_\theta, t) + A_t [\frac{\partial \tilde{V}}{\partial t}(\tilde{a}_\theta, t) - g_t \frac{\partial \tilde{V}}{\partial \tilde{a}}(\tilde{a}_\theta, t)] \Big\}
\end{aligned}
$$

Note I have slightly abused the notation by reconsidering λ_i as a function of normalized log productivity. Dividing A_t on both sides and rearranging the terms, we get the HJB with normalized notation:

$$
\begin{aligned}
(r_t - g_t)\tilde{V}(\tilde{a}_\theta, t) = \max_{\lambda_i} \Big\{ & \Pi(\tilde{a}_\theta; \tilde{\phi}_t^M) - f - \tilde{R}_t(\lambda_i; \tilde{a}_\theta) - [\frac{1}{2}\nu^2 + g_t] \frac{\partial \tilde{V}}{\partial \tilde{a}}(\tilde{a}_\theta, t) + \frac{1}{2}\nu^2 \frac{\partial^2 \tilde{V}}{\partial \tilde{a}^2}(\tilde{a}_\theta, t) \\
&+ \lambda_i [\tilde{V}(\tilde{a}_\theta + q, t) - \tilde{V}(\tilde{a}_\theta, t)] - \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M)\tilde{V}(\tilde{a}_\theta, t) + \frac{\partial \tilde{V}}{\partial t}(\tilde{a}_\theta, t) \Big\} \quad \text{(A28)}
\end{aligned}
$$

where

$$\tilde{R}_t(\lambda_i;\tilde{a}_\theta) \triangleq \frac{1}{A_t}R_t(\lambda_i;\tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta\tilde{a}_\theta}\lambda_i^2$$

Similarly, the maximization problem of entrants now becomes:

$$\max_{\lambda_{e,t}}\Big\{-\tilde{R}_{e,t}(\lambda_{e,t}) + \lambda_{e,t}\int_{\mathbb{R}}\tilde{V}(\tilde{a}_\psi,t)\psi_e(\tilde{a}_\psi;\tilde{\phi}_t)\,\mathrm{d}\tilde{a}_\psi\Big\}$$

where

$$\tilde{R}_{e,t}(\lambda_{e,t}) \triangleq \frac{1}{A_t}R_{e,t}(\lambda_{e,t}) = \frac{1}{2}\alpha_e\lambda_{e,t}^2$$

For the KFE, first note that $\phi^M(a_\theta,t) = \tilde{\phi}^M(a_\theta - \int_0^t g_s\,\mathrm{d}s, t)$. Thus

$$\frac{\partial\phi^M}{\partial t}(a_\theta,t) = \frac{\partial\tilde{\phi}^M}{\partial t}(\tilde{a}_\theta,t) - \frac{\partial\tilde{\phi}^M}{\partial\tilde{a}}(\tilde{a}_\theta,t)g_t$$

Hence the normalized version is:

$$\begin{aligned}\frac{\partial\tilde{\phi}^M}{\partial t}(\tilde{a}_\theta,t) = -\frac{\partial[-(\frac{1}{2}\nu^2+g_t)\tilde{\phi}^M]}{\partial\tilde{a}}(\tilde{a}_\theta,t)+\frac{1}{2}\nu^2\frac{\partial^2\tilde{\phi}^M}{\partial\tilde{a}^2}(\tilde{a}_\theta,t)+\lambda_{e,t}M_e\psi_e(\tilde{a}_\theta;\tilde{\phi}_t)-\lambda_d(\tilde{a}_\theta;\tilde{\phi}_t^M)\tilde{\phi}^M(\tilde{a}_\theta,t)\\ +\lambda_i(\tilde{a}_\theta-q,t)\tilde{\phi}^M(\tilde{a}_\theta-q,t)-\lambda_i(\tilde{a}_\theta,t)\tilde{\phi}^M(\tilde{a}_\theta,t)\end{aligned}$$

Define the normalized version of the differential operator:

$$\tilde{\mathcal{F}}_t\tilde{V} = -[\frac{1}{2}\nu^2+g_t]\frac{\partial\tilde{V}}{\partial\tilde{a}} + \frac{1}{2}\nu^2\frac{\partial^2\tilde{V}}{\partial\tilde{a}^2} + \lambda_i[\tilde{V}(\tilde{a}_\theta+q)-\tilde{V}(\tilde{a}_\theta)] - \lambda_d(\tilde{a}_\theta;\tilde{\phi}^M)\tilde{V}$$

Then the HJB is:

$$(r_t-g_t)\tilde{V}(\tilde{a}_\theta,t) = \max_{\lambda_i}\Big\{\Pi(\tilde{a}_\theta;\tilde{\phi}_t^M) - f - \tilde{R}_t(\lambda_i;\tilde{a}_\theta) + \tilde{\mathcal{F}}_t\tilde{V} + \frac{\partial\tilde{V}}{\partial t}\Big\} \tag{A29}$$

with boundary condition

$$\tilde{V}(\underline{\tilde{a}}_t,t) = 0$$

The KFE is:

$$\frac{\partial\tilde{\phi}^M}{\partial t} = \tilde{\mathcal{F}}_t^{\intercal}\tilde{\phi}^M + \lambda_{e,t}M_e\psi_e(\cdot;\tilde{\phi}_t) \tag{A30}$$

with boundary condition

$$\tilde{\phi}^M(\underline{\tilde{a}}_t,t) = 0$$

F.7 Travelling wave solution: Proof of normalization in Definition 1

Proof. First note that all the time indices can be dropped in the normalized version of HJB and KFE in the last section, as the travelling wave after normalization is stationary with all the variables being constant.

The HJB under a travelling wave can be easily proved from equation A28. For the KFE, the equation regarding $\tilde{\phi}$ is straightforward: one only needs to divide M from both sides of equation A30.

To prove equation 48, first rewrite the explicit form of KFE under a travelling wave equilibrium:

$$0 = (\frac{1}{2}\nu^2 + g)(\tilde{\phi}^M)' + \frac{1}{2}\nu^2(\tilde{\phi}^M)'' + \lambda_e M_e \psi_e - \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M)\tilde{\phi}^M + \lambda_i(\tilde{a}_\theta - q)\tilde{\phi}^M(\tilde{a}_\theta - q) - \lambda_i(\tilde{a}_\theta)\tilde{\phi}^M(\tilde{a}_\theta)$$

Integrate the above equation over $[\underline{\tilde{a}}, +\infty)$, using the left boundary condition. Then

$$0 = \frac{1}{2}\nu^2(\tilde{\phi}^M)'(\underline{\tilde{a}}) + \lambda_e M_e - \int_{\underline{\tilde{a}}}^{+\infty} \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M)\tilde{\phi}^M(\tilde{a}_\theta)\,\mathrm{d}\tilde{a}_\theta$$

or equivalently

$$\lambda_e M_e = M\Big[\int_{\mathbb{R}} \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M)\tilde{\phi}(\tilde{a}_\theta)\,\mathrm{d}\tilde{a}_\theta + \frac{\nu^2}{2}\tilde{\phi}'(\underline{\tilde{a}})\Big]$$

□

F.8 Indistinguishable economies

Proof of **Proposition 5**

Proof. We prove by verification. Given the same $\tilde{\phi}$ and M across the two economies, the profit function of the economy with $s \neq 1$ (henceforth economy 2) is s times the profit function of the economy with $s = 1$ (henceforth economy 1).

This is because their Z_θ, P/w and A are the same by equations 19, 4 and 17, thus the total welfare is sY in economy 2 and the output of firm θ is $Z_\theta sY$ which is s times large. The scale of production is s times large and the same for the profit. As the fixed cost also changes by s times, the left boundary $\underline{\tilde{a}}$ remains the same.

It is easy to check that $s\tilde{V}$ is the value function of economy 2 (HJB 41), given that α also changes by s. The choice of λ_i and λ_e remains the same, and consequently the equilibrium distribution $\tilde{\phi}$, the measure of firms M and the aggregate growth g remain the same. This feeds into our assumption at the beginning of the proof. □

F.9 HJB-KF system of the extended model

The HJB and KF equations have intuitive interpretations, thus I will not repeat the proofs for the extended model. Interested readers can follow the same procedure for deriving the equations when incumbent learning is introduced. Normalization under a travelling wave equilibrium follows the same procedure and will also be skipped. I shall only prove the adjointness between the HJB operator and the KF operator to show that this is a general feature in a Mean Field Game. The proof is for the non-normalized setting with time index t, which can be easily adapted for the normalized case under a travelling wave equilibrium.

We want to prove that for any $f : [\underline{a}, +\infty) \to \mathbb{R}$ with $f(\underline{a}) = 0$ and any test function $g : [\underline{a}, +\infty) \to \mathbb{R}$ with compact support and $g(\underline{a}) = 0$,

$$< \mathcal{F}_t f, g > = < f, \mathcal{K}_t g > \tag{A31}$$

where

$$\mathcal{F}_t f = -\frac{1}{2}\nu^2 f' + \frac{1}{2}\nu^2 f'' + \lambda_c(a_\theta, t)\int_{\mathbb{R}} \big[f(a_\psi) - f\big]\psi_c(a_\psi; a_\theta)\,\mathrm{d}a_\psi - \lambda_d(a_\theta; \phi_t^M) f \tag{A32}$$

and

$$\begin{aligned}\mathcal{K}_t g = -\mu g' + \frac{1}{2}\nu^2 g'' - \lambda_d(a_\theta; \phi_t^M) g \\ + \int_{(-\infty, a_\theta)} \lambda_{c,t}(a_\psi) g(a_\psi)\psi_c(a_\theta; a_\psi)\,\mathrm{d}a_\psi - g(a_\theta)\lambda_{c,t}(a_\theta)\int_{(a_\theta, +\infty)} \psi_c(a_\psi; a_\theta)\,\mathrm{d}a_\psi\end{aligned} \tag{A33}$$

Note that in A33, c in the first integral is chosen based on a_ψ, while in the second integral it is chosen based on a_θ. I have dropped ϕ_t from the expression of ψ_c for notational simplicity.

In fact,

$$\begin{aligned}< \mathcal{F}_t f, g > = \int_{\underline{a}}^{+\infty} g(a_\theta)\Big\{\mu f'(a_\theta) + \frac{1}{2}\nu^2 f''(a_\theta) \\ + \lambda_{c(a_\theta)}(a_\theta, t)\int_{a_\theta}^{+\infty} \big[f(a_\psi) - f(a_\theta)\big]\psi_{c(a_\theta)}(a_\psi; a_\theta)\,\mathrm{d}a_\psi - \lambda_d(a_\theta; \phi_t^M) f(a_\theta)\Big\}\,\mathrm{d}a_\theta\end{aligned} \tag{A34}$$

where

$$\int_{\underline{a}}^{+\infty} g(a_\theta)\Big\{\mu f'(a_\theta) + \frac{1}{2}\nu^2 f''(a_\theta)\Big\}\,\mathrm{d}a_\theta = -\int_{\underline{a}}^{+\infty} \mu f(a_\theta) g'(a_\theta)\,\mathrm{d}a_\theta + \int_{\underline{a}}^{+\infty} \frac{1}{2}\nu^2 f(a_\theta) g''(a_\theta)\,\mathrm{d}a_\theta \tag{A35}$$

according to equation A26, and

$$
\begin{aligned}
&\int_{\underline{a}}^{+\infty} g(a_\theta)\lambda_{c(a_\theta)}(a_\theta,t)\int_{a_\theta}^{+\infty}\big[f(a_\psi)-f(a_\theta)\big]\psi_{c(a_\theta)}(a_\psi;a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta \\
=&\int_{\underline{a}}^{+\infty}\int_{a_\theta}^{+\infty} g(a_\theta)\lambda_{c(a_\theta)}(a_\theta,t)f(a_\psi)\psi_{c(a_\theta)}(a_\psi;a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta \\
&-\int_{\underline{a}}^{+\infty} g(a_\theta)\lambda_{c(a_\theta)}(a_\theta,t)f(a_\theta)\int_{a_\theta}^{+\infty}\psi_{c(a_\theta)}(a_\psi;a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta && \text{(A36)}\\
=&\int_{\underline{a}}^{+\infty}\int_{\underline{a}}^{a_\psi} g(a_\theta)\lambda_{c(a_\theta)}(a_\theta,t)f(a_\psi)\psi_{c(a_\theta)}(a_\psi;a_\theta)\,\mathrm{d}a_\theta\,\mathrm{d}a_\psi \\
&-\int_{\underline{a}}^{+\infty} g(a_\theta)\lambda_{c(a_\theta)}(a_\theta,t)f(a_\theta)\int_{a_\theta}^{+\infty}\psi_{c(a_\theta)}(a_\psi;a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta && \text{(A37)}\\
=&\int_{\underline{a}}^{+\infty} f(a_\theta)\Big[\int_{\underline{a}}^{a_\theta} g(a_\psi)\lambda_{c(a_\psi)}(a_\psi,t)\psi_{c(a_\psi)}(a_\theta;a_\psi)\,\mathrm{d}a_\psi \\
&-g(a_\theta)\lambda_{c(a_\theta)}(a_\theta,t)\int_{a_\theta}^{+\infty}\psi_{c(a_\theta)}(a_\psi;a_\theta)\,\mathrm{d}a_\psi\Big]\,\mathrm{d}a_\theta && \text{(A38)}
\end{aligned}
$$

.

We have changed the order of integration for the first double integral from A36 to A37. From A37 to A38 we have simply shifted the indexing between a_θ and a_ψ for the first double integral.

Plugging A38 and A35 into A34,

$$
\begin{aligned}
&<\mathcal{F}_t f, g> \\
=&\int_{\underline{a}}^{+\infty} f(a_\theta)\Big[-\mu g'(a_\theta)+\frac{1}{2}\nu^2 g''(a_\theta)+\int_{\underline{a}}^{a_\theta} g(a_\psi)\lambda_{c(a_\psi)}(a_\psi,t)\psi_{c(a_\psi)}(a_\theta;a_\psi)\,\mathrm{d}a_\psi \\
&-g(a_\theta)\lambda_{c(a_\theta)}(a_\theta,t)\int_{a_\theta}^{+\infty}\psi_{c(a_\theta)}(a_\psi;a_\theta)\,\mathrm{d}a_\psi\Big]\,\mathrm{d}a_\theta \\
=&< f,\mathcal{K}_t g>
\end{aligned}
$$

meaning that the KF operator $\mathcal{K}_t$ is the adjoint operator of the HJB operator $\mathcal{F}_t$.

G Proof of the unique tail of KF equation

The non-interference condition of entry will play a key role in proving the uniqueness of the equilibrium tail, regardless of the initial condition of ϕ. To grasp the intuition, consider a dynamic process in which individual firms constantly evolve over time. If $\phi(a)\sim C_1e^{-k_\phi a}$, then $\psi_e(a)\sim C_2e^{-(k_\phi+k_\psi)a}$. Thus ψ_e has a lighter tail than ϕ if $k_\psi>0$, and the non-interference condition holds. When two exponential tails are linearly combined, one with a heavier tail than the other, the heavier tail dominates. For instance, $C_4\phi(a)+C_5\phi_e(a)\sim C_6e^{-k_\phi a}$. The non-interference condition thus ensures that entrants do not change the tail of ϕ when they are injected into it. This is supported by the data in which the entrant distribution has a lighter tail than the incumbent distribution. Consequently, the tail of ϕ is determined by the drift term ($\mathrm{d}t$ term) and idiosyncratic shocks ($\mathrm{d}B_{\theta,t}$ term). Entry affects the tail of ϕ mainly through a change in the growth rate which appears in the drift term.

It is useful to compare with the case in which $k_\psi = 0$ and the non-interference conditions fail. Denote k_ϕ to be the tail index *if only the drift term and the idiosyncratic shocks shape the distribution*. Then if the initial condition of ϕ is very thin-tailed, for instance a Dirac mass point, then the injection of ψ_e does not interfere with the process of converging towards the tail k_ϕ, as they are parameterized by past values of ϕ which are more thin-tailed. This sheds light on why Luttmer (2012) and Perla et al. (2021) assume an initial condition with a lighter tail. On the other hand, if the initial condition is heavier-tailed than k_ϕ, then ψ_e keeps interfering with the incumbent distribution every period so that the equilibrium distribution depends on the initial distribution.

The mild assumption of imperfect learning and its implication of non-interference can thus establish the tail uniqueness in a powerful way. The proofs below demonstrate the intuition in a rigorous fashion. Moreover, they show that the log-productivity distribution must be exponential-tailed and thus the productivity distribution must be Pareto-tailed.

The proof will be based on the Laplace transform defined by:

$$\hat{f}(\xi) = \int_{\underline{a}}^{+\infty} e^{-\xi a} f(a)\,\mathrm{d}a \tag{A39}$$

where $\xi = \xi^r + \iota\xi^\iota \in \mathbb{C}$, $\iota^2 = -1$, ξ^r and ξ^ι are respectively the real and imaginary part of ξ. While the Laplace transform is often used with real ξ in practice, we need a complex ξ to define a complex function $\hat{f}(\xi)$. This allows us to take advantage of holomorphic (i.e. analytic) and meromorphic functions which will be key for the proof.[87] As in complex analysis, $e^{-\xi a} = e^{-\xi^r a} e^{-\iota\xi^\iota a} = e^{-\xi^r a}[\cos(\xi^\iota a) - \iota\sin(\xi^\iota a)]$.

Definition A2 (At Least Exponential Decay)**.**
A non-negative function $f(x)$ is said to decay at least exponentially to $f(\infty) \geq 0$ if there exists an $s > 0$ such that $|f(x) - f(\infty)| \sim o(e^{-sx})$.

Remark: The exponential decay of e^{-kx} is obviously a case of at least exponential decay by setting $s \in (0, k)$.

[87] See any textbook on complex analysis for holomorphic and meromorphic functions, for instance Stein and Shakarchi (2010).

Proposition A10 (Widder (1946) Corollary 1a 1b)**.**

1. *If the integral A39 converges for* $\xi = \xi^r + \iota\xi^\iota$*, it converges for all* $\zeta = \zeta^r + \iota\zeta^\iota$ *for which* $\zeta^r > \xi^r$*.*

2. *If the integral A39 diverges for* $\xi = \xi^r + \iota\xi^\iota$*, it diverges for all* $\zeta = \zeta^r + \iota\zeta^\iota$ *for which* $\zeta^r < \xi^r$*.*

Based on the Proposition, there are three possible cases:

- The integral converges for any $\xi \in \mathbb{C}$
- The integral diverges for any $\xi \in \mathbb{C}$
- There exists a $\xi_0^r \in \mathbb{R}$ such that the integral converges for any $\xi \in \mathbb{C}$ with $\xi^r > \xi_0^r$, and diverges for any $\xi \in \mathbb{C}$ with $\xi^r < \xi_0^r$

$\xi_0^r \in \mathbb{R}$ is called the abscissa of convergence. We harmonize the notation by denoting $-\infty$ as the abscissa of convergence for the first case and $+\infty$ for the second. In cases where f is a PDF, the abscissa of convergence must fall within the range of $[-\infty, 0]$, as $\hat{f}(0) = \int_{\underline{a}}^{+\infty} f(a)\,\mathrm{d}a = 1$ converges. As examples, $\xi_0^r = 0$ if f is Pareto-tailed; $\xi_0^r \in (-\infty, 0)$ if f is exponential-tailed; $\xi_0^r = -\infty$ if f is normal-tailed.

To give an idea of the proof, consider an exponentially-distributed $f(a) = Ce^{-sa}$, where $s > 0$. Then $\hat{f}(\xi) = \frac{C}{\xi+s}e^{-(\xi+s)\underline{a}}$, which is analytic for $\xi > -s$ and has a pole at $\xi = -s$. Intuitively, the Laplace transform allows us to gauge the tail index of the distribution (s in this case) by focusing on its singularities. The proof follows a similar idea as that in Gabaix et al. (2016) but allows for heterogeneous growth and death across firms. The following propositions will be used for the proof. In particular, Mimica (2016) allows us to establish the tail of f from its Laplace transform.

Proposition A11 (Widder (1946) Theorem 5a)**.**
$\hat{f}(\xi)$ *as defined by equation A39 is analytic for any* ξ *whose real part* $\xi^r > \xi_0^r$*.*

Proposition A12 (Widder (1946) Theorem 5b)**.**
If $\mathrm{d}F(a) = f(a)\mathrm{d}a$ *and if* F *is monotonic, then* ξ_0^r *is a singularity of* $\hat{f}(\xi)$*.*

Remark: In cases where f is a PDF and its abscissa of convergence is 0, even though $\hat{f}(0)$ converges, 0 is still a singular point of $\hat{f}$ in the sense of complex analysis. This arises from the fact the definition of analyticity requires a function to have Taylor expansions *in a neighborhood* of a point in order to be analytic at that point. As $\hat{f}$ cannot be defined for any $\xi < 0$ if $\xi_0^r = 0$, 0 is a singular point of $\hat{f}$.

Proposition A13 (Widder (1946) Theorem 2.4a)**.** *If* $\limsup_{a\to+\infty} \frac{\log|f(a)|}{a} = \xi_0^r \neq 0$*, then* ξ_0^r *is the abscissa of convergence of equation A39.*

Proposition A14 (Widder (1946) Theorem 2.4b)**.** *If equation A39 has non-negative abscissa of convergence* ξ_0^r*, then* $\limsup_{a\to+\infty}\frac{\log|f(a)|}{a}=\xi_0^r$.

Proposition A15 (Mimica (2016) Corollary 1.4)**.**
If f *is a PDF and if* ξ_0^r *is a pole of* $\hat{f}$*, then*

$$\lim_{a\to+\infty}\frac{\log(\bar{F}(a))}{a}=\xi_0^r$$

where $\bar{F}(a)=1-F(a)$ *and* F *is the CDF of* f*.*

G.1 Proof for the KF equation in the baseline model

To simplify the proofs, first note that the baseline model can be approximated by an alternative model in which each incumbent evolves continuously without jumps. In fact, the HJB equation 41 and the KF equation 46 can be approximated respectively by:

$$(r-g)\tilde{V}=\max_{\lambda_i}\Big\{\Pi(\tilde{a}_\theta;\tilde{\phi}^M)-f-\tilde{R}(\lambda_i;\tilde{a}_\theta)+\Big[-\frac{1}{2}\nu^2-g+\lambda_i q\Big]\tilde{V}'(\tilde{a}_\theta)+\frac{1}{2}\nu^2\tilde{V}''(\tilde{a}_\theta)-\lambda_d(\tilde{a}_\theta;\tilde{\phi}^M)\tilde{V}\Big\} \tag{A40}$$

$$0=(\frac{1}{2}\nu^2+g)\tilde{\phi}'+\frac{1}{2}\nu^2\tilde{\phi}''+\frac{\lambda_e M_e}{M}\psi_e-\lambda_d(\tilde{a}_\theta;\tilde{\phi}^M)\tilde{\phi}-q\frac{\mathrm{d}[\lambda_i(\tilde{a}_\theta)\tilde{\phi}(\tilde{a}_\theta)]}{\mathrm{d}\tilde{a}_\theta} \tag{A41}$$

which are HJB and KF equations of an alternative model in which the normalized log productivity evolves according to

$$\mathrm{d}\tilde{a}_{\theta,t}=\Big[-\frac{1}{2}\nu^2-g+\lambda_i(\tilde{a}_{\theta,t})q\Big]\mathrm{d}t+\nu\mathrm{d}B_{\theta,t}$$

The only difference in the alternative model is that log productivity now improves with certainty with per-period rate $\lambda_i q$ after incurring the innovation cost. Firm dynamics is thus different at the micro level. However, macro level variables including the productivity distribution are approximately the same under the two models. Both models can use exactly the same numerical scheme in computations. We will stick with the continuous specification as it simplifies some of the proofs.

The proof follows the following steps:

1. Prove the death rate declines at least exponentially with log productivity to the asymptotic death rate, thus a similar non-interference condition is satisfied for death.

2. Prove the asymptotic property of the value function.

3. Prove the monotonicity of the value function.

4. Prove the at least exponential decay of the innovation intensity.

5. Prove the ultimate monotonicity of the equilibrium log productivity PDF

6. Prove the uniqueness of the tail of the KF equation, using the previous results.

For notational conciseness, I will drop the tilde notation in this proof even though the HJB and KF equations are the normalized version. I will also drop notations of the dependence on ϕ whenever it does not cause confusion.

Proposition A16 (Decay of Death Function)**.**
$\lambda_d(a_\theta)$ decreases at least exponentially to $\lambda_d(\infty)$.

Proof. Using equations A3 and A4, we have:

$$
\begin{aligned}
\frac{\partial \log(P_\theta Y_\theta)}{\partial a_\theta} &= \frac{\partial z_\theta}{\partial a_\theta} - \frac{\partial \log(P_\theta)}{\partial MC_\theta} \\
&= \frac{(\sigma_\theta - 1)^2}{\sigma_\theta - 1 - \frac{\partial \sigma}{\partial z_\theta}} \\
&< \sigma_\theta - 1
\end{aligned}
$$

Thus

$$
\begin{aligned}
\lambda_d(a_\theta) &= \mathbb{P}\Big(\frac{\partial \log(P_\theta Y_\theta)}{\partial a_\theta} \cdot \nu B_1 < -\kappa\Big) \\
&= \mathbb{P}\Big(B_1 < -\frac{\kappa}{\nu}[\frac{\partial \log(P_\theta Y_\theta)}{\partial a_\theta}]^{-1}\Big) \\
&< \mathbb{P}\Big(B_1 < -\frac{\kappa}{\nu(\sigma_\theta - 1)}\Big) \\
&= \Phi_{\mathcal{N}}\Big(-\frac{\kappa}{\nu(\sigma_\theta - 1)}\Big) \\
&\triangleq \overline{\lambda}_d(a_\theta)
\end{aligned}
\tag{A42}
$$

where $\Phi_{\mathcal{N}}$ is the CDF of $\mathcal{N}(0,1)$.

It is easy to see that

$$
\lim_{a_\theta \to +\infty} \overline{\lambda}_d(a_\theta) = \Phi_{\mathcal{N}}\Big(-\frac{\kappa}{\nu(\underline{\sigma} - 1)}\Big) \triangleq \overline{\lambda}_d(\infty)
$$

and that $\overline{\lambda}_d(a_\theta)$ is monotonically decreasing in a_θ. We want to describe the speed of decrease of $\overline{\lambda}_d(a_\theta)$ in order to bound the decrease of $\lambda_d(a_\theta)$. This is achieved by using

$$
\overline{\lambda}_d(a_\theta) = -\int_{a_\theta}^{+\infty} \frac{\partial \overline{\lambda}_d}{\partial a_\theta} \, \mathrm{d}a_\theta + \overline{\lambda}_d(\infty) \tag{A43}
$$

where

$$
\frac{\partial \overline{\lambda}_d}{\partial a_\theta} = \phi_{\mathcal{N}}\Big(-\frac{\kappa}{\nu(\sigma_\theta - 1)}\Big) \frac{\kappa}{\nu(\sigma_\theta - 1)^2} \frac{\partial \sigma}{\partial z_\theta} \frac{\partial z_\theta}{\partial a_\theta} \tag{A44}
$$

The terms in A44 shall be analysed one by one. First note that the first and second terms are positive and bounded away from 0 and $+\infty$. For the fourth term, using equation A3,

$$\frac{\partial z_\theta}{\partial a_\theta} < \frac{\sigma_\theta(\sigma_\theta - 1)}{\sigma_\theta - 1} = \sigma_\theta < \overline{\sigma} \tag{A45}$$

For the third term $\frac{\partial \sigma}{\partial z_\theta}$, we first note by 11 that

$$\begin{aligned}\frac{\partial \sigma}{\partial z_\theta} &= -\frac{k(\overline{\sigma} - \underline{\sigma})}{(1 + e^{-kz_\theta})^2} e^{-kz_\theta} \\ &\geq -C_1 e^{-kz_\theta}\end{aligned}$$

Moreover, according to A5, $\frac{\partial z_\theta}{\partial a_\theta} \geq C_2$ where $C_2 = \frac{\underline{\sigma}}{1 + \frac{\overline{\sigma} - \underline{\sigma}}{\underline{\sigma} - 1} k} > 0$ is a constant. Hence

$$\begin{aligned}z_\theta &= \int_{\underline{a}}^{a_\theta} \frac{\partial z_\theta}{\partial a_\theta} \, \mathrm{d}a_\theta + z(\underline{a}) \\ &\geq C_2(a_\theta - \underline{a}) + z(\underline{a}) \\ &= C_2 a_\theta + C_3\end{aligned}$$

where $C_3 \in \mathbb{R}$ is a constant. Consequently,

$$\begin{aligned}\frac{\partial \sigma}{\partial z_\theta} &\geq -C_1 e^{-k(C_2 a_\theta + C_3)} \\ &= -C_4 e^{-kC_2 a_\theta}\end{aligned} \tag{A46}$$

where $C_4 > 0$ is a constant.

Combining A45, A46 and the fact that the first two terms of A44 are bounded away from $+\infty$, we get:

$$\frac{\partial \overline{\lambda}_d}{\partial a_\theta} \geq -C_5 e^{-kC_2 a_\theta}$$

for some positive constants C_2 and C_5. Plugging this into A43, we get:

$$\begin{aligned}\overline{\lambda}_d(a_\theta) &\leq \int_{a_\theta}^{+\infty} C_5 e^{-kC_2 a_\theta} \, \mathrm{d}a_\theta + \overline{\lambda}_d(\infty) \\ &= C_6 e^{-kC_2 a_\theta} + \overline{\lambda}_d(\infty)\end{aligned} \tag{A47}$$

where $C_6 > 0$ is a constant.

Combining A47 and A42,

$$\begin{aligned}\lambda_d(a_\theta) &\leq C_6 e^{-kC_2 a_\theta} + \overline{\lambda}_d(\infty) \\ &= C_6 e^{-kC_2 a_\theta} + \lambda_d(\infty)\end{aligned}$$

where the last equality arises from the fact that $\lambda_d(\infty) = \overline{\lambda}_d(\infty)$. This means that $\lambda_d(a_\theta) - \lambda_d(\infty)$ decreases at least exponentially, which concludes the proof.

$\square$

Proposition A17.
Any solution $V(a_\theta)$ to the HJB equation 41 must have an abscissa of convergence of $\underline{\sigma} - 1$.

Proof.

Recall the HJB:

$$(r-g)V = \max_{\lambda_i} \Big\{ \Pi(a_\theta) - f - R(\lambda_i; a_\theta) - \lambda_d(a_\theta)V + \Big[-\frac{1}{2}\nu^2 - g + \lambda_i q \Big] V'(a_\theta) + \frac{1}{2}\nu^2 V''(a_\theta) \Big\}$$

For the lower bound, set $\lambda_i = 0$:

$$(r-g)V \geq \Pi(a_\theta) - f - \lambda_d(a_\theta)V + \Big[-\frac{1}{2}\nu^2 - g \Big] V'(a_\theta) + \frac{1}{2}\nu^2 V''(a_\theta) \tag{A48}$$

Denote the abscissa of convergence of V as ξ_0^r. Suppose by contradiction that $\xi_0^r < \underline{\sigma} - 1$, then all the terms in the inequality A48 have abscissas of convergence strictly smaller than $\underline{\sigma} - 1$ except for $\Pi(a_\theta)$. The last increases asymptotically at the speed of $e^{(\underline{\sigma}-1)a_\theta}$ according to Proposition 2. It suffices to take the Laplace transform on the two sides of A48 and take the limit of $\xi \to (\underline{\sigma} - 1)+$ to see the contradiction. In fact, the Laplace transform of all the other terms converge to finite numbers except for $\hat{\Pi}(\xi)$ which tends to $+\infty$ as $\xi \to (\underline{\sigma} - 1)+$. This breaks the inequality and hence $\xi_0^r \geq \underline{\sigma} - 1$.

We now prove $\xi_0^r \leq \underline{\sigma} - 1$ so that it must be $\underline{\sigma} - 1$. Consider another dynamic control problem in which each firm can choose from a second type of growth strategy: by investing $R_s(\lambda_s; a_\theta) = \frac{1}{2}\alpha A_\theta^{\beta_s} \lambda_s^2$ in research where $\beta_s > \underline{\sigma} - 1$, the firm's productivity can grow with certainty λ_s per period. Following the notational convention in this section, I have removed the tildes. The HJB of this alternative dynamic control problem is:

$$\begin{aligned}(r-g)V_E = \max_{c,\,\lambda_c} \Big\{ &\Pi(a_\theta) - f - \lambda_d(a_\theta)V_E + \Big[-\frac{1}{2}\nu^2 - g \Big] V_E'(a_\theta) + \frac{1}{2}\nu^2 V_E''(a_\theta) \\ &+ \mathbb{1}_{c=i}\Big[- R(\lambda_c; a_\theta) + \lambda_c q V_E'(a_\theta) \Big] + \mathbb{1}_{c=s}\Big[- R_s(\lambda_c; a_\theta) + \lambda_c V_E'(a_\theta) \Big] \Big\}\end{aligned}$$

where $c \in \{i, s\}$ is the choice between innovation and the second type of growth strategy. First note that $V_E \geq V$, as each firm has more choices in the second dynamic control problem. Moreover,

$$\begin{aligned}(r-g)V_E &\geq \max_{\lambda_s} \Big\{ \Pi(a_\theta) - f - \lambda_d(a_\theta)V_E + \Big[-\frac{1}{2}\nu^2 - g \Big] V_E'(a_\theta) + \frac{1}{2}\nu^2 V_E''(a_\theta) - R_s(\lambda_s; a_\theta) + \lambda_s V_E'(a_\theta) \Big\} \\ &= \Pi(a_\theta) - f - \lambda_d(a_\theta)V_E + \Big[-\frac{1}{2}\nu^2 - g \Big] V_E'(a_\theta) + \frac{1}{2}\nu^2 V_E''(a_\theta) + \frac{(V_E')^2}{2\alpha e^{\beta_s a_\theta}} \end{aligned} \tag{A49}$$

where the inequality arises from the fact that the firm is only choosing λ_s on the right hand side of the inequality, while V_E on the left hand side is based on all possible choices. Denote ξ_E^r to be the abscissa of convergence of V_E, which is also the abscissa for V_E' and V_E''. According to Proposition A14,

$$\limsup_{a\to+\infty} \frac{\log(|V_E'(a)|)}{a} = \xi_E^r$$

Thus for the last term of equation A49,

$$\limsup_{a\to+\infty} \frac{1}{a} \log\left[\frac{(V_E')^2}{2\alpha e^{\beta_s a}}\right] = 2\xi_E^r - \beta_s$$

which implies, according to Proposition A13, that $2\xi_E^r - \beta_s$ is the abscissa of convergence of $\frac{(V_E')^2}{2\alpha e^{\beta_s a_\theta}}$. It is easy to see that $\xi_E^r \leq \beta_s$. Otherwise $2\xi_E^r - \beta_s > \xi_E^r > \beta_s > \underline{\sigma} - 1$: the LHS of equation A49 has an abscissa of ξ_E^r but the RHS has an abscissa of $2\xi_E^r - \beta_s$, which contradicts the inequality. Based on $V \leq V_E$, $\xi_0^r \leq \beta_s$. The argument is true for any $\beta_s > \underline{\sigma} - 1$, hence $\xi_0^r \leq \underline{\sigma} - 1$ and we can conclude. □

Proposition A18.
$V(a_\theta) > 0$ ($\forall a_\theta > \underline{a}$), and is monotonically increasing.

Proof.
To see $V(a_\theta) > 0$ ($\forall a_\theta > \underline{a}$), consider an alternative stochastic process in which λ_i is always set to 0. Value function V_{NC} of this alternative process must be positive for all $a_\theta > \underline{a}$, as per-period profit net of fixed cost is positive before death so that the discounted future cash flow is positive. Moreover, $V(a_\theta) \geq V_{NC}(a_\theta)$, as the choice set of the baseline model includes that of the alternative model. Thus $V > 0$ for all $a_\theta > \underline{a}$.

To prove the monotonicity of V, we use the idea of coupling. Given a specific travelling wave, i.e. ϕ^M with speed g, we want to show that the dynamic control problem starting from a_k has a higher value than the problem starting from a_j if $a_k > a_j$. The general idea is that we want to compare stochastic processes with the two initial starting points. Such a comparison is much easier if the randomness across the two processes are not independent but "coupled", for instance if the realizations of Brownian motion are the same. Note that coupling does *not* modify the value function, as the latter depends on the marginal distribution of each of the stochastic processes: introducing correlations between the two stochastic processes does not modify the marginal distribution.

In particular, consider a dynamic control problem starting from a_j. Consider another problem starting from a_k with modifications from the baseline assumptions. We assume that the death rate of the firm k is always the same as firm j, and that the realization of sales death happens at the same time for both firms. Firms j and k may meet each other at some time τ. We assume that the realization of randomness is the same after meeting so that the two processes are identical after time τ. Assume in addition that firm k always sets $\lambda_i = 0$ before meeting. There are three possibilities:

- The two firms meet each other for the first time at time τ.
- The two firms die at the same time due to the realization of λ_d before meeting.
- Firm j hits the left boundary and dies. Firm k keeps moving with the death rate of the baseline model and with $\lambda_i = 0$.

Before meeting, the stochastic process of firm k is simply a Brownian motion with negative drift so that it reaches any point $a_\theta \in [\underline{a}, a_k)$ within a finite time. Thus firm k must meet with firm j within a finite time conditional on the survival of both, as both processes are continuous. The last two possibilities correspond to two scenarios of death before meeting.

The value function of firm j's problem is $V(a_j)$. Denote firm k's value as $V_m(a_k)$ where m stands for "modified". It is easy to see $V_m(a_k) > V(a_j)$, as cash flows after meeting are the same across both firms, and are strictly higher in firm k than in firm j before meeting. The latter arises from the fact that firm k's operational profit net of fixed cost at any time is strictly higher than that of firm j before meeting, and that firm j spends on innovation so that its cash flow is further reduced by innovation costs.

Moreover, $V(a_k) \geq V_m(a_k)$, as choice is more restricted in the modified process and the death rate is higher. Thus $V(a_k) > V(a_j)$ and we have reached the conclusion.

□

Proposition A19.
$\lambda_i(a_\theta) > 0\ \forall a_\theta > \underline{a}$. Moreover, if $\beta > \underline{\sigma} - 1$, then λ_i decreases at least exponentially to 0.

Proof. Recall that

$$\lambda_i(a_\theta) = \frac{1}{\alpha e^{\beta a_\theta}} q V'(a_\theta)$$

The positiveness of λ_i thus follows directly from the monotonicity of V.

We have seen in Proposition A17 that the abscissa of convergence of V, $\xi_0^r = \underline{\sigma} - 1$. The same is true for the abscissa of V'. Denote $\epsilon = \frac{1}{2}(\beta - \xi_0^r) > 0$, then $V(a_\theta) = o(e^{(\xi_0^r + \epsilon)a_\theta})$. Rewrite λ_i as:

$$\lambda_i(a_\theta) = \frac{qV'(a_\theta)}{e^{(\xi_0^r + \epsilon)a_\theta}} \cdot e^{-(\beta - (\xi_0^r + \epsilon))a_\theta}$$

The first term on the right hand size is bounded, while the second term converges exponentially to 0 as $\beta > \xi_0^r + \epsilon$. Thus $\lambda_i(a_\theta)$ converges exponentially to 0.

□

Proposition A20.
If $\beta > \underline{\sigma} - 1$, then any solution ϕ of the KF equation 46 must be ultimately decreasing.

Proof.
Recall the KF equation:

$$0 = (\frac{1}{2}\nu^2 + g)\phi' + \frac{1}{2}\nu^2\phi'' + \frac{\lambda_e M_e}{M}\psi_e - \lambda_d(a;\phi^M)\phi - q\frac{\mathrm{d}[\lambda_i(a)\phi(a)]}{\mathrm{d}a} \tag{A50}$$

Integrating both sides on $[\underline{a}, a]$ and using the boundary condition, we get:

$$0 = (\frac{1}{2}\nu^2 + g)\phi(a) + \frac{1}{2}\nu^2(\phi'(a) - \phi'(\underline{a})) - q\lambda_i(a)\phi(a) - \int_{\underline{a}}^{a}\lambda_d(a)\phi(a)\,\mathrm{d}a + \frac{\lambda_e M_e}{M}\int_{\underline{a}}^{a}\psi_e(a)\,\mathrm{d}a$$

Rearranging the terms and using the balanced entry/exit condition:

$$\begin{aligned}
&\frac{1}{2}\nu^2\phi'(a)\\
=&-(\frac{1}{2}\nu^2 + g)\phi(a) + \Big[\frac{\lambda_e M_e}{M} - \int_{\underline{a}}^{+\infty}\lambda_d(a)\phi(a)\,\mathrm{d}a\Big] + q\lambda_i(a)\phi(a) + \int_{\underline{a}}^{a}\lambda_d(a)\phi(a)\,\mathrm{d}a - \frac{\lambda_e M_e}{M}\int_{\underline{a}}^{a}\psi_e(a)\,\mathrm{d}a\\
=&-(\frac{1}{2}\nu^2 + g)\phi(a) + q\lambda_i(a)\phi(a) + \frac{\lambda_e M_e}{M}\int_{a}^{+\infty}\psi_e(a)\,\mathrm{d}a - \int_{a}^{+\infty}\lambda_d(a)\phi(a)\,\mathrm{d}a\\
\leq&\Big[-\frac{1}{2}\nu^2 - g + q\lambda_i(a)\Big]\phi(a) + \frac{\lambda_e M_e}{M}\int_{a}^{+\infty}\psi_e(a)\,\mathrm{d}a - \int_{a}^{+\infty}\lambda_d(\infty)\phi(a)\,\mathrm{d}a\\
=&\Big[-\frac{1}{2}\nu^2 - g + q\lambda_i(a)\Big]\phi(a) + \Big[\frac{\lambda_e M_e}{M}e^{-k_\psi(a-\underline{a})} - \lambda_d(\infty)\Big]\overline{\Phi}(a)
\end{aligned} \tag{A51}$$

We have seen in Proposition A19 that λ_i declines at least exponentially to 0. Thus $-\frac{1}{2}\nu^2 - g + q\lambda_i(a) < 0$ for a sufficiently large a. The same is true for $\frac{\lambda_e M_e}{M}e^{-k_\psi(a-\underline{a})} - \lambda_d(\infty)$ as $\lambda_d(\infty) > 0$. Thus $\phi'(a) < 0$ for a sufficiently large a, meaning that ϕ is ultimately decreasing.

□

We are now ready to prove the tail uniqueness in **Proposition 4**.

Proof.
Apply the Laplace transform to the KF equation A50:

$$0 = (\frac{1}{2}\nu^2 + g)\xi\widehat{\phi}(\xi) + \frac{1}{2}\nu^2\Big[\xi^2\widehat{\phi}(\xi) - e^{-\xi\underline{a}}\phi'(\underline{a})\Big] + \frac{\lambda_e M_e}{M}\widehat{\psi_e} - \widehat{\lambda_d\phi} - q\xi\widehat{\lambda_i\phi} \tag{A52}$$

where we have used the facts that $\widehat{\phi'}(\xi) = \xi\widehat{\phi}(\xi)$, $\widehat{\phi''}(\xi) = \xi^2\widehat{\phi}(\xi) - e^{-\xi\underline{a}}\phi'(\underline{a})$, and $\widehat{\mathrm{d}[\lambda_i(a)\phi(a)]/\mathrm{d}a}(\xi) = \xi\widehat{\lambda_i\phi}$.

Decompose $\widehat{\lambda_d\phi}(\xi)$ into

$$\widehat{\lambda_d\phi}(\xi) = \widehat{(\lambda_d - \lambda_d(\infty))\phi}(\xi) + \lambda_d(\infty)\widehat{\phi}(\xi)$$

Thus equation A52 can be rearranged as:

$$\widehat{\phi} = \frac{-\frac{1}{2}\nu^2 e^{-\xi \underline{a}}\phi'(\underline{a}) - \widehat{(\lambda_d - \lambda_d(\infty))\phi} - q\xi\widehat{\lambda_i\phi} + \frac{\lambda_e M_e}{M}\widehat{\psi_e}}{-(\frac{1}{2}\nu^2 + g)\xi - \frac{1}{2}\nu^2\xi^2 + \lambda_d(\infty)} \tag{A53}$$

We want to find the pole of $\hat{\phi}$ in order to apply Proposition A15. This takes a few steps. Denote $\xi_0^r \in [-\infty, 0]$ to be the abscissa of convergence for the Laplace transform of ϕ.

Step 1. *$\exists$ some constant $s > 0$, such that the abscissas of convergence of the hat terms in the numerator of equation A53 are at most $\xi_0^r - s$.*

This is obvious for $\widehat{(\lambda_d - \lambda_d(\infty))\phi}$ and $\widehat{\lambda_i\phi}$, as $\lambda_d(a_\theta) - \lambda_d(\infty)$ and $\lambda_i(a_\theta)$ decrease at least exponentially to 0 as $a_\theta \to +\infty$ (Proposition A16 and A19).
We have shown in Proposition A20 that ϕ must be ultimately monotone. Together with the assumption of $k_\psi > 0$, Proposition 3 says that there is an exponential wedge between ψ_e and ϕ. Hence the result is also true for $\widehat{\psi_e}$.

Step 2. $\xi_0^r > -\infty$.

Suppose $\xi_0^r = -\infty$ by contradiction. We want to show that this implies $\hat{\phi}(\xi) < 0$ as $\xi \to -\infty$ and hence a contradiction.

We analyse term by term for equation A52. First, $-q\xi\widehat{\lambda_i\phi}(\xi) \to +\infty$ as $\xi \to -\infty$. This is because $-q\xi \to +\infty$ as $\xi \to -\infty$ and

$$\begin{aligned}
\widehat{\lambda_i\phi}(\xi) &= \int_{\underline{a}}^{+\infty} e^{-\xi a}\lambda_i(a)\phi(a)\,\mathrm{d}a \\
&\geq \int_0^{+\infty} e^{-\xi a}\lambda_i(a)\phi(a)\,\mathrm{d}a \\
&\geq \int_0^{+\infty} \lambda_i(a)\phi(a)\,\mathrm{d}a > 0, \qquad \forall \xi < 0.
\end{aligned}$$

Second, $\lambda_d(a_\theta)$ is obviously bounded, hence

$$\begin{aligned}
\widehat{\lambda_d\phi}(\xi) &= \int_{\underline{a}}^{+\infty} e^{-\xi a_\theta}\lambda_d(a_\theta)\phi(a_\theta)\,\mathrm{d}a_\theta \\
&\leq C_1\hat{\phi}(\xi)
\end{aligned}$$

for some positive constant C_1.

Third, using the balanced entry/exit condition 48,

$$
\begin{aligned}
-\frac{1}{2}\nu^2 e^{-\xi\underline{a}}\phi'(\underline{a}) + \frac{\lambda_e M_e}{M}\widehat{\psi_e}(\xi) &= -e^{-\xi\underline{a}}\Big[\frac{\lambda_e M_e}{M} - \int_{\underline{a}}^{+\infty} \lambda_d(a_\theta)\phi(a_\theta)\,\mathrm{d}a_\theta\Big] + \frac{\lambda_e M_e}{M}\widehat{\psi_e}(\xi) \\
&= \frac{\lambda_e M_e}{M}\Big[\widehat{\psi_e}(\xi) - e^{-\xi\underline{a}}\Big] + e^{-\xi\underline{a}}\int_{\underline{a}}^{+\infty} \lambda_d(a_\theta)\phi(a_\theta)\,\mathrm{d}a_\theta
\end{aligned}
$$

where

$$
\widehat{\psi_e}(\xi) = \int_{\underline{a}}^{+\infty} e^{-\xi a}\psi_e(a)\,\mathrm{d}a \geq \int_{\underline{a}}^{+\infty} e^{-\xi\underline{a}}\psi_e(a)\,\mathrm{d}a = e^{-\xi\underline{a}}, \quad \forall \xi \leq 0
$$

Combining these elements with equation A52, we get:

$$
\begin{aligned}
&\frac{\lambda_e M_e}{M}\Big[\widehat{\psi_e}(\xi) - e^{-\xi\underline{a}}\Big] + e^{-\xi\underline{a}}\int_{\underline{a}}^{+\infty} \lambda_d(a_\theta)\phi(a_\theta)\,\mathrm{d}a_\theta - q\xi\widehat{\lambda_i\phi} \\
=& -(\frac{1}{2}\nu^2 + g)\xi\hat{\phi}(\xi) - \frac{1}{2}\nu^2\xi^2\hat{\phi}(\xi) + \widehat{\lambda_d\phi}(\xi) \\
\leq& \Big[-(\frac{1}{2}\nu^2 + g)\xi - \frac{1}{2}\nu^2\xi^2 + C_1\Big]\hat{\phi}(\xi)
\end{aligned} \tag{A54}
$$

By our analysis above, the first line of the above equation tends to $+\infty$ as $\xi \to -\infty$. Moreover, $-(\frac{1}{2}\nu^2 + g)\xi - \frac{1}{2}\nu^2\xi^2 + C_1 \to -\infty$ as $\xi \to -\infty$. This implies that $\hat{\phi}(\xi) < 0$ for sufficiently negative ξ, which contradicts the fact that $\hat{\phi}(\xi)$ must be positive.

Step 3. We want to show that $\xi_0^r \leq \xi_-$, where $\xi_- \in \mathbb{R}$ is the only negative root of

$$
f(\xi) \triangleq -(\frac{1}{2}\nu^2 + g)\xi - \frac{1}{2}\nu^2\xi^2 + \lambda_d(\infty) = 0
$$

The above equation obviously has one negative root $\xi_- = -(\frac{1}{2} + \frac{g}{\nu^2}) - \sqrt{(\frac{1}{2} + \frac{g}{\nu^2})^2 + \frac{2\lambda_d(\infty)}{\nu^2}}$ and one positive root $\xi_+ = -(\frac{1}{2} + \frac{g}{\nu^2}) + \sqrt{(\frac{1}{2} + \frac{g}{\nu^2})^2 + \frac{2\lambda_d(\infty)}{\nu^2}}$.

ξ_+ is a removable singularity of A53. This is because $\xi_0^r \leq 0$ and hence $\widehat{\phi}$ must be well defined in the neighborhood of ξ_+. This implies that the numerator of A53 must take the value of 0 at ξ_+, i.e. it has the Taylor expansion of $\sum_{n=1}^{\infty} a_n(\xi - \xi_+)^n$ around the neighborhood of ξ_+. Segregating the denominator of A53 as $-(\xi - \xi_-)(\xi - \xi_+)$, equation A53 can be written as $\widehat{\phi}(\xi) = -\frac{1}{\xi - \xi_-}\sum_{n=1}^{\infty} a_n(\xi - \xi_+)^{n-1}$ around the neighborhood of ξ_+, meaning that it is analytic at ξ_+.

Given that $\xi_0^r \leq 0$, according to Step 1 the terms in the numerator converge for any ξ whose real part ξ^r is larger than $-s$. As long as $-s > \xi_-$, $\hat{\phi}$ is analytic for the same ξ with $\xi^r > -s$. As $\hat{\phi}$ is analytic for any ξ with $\xi^r > -s$, its abscissa of convergence ξ_0^r is at most $-s$. Otherwise it contradicts Proposition A12. By the same token, the numerator of A53 converges and hence is analytic for any ξ with $\xi^r > -2s$, and ξ_0^r is thus at most $-2s$. The process repeats until we touch ξ_- so that $\xi_0^r \leq \xi_-$. Figure A3 illustrates the successive coverage.

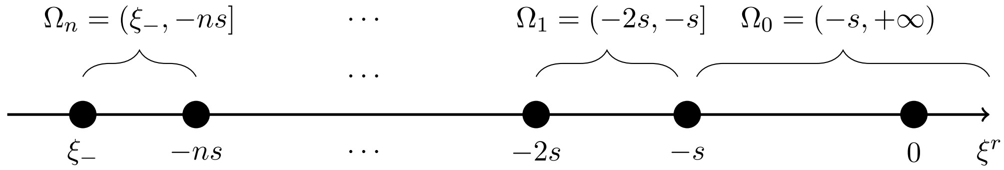


Figure A3: Successive coverage in Step 3 of the proof

Step 4. $\lim_{x\to+\infty} \frac{\log(\bar{\Phi}(x))}{x} = \xi_-$, where $\bar{\Phi}(x) = \int_x^{+\infty} \phi(x)\mathrm{d}x$.

$\xi = \xi_-$ cannot be a removable singularity of $\phi(\xi)$, otherwise we could perpetually continue the process in Step 3 and violate the result of Step 2. It is easy to see that $\xi = \xi_-$ is a pole, as the numerator of $\hat{\phi}(\xi)$ is analytic in its neighborhood and the denominator is analytic and 0. We can now apply Proposition A15 to conclude that $\lim_{x\to+\infty} \frac{\log(\bar{\Phi}(x))}{x} = \xi_-$.

□

G.2 Proof for the KF equation in the extended model

The proof follows a similar procedure as that of the baseline model. Proposition A16 remains valid as it depends on the static setting. Proposition A17 can be easily modified for the extended model and the result remains the same. The strategy for proving Proposition A18, which relies on the stochastic process being continuous, is no longer valid in the extended model in which incumbent learning allows large jumps. As the monotonicity of V seems natural, and numerical solutions consistently find it to be the case, I assume it for convenience. For Proposition A19, readers can easily prove that $\lambda_i \geq 0$, $\lambda_l \geq 0$ and that both decline at least exponentially to 0. The equalities accommodate cases in which i or l is not chosen. I shall prove a counterpart of A20 in the extended model and then the tail uniqueness.

Like in the previous section, for notational conciseness, I will drop the tilde notation in this proof even though the HJB and KF equations are the normalized version. I will also drop notations of the dependence on ϕ whenever it does not cause confusion.

Recall the KF equation of the extended model:

$$0 = (\frac{1}{2}\nu^2 + g)\phi'(a_\theta) + \frac{1}{2}\nu^2\phi''(a_\theta) - \lambda_d(a_\theta)\phi(a_\theta) + \lambda_i(a_\theta - q)\phi(a_\theta - q) - \lambda_i(a_\theta)\phi(a_\theta)$$
$$\int_{\underline{a}}^{a_\theta} \lambda_l(a_\psi)\phi(a_\psi)\psi_l(a_\theta; a_\psi)\,\mathrm{d}a_\psi - \phi(a_\theta)\lambda_l(a_\theta)\int_{a_\theta}^{+\infty} \psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi + \frac{\lambda_e M_e}{M}\psi_e(a_\theta) \quad \text{(A55)}$$

where λ_i and λ_l take the value of 0 if not chosen.

Proposition A21.
If $\beta > \underline{\sigma} - 1$, then any solution ϕ to the KF equation 61 must be ultimately decreasing.

Proof.
Like in the baseline model, the proof shall be based on an approximated version of equation 61:

$$0 = (\frac{1}{2}\nu^2 + g)\phi'(a_\theta) + \frac{1}{2}\nu^2\phi''(a_\theta) - \lambda_d(a_\theta)\phi(a_\theta) - q\frac{\mathrm{d}[\lambda_i(a_\theta)\phi(a_\theta)]}{\mathrm{d}a_\theta}$$
$$\int_{\underline{a}}^{a_\theta} \lambda_l(a_\psi)\phi(a_\psi)\psi_l(a_\theta; a_\psi)\,\mathrm{d}a_\psi - \phi(a_\theta)\lambda_l(a_\theta)\int_{a_\theta}^{+\infty}\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi + \frac{\lambda_e M_e}{M}\psi_e(a_\theta) \quad \text{(A56)}$$

Integrating the equation on $[\underline{a}, a]$, we get:

$$0 = (\frac{1}{2}\nu^2 + g)\phi(a) + \frac{1}{2}\nu^2(\phi'(a) - \phi'(\underline{a})) - q\lambda_i(a)\phi(a) - \int_{\underline{a}}^{a}\lambda_d(a)\phi(a)\,\mathrm{d}a + \frac{\lambda_e M_e}{M}\int_{\underline{a}}^{a}\psi_e(a)\,\mathrm{d}a$$
$$+ \int_{\underline{a}}^{a}\int_{\underline{a}}^{a_\theta}\lambda_l(a_\psi)\phi(a_\psi)\psi_l(a_\theta; a_\psi)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta - \int_{\underline{a}}^{a}\phi(a_\theta)\lambda_l(a_\theta)\int_{a_\theta}^{+\infty}\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta \quad \text{(A57)}$$

where the last two terms are:

$$\begin{aligned}
&\int_{\underline{a}}^{a}\int_{\underline{a}}^{a_\theta}\lambda_l(a_\psi)\phi(a_\psi)\psi_l(a_\theta; a_\psi)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta - \int_{\underline{a}}^{a}\phi(a_\theta)\lambda_l(a_\theta)\int_{a_\theta}^{+\infty}\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta \\
=&\int_{\underline{a}}^{a}\int_{\underline{a}}^{a_\psi}\lambda_l(a_\theta)\phi(a_\theta)\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\theta\,\mathrm{d}a_\psi - \int_{\underline{a}}^{a}\int_{a_\theta}^{+\infty}\phi(a_\theta)\lambda_l(a_\theta)\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta \\
=&-\int_{\underline{a}}^{a}\int_{a}^{+\infty}\phi(a_\theta)\lambda_l(a_\theta)\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta \\
=&-\int_{\underline{a}}^{a}\phi(a_\theta)\lambda_l(a_\theta)e^{-k_\psi(a-a_\theta)}\overline{\Phi}(a)\,\mathrm{d}a_\theta \\
=&-e^{-k_\psi a}\overline{\Phi}(a)\int_{\underline{a}}^{a}\phi(a_\theta)\lambda_l(a_\theta)e^{k_\psi a_\theta}\,\mathrm{d}a_\theta
\end{aligned} \quad \text{(A58)}$$

Inserting A58 into A57, and using the balanced entry/exit condition to replace $\frac{1}{2}\nu^2\phi'(\underline{a})$, we get:

$$\begin{aligned}
&\frac{1}{2}\nu^2\phi'(a) \\
=&-(\frac{1}{2}\nu^2 + g)\phi(a) + \Big[\frac{\lambda_e M_e}{M} - \int_{\underline{a}}^{+\infty}\lambda_d(a)\phi(a)\,\mathrm{d}a\Big] + q\lambda_i(a)\phi(a) + \int_{\underline{a}}^{a}\lambda_d(a)\phi(a)\,\mathrm{d}a - \frac{\lambda_e M_e}{M}\int_{\underline{a}}^{a}\psi_e(a)\,\mathrm{d}a \\
&+ e^{-k_\psi a}\overline{\Phi}(a)\int_{\underline{a}}^{a}\phi(a_\theta)\lambda_l(a_\theta)e^{k_\psi a_\theta}\,\mathrm{d}a_\theta \\
\leq&\Big[-\frac{1}{2}\nu^2 - g + q\lambda_i(a)\Big]\phi(a) + \Big[\frac{\lambda_e M_e}{M}e^{-k_\psi(a-\underline{a})} - \lambda_d(\infty)\Big]\overline{\Phi}(a) + e^{-k_\psi a}\overline{\Phi}(a)\int_{\underline{a}}^{a}\phi(a_\theta)\lambda_l(a_\theta)e^{k_\psi a_\theta}\,\mathrm{d}a_\theta
\end{aligned}$$
$$\text{(A59)}$$

where we have followed similar procedures as in equation A51.

As λ_l decreases at least exponentially to 0 if $\beta > \underline{\sigma} - 1$, there exists some $s \in (0, k_\psi)$ and $C_1 > 0$ such that $\lambda_l(a_\theta) \leq C_1 e^{-sa_\theta}$, $\forall a_\theta \geq \underline{a}$. Moreover, there exists some $C_2 > 0$ such that $\phi(a_\theta) \leq C_2$, $\forall a_\theta \geq \underline{a}$. Thus

$$\int_{\underline{a}}^{a} \phi(a_\theta)\lambda_l(a_\theta)e^{k_\psi a_\theta}\,\mathrm{d}a_\theta \leq \int_{\underline{a}}^{a} C_1 C_2 e^{(k_\psi - s)a_\theta}\,\mathrm{d}a_\theta = \frac{C_1 C_2}{k_\psi - s}\Big[e^{(k_\psi - s)a} - e^{(k_\psi - s)\underline{a}}\Big] \leq Ce^{(k_\psi - s)a} \tag{A60}$$

where $C > 0$ is a constant. Plugging A60 into A59:

$$\frac{1}{2}\nu^2\phi'(a) \leq \Big[-\frac{1}{2}\nu^2 - g + q\lambda_i(a)\Big]\phi(a) + \Big[\frac{\lambda_e M_e}{M}e^{-k_\psi(a-\underline{a})} + Ce^{-sa} - \lambda_d(\infty)\Big]\overline{\Phi}(a)$$

We have seen that λ_i declines at least exponentially to 0 if $\beta > \underline{\sigma} - 1$. Thus $-\frac{1}{2}\nu^2 - g + q\lambda_i(a) < 0$ for a sufficiently large a. The same is true for $\frac{\lambda_e M_e}{M}e^{-k_\psi(a-\underline{a})} + Ce^{-sa} - \lambda_d(\infty)$ as $\lambda_d(\infty) > 0$. Thus $\phi'(a) < 0$ for a sufficiently large a, meaning that ϕ is ultimately decreasing.

□

We are now ready to prove **a counterpart of Proposition 4 under the extended model**.

Proof.
The proof follows a similar procedure as that of the baseline model so I shall only highlight the differences. The only difference between the KF equation A56 in the extended model and that in the baseline model is the two terms of incumbent learning. We have already denoted one of them $\psi_l^{\text{in}}(a_\psi; \phi)$ in equation A11. Denote the other one as:

$$\psi_l^{\text{out}}(a_\theta) = \phi(a_\theta)\lambda_l(a_\theta)\int_{a_\theta}^{+\infty} \psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi \tag{A61}$$

The counterpart of A52 under the extended model is thus:

$$0 = (\frac{1}{2}\nu^2 + g)\xi\widehat{\phi}(\xi) + \frac{1}{2}\nu^2\Big[\xi^2\widehat{\phi}(\xi) - e^{-\xi\underline{a}}\phi'(\underline{a})\Big] + \frac{\lambda_e M_e}{M}\widehat{\psi_e} - \widehat{\lambda_d\phi} - q\xi\widehat{\lambda_i\phi} + \widehat{\psi_l^{\text{in}}}(\xi) - \widehat{\psi_l^{\text{out}}}(\xi) \tag{A62}$$

with an additional $\widehat{\psi_l^{\text{in}}}(\xi) - \widehat{\psi_l^{\text{out}}}(\xi)$ term compared to A52. The counterpart of A53 under the extended model is

$$\widehat{\phi} = \frac{-\frac{1}{2}\nu^2 e^{-\xi\underline{a}}\phi'(\underline{a}) - \widehat{(\lambda_d - \lambda_d(\infty))\phi} - q\xi\widehat{\lambda_i\phi} + \frac{\lambda_e M_e}{M}\widehat{\psi_e} + \widehat{\psi_l^{\text{in}}}(\xi) - \widehat{\psi_l^{\text{out}}}(\xi)}{-(\frac{1}{2}\nu^2 + g)\xi - \frac{1}{2}\nu^2\xi^2 + \lambda_d(\infty)} \tag{A63}$$

Step 1.
As $k_\psi > 0$, ϕ is ultimately monotone and $\lambda_l(a_\theta)$ converges to 0 at least exponentially, Proposition A8 shows that $\exists k_s > 0$ s.t. $\psi_l^{\text{in}}(a_\theta)$ converges to 0 more quickly than $\phi(a_\theta)e^{-k_s a_\theta}$. For ψ_l^{out},

$$\psi_l^{\text{out}}(a_\theta) \leq \phi(a_\theta)\lambda_l(a_\theta)$$

where λ_l creates an at least exponential wedge between ψ_l^{out} and ϕ.

Step 2.
We need to additionally analyze $\widehat{\psi_l^{\text{in}}}(\xi) - \widehat{\psi_l^{\text{out}}}(\xi)$:

$$
\begin{aligned}
&\widehat{\psi_l^{\text{in}}}(\xi) - \widehat{\psi_l^{\text{out}}}(\xi) \\
&= \int_{\underline{a}}^{+\infty} e^{-\xi a_\theta} \Big[\int_{\underline{a}}^{a_\theta} \lambda_l(a_\psi)\phi(a_\psi)\psi_l(a_\theta; a_\psi)\,\mathrm{d}a_\psi - \phi(a_\theta)\lambda_l(a_\theta) \int_{a_\theta}^{+\infty} \psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi \Big]\,\mathrm{d}a_\theta \\
&= \int_{\underline{a}}^{+\infty} e^{-\xi a_\theta} \Big[\int_{\underline{a}}^{+\infty} \lambda_l(a_\psi)\phi(a_\psi)\psi_l(a_\theta; a_\psi)\,\mathrm{d}a_\psi - \phi(a_\theta)\lambda_l(a_\theta) \int_{\underline{a}}^{+\infty} \psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi \Big]\,\mathrm{d}a_\theta \\
&= \int_{\underline{a}}^{+\infty} \int_{\underline{a}}^{+\infty} e^{-\xi a_\theta} \lambda_l(a_\psi)\phi(a_\psi)\psi_l(a_\theta; a_\psi)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta - \int_{\underline{a}}^{+\infty} \int_{\underline{a}}^{+\infty} e^{-\xi a_\theta} \phi(a_\theta)\lambda_l(a_\theta)\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta \\
&= \int_{\underline{a}}^{+\infty} \int_{\underline{a}}^{+\infty} e^{-\xi a_\psi} \lambda_l(a_\theta)\phi(a_\theta)\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta - \int_{\underline{a}}^{+\infty} \int_{\underline{a}}^{+\infty} e^{-\xi a_\theta} \phi(a_\theta)\lambda_l(a_\theta)\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta \\
&= \int_{\underline{a}}^{+\infty} \int_{\underline{a}}^{+\infty} (e^{-\xi a_\psi} - e^{-\xi a_\theta})\lambda_l(a_\theta)\phi(a_\theta)\psi_l(a_\psi; a_\theta)\,\mathrm{d}a_\psi\,\mathrm{d}a_\theta
\end{aligned}
$$

Note that when $\xi < 0$,

$$
(e^{-\xi a_\psi} - e^{-\xi a_\theta})\psi_l(a_\psi; a_\theta) \begin{cases} \geq 0, & \text{if } a_\psi \geq a_\theta \\ = 0, & \text{if } a_\psi < a_\theta \end{cases}
$$

Hence $\widehat{\psi_l^{\text{in}}}(\xi) - \widehat{\psi_l^{\text{out}}}(\xi) > 0$ when $\xi \to -\infty$.

For the extended model, there is an additional term $\widehat{\psi_l^{\text{in}}}(\xi) - \widehat{\psi_l^{\text{out}}}(\xi)$ on the left hand side of A54. It is easy to see that the logic still goes through under the extended model.

Step 3. and **Step 4.**
Same as in the baseline model.

□

H Computation

H.1 Numerical calculation of γ

γ generally does not have an explicit form and needs to be recovered from γ' by:

$$
\gamma(Z_\theta) = \int_0^{Z_\theta} \gamma'(Z)\,\mathrm{d}Z
$$

using the fact that $\gamma(0) = 0$.

Note, however, that $\gamma'(Z_\theta) \sim \mathcal{O}(Z_\theta^{-\frac{1}{\sigma}})$ as $Z_\theta \to 0+$ so that the integral above is improper around 0.

The improper integral is theoretically well-defined as $\frac{1}{\overline{\sigma}} < 1$ but can lead to inaccurate numerical results. To resolve the issue, do a change of variable of $Z = \xi^s$ with $s = \frac{\overline{\sigma}}{\overline{\sigma}-1}$ for the integral:

$$\begin{aligned}\gamma(Z_\theta) &= \int_0^{Z_\theta^{1/s}} \gamma'(\xi^s)\,\mathrm{d}(\xi^s) \\ &= \int_0^{Z_\theta^{1/s}} sC\Big[1 + \frac{1}{\overline{\sigma}}\xi^{ks}\Big]^{-\frac{\overline{\sigma}-\sigma}{k\overline{\sigma}\sigma}}\,\mathrm{d}\xi \end{aligned} \tag{A64}$$

The integrand tends to sC as $\xi \to 0+$ so that the integral is proper. Equation A64 is used for numerically calculating γ.

H.2 Overall algorithm for travelling wave

This section summarizes the computational algorithm at a high level. Incumbent learning is only present under the extended model: I will use square brackets $\Big[\;\Big]$ to single out these parts in this and subsequent sections.

Algorithm 1 (Travelling Wave)**.**

`Initialize` $\underline{a}$, g, M `and` ϕ. `Initialize tolerance and some` $\epsilon >$ tolerance. `Choose a dampening parameter` $\in (0,1)$ `for the updates.`
`While` $\epsilon >$ tolerance

1. `Discretize the grid between` $\underline{a}$ `and` $\overline{a}$.
2. `Solve the static problem per period based on` M `and` ϕ; `get sales, profit and death rate for each firm.`
3. `Use` ϕ `to parameterize` $\Big[\psi_l$ `and`$\Big]$ ψ_e.
4. `Solve the HJB to get the innovation`$\Big[$`/learning`$\Big]$ `decision of incumbents and the value function` V.
5. `Solve the entry decision based on` ψ_e `and` V.
6. (a) `Solve the KFE to get an updated distribution` $\tilde{\phi}$.
   (b) `If` $\tilde{\phi}$ `implies` $A = 1$, `continue to the next step. Otherwise, adjust` g `and return to step 6a. The iteration gives an updated growth rate` g.
7. `Update` $\underline{a}$ `based on` $\tilde{\phi}$, M `and the break-even condition 15.`
8. `Update` M `by the balanced entry/exit condition 48.`

```
9. ϵ = max|φ̃ − φ|.

end
```

H.3 Algorithm for solving the HJB

For the baseline model, the HJB and KF equations are solved using the finite difference method with an implicit scheme à la Barles and Souganidis (1991).[88] In the extended model with large jumps due to incumbent learning, I use instead the explicit-implicit scheme à la Cont and Voltchkova (2005) which can be seen as an extension of Barles and Souganidis (1991).[89] When there are large jumps, an implicit scheme needs to invert a high-dimensional non-sparse matrix at each step of the update and is thus computationally costly. The explicit-implicit scheme deals with the jumps in a computationally efficient way, while ensuring unconditional stability.

More specifically, denote $n = 0, \ldots, N$ to be the index of grid points such that $\tilde{a}_0 = \underline{\tilde{a}}$ and $\tilde{a}_N = \bar{\tilde{a}}$. The grid is equally spaced with $\Delta\tilde{a}$ between two adjacent points. I have slightly abused the notation by putting n on the subscript: they should not be understood as θ in the previous sections. Similarly, denote $\tilde{V}_n$ to be the value function at these points, i.e. $\tilde{V}_n = \tilde{V}(\tilde{a}_n)$. Use the finite difference for approximating the first and second derivatives:

$$\tilde{V}'(\tilde{a}_n) \approx \frac{\tilde{V}_n - \tilde{V}_{n-1}}{-\Delta\tilde{a}} \tag{A65}$$

$$\tilde{V}''(\tilde{a}_n) \approx \frac{\tilde{V}_{n+1} - 2\tilde{V}_n + \tilde{V}_{n-1}}{(\Delta\tilde{a})^2} \tag{A66}$$

I have used the backward difference for A65 as the drift term in the HJB, $-\frac{1}{2}\nu^2 - g$, is always negative. This "upwind" approach ensures the monotonicity of the numerical scheme and is a precondition for its convergence. The boundary conditions are incorporated as:

$$\tilde{V}(\tilde{a}_0) = 0$$
$$\tilde{V}(\tilde{a}_N) = \tilde{V}(\tilde{a}_{N+1})$$

Thus the term $[-\frac{1}{2}\nu^2 - g\Big]\tilde{V}'(\tilde{a}) + \frac{1}{2}\nu^2\tilde{V}''(\tilde{a})$ can be represented as $(F_1 + F_2)\tilde{V}$ in matrix forms,

[88] See Achdou et al. (2022) for a discussion on Barles and Souganidis (1991) and its application in the economics literature.

[89] See also Cont and Tankov (2003).

where $\tilde{V} = [\tilde{V}_0, \cdots, \tilde{V}_N]^\intercal$ and

$$F_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{\nu^2/2+g}{\Delta\tilde{a}} & -\frac{\nu^2/2+g}{\Delta\tilde{a}} & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \frac{\nu^2/2+g}{\Delta\tilde{a}} & -\frac{\nu^2/2+g}{\Delta\tilde{a}} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\frac{\nu^2/2+g}{\Delta\tilde{a}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \frac{\nu^2/2+g}{\Delta\tilde{a}} & -\frac{\nu^2/2+g}{\Delta\tilde{a}} & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & \frac{\nu^2/2+g}{\Delta\tilde{a}} & -\frac{\nu^2/2+g}{\Delta\tilde{a}} \end{pmatrix}$$

$$F_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\frac{\nu^2}{(\Delta\tilde{a})^2} & \frac{\nu^2}{2(\Delta\tilde{a})^2} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{\nu^2}{2(\Delta\tilde{a})^2} & -\frac{\nu^2}{(\Delta\tilde{a})^2} & \frac{\nu^2}{2(\Delta\tilde{a})^2} & \dots & 0 & 0 & 0 \\ 0 & 0 & \frac{\nu^2}{2(\Delta\tilde{a})^2} & -\frac{\nu^2}{(\Delta\tilde{a})^2} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\frac{\nu^2}{(\Delta\tilde{a})^2} & \frac{\nu^2}{2(\Delta\tilde{a})^2} & 0 \\ 0 & 0 & 0 & 0 & \dots & \frac{\nu^2}{2(\Delta\tilde{a})^2} & -\frac{\nu^2}{(\Delta\tilde{a})^2} & \frac{\nu^2}{2(\Delta\tilde{a})^2} \\ 0 & 0 & 0 & 0 & \dots & 0 & \frac{\nu^2}{2(\Delta\tilde{a})^2} & -\frac{\nu^2}{2(\Delta\tilde{a})^2} \end{pmatrix}$$

Note that firms at position $n = 1$ die from the left boundary at the rate of $\frac{\nu^2}{2(\Delta\tilde{a})^2}$ due to the Brownian motion, but not due to the drift term. This comes from the theoretical result of Section F.2: the Brownian motion dominates the drift as $\Delta t \to 0+$.

The innovation term $\lambda_i(\tilde{a})[\tilde{V}(\tilde{a}+q) - \tilde{V}(\tilde{a})]$, where $\lambda_i = 0$ if not chosen in the extended model, is approximated by $\lambda_i(\tilde{a})q\tilde{V}'(\tilde{a})$ and discretized by $F_3^{i,D}\tilde{V}$, where

$$F_3^{i,D} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\frac{\lambda_{i,1}q}{\Delta\tilde{a}} & \frac{\lambda_{i,1}q}{\Delta\tilde{a}} & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -\frac{\lambda_{i,2}q}{\Delta\tilde{a}} & \frac{\lambda_{i,2}q}{\Delta\tilde{a}} & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{\lambda_{i,3}q}{\Delta\tilde{a}} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\frac{\lambda_{i,N-2}q}{\Delta\tilde{a}} & \frac{\lambda_{i,N-2}q}{\Delta\tilde{a}} & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -\frac{\lambda_{i,N-1}q}{\Delta\tilde{a}} & \frac{\lambda_{i,N-1}q}{\Delta\tilde{a}} \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix} \tag{A67}$$

and $\lambda_{i,n} = \lambda_i(\tilde{a}_n)$.

$\Big[$Incumbent learning corresponds to the integral term $\lambda_l(\tilde{a}) \int_{(\tilde{a},+\infty)} \big[\tilde{V}(\tilde{a}_\psi) - \tilde{V}(\tilde{a})\big] \psi_l(\tilde{a}_\psi; \tilde{a}, \tilde{\phi})\, \mathrm{d}\tilde{a}_\psi$

in the HJB, where $\lambda_l = 0$ if not chosen. Note that

$$\begin{aligned}
&\lambda_l(\tilde{a}) \int_{(\tilde{a},+\infty)} \left[\tilde{V}(\tilde{a}_\psi) - \tilde{V}(\tilde{a})\right] \psi_l(\tilde{a}_\psi; \tilde{a}, \tilde{\phi})\, \mathrm{d}\tilde{a}_\psi \\
=&\lambda_l(\tilde{a}) \int_{(\tilde{a},+\infty)} \tilde{V}(\tilde{a}_\psi)\psi_l(\tilde{a}_\psi; \tilde{a}, \tilde{\phi})\, \mathrm{d}\tilde{a}_\psi - \lambda_l(\tilde{a}) \int_{(\tilde{a},+\infty)} \psi_l(\tilde{a}_\psi; \tilde{a}, \tilde{\phi})\, \mathrm{d}\tilde{a}_\psi \tilde{V}(\tilde{a})
\end{aligned} \tag{A68}$$

Denote $\psi_{l,nk} = \psi_l(\tilde{a}_k; \tilde{a}_n)$, $\lambda_{l,n} = \lambda_l(\tilde{a}_n)$, $p_{l,nk} = \lambda_{l,n}\psi_{l,nk}\Delta\tilde{a}$ and $p_{l,n} = \sum_{k=n+1}^{N} p_{l,nk}$. Then the first term in the last line is discretized as $F_3^{l,ND}\tilde{V}$, where

$$F_3^{l,ND} = \begin{pmatrix}
0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\
0 & 0 & p_{l,12} & p_{l,13} & \dots & p_{l,1\,N-2} & p_{l,1\,N-1} & p_{l,1\,N} \\
0 & 0 & 0 & p_{l,23} & \dots & p_{l,2\,N-2} & p_{l,2\,N-1} & p_{l,2\,N} \\
0 & 0 & 0 & 0 & \dots & p_{l,3\,N-2} & p_{l,3\,N-1} & p_{l,3\,N} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & 0 & p_{l,N-2\,N-1} & p_{l,N-2\,N} \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & p_{l,N-1\,N} \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 0
\end{pmatrix} \tag{A69}$$

and the second term in A68 is discretized as $F_3^{l,D}\tilde{V}$, where

$$F_3^{l,D} = \begin{pmatrix}
0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\
0 & -p_{l,1} & 0 & 0 & \dots & 0 & 0 & 0 \\
0 & 0 & -p_{l,2} & 0 & \dots & 0 & 0 & 0 \\
0 & 0 & 0 & -p_{l,3} & \dots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & -p_{l,N-2} & 0 & 0 \\
0 & 0 & 0 & 0 & \dots & 0 & -p_{l,N-1} & 0 \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 0
\end{pmatrix} \tag{A70}$$

Denote $F_3^l = F_3^{l,D} + F_3^{l,ND}$, then F_3^l is the discrete version of the incumbent learning's infinitesimal generator. Because jumps are non-local, the non-zero elements of F_3^l lie outside the tridiagonal area of the matrix. For computational efficiency, I have separated it into two parts, the diagonal part $F_3^{l,D}$ and the non-diagonal part $F_3^{l,ND}$ where ND stands for "non-diagonal".[90] Such a separation will be the basis of the explicit-implicit scheme. $\rfloor$

[90]One can also put the first upper diagonal in $F_3^{l,D}$ instead of in $F_3^{l,ND}$.

Finally, denote $\lambda_{d,n} = \lambda_d(\tilde{a}_n)$ for firm death. Then $-\lambda_d(\tilde{a}_\theta; \tilde{\phi}^M)\tilde{V}$ is discretized as $F_4\tilde{V}$, where

$$F_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\lambda_{d,1} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -\lambda_{d,2} & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & -\lambda_{d,3} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\lambda_{d,N-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -\lambda_{d,N-1} & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & -\lambda_{d,N} \end{pmatrix}$$

I use an implicit scheme for solving the HJB in the baseline model:

$$\begin{aligned} (r-g)\tilde{V}^{m+1} &= \Pi - f - \tilde{R} + F\tilde{V}^{m+1} + \frac{\tilde{V}^m - \tilde{V}^{m+1}}{\Delta t} \\ \Rightarrow \tilde{V}^{m+1} &= \Big[(r - g + \frac{1}{\Delta t})I - F\Big]^{-1}\Big[\Pi - f - \tilde{R} + \frac{1}{\Delta t}\tilde{V}^m\Big] \end{aligned} \tag{A71}$$

where I is the identity matrix and $F = F_1 + F_2 + F_3^{i,D} + F_4$.

[In the extended model, $F = F_1 + F_2 + F_3^{i,D} + F_3^{l,D} + F_3^{l,ND} + F_4$ and is non-sparse. Each update of $\tilde{V}^m$ in the implicit scheme requires an inversion of the matrix $(r - g + \frac{1}{\Delta t})I - F$ which is non-sparse. Such an inversion is costly, especially when N is high.

The explicit-implicit scheme resolves the issue by separating F into two parts, $F^D = F_1 + F_2 + F_3^{i,D} + F_3^{l,D} + F_4$ and $F^{ND} = F_3^{l,ND}$. All the elements of F^D lie on the tridiagonal area, making $(r - g + \frac{1}{\Delta t})I - F^D$ a sparse matrix and its inversion efficient. More specifically, the explicit-implicit scheme is:

$$\begin{aligned} (r-g)\tilde{V}^{m+1} &= \Pi - f - \tilde{R} + F^D\tilde{V}^{m+1} + F^{ND}\tilde{V}^m + \frac{\tilde{V}^m - \tilde{V}^{m+1}}{\Delta t} \\ \Rightarrow \tilde{V}^{m+1} &= \Big[(r - g + \frac{1}{\Delta t})I - F^D\Big]^{-1}\Big[\Pi - f - \tilde{R} + (\frac{1}{\Delta t}I + F^{ND})\tilde{V}^m\Big] \end{aligned} \tag{A72}$$

Importantly, like the implicit scheme, the explicit-implicit scheme is unconditionally stable, meaning that we can choose a large Δt which speeds up the convergence.]

To summarize, in step 4 of Algorithm 1, the HJB equation is solved by the following algorithm:

Algorithm 2 (HJB)**.**

`Initialize` $\tilde{V}^m$ `with` $m = 0$. `Initialize tolerance and some` $\epsilon >$ tolerance. `Fix`

Δt.

`While` $\epsilon >$ tolerance

1. $\Big[$`Choose between innovation and learning based on 60.`$\Big]$
2. `Choose the optimal intensity of innovation`$\Big[$`/learning`$\Big]$ `based on the first order condition.`
3. `Construct` $F_3^{i,D}$ `defined in A67`$\Big[$`,` $F_3^{l,D}$ `in A70 and` $F_3^{l,ND}$ `in A69`$\Big]$ `based on optimal choices. Construct` F `in the baseline model` $\Big[$`or` F^D `and` F^{ND} `in the extended model`$\Big]$`.` F `in the baseline model` $\Big[$`or` F^D `in the extended model`$\Big]$ `is specified as a sparse matrix.`
4. `Solve the linear system of the implicit scheme A71 in the baseline model` $\Big[$`or the explicit-implicit scheme A72 in the extended model`$\Big]$ `to get` $\tilde{V}^{m+1}$.
5. $\epsilon = \max |\tilde{V}^{m+1} - \tilde{V}^m|$.

`end`

$\Big[$The discrete choice between innovation and learning introduces a discontinuity in the extended model, i.e. the argmax operator is discontinuous. Even though it almost never causes a problem in practice, using softmax instead of argmax can make the algorithm more robust, as softmax smooths the transition from one choice to another. The smoothing only occurs in a small neighborhood of the discontinuous point so that there is little difference in the model solution.$\Big]$

H.4 Algorithm for solving the KF

A similar implicit scheme is used for solving the KF equation in the baseline model:

$$\begin{aligned} \frac{\tilde{\phi}^{m+1} - \tilde{\phi}^m}{\Delta t} &= F^\intercal \tilde{\phi}^{m+1} + \frac{\lambda_e M_e}{M}\psi_e \\ \Rightarrow \tilde{\phi}^{m+1} &= \left[\frac{1}{\Delta t} I - F^\intercal\right]^{-1}\left[\frac{1}{\Delta t}\tilde{\phi}^m + \frac{\lambda_e M_e}{M}\psi_e\right] \end{aligned} \tag{A73}$$

Note that the superscript m denotes the number of iterations in computations, not the measure of firms.

$\Big[$A similar explicit-implicit scheme is used for solving the KF equation in the extended model:

$$\begin{aligned}&\frac{\tilde{\phi}^{m+1}-\tilde{\phi}^{m}}{\Delta t}=(F^{D})^{\intercal}\tilde{\phi}^{m+1}+(F^{ND})^{\intercal}\tilde{\phi}^{m}+\frac{\lambda_e M_e}{M}\psi_e\\ \Rightarrow&\tilde{\phi}^{m+1}=\Big[\frac{1}{\Delta t}I-(F^{D})^{\intercal}\Big]^{-1}\Big[\Big(\frac{1}{\Delta t}I+(F^{ND})^{\intercal}\Big)\tilde{\phi}^{m}+\frac{\lambda_e M_e}{M}\psi_e\Big]\end{aligned} \tag{A74}$$

$\Big]$

We normalize $\tilde{\phi}$ each update to make sure it conserves its mass and integrates into 1.

To summarize, in step 6a of Algorithm 1, the KF equation is solved by the following algorithm:

Algorithm 3 (KF)**.**

`Initialize` $\tilde{\phi}^m$ `with` $m = 0$. `Initialize tolerance and some` $\epsilon >$ tolerance. `Fix` Δt.

`While` $\epsilon >$ tolerance

1. `From Algorithm 2, take` F `in the baseline model` $\Big[$`or` F^D `and` F^{ND} `in the extended model`$\Big]$. `Solve` $\tilde{\phi}^{m+1}$ `by the implicit scheme A73 in the baseline model` $\Big[$`or the explicit-implicit scheme A74 in the extended model`$\Big]$.
2. `Normalize` $\tilde{\phi}^{m+1}$ `so that it integrates into` 1, `i.e. it is a PDF.`
3. $\epsilon = \max|\tilde{\phi}^{m+1}-\tilde{\phi}^{m}|$.

`end`

H.5 Algorithm for solving the growth rate

Given a specific growth rate, the previous section solves the equilibrium distribution which implies an aggregate productivity. The actual growth rate is the one which makes the aggregate productivity to be 1 according to Definition 1. Thus the algorithm for solving the aggregate growth rate, i.e. step 6 of Algorithm 1, is the following:

Algorithm 4 (Growth)**.**

`Take` g `from the previous steps of Algorithm 1.`

1. Solve the KF based on Algorithm 3 to get $\tilde{\phi}$.

2. Solve the static problem based on $\tilde{\phi}$ and M to get the aggregate log TFP $\log(A)$.

   (a) If $|\log(A)| < \text{tolerance}$, the problem is solved.

   (b) If $\log(A) > \text{tolerance}$, increase g; if $\log(A) < -\text{tolerance}$, decrease g. Update F_1 and consequently F $\left[\text{or } F^D\right]$ based on the updated g. Go back to Step 1.

H.6 Balanced entry/exit condition

The discretized version of the balanced entry/exit condition 48 is:

$$\lambda_e M_e = M\Big[\sum_{n=1}^{N} \lambda_{d,n}\phi_n \Delta\tilde{a} + \frac{\nu^2}{2\Delta\tilde{a}}\phi_1\Big]$$

where $\phi_n = \tilde{\phi}(\tilde{a}_n)$.

H.7 Job creation and job destruction due to idiosyncratic shocks

Denote X to be a random variable with distribution $\mathcal{N}(-\frac{1}{2}\nu^2 - g, \nu^2)$. Job creation due to the drift term and the Brownian motion of incumbents at position θ is:

$$\begin{aligned}\Delta L_\theta^{c,b} &= \mathbb{E}\Big[[L(a_\theta + X) - L(a_\theta)]\mathbb{1}_{\{X>0\}}\Big] \\ &= \int_0^{+\infty} [L(a_\theta + x) - L(a_\theta)]\frac{1}{\sqrt{2\pi}\nu} \exp\Big[-\frac{1}{2}\Big(\frac{x + \frac{1}{2}\nu^2 + g}{\nu}\Big)^2\Big] \,\mathrm{d}x\end{aligned}$$

Do a change of variable of $x = \nu\sqrt{2y}$, then

$$\Delta L_\theta^{c,b} = \int_0^{+\infty} \frac{1}{2\sqrt{\pi}}[L(a_\theta + \nu\sqrt{2y}) - L(a_\theta)] \exp\Big[-(\frac{1}{2}\nu^2 + g)\frac{1}{\nu}\sqrt{2y} - \frac{1}{2\nu^2}(\frac{1}{2}\nu^2 + g)^2\Big] \exp(-y)y^{-\frac{1}{2}} \,\mathrm{d}y$$

The integration can be calculated using the Generalized Gauss–Laguerre quadrature with the form

$$\int_0^{+\infty} y^s e^{-y} f(y) \,\mathrm{d}y$$

by setting $s = -\frac{1}{2}$ and

$$f(y) = \frac{1}{2\sqrt{\pi}}[L(a_\theta + \nu\sqrt{2y}) - L(a_\theta)] \exp\Big[-(\frac{1}{2}\nu^2 + g)\frac{1}{\nu}\sqrt{2y} - \frac{1}{2\nu^2}(\frac{1}{2}\nu^2 + g)^2\Big]$$

Similarly, job destruction due to the drift and the Brownian motion of incumbents at position θ is:

$$\begin{aligned}\Delta L_\theta^{d,b} &= \mathbb{E}\Big[[L(a_\theta) - L(a_\theta + X)]\mathbb{1}_{\{X<0\}}\Big] \\ &= \int_{-\infty}^{0} [L(a_\theta) - L(a_\theta + x)] \frac{1}{\sqrt{2\pi}\nu} \exp\Big[-\frac{1}{2}\Big(\frac{x + \frac{1}{2}\nu^2 + g}{\nu}\Big)^2\Big] \,\mathrm{d}x\end{aligned}$$

Do a change of variable of $x = -\nu\sqrt{2y}$, then

$$\Delta L_\theta^{d,b} = \int_0^{+\infty} \frac{1}{2\sqrt{\pi}} [L(a_\theta) - L(a_\theta - \nu\sqrt{2y})] \exp\Big[(\frac{1}{2}\nu^2 + g)\frac{1}{\nu}\sqrt{2y} - \frac{1}{2\nu^2}(\frac{1}{2}\nu^2 + g)^2\Big] \exp(-y) y^{-\frac{1}{2}} \,\mathrm{d}y$$

which can similarly be calculated by the Generalized Gauss–Laguerre quadrature with $s = -\frac{1}{2}$ and

$$f(y) = \frac{1}{2\sqrt{\pi}} [L(a_\theta) - L(a_\theta - \nu\sqrt{2y})] \exp\Big[(\frac{1}{2}\nu^2 + g)\frac{1}{\nu}\sqrt{2y} - \frac{1}{2\nu^2}(\frac{1}{2}\nu^2 + g)^2\Big]$$

H.8 Growth contribution of net entry

From the proof of Proposition 1, it is clear that P/w is a functional of ϕ^M which can be denoted by

$$P/w = \mathcal{P}(\phi^M)$$

Hence by equations 17 and 19, A is also a functional of ϕ^M which can be denoted by

$$A = \mathcal{A}(\phi^M)$$

Note that the definition of $\mathcal{A}$ differs slightly from the one in Proposition 1: it is now expressed as a functional of ϕ^M. The Fréchet derivative of $\mathcal{A}$, as well as its application in defining the growth contribution of net entry, is summarized in the following proposition:

Proposition A22 (Growth Contribution of Net Entry)**.**

1.
$$\mathcal{A}'(\phi^M)h = -A^2\Big[\int_{\mathbb{R}} Z_\theta e^{-a_\theta} h(a_\theta)\,\mathrm{d}a_\theta + \mathcal{P}'(\phi^M)h \cdot \int_{\mathbb{R}} \frac{\partial Z}{\partial(P/w)}(a_\theta, P/w) e^{-a_\theta} \phi^M(a_\theta)\,\mathrm{d}a_\theta\Big] \tag{A75}$$

   where

$$\mathcal{P}'(\phi^M)h = -\Big[\int_{\mathbb{R}} \gamma'(Z_\theta)\frac{\partial Z}{\partial(P/w)}(a_\theta, P/w)\phi^M(a_\theta)\,\mathrm{d}a_\theta\Big]^{-1} \int_{\mathbb{R}} \gamma(Z_\theta) h(a_\theta)\,\mathrm{d}a_\theta \tag{A76}$$

   and $Z(a_\theta, P/w)$ is defined implicitly in equation 19.

2. *The growth contribution of net entry is*

$$\frac{1}{A}\mathcal{A}'(\phi^M)\Big[\lambda_e M_e \psi_e - \lambda_d(a_\theta; \phi^M)\phi^M\Big] \tag{A77}$$

where $\mathcal{A}'$ is given in A75.

Proof.

1. Write equation 4 as:
$$\int_{\mathbb{R}} \gamma\Big(Z(a_\theta, \mathcal{P}(\phi^M))\Big)\phi^M(a_\theta)\,\mathrm{d}a_\theta = 1$$
The LHS is a functional of ϕ^M. Take the Fréchet derivative with respect to ϕ^M on both sides, using the chain rule for Fréchet derivatives:
$$\int_{\mathbb{R}} \gamma'(Z_\theta)\frac{\partial Z}{\partial(P/w)}(a_\theta, P/w)\cdot \mathcal{P}'(\phi^M)h\cdot\phi^M(a_\theta)\,\mathrm{d}a_\theta + \int_{\mathbb{R}}\gamma(Z_\theta)h(a_\theta)\,\mathrm{d}a_\theta = 0$$
which gives equation A76.

   For the Fréchet derivative of $\mathcal{A}$, rewrite 17 as:
$$A = \mathcal{A}(\phi^M) = \Big[\int_{\mathbb{R}} Z(a_\theta, \mathcal{P}(\phi^M))e^{-a_\theta}\phi^M(a_\theta)\,\mathrm{d}a_\theta\Big]^{-1}$$
Hence equation A75 can be obtained by taking the Fréchet derivative of the above equation.

2. Obvious.

□

I Calibration

I.1 ω as index

Denote $\omega = -\log(1-\theta) \in [0, +\infty)$, where θ as before is the CDF of the firm along some distribution, for instance the value-added distribution. The incentive to use ω instead of θ to index firms comes from the Pareto tail. As a simple example, consider the case in which the value added x follows a Pareto distribution with Parameter η. The notations in equations A78-A79 are only temporary and should not be confused with other parts of the paper.

$$\mathbb{P}(x \le x) = 1 - (\frac{x_m}{x})^\eta, \quad \forall x \ge x_m \tag{A78}$$

with PDF $\frac{\eta x_m^\eta}{x^{\eta+1}}$. Then
$$\theta = 1 - (\frac{x_m}{x_\theta})^\eta$$

The complementary cumulative value-added share Ξ^r_θ, which is the value-added share of firms between $[\theta, 1]$ among the total value added, is

$$\begin{aligned}\Xi^r_\theta &= \frac{\int_{x_\theta}^{+\infty} x \frac{\eta x_m^\eta}{x^{\eta+1}}\,\mathrm{d}x}{\int_{x_m}^{+\infty} x \frac{\eta x_m^\eta}{x^{\eta+1}}\,\mathrm{d}x} \\ &= (\frac{x_m}{x_\theta})^{\eta-1}\end{aligned}$$

Hence

$$\log \Xi^r_\theta = -\frac{\eta-1}{\eta}\omega \tag{A79}$$

which is a linear function between ω and $\log \Xi^r_\theta$. In reality, the value-added distribution is Pareto at the right tail but not overall. Non-linearity would thus occur for ω close to 0. Nonetheless, the functional form is still globally close to a polynomial, which makes it easy to fit the data points.

I.2 Value added

The US Census Bureau publishes data on value added, material cost and number of establishments for each employment size range of manufacturing establishments. The value added is, however, overstated as the Census Bureau does not collect information on service input and thus does not deduce it from sales when calculating value added. I use the KLEMS data which reports material input and service input at the sectoral level to adjust for this bias. Assuming the same proportion of material input versus service input across size bins, service input can be inferred from material input in the Census of Manufacturing for each size bin. It is then subtracted from the published value added to obtain the corrected version to be used subsequently. See Table A9 for the adjusted data.

I fit a 3rd-order polynomial of $\log \Xi^r(\omega)$ on ω, i.e. $\log \Xi^r(\omega) = c_1\omega + c_2\omega^2 + c_3\omega^3$. There is no intercept in order to satisfy $\Xi^r(0) = 1$. See Figure A4a for the data and fitted curve. Denote ξ_θ as the value-added density of firm θ, i.e. $\Xi^r_\theta = \int_\theta^1 \xi_\theta\,\mathrm{d}\theta$. With a change of variable of $\theta = 1 - e^{-\omega}$,

$$\log \Xi^r(\omega) = \log \int_\omega^{+\infty} \xi(\omega) e^{-\omega}\,\mathrm{d}\omega$$

Take the derivative w.r.t. ω and rearrange the terms:

$$\xi(\omega) = -e^\omega \Xi^r(\omega) \frac{\mathrm{d}\log \Xi^r(\omega)}{\mathrm{d}\omega}$$

I.3 Labor share

Data on labor share across employment size bins also come from the Census of Manufacturing. A few adjustments are needed: (1) Payroll is adjusted by fringe benefits, as the former is narrowly defined in the Census of Manufacturing. The fringe/narrow payroll ratio comes from the Annual Survey of Manufacturing and is assumed to be constant across employment size bins. (2) Since the model does not feature heterogeneous wages, the average payroll per employee of each size bin is adjusted to the sectoral average. (3) The model does not feature capital, so the labor share in the data is adjusted to be consistent with the model. See Table A9 for the adjusted data.

For the adjustment regarding capital, consider the following production function with capital:

$$Y_\theta = A_\theta K_\theta^{1-s} L_\theta^s \tag{A80}$$

The optimal production decision per period is given by:

$$\begin{aligned}&\max_{K_\theta, L_\theta} P_\theta Y_\theta - wL_\theta - rK_\theta \\ &\text{s.t. A80 and 6}\end{aligned}$$

The first order conditions are:

$$\frac{\partial P_\theta}{\partial Y_\theta}\frac{\partial Y_\theta}{\partial L_\theta} Y_\theta + P_\theta \frac{\partial Y_\theta}{\partial L_\theta} = w \tag{A81}$$

$$\frac{\partial P_\theta}{\partial Y_\theta}\frac{\partial Y_\theta}{\partial K_\theta} Y_\theta + P_\theta \frac{\partial Y_\theta}{\partial K_\theta} = r \tag{A82}$$

Rewrite equation A81:

$$\begin{aligned}w &= \frac{\partial Y_\theta}{\partial L_\theta}\Big[\frac{\partial(P_\theta/P)}{\partial(Y_\theta/Y)}\frac{Y_\theta}{Y}P + P_\theta\Big] \\ &= \frac{\partial Y_\theta}{\partial L_\theta} P\Big[\gamma''(Z_\theta)Z_\theta + \gamma'(Z_\theta)\Big] \\ &= \frac{\partial Y_\theta}{\partial L_\theta} P\gamma'(Z_\theta)\Big[1 - \frac{1}{\sigma_\theta}\Big]\end{aligned}$$

or equivalently,

$$\frac{P_\theta}{w/\frac{\partial Y_\theta}{\partial L_\theta}} = \frac{\sigma_\theta}{\sigma_\theta - 1} \tag{A83}$$

Similarly, A82 can be rewritten as:

$$\frac{P_\theta}{r/\frac{\partial Y_\theta}{\partial K_\theta}} = \frac{\sigma_\theta}{\sigma_\theta - 1} \tag{A84}$$

The labor share of firm θ is then:

$$\chi_\theta = \frac{wL_\theta}{P_\theta Y_\theta} = \frac{w}{P_\theta A_\theta K_\theta^{1-s} L_\theta^{s-1}} = s\frac{w/\frac{\partial Y_\theta}{\partial L_\theta}}{P_\theta} = s\frac{\sigma_\theta - 1}{\sigma_\theta}$$

where the last step uses equation A83.

Employment smaller than	Number of establishments	θ	ω	Value added	Ξ_r	Labor share adjusted	Labor share final
5	141992	0.405	0.519	17.4	0.989	0.649	0.853
10	49284	0.545	0.788	22.7	0.973	0.627	0.824
20	50824	0.690	1.171	46.6	0.943	0.638	0.854
50	51660	0.837	1.816	112.0	0.867	0.632	0.846
100	25883	0.911	2.420	139.0	0.771	0.587	0.780
250	20346	0.969	3.477	269.0	0.581	0.538	0.746
500	6853	0.989	4.478	220.0	0.424	0.505	0.720
1000	2720	0.996	5.624	214.0	0.273	0.435	0.637
2500	1025	0.999	7.283	195.0	0.134	0.436	0.578
$+\infty$	241	1.000	$+\infty$	185.0	0.000	0.459	0.482

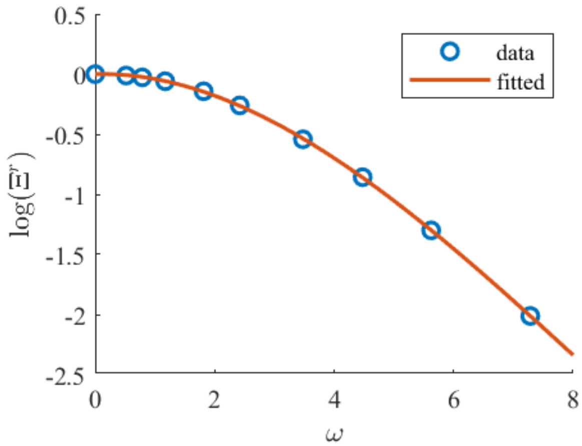
Table A9: Value added and labor share in the US Census of Manufacturing. Value added is in billions of USD. "Labor share adjusted data" is after the adjustments of value added and fringe benefits. "Labor share final" reports the final adjusted version.

Hence the labor share in the data should be divided by s to speak to the model. Intuitively, the production function of the data exhibits constant returns to scale for the combination of capital and labor, while its counterpart in the model exhibits constant returns to scale for labor. The labor share in the data should thus be adjusted by s to recover the labor share in the model. s can be estimated by the ratio between capital expenditure and payroll for each employment size bin. In particular, A83 and A84 imply:

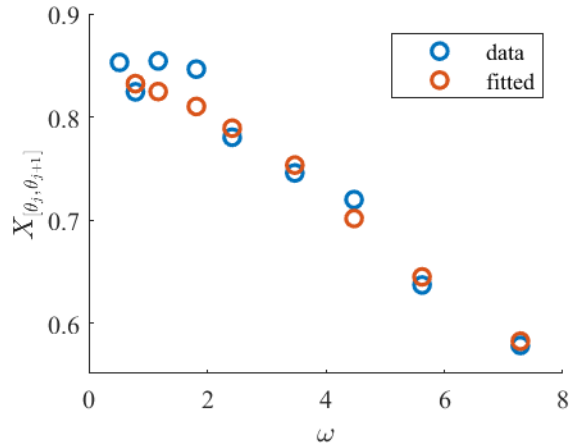
$$\frac{s}{1-s} = \frac{wL_\theta}{rK_\theta}$$

where wL_θ is the payroll after the first two steps of adjustment and rK_θ is capital expenditure. The latter needs to be imputed as the "total capital expenditures" in the census data does not include equity cost. The census variable does, however, provide an estimate of the share of capital stock across size bins among the total capital stock, assuming that the per unit capital cost and depreciation rate are the same across size bins. I estimate the capital stock of each size bin by multiplying the share with the total manufacturing capital stock in BEA fixed assets tables. It is subsequently multiplied by the required return of capital estimated by Barkai (2020) to obtain rK_θ for each size bin.

A final adjustment (4) is sometimes needed for the upper end of the labor share. When the labor share is close to 1, demand elasticity changes dramatically with the labor share, making $\overline{\sigma}$ sensitive to the data. To resolve this issue, it is useful to calibrate $\overline{\sigma}$ with respect to other papers and then adjust the data points accordingly. I calibrate $\overline{\sigma}$ as 6.5, which is in the middle of the range used in other papers with a nested CES preference under oligopolistic competition (see Atkeson and Burstein (2008), De Loecker et al. (2021), Gaubert and Itskhoki (2021) and Burstein et al. (2020)).



(a) Log cumulative value added share



(b) Bin-wise labor share

Figure A4: Estimating k.

I.4 Estimation of k and $\underline{\sigma}$

Denote $\chi(\omega) = 1 - \frac{1}{\sigma(\omega)}$ to be the labor share of firm at position ω, and $\overline{\chi} = 1 - \frac{1}{\overline{\sigma}}$ and $\underline{\chi} = 1 - \frac{1}{\underline{\sigma}}$ are the highest and lowest labor share respectively. We have set $\overline{\sigma} = 6.5$ in the last section. $\underline{\sigma}$ is set to 1.93 to match the lowest labor share along size bins in the census data. It remains to estimate k which governs the transition from $\overline{\sigma}$ to $\underline{\sigma}$.

Denote ξ_θ as the value-added of firm θ over sectoral value added, i.e. firm θ's market share in terms of value added. Denote χ_θ as the labor share of firm θ. The data gives the average labor share for a few size bins of establishments. Formally, for $j \in \{1, \cdots, n-1\}$, we have the bin-wise labor share

$$X_{[\theta_j,\theta_{j+1}]} = \frac{\int_{\theta_j}^{\theta_{j+1}} \chi_\theta \xi_\theta \, \mathrm{d}\theta}{\int_{\theta_j}^{\theta_{j+1}} \xi_\theta \, \mathrm{d}\theta}$$

where $0 = \theta_1 < \theta_2 < \cdots < \theta_n = 1$.

ξ_θ is recovered from the value added across size bins. To take care of the Pareto tail, I fit the cumulated value added share from the right tail as a function of $\omega = -\log(1-\theta)$. Formally, denote $\Xi^r_\theta = \int_\theta^1 \xi_\theta \, \mathrm{d}\theta$ as the cumulated value added share from the right tail. Figure A4a fits $\log(\Xi^r)$ as a 3-order polynomial of ω.

Note that

$$\begin{aligned}
\frac{\mathrm{d}\log\xi(\omega)}{\mathrm{d}\omega} &= \frac{\mathrm{d}\log\frac{P(\omega)Y(\omega)}{P/\zeta\cdot Y}}{\mathrm{d}\omega} \\
&= \frac{\mathrm{d}\log\frac{P(\omega)}{P}}{\mathrm{d}\omega} + \frac{\mathrm{d}\log\frac{Y(\omega)}{Y}}{\mathrm{d}\omega} \\
&= \chi(z(\omega))\frac{\mathrm{d}z(\omega)}{\mathrm{d}\omega}
\end{aligned}$$

where

$$\chi(z(\omega)) = \overline{\chi} + \frac{\underline{\chi} - \overline{\chi}}{\exp[-kz(\omega) + \log(\overline{\sigma}/\underline{\sigma})] + 1} \tag{A85}$$

Denote $X(z)$ to be a primitive of $\chi(z)$, then

$$\begin{aligned}
\frac{\mathrm{d}\log\xi(\omega)}{\mathrm{d}\omega} &= \frac{\mathrm{d}X}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}\omega} = \frac{\mathrm{d}X}{\mathrm{d}\omega} \\
\Rightarrow \log\xi &= X + C
\end{aligned} \tag{A86}$$

It is easy to check that $X(z)$ is

$$X(z) = \overline{\chi}z + \frac{\underline{\chi} - \overline{\chi}}{k}\log\Big[1 + \exp[kz(\omega) - \log(\overline{\sigma}/\underline{\sigma})]\Big]$$

Thus given $\xi(\omega)$ estimated in Section I.2, $z(\omega)$ can be obtained by solving the function A86. Plugging $z(\omega)$ into A85, we get χ as a function of ω. For $i \in \{1, \cdots, n-1\}$, we get the model counterpart of $X_{[\theta_i,\theta_{i+1}]}$:

$$\hat{X}_{[\theta_i,\theta_{i+1}]} = \frac{\int_{\theta_i}^{\theta_{i+1}} \chi_\theta\xi_\theta\,\mathrm{d}\theta}{\int_{\theta_i}^{\theta_{i+1}} \xi_\theta\,\mathrm{d}\theta} = \frac{\int_{\omega_i}^{\omega_{i+1}} \chi(\omega)\xi(\omega)e^{-\omega}\,\mathrm{d}\omega}{\Xi^r(\omega_i) - \Xi^r(\omega_{i+1})}$$

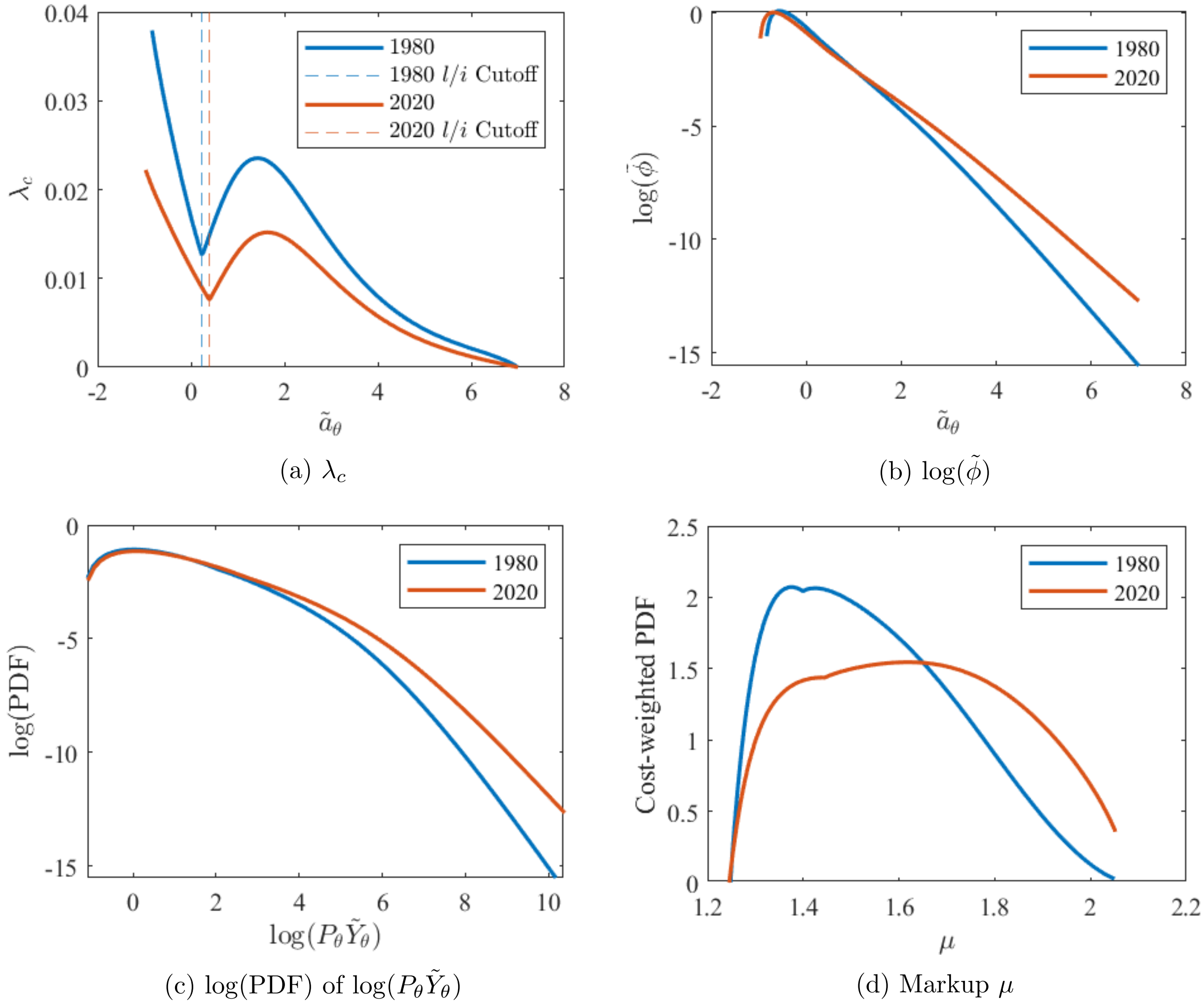
It remains to compare $X_{[\theta_i,\theta_{i+1}]}$ with $\hat{X}_{[\theta_i,\theta_{i+1}]}$, where $i \in \{1, \cdots, n-1\}$, and take the k and C which minimize their distance, i.e.

$$k^*, C^* = \arg\min_{k,C} \sum_{i=1}^{n-1} \Big[\hat{X}_{[\theta_i,\theta_{i+1}]} - X_{[\theta_i,\theta_{i+1}]}\Big]^2$$

See Figure A4b. k is then adjusted by the ratio between the employment Pareto tail index of all firms and that of manufacturing establishments, so that it governs the transition along the firm size distribution. k is estimated to be 0.32.

J Comparative statics under the extended model

Figure A5 shows the changes in innovation/learning intensities, as well as in equilibrium distributions. Table A10 calculates the moment changes as ideas get harder to find, and compares them



(a) λ_c

(b) $\log(\tilde{\phi})$

(c) log(PDF) of $\log(P_\theta \tilde{Y}_\theta)$

(d) Markup μ

Figure A5: Comparison between 1980 and 2020 under the extended model. Panel (a): learning/innovation intensity; Panel (b): log PDF of log productivity; Panel (c): log PDF of log sales; Panel (d): cost-weighted markup distribution.

with the data. Table A11 conducts a Melitz-Polanec decomposition of markup and labor share changes: the great majority of the changes still comes from between-firm reallocation.

Moment	Data Δ	Model Δ
Sales share of top 1% firms	10.07%	9.33%
TFP growth	-0.41%	-0.41%
Cost-weighted markup	11.40%	8.53%
Labor share	-5.39%	-3.38%
Job creation by birth (entry)	-2.70%	-1.79%
Job destruction by death (exit)	-1.67%	-1.66%
Job creation rate	-5.13%	-2.82%
Job destruction rate	-4.06%	-2.87%
Job reallocation rate	-8.12%	-5.69%
R&D over value added	64.50%	62.91%

Table A10: Moment changes between 1980 and 2020, extended model and data. All the changes are non-targeted.

	Within	Between	Net entry	Total
ΔLabor share	-0.30%	-3.05%	-0.03%	-3.38%
ΔMarkup	0.73%	7.82%	-0.02%	8.53%

Table A11: Melitz-Polanec decomposition of markup and labor share changes between 1980 and 2020 under the extended model.