# A Tale of Sunset Industries

*Job market paper*

Jinglun Yao *
London Business School

August 30, 2023
Version 1.0.0

**Abstract**

US corporate concentration has been persistently rising over the past century with flattening Pareto tails, and productivity growth has been concurrently declining. This paper builds a continuous-time Schumpeterian growth model that interprets higher concentration *as a result of* lower growth, which complements the existing Schumpeterian literature that focuses on the opposite direction. In the model, laggards benefit from dynamic growth advantage, while leaders possess the static advantage inherent to their leading position. A *Uniform* decline in research productivity hurts endogenous growth of all firms but in particular that of laggards, increasing the *relative* growth of leaders and fattening the Pareto tail of productivity distribution. With a demand system featuring variable demand elasticities, the proposed mechanism stemming from declining research productivity explains a majority of the changes in productivity growth, corporate concentration, markup, labor share, R&D cost, entry and exit rates, and job creation and desctruction rates of US since 1980s. The model can accommodate increasing concentration with stable markup and labor share in the pre-1980 period by introducing economic integration in addition to declining research productivity.

**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. The paper could not have taken its current form without the continuous and committed support of Hélène Rey. I am indebted to Philippe Aghion and Joseba Martinez for their insights and assistance during repeated discussions, and grateful to Ufuk Akcigit, David Argente, Martin Beraja, Florin Bilbiie, Johannes Boehm, Santiago Caicedo, Julieta Caunedo, Nicolas Coeurdacier, Maarten de Ridder, Gustavo de Souza, Basile Grassi, Chenlin Gu, Stéphane Guibaud, Wei Hua, Anders Humlum, Peter Klenow, Howard Kung, Pierre-Louis Lions, Isabelle Méjean, 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, John Van Reenen, Conor Walsh and seminar participants at Sciences Po, Collège de France/INSEAD, LBS, TADC conference for sharing their insights and valuable suggestions. All errors are my own.

# 1 Introduction

US corporate concentration has been persistently rising in the last 100 years. (Figure 1a, Kwon et al. (2022)). While many explanations have been proposed for the phenomenon since 1980s, its strikingly long-run nature remains elusive. Moreover, rising concentration takes a specific form of flattening Pareto tails (Kwon et al. (2022), Chen (2022)), which lies outside the purview of existing studies. What is the driving force behind the long-run rise in US corporate concentration? This question not only is important for its own sake but also sheds light on other questions since 1980s, including the rise of market power and the decline of labor share[1]. If demand elasticity decreases with firm size, as is assumed in standard macroeconomic model with heterogeneous markups, [2] then understanding higher concentration paves the way for understanding higher market power and lower share associated with lower demand elasticities. Nevertheless, taking a longer historical perspective also raises a puzzle: Despite a similar rise of corporate concentration before 1980, both the cost-weighted markup (Edmond et al. (forthcoming)) and the sales-weighted markup (De Loecker et al. (2020)) are estimated to be stable between 1950 and 1980. Corroborating the stable market power is the stable labor share before 1980s known as one of the Kaldor's facts (Kaldor (1961)). How can we reconcile the flattening tail of firm size distribution and stable market power during the same period?

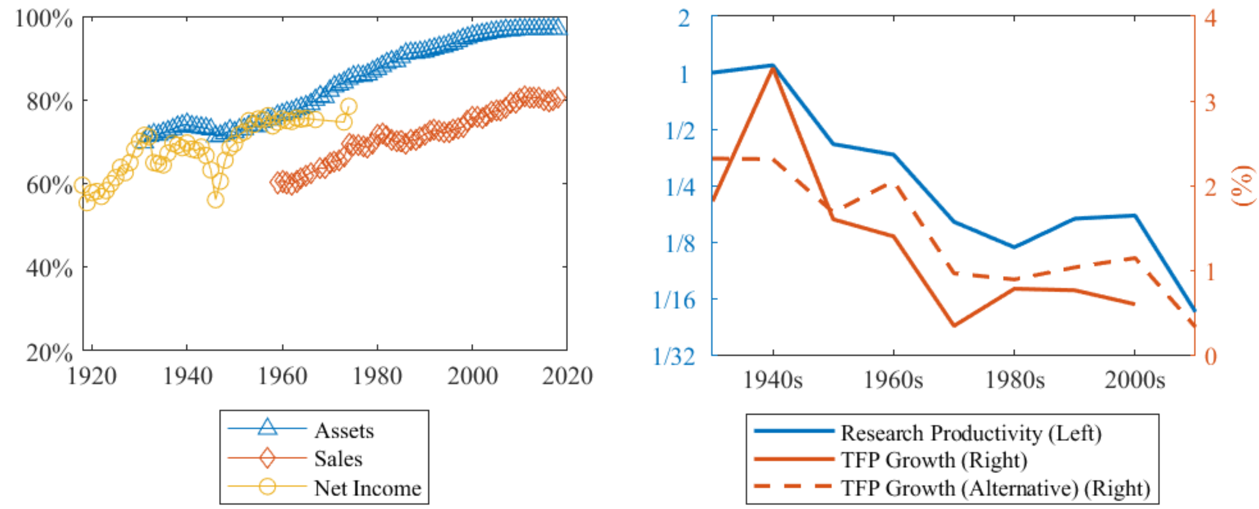
This paper argues that the driving force behind the long-run rise of corporate concentration is increasing research difficulty, i.e. "ideas are getting harder to find" borrowing the phrase from Bloom et al. (2020). Increasing research difficulty, or declining research productivity, has been identified as a widespread feature across scientific fields or industrial classifications (e.g. Gordon (2016), Bloom et al. (2020) and Park et al. (2023)). According to Gordon (2016) and Nordhaus (2021), US Total Factor Productivity (TFP) growth peaked in the 1930s and 1940s with an annual rate 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 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 return to scale in the idea production function. By construction, research productivity falls by a factor of 19 since 1930s with an annual growth rate of -3.67% (Figure 1b). [3]

To establish the link between increasing research difficulty and rising concentration, I build a

[1] See Karabarbounis and Neiman (2014), Autor et al. (2020) and Kehrig and Vincent (2021) for labor share, De Loecker et al. (2020) and Edmond et al. (forthcoming) for markup

[2] For instance, Atkeson and Burstein (2008) in the context of oligopolistic competition and Melitz and Ottaviano (2008) in the context of monopolistic competition.

[3] To be consistent with the model, I always 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 1930s with an annual grow 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.

continuous-time Schumpeterian growth model which endogenizes the growth decisions of individual firms and generates a productivity distribution with Pareto tail. In the model, each firm can either learn from other firms by imitating the technologies of more productive firms, or conduct innovation on a stand-alone basis. Both types of conscious productivity improvement are interpreted as "research", as measured growth rate and R&D cost, and consequently measured research productivity à la Bloom et al. (2020), include both.[4] In other words, both learning and innovation allow firms to improve their productivities via better knowledge, regardless of the specific form of improvement. Declining research productivity reflects more difficult knowledge to acquire along economic growth, as "standing on the shoulders of giants" is hard (Jones (2009)). The dynamic benefit of learning and innovation is especially important for industry laggards, as it is the only way that they can consciously challenge industry leaders. When research productivity declines *uniformly* across all firms, growth declines for all firms but in particular for laggard firms, as declining research productivity hurts the dynamic advantage that laggards heavily rely on. This increases the *relative growth* of leaders versus laggards and allows the former to stretch further into the right tail, with the consequence of flatter tail of productivity distribution and higher concentration. Under model assumptions, I demonstrate that there is a one-to-one correspondence between TFP growth rate and Pareto tail index of firm size distribution, so that the secular rise of

[4]Moreover, even though learning is modelled as immediate adoption of better technology in the form of jump, it can be interpreted as consecutive marginal improvements *with commitment*.

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 later is another empirical regularity documented by Kwon et al. (2022). In the model, higher TFP growth translates into 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.

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 this model, as it couples an Hamilton-Jacobi-Bellman (HJB) equation which describes firms' growth decisions based on profit incentives and hence on concentration, with a Kolmogorov Forward (KF) Equation which determines concentration based on growth decisions. In the model, higher research difficulty reduces growth and hurts the dynamic advantage of laggards, consequently increasing concentration. Instead of reducing growth, higher concentration in fact partially mitigates harder research and encourages growth, as it makes learning from leaders more effective and increases Schumpeterian profits of successful learning or innovation. Compared to the exiting Schumpeterian literature, the emphasis has now been shifted from how market structure determines growth to how growth determines market structure, at the application level if not at the methodology level.

The methodology benefits from the newly developed Mean Field Game which jointly determines the HJB equation and KF equation. Similar to other papers of Mean Field Game [5], I show that the functional operator of the KF equation is the adjoint operator of the operator in the HJB equation, which tightly characterizes the relationship between these two equations and paves the way for efficient computation. In addition to conscious productivity improvements discussed above, the model introduces idiosyncratic productivity shocks, i.e. random growth, in order to generate Pareto tail like many existing papers (e.g. Gabaix (1999), Gabaix et al. (2016)) [6]. When learning is present, however, these papers typically assume an initial productivity distribution with lighter tail to uniquely pin down the equilibrium distribution of the KF equation (e.g. Luttmer (2012), Perla et al. (2021)). While the assumption is mild for the objectives of these papers, it amounts to assuming what the present paper seeks to explain. This paper instead introduces imperfect learning in the sense that when a laggard is matched with a leader in the case of learning, the laggard can improve its productivity to that of the leader with positive Dirac-mass probability *but also* faces positive probability of moving to some productivity in-between. Learning is imperfect as the second probability is positive so that the Dirac probability of jumping to the leader productivity is less than 1. The specification can be seen as the continuous counterpart of the discrete case in

[5] See for instance Achdou et al. (2022).

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

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 distribution and concentration, without resorting to additional assumptions on initial productivity distribution.

To study the implications of higher concentration on markup and labor share, I design a demand system featuring variable demand elasticities and assume firms competing monopolistically with each other. The demand system is a special specification of the Kimball (1995) preference 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 getting rid of strategic interactions of an oligopolistic competition. The last is necessary for the methodology of Mean Field Game, as each firm is infinitesimal in the case of monopolistic competition and does not exert noticeable impact on the aggregate economy by its own. Like in a static setting, fatter tail of the productivity distribution translates into higher markup and lower labor share given other things equal. Both the increase in markup 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 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 1980s, basic infrastructure projects of highways, airline facilities etc. connecting different regions of the US (Gordon (2016)), as well as deregulations such as the 1978 Airline Deregulation Act which remove market access restriction across regions, motivates us to consider economic integration as another first-order economic force during the pre-1980 period. [7] As the productivity distribution is Pareto at the right tail but log-concave in log-productivity overall, economic integration reduces markups and increases labor share through the within-firm component as it brings more firms competing with each other (Autor et al. (2020)). 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 largely stable while concentration similarly increases to a fatter tail. [8]

Higher market power since 1980s interprets concurrent declining job reallocation rate as reduced reaction to idiosyncratic shocks and not as reduced volatility of these shocks. 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 model thus also speaks

[7] Other examples of rising concentration but stable markup and labor share include European countries in the last few decades. 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. The French case is currently under investigation.

[8] 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 flattening Pareto tail.

to diminished business dynamism and the mechanism through reduced reaction is consistent with empirical evidence in Decker et al. (2020). In addition, the model incorporates entry and exit along the firm size distribution in a parsimonious way. Exit rate declines for similar reasons due to reduced reaction to productivity shocks, while entry is modelled as potential entrants learning from incumbent firms and its rate declines as research becomes more difficult. The model thus also speaks to declined entry and exit rate in the data, another facet of diminished business dynamism.

Using sectoral-level data from the US since 1987, I find empirical evidence for the proposed mechanism by which harder research shapes market structure and business dynamism. Research productivity, measured by TFP growth over R&D cost, is significantly negatively correlated with market concentration and profitability of largest firms in the sector, positively correlated with labor share and major indicators of business dynamism in the BDS (job creation rate by entrants, total job creation rate, job destruction rate by death, total job destruction rate, job reallocation rate, and number of entrants/deaths over total number of establishments). To quantify the model, I calibrate the demand system to match heterogeneous labor shares across 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 Bloom et al. (2020)'s methodology with aggregate US data, and a decrease of 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 innovation 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. [9] The increase in R&D expenditure *in conjunction with* harder research in the data needs to be understood in a setting in which compensating forces such as lower interest rate mitigate the growth effects of harder research. [10] [11] Despite the pro-growth effect of lower interest rate and higher concentration discussed above, the large increase in research difficulty dominates the landscape. The model can explain a majority 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 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

[9] 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 $\lambda$ is the rate of success to be optimally chosen. Denote $V > 0$ to be the benefit if $\lambda$ is successfully realized. Then the optimal $\lambda$ 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 later is decreasing in $\alpha$, i.e. decreasing in research difficulty in a partial equilibrium. If, however, compensating forces such as lower interest rate make the change of $\lambda$ less than what is implied by the sole force of harder research, then R&D expenditure can increase in a general equilibrium.

[10] 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.

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

1900s to 1930s and peaked in 1930s and 1940s according to Gordon (2016) and Nordhaus (2021). In Philippon (2022), productivity is best described as "additive" *per historical era*, meaning that growth rate declines during each era but can temporarily increase at some conjunction moments. In Schumpeterian literature, this pattern of rise and fall of growth is termed the "Schumpeterian Wave". [12] 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 our era, the lack of disruptive technologies has been extensively documented by the recent literature. Gordon (2016) provides an 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 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 *wide scope* of applicability.

**Related Literature** This paper first of all situates within the Schumpeterian growth literature and owes its inspirations to seminal works of Aghion and Howitt (1992) and Aghion et al. (2001). While Aghion et al. (2001) does not generate a well-defined productivity distribution[13], this paper contributes by generating a well-defined productivity distribution with Pareto tail. Such a distribution allows us to understand why market concentration has increased via the specific form of flattening Pareto tail. It also allows us to evaluate economic integration like in the international trade literature, and thus to study dynamic growth effects and static market structure changes in a unified framework. At the application level, the paper has shifted from the traditional Schumpeterian emphasis of how market structure determines growth to how growth determines market structure. Akcigit and Ates (2021) is one of the most recent examples of the traditional emphasis which interprets lower growth since 1980s through a change in market structure, i.e. more concentrated markets when exogenous technological diffusion rate decreases. [14] 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 has been shifted to how growth determines market structure, with dramatic implications on whether

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

[13] 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) makes innovation harder for more productive sectors and generates a stationary productivity distribution, but the distribution is thin-tailed.

[14] However, decrease in 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 (forthcoming) calibrates the model to the data, the two forces compensate each other so that growth decreases little.

anti-trust policies should be conducted.

Declining research productivity, the driving force in this paper, draws from fundamental insights of the semi-endogenous growth literature, e.g. Jones (1995), Kortum (1997) and Segerstrom (1998). [15] Bloom et al. (2020) provides recent empirical evidence based on micro data. Existing theoretical frameworks typically assume a positive exogenous population growth which cancels harder research effect by increasing number of researchers, and consequently maintains constant productivity growth. When Jones (2002) brings the model to the data, however, less than 20% of growth is explained by population growth. Thus the present paper assumes a fixed number of population in cases without economic integration, and studies the effect of declining research productivity on declining growth. It contributes to the literature by deciphering the link between research productivity and concentration via growth, with implications on market power and business dynamism.

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 cost so that increasing concentration is a result of higher fixed cost. However, total R&D expenditure over GDP has never exceeded 3.5% in the past century of US. [16] Such a small change in fixed cost 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. [17] 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 can largely increase concentration via a change in the endogenous productivity distribution. 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[18], my paper focuses on equilibrium distribution for each historical period and the transition is understood as moving from one equilibrium to another due to changes in research productivity[19]. While concentration increases and growth keeps constant in Chen (2022),

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

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

[17] 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 *lighter* tail of employment, as the productivity tail remains the same but demand elasticity is lower.

[18] 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.

[19] 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 should be too slow to match the data, as explained by Luttmer (2011), Gabaix et al. (2016) and Jones and Kim (2018). Like these papers, type dependence can be potentially introduced in an extended model to speed up the transition to a reasonable scale.

in the present 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 dynamism since 1980s. Notable works on the theory side include Akcigit and Ates (2021, forthcoming) on technological diffusion, Aghion et al. (forthcoming) and De Ridder (2019) on intangibles, and Liu et al. (2022) on low interest rate. While each theory sheds light on specific mechanism, they face the common challenge of interpreting the long-run trend of declining growth and rising concentration since 1930s. [20] Lower technological diffusion in Akcigit and Ates (2021, forthcoming) is interpreted as a result of more stringent intellectual property right (IPR) enforcement. While this could be the case since 1980s, there is little evidence on the secular increase of IPR enforcement since 1930s. Even if the later is true, its pro-growth effect on leaders largely compensates its anti-growth effect from higher concentration, making it difficult to speak to the secular decline in growth. Despite this, my paper does concur with Akcigit and Ates (2021, forthcoming) in that technological diffusion rate, i.e. learning rate in this paper, declines. Declining diffusion rate is now understood to be *not* due to institutional reasons, but rather to increasing technological difficulty. Aghion et al. (forthcoming) and De Ridder (2019) assume ex ante differences in firms' ability 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 widely relevant before 1980s, making its historical relevance limited. Despite this, my paper does concur with higher intangibles of 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 R&D share of GDP increases in my model. Liu et al. (2022) investigates anti-growth effect of lower interest rate when the last is close to 0 and if there is no quick catch-up of laggards. As the present paper includes learning as quick catch-up, lower interest rate only has a pro-growth effect.[21] In fact, in my paper lower interest rate acts as a compensating force to mitigate the effects of harder research. Closest in spirit to the present paper in the literature are Olmstead-Rumsey (2019) and Cavenaile et al. (2019). Olmstead-Rumsey (2019) decreases the innovative-ness of all firms which corresponds to declining research productivity in this paper, and Cavenaile et al. (2019) compares different mechanisms and argues harder research to be the most likely driver behind lower growth. The present paper contributes by generating well-defined distributions with Pareto tails, and by linking the specific form of rising concentration, i.e. flattening Pareto tail, to lower growth. It also discusses why market power may not change with increasing concentration, making it possible to speak to a longer history and a wider geographical horizon.

[20] 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.

[21] 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 1980s according to Barkai (2020).

While the paper acknowledges the merits of each of these theories, it aims towards forming a synthesis of at least some of the literature. It is a synthesis in the sense that it speaks to not only the key macroeconomic trends to be explained, but also to the drivers in the literature that explain these moments, even though it gives different interpretations to these drivers. These key macro moments include TFP growth (Gordon (2016), Nordhaus (2021)), corporate concentration (Kwon et al. (2022)), markup (De Loecker et al. (2020), Edmond et al. (forthcoming)), labor share (Karabarbounis and Neiman (2014), Autor et al. (2020), Kehrig and Vincent (2021)) and various business dynamism indicators (Decker et al. (2016), Decker et al. (2020)). It explains why market power and labor share may or may not change with flattening Pareto tail depending on different circumstances. The drivers in the literature include at least technological diffusion rate, intangibles, interest rate and innovative-ness of R&D, which the current model can speak to as explained by the previous paragraph. In this sense, it strives to resolve the doubling accounting issue in the literature and hopes to avoid "explaining the labor share decline many times over" as advocated by Grossman and Oberfield (2021).

The framework takes advantages 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 into 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 innovation and learning decisions of individual firms. This paper differs by generating Pareto-tailed distributions and links its flattening tail to lower growth when ideas are getting harder to find. It also studies implications on market power and labor share with a demand system featuring variable demand elasticities.

**Roadmap** Section 2 presents empirical evidence on relationships between research productivity and various indicators of market structure and business dynamism. Section 3 constructs a model with endogenous learning and innovation, generates a travelling wave solution of productivity distribution, and describes business dynamics with entry and exit. The distribution forms the basis of firm heterogeneity and market power in a monopolistic competition with variable demand elasticities. Section 4 calibrates the model parameters based on US data. When ideas are getting harder to find, Section 5 explores its implications on growth, market structure and business dynamism through comparative statics analysis. Section 6 reconciles the earlier US experience of increasing concentration with stable markup and labor share by introducing economic integration alongside harder innovation. Section 7 concludes with suggestive policy implications.

## 2 Empirical Evidence

I use US sectoral level data between 1987 and 2018 to show 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 et al. (2020) and define research productivity as the ratio between TFP growth and effective number of researchers at the sectoral level, taking into account decreasing returns to scale of the idea production function. 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. In line with the model specified later where R&D cost is a quadratic function of growth, in the idea production function multiplying the number of researchers by $\lambda > 0$ changes TFP growth by $\lambda^{\frac{1}{2}}$. 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.[22] Results remain robust with constant returns to scale of the idea production function (see Appendix A). R&D expenditure comes from aggregation of firm-level R&D in Compustat into sectoral level. TFP growth comes from KLEMS data of US Bureau of Labor Statistics (BLS) and Bureau of Economic Analysis (BEA). To remove cyclical variations of TFP growth in a business cycle, I divide the 1987-2018 history into 4 equal periods, each with 8 years. All variables are calculated as per period average. Such a division is also conceptually reasonable as a change in "research productivity" seems only meaningful when a longer horizon is involved. Using divisions other than 8 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.

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 considerable problem for the definition of research productivity as the economic intuition does not coincide with the mathematical definition: Given a negative growth rate, higher research input suggests lower research productivity according to economic reasonings, but the ratio of growth/research input goes *up* because of the negative sign. Moreover, the ratio is negative so that we cannot take 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 regards a growth rate very close to 0 to be indifferent from 0, and is similar to winsorization in the applied microeconomics literature. See Appendix A for details.

Indicator of market concentration is constructed from Business Dynamics Statistics (BDS) which reports employment by firm size category and sector. For each sector*period, I define market concentration as the employment share of firms with more than 5000 employees. EBIT/Sales of the largest 4 or 8 firms by sector in Compustat is used to indicate the profitability of largest firms in the sector. Sectoral labor shares come from KLEMS. For business dynamics, I use entry rate in terms of number of establishments, entry rate in terms of job creation from birth, total job creation rate, exit rate in terms of number of establishments, exit rate in terms of job destruction from

[22] 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 |
|---|---|---|---|---|---|
| logideaprod | 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.

death, total job destruction rate, job reallocation rate, all from BDS.

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

$$y_{s,t} = \beta_0 + \beta_1 \widetilde{\log}(g_{s,t}/R_{s,t}^{1/2}) + \gamma_t + \epsilon_{s,t} \tag{1}$$

where $\gamma_t$ is period fixed effect, $g_{s,t}$ is sectoral TFP growth during period t, $R_{s,t}$ is effective number of researchers, and $\widetilde{\log}(g_{s,t}/R_{s,t}^{1/2}) = \text{glog}(g_{s,t}) - \log(R_{s,t}^{1/2})$. 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, profitability of the largest firms in a sector, and positively correlated with labor share and all indicators of business dynamics. These empirical patterns are consistent with a model in which declining research productivity decreases growth and increases market concentration. Higher market concentration manifests itself via higher market power and lower labor share under 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 model consistent with the empirical findings.

## 3 Model

Time $t$ is continuous. The economy has a representative consumer who consumes differentiated goods per period. Each good is produced by one firm which compete monopolistically among themselves. Firms' productivities are heterogeneous, and each firm can improve its productivity by learning from other firms or by innovating on a stand-alone basis. Incumbent firms can die due to idiosyncratic shocks to their productivities, and entrants keep joining the pool of incumbent firms. For notational simplicity, I shall drop the time index $t$ whenever possible.

| | (1) Entry (num) | (2) Exit (num) | (3) Entry (job) | (4) Exit (job) | (5) Job Creation | (6) Job Destruction | (7) Job Reallocation |
|---|---|---|---|---|---|---|---|
| logideaprod | 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 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 number of jobs destructed by exits and continuers over total jobs. Job reallocation is 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.

## 3.1 Market Structure

### 3.1.1 Preference

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

$$M \int_0^1 \gamma(\frac{Y_\theta}{Y})\mathrm{d}\theta = 1 \tag{2}$$

where $M$ denotes the total measure of varieties, $Y_\theta$ is the consumption of variety $\theta$ and $\gamma$ is an increasing and concave function with $\gamma(0) = 0$ and $\gamma(1) = 1$[23].

Each variety is produced by one firm and firms differ in their productivities. Firms are indexed by $\theta \in [0, 1]$ which is the cumulative probability of the log-productivity distribution with $\theta = \Phi(a_\theta)$, where $a_\theta$ is the log productivity of firm $\theta$ and $\Phi$ is the CDF of log productivities.[24] Thus more productive firms have higher values of $\theta$. It should be emphasized that equation 2 implicitly incorporates the distribution of productivities. To see this more clearly, do a change of variable from $\theta$ to $a_\theta$ to get:

$$M \int_{\mathbb{R}} \gamma(\frac{Y(a_\theta)}{Y})\phi(a_\theta)\mathrm{d}a_\theta = 1 \tag{3}$$

where $\phi$ is the probability distribution function (PDF) of log productivities. 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 for the paper.

[23]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 $\gamma(1) = 1$.

[24]To see why using cumulative probability for indexing firms, consider the discrete counterpart of Equation 2: $M \frac{1}{I} \sum_{i=1}^{I} \gamma(\frac{Y_{i/I}}{Y}) = 1$. In this case, $\theta(a_{i/I}) = \frac{i}{I}$ for the $i$-th firm.

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

$$M \int_0^1 P_\theta Y_\theta \,\mathrm{d}\theta = W \tag{4}$$

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

$$\frac{P_\theta}{P} = \gamma'(\frac{Y_\theta}{Y}) \tag{5}$$

where the price aggregator $P$ satisfies:

$$\frac{P}{\zeta} Y = W \tag{6}$$

and $\zeta$ is defined as

$$\zeta = \big[M \int_0^1 \gamma'(\frac{Y_\theta}{Y}) \frac{Y_\theta}{Y} \,\mathrm{d}\theta\big]^{-1} \tag{7}$$

Note that there are two price indices when the preference is Kimball: $P$ and $\zeta$.[25] The ideal price index, which is used to deflate nominal income to evaluate welfare, is $P/\zeta$ and *not* $P$.

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

$$\sigma_\theta = \frac{\gamma'(Z_\theta)}{-Z_\theta \gamma''(Z_\theta)} \tag{8}$$

Constant Elasticity of Substitutions (CES) preference is a special case of the Kimball preference 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 Kimball aggregator can be seen from equations 5 and 8: by introducing a different functional form of $\gamma$, we can generate heterogeneous demand elasticities and hence heterogeneous markups for heterogeneous firms. Current literature has often adopted the specification of Klenow and Willis (2016) for the Kimball aggregator. However, as Baqaee et al. (2020) 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 of

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

market power.[26] In particular, define $\gamma'$ for the inverse demand curve 5:

$$\gamma'(Z_\theta) = C\Big[\frac{1}{\underline{\sigma}}Z_\theta^{\frac{k\underline{\sigma}}{\overline{\sigma}-\underline{\sigma}}} + \frac{1}{\overline{\sigma}}Z_\theta^{\frac{k\overline{\sigma}}{\overline{\sigma}-\underline{\sigma}}}\Big]^{-\frac{\overline{\sigma}-\underline{\sigma}}{k\overline{\sigma}\underline{\sigma}}} \tag{9}$$

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.[27] $k$ 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 nested-CES of Atkeson and Burstein (2008) as it combines two power functions, each corresponding to one limit demand elasticity. Section 3.1.3 discusses 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 8 for calculating the demand elasticity:

$$\sigma(z_\theta) = \overline{\sigma} + \frac{\underline{\sigma} - \overline{\sigma}}{1 + \exp(-kz_\theta)} \tag{10}$$

i.e. $\sigma$ is a logistic function of $z_\theta$, which decreases monotonically with $z_\theta$ and has limits:

$$\lim_{z_\theta \to +\infty} \sigma(z_\theta) = \underline{\sigma}, \quad \lim_{z_\theta \to -\infty} \sigma(z_\theta) = \overline{\sigma} \tag{11}$$

### 3.1.2 Firms' production decision

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

$$Y_\theta = A_\theta L_\theta \tag{12}$$

where $A_\theta$ is the productivity of the firm and $L_\theta$ the amount of labor used in the production. Firms engage in monopolistic competition with each other and take into account the inverse demand curve 5 for pricing decisions. Optimal pricing implies the markup:

$$\mu_\theta = \frac{P_\theta/P}{MC_\theta/P} = \frac{\sigma(z_\theta)}{\sigma(z_\theta) - 1} \tag{13}$$

where $MC_\theta = \frac{w}{A_\theta}$ is the marginal cost of production and $w$ is the wage rate. Equation 13 is similar to the markup formula under CES, with only the exception that $\sigma$ now depends on $z_\theta$. With given aggregate variales $w$ and $P$, the intersection of the inverse demand curve 5 and the pricing equation 13 solves a firm's static production problem at each point in time. Both price and quantity are functions of the firm's productivity.

[26] In principle, I can also use a non-parametric form of Kimball aggregator following Baqaee et al. (2020). Nonetheless, the parametric form makes numerical calculations easier and more accurate. It also gives a portable demand system for further studies.

[27] See Appendix G.1 for numerically calculating $\gamma$.

Production incurs a fixed cost of $fA$ per period, where $A$ is the aggregate TFP to reflect higher wage as economy grows. 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:

$$\Pi(\underline{a}; \phi^M) = fA \tag{14}$$

#### 3.1.3 Properties of the Demand System

Aggregate productivity is defined by:

$$A = \frac{Y}{L} \tag{15}$$

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

$$A = \left[ M \int_0^1 Z_\theta A_\theta^{-1} \, \mathrm{d}\theta \right]^{-1} \tag{16}$$

i.e. $A$ is a harmonic mean of $A_\theta$ 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 $A = \mathcal{A}(\{A_\theta\}_{0 \le \theta \le 1})$. 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 $\{A_\theta\}_{0 \le \theta \le 1}$, i.e. $\mathcal{A}(\{\kappa A_\theta\}_{0 \le \theta \le 1}) = \kappa \mathcal{A}(\{A_\theta\}_{0 \le \theta \le 1})$, $\forall \kappa > 0$*

*Proof.* Consider the case in which each $A_\theta$ is multiplied by $\kappa$. To show the homotheticity of $A$ with respect to $\{A_\theta\}_{0 \le \theta \le 1}$, by equation 16 we only need to show that $Z_\theta = \frac{Y_\theta}{Y} \quad (\forall \theta)$ do not change.

Combining equations 5 and 13 gives:

$$\frac{\gamma'(Z_\theta)}{MC_\theta / P} = \frac{\sigma(z_\theta)}{\sigma(z_\theta) - 1} \tag{17}$$

which pins down $Z_\theta$ as an implicit function of $A_\theta$ and $P/w$:

$$Z_\theta = Z(a_\theta, P/w) \tag{18}$$

Plugging this function into the Kimball aggregator 2 pins down $P/w$ as a function of $\{A_\theta\}_{0 \le \theta \le 1}$ and $M$. If each $A_\theta$ is multiplied by $\kappa$, it is easy to check that $\{Z_\theta\}_{0 \le \theta \le 1}$ and $\frac{P}{\kappa w}$ is the new solution of this system of equations. Price-wage ratio decreases by $\kappa$ 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 $\kappa$ but relative quantities $\{Z_\theta\}_{0\le\theta\le1}$ do not change. In the special 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 $A_\theta$ 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 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* $a_\theta$ *is decreasing on* $(-\infty, \hat{a}]$ *and increasing on* $[\hat{a}, +\infty)$, *i.e. The passthrough of cost shocks to price decreases with firm's productivity when the last does not exceed a certain threshold* $\hat{a}$.

3. $Y_\theta \sim CA_\theta^{\underline{\sigma}}$, $P_\theta Y_\theta \sim CA_\theta^{\underline{\sigma}-1}$ *and* $L_\theta \sim CA_\theta^{\underline{\sigma}-1}$ *as* $a_\theta \to +\infty$, *where* $C > 0$ *is some constant and is different for each of the variables.*[28] *Moreover, the distributions of these variables are Pareto-tailed if the productivity distribution is Pareto-tailed.*

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

Lower demand elasticities and passthroughs of more productive firms are consistent with empirical evidence, as surveyed by Arkolakis and Morlacco (2017), and are primordial for Schumpeterian incentives of innovation. An improvement in productivity reduces 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 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 much, giving excessively stronger incentive to innovate to high productivity firms. Similarly, if demand elasticity $\sigma$ is mistakenly high for productive firms, their growth rate would also be excessive. In fact, the demand system in 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

[28] As with standard mathematical notations, $f(x) \sim g(x)$ as $x \to +\infty$ means $\lim_{x\to+\infty} \frac{f(x)}{g(x)} = 1$.

(2016) specification of Kimball satisfy the first two properties, they imply 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 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). [29] The oligopolistic competition setting of Atkeson and Burstein (2008) however implies strategic interactions between firms. When growth decision is involved as in my paper, such interactions make even numerical solutions untractable. 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.

Finally, it should be noted that heterogeneous demand elasticities must be introduced even if we want to match *aggregate* markup *at a specific time* and not its change over time. To appreciate this point, consider CES instead. The profit function is of the order of magnitude $CA_\theta^{\sigma-1}$. On the other hand, the R&D cost function is quadratic in growth by empirical estimates. If $\sigma > 3$, firms have infinite incentives to innovate and the model breaks down. However, empirical estimates of the average elasticity suggest it very likely to be in this range, especially if it is estimated within sectors (Broda and Weinstein (2006)). A demand system with variable demand elasticity reconciles the empirical estimate of R&D cost function with that of average demand elasticity. More productive firms face a lower demand elasticity (i.e. $\sigma < 3$), which can be interpreted as market saturation and limits the firm's incentive to innovate. A low productivity firm, though facing a high elasticity at the moment, anticipates the low elasticity / market saturation that it will face once it becomes a market leader. This makes sure that R&D decisions are well-defined for each firm.

The specification is also reasonable from an empirical point view. Figure 2 shows labor shares across US manufacturing establishments, where each point represents a bin of establishment size range. $\omega = -\log(1-\theta)$ is used as index instead of $\theta$ to take into account the Pareto-tail nature of distributions. Labor share decreases with establishment size in the Census of Manufacturing with possible bounds on the two extremities (See Appendix H for the data table). A logistic function is a natural choice to capture this feature.

[29]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 competing with other sectors as if in a monopolistic competition between sectors. Since sectoral aggregator is a CES, passthrough approaches 1 when the firm's productivity tends to $+\infty$.

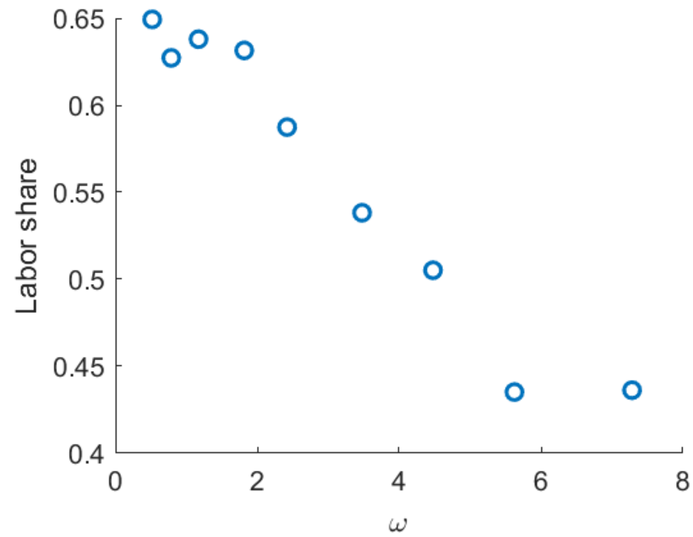



Figure 2: Labor share in a cross-section of establishments, US Census of Manufacturing 2002 corrected by KLEMS and Annual Survey of Manufacturing. $\omega = -\log(1-\theta)$ to take into account the Pareto tail.

## 3.2 Exit and Entry

### 3.2.1 Exit

Incumbents face idiosyncratic shocks of log productivity modelled as Brownian motion with standard deviation of $\nu$, regardless of their current level of productivities. [30] The manifestation of this productivity shock into 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 of sales per period exceeds $\kappa$, the firm dies. Death rate $\lambda_{d,\theta}$ is thus monotonically decreasing in firm size, which is consistent with empirical evidence shown in Figure 3. A demand system with variable demand elasticities, in addition to the benefits discussed in Section 3.1.3, also lends itself to heterogeneous death rates.

Formally[31],

$$\lambda_{d,\theta} = \lambda_d(a_\theta; \phi^M) = \mathbb{P}\Big(\frac{\partial \log(P_\theta Y_\theta)}{\partial a_\theta} \cdot \nu B_1 < -\kappa\Big) \tag{19}$$

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

[30] Put it in discrete terms, during each period a productivity shock of $\mathcal{N}(0, \nu^2)$ is realized at the log level.
[31] To incentivize the definition, I have stayed slightly outside the continuous-time setting and considered its discrete counterpart. Once definition 19 is put down, however, it only depends on variables at time $t$ and is equally suitable for the continuous-time setting.

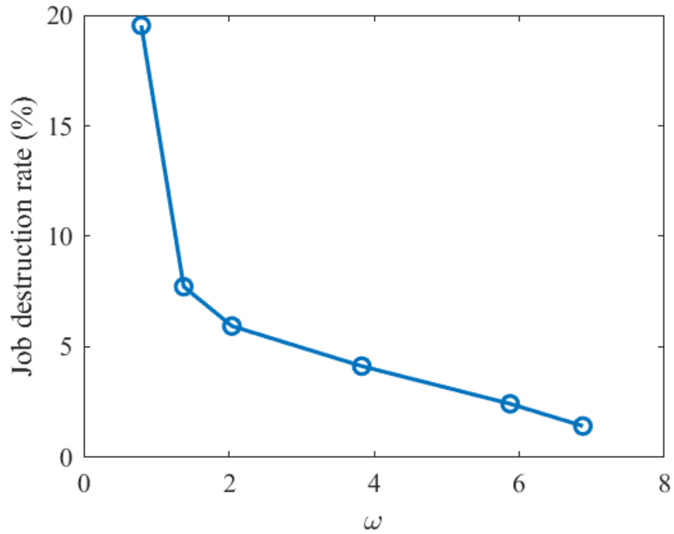



Figure 3: Job destruction rate in a cross-section of establishments, average of US Business Dynamic Statistics 1978-1982, all sectors. $\omega = -\log(1-\theta)$ to take into account the Pareto tail.

productivity can bring these firms to the left of $\underline{a}$ and hence force them to exit. 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 E.2 for proof.

#### 3.2.2 Entry

At any point in time, there are $M_e$ potential entrants or outsiders. Each of them invests in learning and enter the market if the learning is successful. If not successful, the opportunity to enter elapses and a new set of $M_e$ outsiders prepare 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. Learning cost is assumed to be quadratic with respect to the Poisson success rate:

$$R_e = \frac{1}{2}\alpha_e \lambda_e^2 A \tag{20}$$

where $\lambda_e$ is the success rate. If 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 $\phi$. The entrant enters at position $a_\psi \in (\underline{a}, a_j]$ with probability

$$\psi(a_\psi; \underline{a}, a_j) = k_\psi e^{-k_\psi (a_\psi - \underline{a})} + e^{-k_\psi (a_j - \underline{a})} \delta_{a_j}(a_\psi) \tag{21}$$

where $\delta_{a_j}$ is the Dirac mass function at point $a_j$ and $k_\psi > 0$ suggests imperfect learning. Figure 4 illustrates two matchings, each with a different incumbent. Section 3.2.3 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 initial position $\underline{a}$.

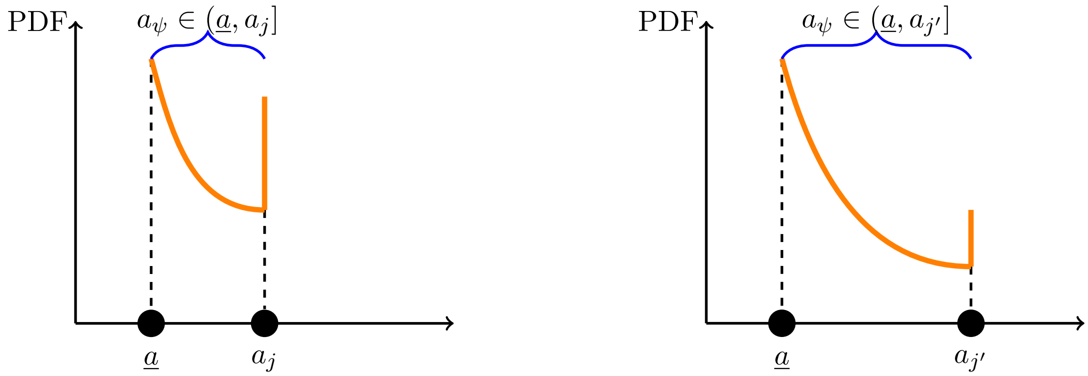



Figure 4: Illustration of an entrant being matched with an incumbent $a_j$ or $a_j'$. Vertical line at $a_j$ or $a_j'$ illustrates the Dirac mass.

Taking into account all possible matchings, entry occurs at position $a_\psi$ with probability distribution:

$$\psi_e(a_\psi; \phi) = \int_{\underline{a}}^{+\infty} \psi(a_\psi; \underline{a}, a_j)\phi(a_j)\, \mathrm{d}a_j \tag{22}$$

$$= [k_\psi \overline{\Phi}(a_\psi) + \phi(a_\psi)]e^{-k_\psi(a_\psi - \underline{a})} \tag{23}$$

where $\overline{\Phi}(a_\psi) = 1 - \Phi(a_\psi) = \mathbb{P}(a_\theta > a_\psi)$ is the survival function of the log productivity distribution.

### 3.2.3 Properties of the Entry Specification

The functional form of 21 looks strange at first sight but is actually very natural. It is simply a succinct description of step-by-step learning under the continuous setting. Consider its discrete counterpart illustrated in Figure 5 to appreciate this point. [32] 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, 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

$$p^n(1-p) = e^{-k_\psi(a_\psi - \underline{a})}(1 - e^{-\frac{1}{N}k_\psi(a_j - \underline{a})}) \tag{24}$$

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

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

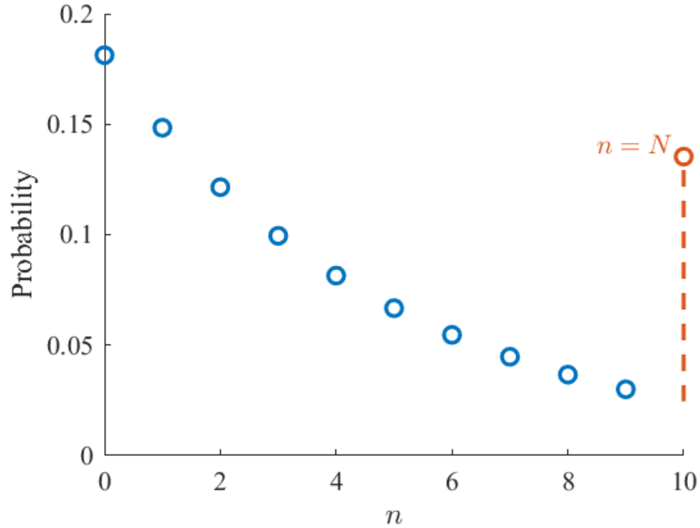



Figure 5: Discrete counterpart of an entrant learning from $a_j$. For illustration purposes with $\underline{a} = -1$, $k_\psi = 1$, $a_j = 1$ and $N = 10$.

which is just the continuous part of 21 times the size of step. The Dirac mass of 21 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.

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 21. 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 $\phi$ if $k_\psi = 0$ and thinner tail if $k_\psi > 0$. See Figure 6 of an illustration of this example. Thinking about the evolution of productivity distribution over time, during any time period d$t$, a total mass of $M_e\lambda_e\mathrm{d}t$ entrants with PDF $\psi_e$ are injected into the market, so that the incumbent distribution $\phi$ 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 sufficiently heavy tail, the injected distribution keeps interfering with the incumbent distribution, preventing the later from evolving to a lighter tail. The model then features multiple equilibria with the equilibrium distribution depending on the initial condition of $\phi$ like in Lucas and Moll (2014). Adding the Brownian motion as in Luttmer (2012) and Perla et al. (2021) does not essentially

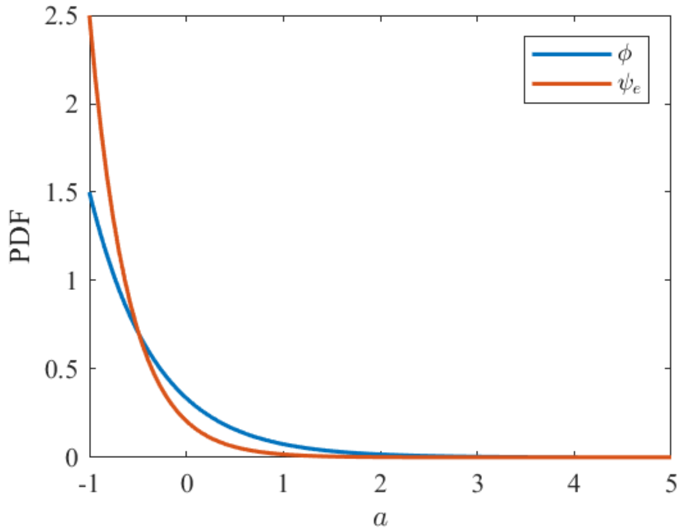



Figure 6: Illustration of non-interference condition of entrants with $\phi(a) = k_\phi e^{-k_\phi(a-\underline{a})}$, $\underline{a} = -1$, $k_\phi = 1.5$ and $k_\psi = 1$

make the equilibrium distribution unique, as both papers requires the initial condition to be thin-tailed enough to avoid the interference. [33] On the other hand, if $k_\psi > 0$, entrant distribution has a lighter tail than the incumbent distribution, making sure that it does not interfere with the tail of the later 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 in empirics.

In fact, the non-interference condition under $k_\psi > 0$ is valid for virtually any distribution $\phi$ one can think of with infinite right endpoint, proved in the following proposition. It will be 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 $\phi$.[34]

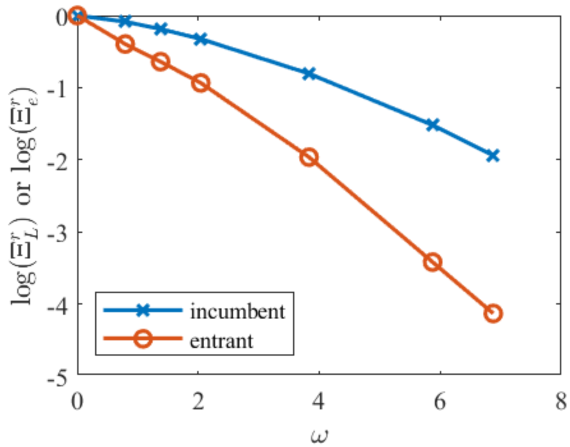
**Proposition 3** (Non-interference Condition of Entry)**.** *If $\phi$ is ultimately monotone (i.e. $\phi$ is monotone on $[\hat{a}, +\infty)$ for some $\hat{a}$), then $\forall k_s \in (-\infty, k_\psi)$, $\lim_{a_\psi \to +\infty} \frac{\psi_e(a_\psi;\phi)}{\phi(a_\psi)} e^{k_s a_\psi} = 0$.*
*In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

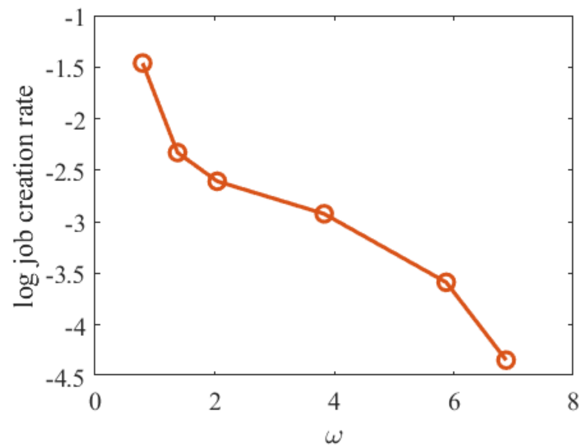
$$\frac{\psi_e(a_\psi;\phi)}{\phi(a_\psi)} e^{k_s a_\psi} = \left[k_\psi \frac{\overline{\Phi}(a_\psi)}{\phi(a_\psi)} + 1\right] e^{-(k_\psi - k_s)a_\psi} e^{k_\psi \underline{a}} \tag{26}$$

[33] 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.

[34] Luttmer (2020) discusses how bounded learning can ensure uniqueness of solution to the KF equation. In the present model, this means $\psi_e(a_\psi;\phi) = 0$ if $a_\psi$ 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".




(a)




(b)

Figure 7: 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 $\omega$ is used for incumbents and entrants of the same bin. Right panel: log job creation rate as a function of $\omega$. Job creation rate per bin is number of new jobs created by entrants divided by number of jobs employed by incumbents for the bin. Data source: Average of US Business Dynamics Statistics 1978-1982, all sectors.

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 D for the proof. $\square$

The ultimate-monotonicity condition is a very weak condition whose violation virtually never appears in practice. [35] In addition to its theoretical convenience, the assumption of $k_\psi > 0$ is also empirically plausible: it simply says that learning has a positive probability of failure, no matter how small it is. US BDS data reports number of establishments, incumbent employment and job creation from entrants by 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 values of $\omega$. A flatter line suggests heavier tail.

Figure 7a shows that entrants indeed exhibit lighter tail than incumbents. Another way to see this is to look at job creation rate from birth by size bin of establishments. If the exponential specification of in equation 22 is true and $k_\psi > 0$, then log job creation rate from birth should be a linear and decreasing function of $\omega$ at the right tail. Figure 7b shows this is indeed the case.

[35] Moreover, it is probable that any solution of the KF equation specified below should be ultimately monotone, even though a rigorous proof is wanting.

## 3.3 Evolution of incumbents

Apart from idiosyncratic shocks, incumbents can consciously improve their productivities via innovation and learning. Innovation takes place on a stand-alone basis, while learning happens when a firm is matched with another more productive firm. At time $t$, a firm chooses only innovation or learning based on whichever yields higher expected value gain.

### 3.3.1 Innovation and learning

A firm with log-productivity $a_\theta$ can conduct innovation on a stand-alone basis and improves its log productivity by $q > 0$ by incurring a research cost $R_{i,\theta}$. Figure 8a illustrates the change in log productivity after a successful innovation. Denote the destination PDF of the improvement as

$$\psi_i(a_\psi; a_\theta) = \delta_{a_\theta+q}(a_\psi) \tag{27}$$

where $\delta_{a_\theta+q}$ is the Dirac mass function at $a_\theta + q$. This slightly advanced notation is harmonious with the one introduced in learning below. By construction, the improvement is gradual and allows the firm to slowly improve its productivity. Denote $\lambda_{i,\theta}$ the Poisson rate of innovation success and assume the innovation cost to be quadratic in it:

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

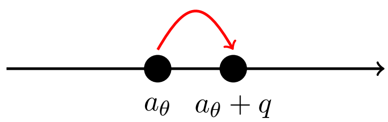
where $\tilde{a}_\theta = a_\theta - a$ is relative log TFP of firm $\theta$ compared to the aggregate economy. Unless otherwise stated, a tilde will always denote normalization by the aggregate TFP.[36] $\beta > 0$ captures the fact that ideas are harder to find at higher levels of productivity (Jones (1995), Bloom et al. (2020)). It governs heterogeneous research difficulties in a cross section of firms at a specific time, while $\alpha$ governs the general difficulty for all firms over time. During each period, $\alpha$ 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 higher value of $\alpha$ in the more recent period to reflect harder research.

Learning improves a firm's productivity in a quicker fashion. It captures learning-by-doing (Arrow (1962)), technology diffusion (Aghion et al. (2001)), technology adoption (Comin and Gertler (2006)) or imitation (König et al. (2022)). Denote $\lambda_{l,\theta}$ the Poisson rate of learning success and assume the learning cost $R_{l,\theta}$ to be the same quadratic function as 28:

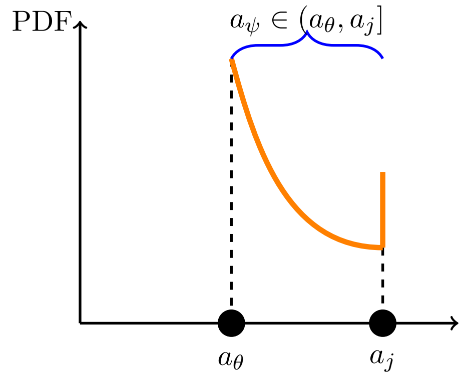
$$R_{l,\theta} = R(\lambda_{l,\theta}; \tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta\tilde{a}_\theta}\lambda_{l,\theta}^2 A \tag{29}$$

See Appendix C for a micro foundation of the adjustment factor $e^{\beta\tilde{a}_\theta}$ in the learning cost function. If learning is successfully realized with Poisson rate $\lambda_{l,\theta}$, the firm's log productivity jumps to a

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




(a) Innovation




(b) Learning (one matching)

Figure 8: Illustration of incumbent innovation and learning.

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 $a_\theta$ is randomly matched with an incumbent firm with log-productivity $a_j$. If $a_j \leq a_\theta$, no learning occurs. If $a_j > a_\theta$, then firm $\theta$ jumps to position $a_\psi \in [a_\theta, a_j]$ according to the PDF $\psi(a_\psi; a_\theta, a_j)$:

$$\psi(a_\psi; a_\theta, a_j) = k_\psi e^{-k_\psi(a_\psi - a_\theta)} + e^{-k_\psi(a_j - a_\theta)} \delta_{a_j}(a_\psi) \tag{30}$$

Figure 8b illustrates such matching with a more productive firm. The only difference from entry is that the firm now starts from position $a_\theta$, instead of from $\underline{a}$. Integrating over all possible draws of $a_j$ to get the learning outcome distribution $\psi_l$:

$$\psi_l(a_\psi; a_\theta, \phi) = \int_{a_\theta}^{+\infty} \psi(a_\psi; a_\theta, a_j)\phi(a_j)\,\mathrm{d}a_j \tag{31}$$

$$= [k_\psi \overline{\Phi}(a_\psi) + \phi(a_\psi)] e^{-k_\psi(a_\psi - a_\theta)}, \quad \forall a_\psi > a_\theta \tag{32}$$

I have included $a_\theta$ and $\phi$ in the parentheses of $\psi_l$ 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 $\delta_{a_\theta}$ is the Dirac mass function at point $a_\theta$. The latter comes from the fact that all draws from the left of $a_\theta$ result in firm $\theta$ staying at $a_\theta$. Taken together,

$$\psi_l(a_\psi; a_\theta, \phi) = [k_\psi \overline{\Phi}(a_\psi) + \phi(a_\psi)] e^{-k_\psi(a_\psi - a_\theta)} \mathbb{1}_{(a_\theta, +\infty)}(a_\psi) + \Phi(a_\theta)\delta_{a_\theta}(a_\psi), \quad \forall a_\psi \geq \underline{a} \tag{33}$$

where $\mathbb{1}$ is the indicator function. It is easy to check that 33 integrates into 1 so that it is a well-defined PDF.

If the log-productivity distribution $\phi$ 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 33 also makes clear the drawback of learning despite its quick speed: a productive firm finds it difficult to be matched with another one even more productive . The probability of such 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 the cost functions 28 and 29 are the same, the firm chooses the strategy with higher expected improvement of the firm's value. [37]Based on the setting, a cutoff feature can be envisaged: More productive firms choose innovation and less productive firms choose learning. This will be apparent in the numerical results.

Taken together, firm $\theta$'s log productivity $a_\theta$ evolves according to the following stochastic process:

$$\mathrm{d}a_{\theta,t} = -\frac{1}{2}\nu^2 \mathrm{d}t + \nu \mathrm{d}B_{\theta,t} + \mathrm{d}J_{\theta,t} \tag{34}$$

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, $B_{\theta,t}$ is Brownian motion independent across $\theta$. $J_{\theta,t}$ is a jump term and consists of two parts: death and innovation/learning. If death is realized with Poisson intensity $\lambda_{d,\theta,t}$, then $a_{\theta,t}$ jumps directly to $-\infty$ and firm $\theta$ exits. If innovation/learning is realized with intensity $\lambda_{c,\theta,t}$, where $c \in \{i, l\}$ to denote innovation/learning, then the firm jumps according to PDF $\psi_{c,\theta,t}$. The firm chooses either innovation or learning at time $t$ based on expected value gains.

### 3.3.2 Properties of innovation and learning

As will be apparent in numerical results, innovation intensity features an inverted-U relationship with log-productivity if all firms are forced to choose innovation, which is typically found in Schumpeterian growth models. Since per-period profit net of fixed cost is non-negative, value function has a lower bound of 0 and is extremely flat for lowest productivities. Consequently, marginal value function is close to 0 for the least productive firms so that they have little incentive to innovate. [38] Thus laggards are disadvantaged not only in the static sense as laggards but also in the dynamic sense with lower growth, implying that an equilibrium may not exist. To allow an equilibrium, Schumpeterian growth models following Aghion et al. (2001) and Aghion et al. (2005) often introduce an exogenous technological diffusion in addition to innovation. In the present model, the diffusion is endogenized as active learning.

[37]Despite the same cost function, $q$ and $k_\psi$ serve to adjust the efficiency of innovation versus learning.

[38]Even though a positive $\beta$ make innovation easier for the least productive firms, the effect is not sufficiently large to overturn the near-zero marginal value effect as long as $\beta$ is not too high. When $\beta$ is calibrated to micro-data, inverted-U remains, testifying the robustness of Aghion et al. (2005)'s insight.

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 $a_\psi$ due to learning:

$$\psi_l^{\text{in}}(a_\psi; \phi) \triangleq \int_{(\underline{a}, a_\psi)} \psi_l(a_\psi; a_\theta, \phi)\phi(a_\theta)\lambda_l(a_\theta)\,\mathrm{d}a_\theta \tag{35}$$

**Proposition 4** (Non-interference Condition of Incumbent Learning)**.**
*If $\phi$ 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 $\lim_{a_\psi \to +\infty} \frac{\psi_l^{in}(a_\psi;\phi)}{\phi(a_\psi)} e^{k_s x} = 0$, $\forall k_s < \min\{k', k_\psi\}$.*
*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 $k_\psi$.
2. In 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, is a specially 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 $\min\{k', k_\psi\}$.

See Appendix D for the proof. The condition that $\lambda_l$ decreases at least exponentially will be established when proving Proposition 5.

### 3.4 Innovation/Learning Decisions and Evolution of Productivity Distribution

Given the building blocks in previous sections, each incumbent firm solves its optimal innovation/learning 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:

$$\begin{aligned} r_t V(a_\theta, t) = \max_{c,\,\lambda_c} \Big\{ &\Pi(a_\theta; \phi_t^M) - f A_t - R_t(\lambda_c; \tilde{a}_\theta) - \frac{1}{2}\nu^2 \frac{\partial V}{\partial a}(a_\theta, t) + \frac{1}{2}\nu^2 \frac{\partial^2 V}{\partial a^2}(a_\theta, t) \\ &+ \lambda_c \int_{\mathbb{R}} \big[V(a_\psi, t) - V(a_\theta, t)\big] \psi_c(a_\psi; a_\theta, \phi_t)\,\mathrm{d}a_\psi - \lambda_d(a_\theta; \phi_t^M) V(a_\theta, t) + \frac{\partial V}{\partial t}(a_\theta, t) \Big\} \end{aligned} \tag{36}$$

where

$$R_t(\lambda_c; \tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta \tilde{a}_\theta} \lambda_c^2 A_t, \tag{37}$$

$r_t$ is interest rate, $c(a_\theta, t) \in \{i, l\}$ is the choice between innovation $i$ and learning $l$ and I have used the shorthand $c$ to simplify the notation, $\phi_t^M = M_t \phi_t$ is generalized PDF of log productivity at time $t$, $\Pi$ is per-period production profit which depends on the firm's log-productivity $a_\theta$ and the market structure $\phi_t^M$, $\lambda_d(a_\theta; \phi_t^M)$ is death rate defined in equation 19, and $\psi_c(a_\psi; a_\theta, \phi_t)$ is the destination PDF of innovation ($c = i$) or learning ($c = l$) for firm $\theta$. In order to make the notation succinct, I have slightly abused the notation and used $\psi_i(a_\psi; a_\theta, \phi_t)$ to represent $\psi_i(a_\psi; a_\theta)$, even though $\psi_i$ does not depend on $\phi_t$. See Appendix E.3 for derivations of the HJB equation.

Equation 36 has an intuitive interpretation: per period flow value should be equal to the sum of current-period profit net of fixed cost and innovation/learning 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/learning, plus a decrease in value due to death and plus a change of value function over time.

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

$$c^*(a_\theta, t) = \begin{cases} i, & \text{if } \int_{\mathbb{R}} V(a_\psi, t)\psi_i(a_\psi; a_\theta)\,\mathrm{d}a_\psi \geq \int_{\mathbb{R}} V(a_\psi, t)\psi_l(a_\psi; a_\theta, \phi_t)\,\mathrm{d}a_\psi \\ l, & \text{otherwise.} \end{cases} \tag{38}$$

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

$$\lambda_c^*(a_\theta, t) = \frac{1}{\alpha A_t} e^{-\beta \tilde{a}_\theta} \int_{\mathbb{R}} \big[V(a_\psi, t) - V(a_\theta, t)\big]\psi_c(a_\psi; a_\theta, \phi_t)\,\mathrm{d}a_\psi \tag{39}$$

I have already introduced the left boundary $\underline{a}_t$ below which a firm dies immediately, i.e. "absorbing barrier" on the left. Solving the HJB 36 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 "reflecting barrier", i.e. $a_{\theta,t}$ is brought back to $\overline{a}_t$ if it goes above $\overline{a}_t$. 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. Formally, the two boundary conditions are:

$$V(\underline{a}_t, t) = 0 \tag{40}$$

$$\frac{\partial V}{\partial a}(\overline{a}_t, t) = 0 \tag{41}$$

For entry, each potential entrant solves the maximisation problem:

$$\max_{\lambda_e} \Big\{ -R_{e,t}(\lambda_e) + \lambda_e \int_{\mathbb{R}} V(a_\psi, t)\psi_e(a_\psi; \phi_t)\,\mathrm{d}a_\psi \Big\} \tag{42}$$

where

$$R_{e,t}(\lambda_e) = \frac{1}{2}\alpha_e\lambda_e^2 A_t \tag{43}$$

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

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

$$\lambda_e^*(t) = \frac{1}{\alpha_e A_t}\int_{\mathbb{R}} V(a_\psi, t)\psi_e(a_\psi;\phi_t)\,\mathrm{d}a_\psi \tag{44}$$

Given the choice of $c$, $\lambda_c$ and $\lambda_e$, the Kolmogorov Forward (KF) Equation describes the evolution of the log-productivity distribution $\phi^M(a_\theta, t)$:

$$\begin{aligned}\frac{\partial\phi^M}{\partial t}(a_\theta,t) = &-\frac{\partial[-\frac{1}{2}\nu^2\phi^M]}{\partial a}(a_\theta,t) + \frac{1}{2}\nu^2\frac{\partial^2\phi^M}{\partial a^2}(a_\theta,t) + \lambda_e(t)M_e\psi_e(a_\theta;\phi_t) - \lambda_d(a_\theta;\phi_t^M)\phi^M(a_\theta,t)\\ &+\int_{(-\infty,a_\theta)}\lambda_c(a_\psi,t)\phi^M(a_\psi,t)\psi_c(a_\theta;a_\psi,\phi_t)\,\mathrm{d}a_\psi - \phi^M(a_\theta,t)\lambda_c(a_\theta,t)\int_{(a_\theta,+\infty)}\psi_c(a_\psi;a_\theta,\phi_t)\,\mathrm{d}a_\psi\end{aligned} \tag{45}$$

See Appendix E.4 for a proof of KFE based on the law of motion of firms. The terms have intuitive interpretations: apart from the usual terms about 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 [39], and the integrals describe the change of distribution due to innovation/learning. The first integral tells the inflow of firms into position $a_\theta$ and the second one tells the outflow of firms from position $a_\theta$. For the outflow, $\phi^M(a_\theta,t)$ firms each with innovation/learning intensity $\lambda_c(a_\theta,t)$ goes up if $\lambda_c(a_\theta,t)$ is realized with a good draw. The probability of having such a good draw is $\int_{(a_\theta,+\infty)}\psi_c(a_\psi;a_\theta,\phi_t)\,\mathrm{d}a_\psi$. The inflow can be similarly analysed by considering jumps from positions $a_\psi < a_\theta$ into $a_\theta$. I have integrated over $(-\infty,a_\theta)$ for inflows and $(a_\theta,+\infty)$ for outflows to emphasize inflow from below and outflow towards above. It is easy to check that they can be equally written as both over $\mathbb{R}$. Two boundary conditions, one for the left "absorbing barrier" and another for the right "reflecting barrier", are:

$$\phi^M(\underline{a}_t, t) = 0 \tag{46}$$

$$\frac{1}{2}\nu^2\phi^M(\overline{a}_t,t) + \frac{1}{2}\nu^2\frac{\partial\phi^M}{\partial a}(\overline{a}_t,t) = 0^{40} \tag{47}$$

For notational elegance and computational simplicity, it is useful to define a functional operator $\mathcal{F}_t$

[39] The second type of firm death near the left boundary is incorporated into the boundary condition.
[40] I have not removed $\frac{1}{2}\nu^2$ to make it consistent with the general form in the case of reflecting barriers.

and write both the HJB and the KF in terms of it:

$$\mathcal{F}_t V(a_\theta, t) = -\frac{1}{2}\nu^2 \frac{\partial V}{\partial a}(a_\theta, t) + \frac{1}{2}\nu^2 \frac{\partial^2 V}{\partial a^2}(a_\theta, t) + \lambda_c(a_\theta, t) \int_{\mathbb{R}} \big[V(a_\psi, t) - V(a_\theta, t)\big] \psi_c(a_\psi; a_\theta, \phi_t) \, \mathrm{d}a_\psi \\ - \lambda_d(a_\theta; \phi_t^M) V(a_\theta, t) \quad (48)$$

The HJB 36 and the KF 45 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 E.5 for the proof of adjoint-ness. 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 adjoint-ness.

### 3.5 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/learning cost parameter $\alpha$, and entry cost parameter $\alpha_e$ are kept constant. 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:*

$$\tilde{\mathcal{F}} \tilde{V} = \Big[-\frac{1}{2}\nu^2 - g\Big] \tilde{V}'(\tilde{a}_\theta) + \frac{1}{2}\nu^2 \tilde{V}''(\tilde{a}_\theta) + \lambda_c(\tilde{a}_\theta) \int_{\mathbb{R}} \big[\tilde{V}(\tilde{a}_\psi) - \tilde{V}\big] \psi_c(\tilde{a}_\psi; \tilde{a}_\theta, \tilde{\phi}) \, \mathrm{d}\tilde{a}_\psi - \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M) \tilde{V} \quad (49)$$

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

$$(r - g)\tilde{V} = \max_{c, \lambda_c} \Big\{ \Pi(\tilde{a}_\theta; \tilde{\phi}^M) - f - \tilde{R}(\lambda_c; \tilde{a}_\theta) + \tilde{\mathcal{F}} \tilde{V} \Big\} \quad (50)$$

*with boundary conditions*

$$\tilde{V}(\underline{\tilde{a}}) = 0 \tag{51}$$

$$\tilde{V}'(\bar{\tilde{a}}) = 0, \tag{52}$$

*where*

$$\tilde{R}(\lambda_c; \tilde{a}_\theta) = \frac{1}{2}\alpha e^{\beta \tilde{a}_\theta} \lambda_c^2. \tag{53}$$

*Entry decision is:*

$$\max_{\lambda_e} \Big\{ -\tilde{C}_e(\lambda_e) + \lambda_e \int_{\mathbb{R}} \tilde{V}(\tilde{a}_\psi)\psi_e(\tilde{a}_\psi; \tilde{\phi})\,\mathrm{d}\tilde{a}_\psi \Big\} \tag{54}$$

*where*

$$\tilde{C}_e(\lambda_e) = \frac{1}{2}\alpha_e \lambda_e^2. \tag{55}$$

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

$$0 = \tilde{\mathcal{F}}^\intercal \tilde{\phi}(\tilde{a}_\theta) + \frac{\lambda_e M_e}{M}\psi_e(\tilde{a}_\theta; \tilde{\phi}), \tag{56}$$

*with boundary conditions*

$$\tilde{\phi}(\underline{\tilde{a}}) = 0, \tag{57}$$

$$\frac{1}{2}\nu^2 \tilde{\phi}(\bar{\tilde{a}}) + \frac{1}{2}\nu^2 \tilde{\phi}'(\bar{\tilde{a}}) = 0, \tag{58}$$

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

$$\lambda_e M_e = M\Big[\int_{\mathbb{R}} \lambda_d(\tilde{a}_\theta; \tilde{\phi}^M)\tilde{\phi}(\tilde{a}_\theta)\,\mathrm{d}\tilde{a}_\theta + \frac{\nu^2}{2}\tilde{\phi}'(\underline{\tilde{a}})\Big]. \tag{59}$$

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

$$\Pi(\underline{\tilde{a}}; \tilde{\phi}^M) = f. \tag{60}$$

*The travelling wave equilibrium (i.e. balanced growth path) is defined by equations 49-60, and aggregate growth* $g$ *which, after solving the KF equation, normalizes aggregate productivity* $A$ *to* 1.

See Appendix E.7 for the proof of normalizations and Appendix G for 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 appears in HJB 50, 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 $-g$ terms comes from the fact we are now reasoning in term of the normalized log-productivity, whose stochastic process features an additional $-g dt$ term compared to equation 34.

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, employment and output are all Pareto-tailed with unique Pareto index. Moreover, there is a one-to-one correspondence between lower aggregate growth and fatter Pareto tail, regardless of market power.

**Proposition 5** (Unique Tail of Equilibrium Distribution)**.**
*If the following conditions are satisfied:*

1. $k_\psi > 0$,
2. $\beta > \underline{\sigma} - 1$,
3. $V(\tilde{a}_\theta)$ *is non-decreasing and* $\limsup_{a \to +\infty} \frac{\log(V(a))}{a} \neq \beta$,

*then*

1. *The solution to the Kolmogorov forward equation 56, if it exists, has an exponential tail with unique tail index* $-\xi$*, where* $\xi$ *is the unique negative root of*

$$-\frac{1}{2}\nu^2\xi^2 - (\frac{1}{2}\nu^2 + g)\xi + \lambda_d(\infty) = 0 \tag{61}$$

   *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 order words, the equilibrium productivity distribution has unique Pareto tail index of* $-\xi$*. Moreover, equilibrium sales and employment distributions have unique Pareto tail index of* $-\xi/(\underline{\sigma}-1)$*, and output distribution has unique Pareto tail index of* $-\xi/\underline{\sigma}$*.*

2. $-\xi$ *is strictly increasing in* $g$*, i.e. lower growth is associated with fatter Pareto tails.* [41]

See Appendix F for the proof. Non-interference conditions under the imperfect learning assumption $k_\psi > 0$ (Propositions 3 and 4) are key for ensuring the uniqueness, as the proof reveals. [42] In the calibrated model, a uniform increase of research difficulty $\alpha$ across all firms decreases aggregate growth and Pareto-tail indices, i.e. tails become fatter. I will explain the detailed mechanism when presenting comparative statics with numerical results in Section 5.

For the period before 1980, we will consider economic integration as an additional force to harder research. The following proposition shows that expanding the market size $L$ can annihilate harder research effect:

[41] As a reminder, a smaller Pareto tail index means fatter tail.

[42] Note that Proposition 5 only ensures 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 and requires future work. Moreover, the existence of a solution to KF equation requires future work, though the numerical algorithm has consistently found solutions.

**Proposition 6** (Indistinguishable Economies)**.** *An economy with labor $L$, fixed cost $f$, innovation/learning cost parameter $\alpha$ and entry cost parameter $\alpha_e$ 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 E.8 for a proof by verification. [43] Entry cost parameter $\alpha_e$ will be calibrated to $\alpha e^{\beta \underline{\tilde{a}}}$ to be consistent with incumbent learning cost, so that its change is tied to that of research difficulty $\alpha$. 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 and learning. While the proposition assumes an equal change in fixed cost to yield a clean result, fixed cost can remain constant like in international trade literature and firms benefit from economies of scale when market size increases. The pro-growth effect of larger market size is reminiscent of Chandler (1990)'s argument that firms expand their geographical reach as a growth strategy. Nevertheless, economic integration involves not only an increase in market size $L$ but also an increase in the measure of firms $M$. The later is pinned down from the balanced entry/exit condition with an increase in the potential number of entrant $M_e$. The implications of economic integration in addition to harder research will be discussed in Section 6.

# 4 Calibration

Comparative statics in Section 5 will be between 1980 and 2020, and in Section 6 between 1960 and 1980. The choice of 1980 as the watershed is incentivised 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 of the literature is partly driven by data availability, as US Business Dynamics Statistics (BDS) is only available since 1978, and partly by data patterns, as declining labor share and increasing markup are most apparent since 1980s. Before 1980, Labor share according to US Bureau of Economic Analysis (BEA) had no apparent trend, both sales-weighted markup (De Loecker et al. (2020)) and cost-weighted markup (Edmond et al. (forthcoming)) had similar value 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 measured by Kwon et al. (2022) is only available since that date. [44] Moreover, US experienced high inflation before Volcker's disinflation policies around 1980, raising the question of what was the relevant interest rate to use before 1980 for discounting cash flows. Despite the increase of measured real interest rate from 1970s to 1980s, natural interest rate declined persistently from 1960 onwards (Holston et al. (2017)). As the present paper takes a secular perspective, it views the decline in natural interest rate from 1960 to 1980 as the change in firms' discount rate during the

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

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

| 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 |
| $\alpha_{t=1980}/\alpha_{t=1960}$ | Change of research cost | 7.93 | Estimated from US aggregate |
| $\alpha_{t=2020}/\alpha_{t=1980}$ | Change of research cost | 6.84 | Estimated from US aggregate |
| $\beta$ | Harder research for higher TFP | 1.6 | Estimated from Compustat |
| $r_{t=1960}$ | Interest rate | 0.0669 | External |
| $r_{t=1980}$ | Interest rate | 0.0469 | External |
| $r_{t=2020}$ | Interest rate | 0.0109 | External |
| $L_{t=1980}$, $L_{t=2020}$ | 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 | 543 | Structural Estimation |
| $q$ | Innovation step | 0.166 | Structural Estimation |
| $\nu$ | std of idiosyncratic shocks | 0.114 | Structural Estimation |
| $\kappa/\nu$ | Relative death threshold | 3.03 | Structural Estimation |
| $M_{e,t=1980}$, $M_{e,t=2020}$ | Mass of potential entrants | 0.89 | Structural Estimation |
| $f$ | fixed cost | 0.0573 | Structural Estimation |
| $k_{\psi}$ | Learning parameter | 1.25 | Structural Estimation |
| $L_{t=1980}/L_{t=1960}$ | Magnitude of economic integration | 3.74 | Structural Estimation (Section 6) |
| $M_{e,t=1980}/M_{e,t=1960}$ | Magnitude of economic integration | 3.74 | Structural Estimation (Section 6) |

Table 3: Summary of parameters. Parameters that can change over time (i.e. 1960, 1980 and 2020) are indexed by $t$.

period. Measured natural interest rate is only available since 1960, which again restricts the study to the post-1960 period.

Most parameters remain unchanged across time, except interest rate $r$, research cost parameter $\alpha$, and entry cost parameter $\alpha_e$ which is intimately tied to $\alpha$. Labor force $L$ and potential number of entrants $M_e$ are also allowed to change in the case of economic integration in Section 6. Table 3 summarizes all the parameters.

## 4.1 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. (2020), except that they use pass-through instead of labor share, and that 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. [45] A few adjustments of the labor shares are needed before the estimation: (1) The

[45] Using the 2012 vintage, another one in which the Census Bureau publishes the information, yields similar results.

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 excluding fringe benefits; (3) The model does not feature heterogeneous wages, so average payroll per employee of each size bin is adjusted to the sectoral average; (4) The model does not feature capital, so labor share in the data is adjusted to be consistent with the model. See Appendix H.1 for detailed discussions. [46]

The adjusted labor share is close to 1 for the smallest firms. As $\overline{\sigma} = \frac{1}{1-\overline{\chi}}$, where $\overline{\chi}$ is upper bound of labor share corresponding to the smallest firms, $\overline{\sigma}$ is sensitive to a small change in $\overline{\chi}$ when the later 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 preference 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. Denote $\xi_\theta$ the value-added of firm $\theta$ over sectoral value-added, i.e. firm $\theta$'s market share in terms of value-added. Denote $\chi_\theta$ the labor share of firm $\theta$. The data gives average labor share for a few size bins of establishments. Formally, for $j \in \{1, \cdots, n-1\}$, we have 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 $0 = \theta_1 < \theta_2 < \cdots < \theta_n = 1$.

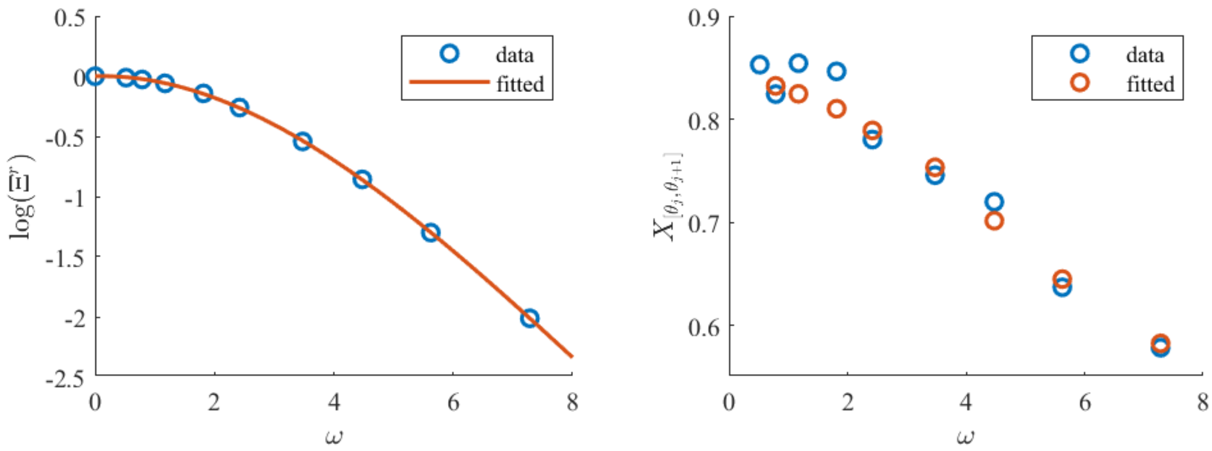
$\xi_\theta$ is recovered from value added across size bins. To take care of the Pareto tail, I fit cumulated value added share from right tail as a function of $\omega = -\log(1-\theta)$. Formally, denote $\Xi^r_\theta = \int_\theta^1 \xi_\theta \, \mathrm{d}\theta$ as the cumulated value added share from right tail. Figure 9a fits $\log(\Xi^r)$ as a 3-order polynomial of $\omega$.[47] $k$ is estimated to minimize the distance between $(X_{[\theta_1,\theta_2]}, X_{[\theta_2,\theta_3]}, \cdots, X_{[\theta_{n-1},\theta_n]})$ and its model counterpart, as shown in Figure 9b. It 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 firm size distribution. $k$ is estimated to be 0.32. See Appendix H.1 for details.

## 4.2 Harder research parameters

The research cost function 53 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 Bloom et al. (2020)'s

[46] See also Autor et al. (2020) for detailed discussions on the first two points. Figure 2 presents labor share after the first two adjustments.

[47] See Appendix H.1.1 on why the Pareto tail incentivise it. In particular, if valued added is exactly Pareto-distributed, $\log(\Xi^r)$ is a linear function of $\omega$.




(a) Log cumulative value added share

(b) Bin-wise labor share

Figure 9: Estimating $k$

methodology with decreasing returns for estimating research productivity, i.e.

$$\text{Research productivity} = \text{Growth}/\text{Effective number of researchers}^{\frac{1}{2}} \tag{62}$$

where effective number of researchers captures research cost and is measured as R&D expenditures deflated by average male wage with 4 years or more college education. With aggregate data, the decrease in research productivity over time reflects a general increase in $\alpha$ across all firms. $\alpha$ is estimated to increase by 7.93-fold from 1960 to 1980, and by 6.84-fold from 1980 to 2020. For $\beta$ which governs 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), $\beta$ is estimated as 1.6.

### 4.3 Interest rates and structurally estimated parameters

I follow Liu et al. (2022) to use US AA corporate bond rate net of current inflation for calibrating interest rates in 1980 and 2020, as it is more relevant than 10-year treasury rate as firms' discount rate. They are respectively 0.0469 for 1980 and 0.0109 for 2020. Changes in natural interest rate from 1960 to 1980 is used to calibrate the decrease of $r$ during the period 1960-1980. According to Holston et al. (2017), natural interest rate changes by about -0.02 from 1960 to 1980 so that I calibrate the interest rate in 1960 as 0.0669.

According to Proposition 6, we can normalize the labor force in 1980 as 1 so that other parameters can be identified. I set $\alpha_e$ to be $\alpha e^{\beta \underline{\tilde{a}}}$ so that entrants' learning cost function is consistent with that of incumbents'.

The remaining parameters are structurally estimated by targeting data moments in 1980. While the change of one parameter moves virtually all the moments, some moments are especially useful for

| 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% |
| $\nu$ | Job reallocation rate | 29% | 28% |
| $\kappa/\nu$ | Exit rate | 5.6% | 5.6% |
| $M_{e,t=1980}$ | Entry rate | 5.6% | 5.6% |
| $f$ | Bottom 55% firms exit rate | 20% | 20% |
| $k_\psi$ | Growth contribution of net entry | 0% | -0.08% |

Table 4: Estimated parameters for 1980 by moment matching

identifying specific parameters. The structurally estimated parameters and their main identifying moments are summarized in Table 4. $\alpha_{t=1980}$ targets average TFP growth of 0.95% in 1970s and 1980s according to BLS-KLEMS data, and $q$ targets 2.4% share of R&D expenditure over GDP in National Income and Product Accounts (NIPA) around 1980. Idiosyncratic shock is the predominant reason of incumbent's job creation and job destruction: $\nu$ is thus identified by the job reallocation rate of 29% around 1980 according to BDS. [48] The travelling wave equilibrium features balanced entry/exit, so I take the average of entry rate and exit rate around 1980 to be the targets for both moments. Entry rate is measured as job creation rate from birth and exit rate as job destruction rate from death. The average of them around 1980 is 5.6% according to BDS. By the definition of first kind of death 19, which is the predominant form of death, the threshold-volatility ratio $\kappa/\nu$ rather than $\kappa$ is what ultimately matters for death rate. Thus I use $\kappa/\nu$ to target death rate of 5.6%. Number of potential entrants $M_e$ targets entry rate of 5.6%. Per-period fixed cost $f$ determines the lower-bound productivity through the zero-profit condition 60, 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 dying firms' size distribution in BDS, the contribution of net entry to growth should be tiny from this point of view and for being consistent with the model. Without a better measure of the contribution, I estimate $k_\psi$ to match a 0 contribution, as $k_\psi$ shapes the entry distribution.[49]

The parameters are estimated using 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.

[48]See Appendix G.7 for computing job creation rate and job destruction rate.
[49]See Appendix G.8 for computing net entry contribution to growth.

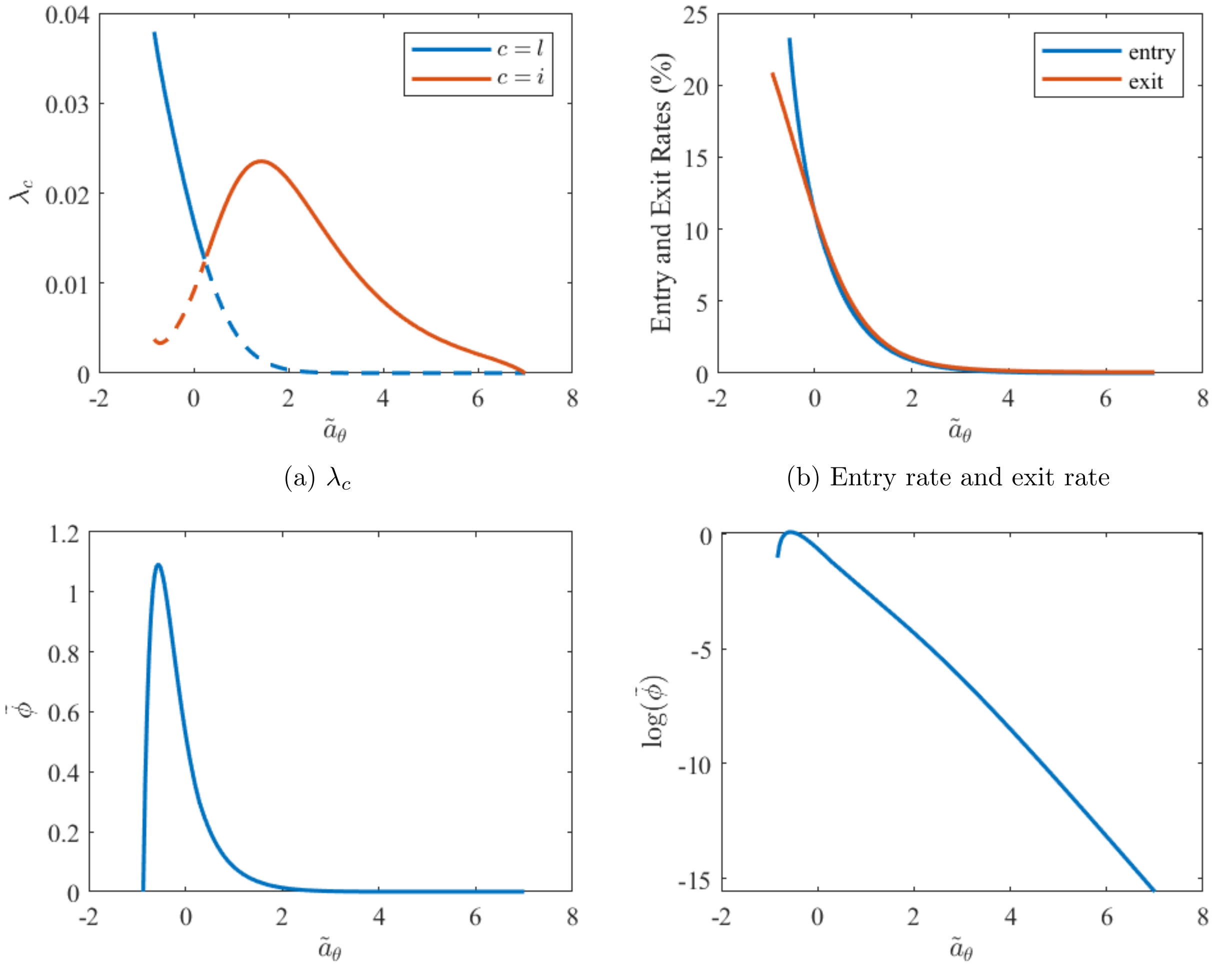
## 4.4 Traveling wave equilibrium of 1980

Figure 10 shows the traveling wave equilibrium of 1980 based on calibrated parameters. Figure 10a displays the optimal learning ($c = l$) and innovation ($c = i$) intensities of incumbent firms where solid lines are the actual choices of firms. There is a cutoff log-productivity above which firms choose innovation over learning and vice versa. Dotted lines are hypothetical optimal intensities if firms are forced to use a strategy (i.e. innovation or learning) when the strategy is not optimally chosen. The overall innovation intensity, i.e. hypothetical for laggards and actual for leaders, shows an inverted-U relationship with log-productivity, which is typical in the Schumpeterian growth literature. As elucidated by Aghion et al. (2005), innovation intensity depends not so much on post-innovation rents as on marginal rents, i.e. the difference between post-innovation rents and pre-innovation rents. Product market competition becomes more fierce as a firm moves down the productivity distribution. Competition reduces more the pre-innovation rent than it does for post-innovation rent when the firm is at the upper side of the productivity distribution, encouraging innovation aimed at "escaping competition". On the other hand, laggards have low levels of pre-innovation rent, so that competition reduces more the post-innovation rent. The "Schumpeterian effect" dominates the "escape competition effect" in this case, discouraging laggards' incentive to innovate. Combining the two cases, innovation intensity features an invert-U relationship along firm size distribution. In addition to these traditional forces, a positive $\beta$ in the present paper makes innovation difficulty increasing with productivity, which reinforces the mechanism at the upper end of the distribution. As long as $\beta$ is not very large, which is the case for its estimated value, it cannot revert the monotonicity at the lower end of the distribution and inverted-U remains robust.

The overall learning intensity is decreasing in firm's productivity, as productive firms find it difficult to be matched more productive firms, and as learning becomes more difficult with more harder knowledge. Since I assume higher cost for both learning and innovation of more productive firms, the optimal choice of innovation over learning of the last 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.

Firm's growth from either learning or innovation decreases with firm size, which is departure from Gibrat's law. The departure is however slight as only a small proportion of firms successfully realize learning or innovation at a specific time. I interpret these firms as high-growth firms when investigating the data and find them to more prevalent in laggards than in leaders (see Section 5). The great majority of firms are subjected to idiosyncratic shocks at any specific time so that their average growth is insensitive to size. [50]

[50] Departure from Gibrat's law with smaller firms growing more rapidly is consistent with empirical evidence in Akcigit and Kerr (2018).




(a) $\lambda_c$

(b) Entry rate and exit rate

(c) $\tilde{\phi}$

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

Figure 10: Travelling wave equilibrium of 1980

Notes: $\tilde{a}_\theta$ is 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 choice between learning and innovation, as well as their intensity; dotted lines means the growth strategy ($l$ or $i$) is not chosen but show the intensity had the firm being forced to use the strategy. Panel (b): entry rate $\frac{\lambda_e M_e \psi_e}{M \tilde{\phi}}$ and exit rate $\lambda_d$ along the distribution of firms. Entry rate close to $\underline{\tilde{a}}$ is now 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

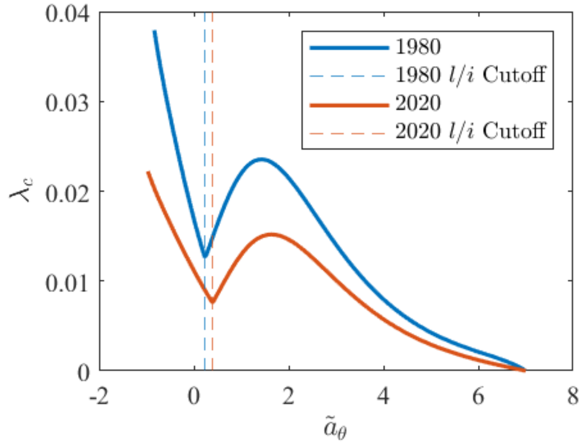
Because the model assumes imperfect learning, entry rate declines exponentially with log-productivity along the distribution in Figure 10b. Exit rate declines in a similar way, as demand elasticity declines with firm size: to knock out a larger firm by sales shock, a larger productivity shock is needed, whose probability of occurrence is lower. Both entry rate and exit rate closely resemble the pattern in the data.

Figure 10c 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 10d plots the distribution again in log scale, with the right tail being linear. The exponential tail of log-productivity distribution means Pareto tail of productivity distribution, which then translates into Pareto tails of output, 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 5. 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 means that within-firm markup decrease dominates between-firm reallocation of market share towards high markup firms in the case of market size expansion. [51] 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 compensates higher markup of fatter tail, as will be discussed in Section 6.

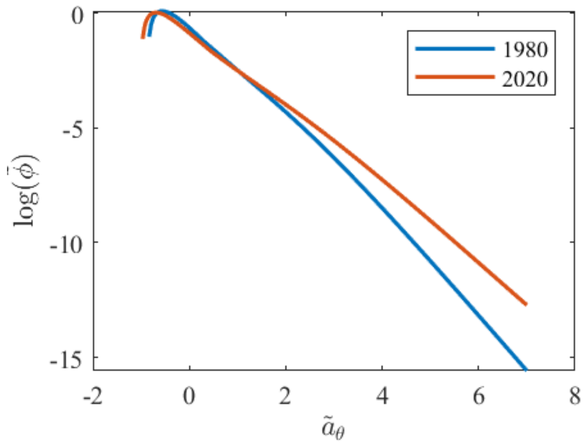
# 5 Comparative statics when ideas are getting harder to find

What happens to growth, concentration, market power, labor share and business dynamics when ideas get harder to find? As explained in Section 4, I calibrate a 6.84-fold increase of $\alpha$ from 1980 to 2020 using Bloom et al. (2020)'s methodology. Entrants' learning cost parameter $\alpha_e$ increases accordingly due to its tight relationship with $\alpha$. 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 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 growth rate. As interest rate can decline for reasons other than growth rate, for instance the global saving glut

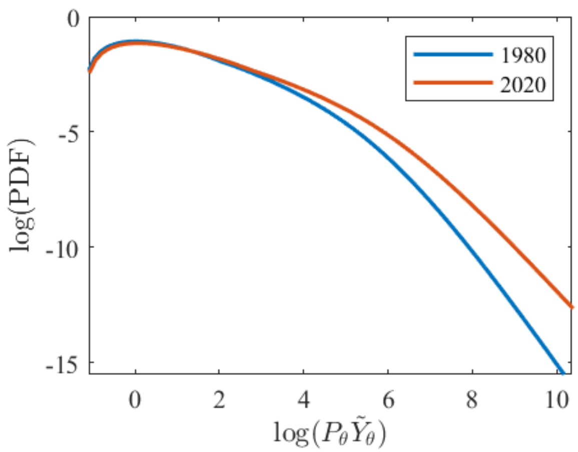
[51] If the log-productivity distribution is log linear, i.e. the productivity distribution is exactly Pareto, then the two forces exactly cancel so that aggregate markup does not change. If it is log convex, then between-firm reallocation dominates so that markup increases in a market size expansion. The results hold for a wide range of demand systems with variable demand elasticities, see the appendix of Autor et al. (2020). To the best of my knowledge, I have not seen any log convex assumption in 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 typically assumed distributions in static models.



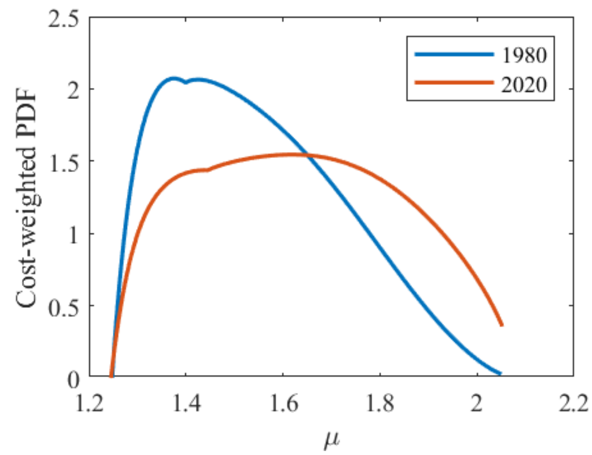

(a) $\lambda_c$




(b) $\log(\phi)$




(c) log(PDF) of $\log(P_\theta Y_\theta)$




(d) Markup $\mu$

Figure 11: Comparison between 1980 and 2020. 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

(Bernanke (2005)), I remain agnostic about the source of interest rate decline and calibrate an exogenous change.

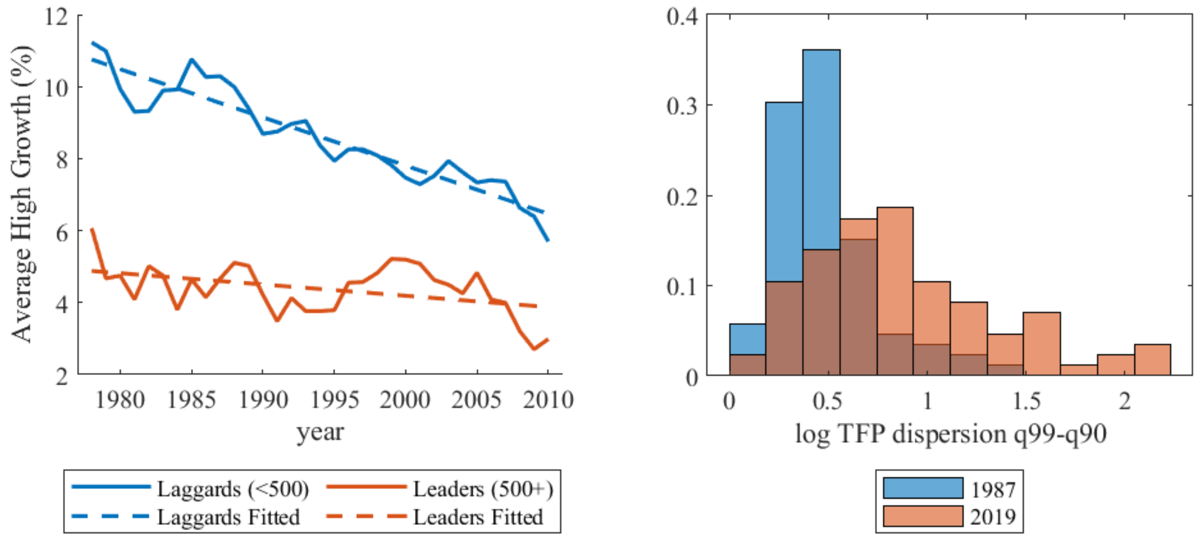
Figure 11a compares innovation and learning intensities in both periods. Given the *uniform* increase in research difficulty across all firms, the intensity curve shifts down so that aggregate growth declines. The decrease in growth is however larger for laggards as they enjoy higher growth at the earlier period and because there is a lower bound of conscious growth of 0. In economic terms, laggards have dynamic advantage in terms of catch-up growth, while leaders have static advantage of being a leader in the existing market structure. Laggards rely on growth, regardless of the specific form of innovation or learning, to challenge existing leaders. Harder research hurts the dynamic advantage of laggards, making them more difficult to catch up with leaders and challenge them. Thus, even though the growth rate of leaders decline in absolute terms, it increases *in relative terms* compared to that of laggards. This allows leaders to stretch out further to the right tail *relative to* laggards, implying a fatter tail of the productivity distribution and firm size distributions. To emphasize the heaviness of tail, Figure 11b and 11c plot 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 higher markup, as shown in the cost-weighted markup distributions of Figure 11d.

To see whether the model's mechanism is consistent with data, I categorize firms into leaders and laggards based on number of employment. Following Decker et al. (2016), high growth firms are defined as those with more than 30% annual growth of employment. I calculate 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 12a shows that the high growth firms among 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 these empirical patterns are consistent with the model in Figure 11a. Moreover, Figure 12b shows the histograms of log TFP dispersion between the 99th quantile and the 90th quantile, where each observation is a 4-digit NAICS sector. Due to data limitation, sectors are only manufacturing ones. The productivity dispersion at tail has in general increased,[52] which is consistent with the model's prediction that Pareto tail of productivity distribution flattens.

Table 5 summarizes 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. Sales share of top 1% firms increases as the Pareto tail of sales distribution fattens. TFP growth declines as harder research dominates all other mitigating forces including lower interest rate. Compared to existing literature of Schumpeterian growth, growth does *not* decline due to a more concentrated market structure in the present paper. Higher concentration actually promotes growth in the model and acts as a mitigating force against harder research, as heavier tail makes learning more effective and higher profits incentivize innovation and learning. Higher concentration is a result of lower growth, *not* a reason for it.

Because we have assumed a demand system with variable demand elasticities, aggregate markup increases with tail fatness of the productivity distribution *other things being equal*. Aggregate labor share decreases, as labor is the only factor of production in the model so that labor share is simply the inverse of markup. 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. Similarly analysis is done for labor share. Both the increase in markup

[52] 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 12: 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 employment-weighted average growth of high growth firms among leaders and laggards respectively. Data source: Decker et al. (2016). Panel (b): The histogram shows 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).

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

| | Within | Between | Net entry | Total |
|---|---|---|---|---|
| $\Delta$Labor share | $-0.30\%$ | $-3.05\%$ | $-0.03\%$ | $-3.38\%$ |
| $\Delta$Markup | $0.73\%$ | $7.82\%$ | $-0.02\%$ | $8.53\%$ |

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

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 are consistent with micro-level empirical evidence (see De Loecker et al. (2020) for markup, Autor et al. (2020) and Kehrig and Vincent (2021) for labor share).

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 Decker et al. (2020)'s evidence that declining job fluctuations come not from decreasing volatility of idiosyncratic shocks but from weaker responsiveness of firms to these shocks. Incumbent job creation and job destruction rates decline, so does their death rate. Entry rate declines as learning becomes more difficult for potential entrants. Combining them together, job reallocation rate decreases.

# 6 Introducing Economic Integration

While the previous section well accounts for the key macroeconomic moments since 1980s, a major puzzle remains: before 1980s, how could concentration increase with flattening Pareto tail without trend in markup or labor share? 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), 1920s-1970s saw the heyday of US government spending on highways, with only the interruption of World War II. Inflation-adjusted domestic airfare per mile decreased at -4.55% per year between 1940 and 1980, and only -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 have also played a role in facilitating integration. One of the most notable examples is 1978 Airline Deregulation Act which removed restrictions on 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 a key. The above historical examples thus prompt us to consider economic integration as another importance force shaping the economic landscape of earlier periods. Stable labor share is thus not understood as an economic regularity in Kaldor (1961), but rather as a historical coincidence.

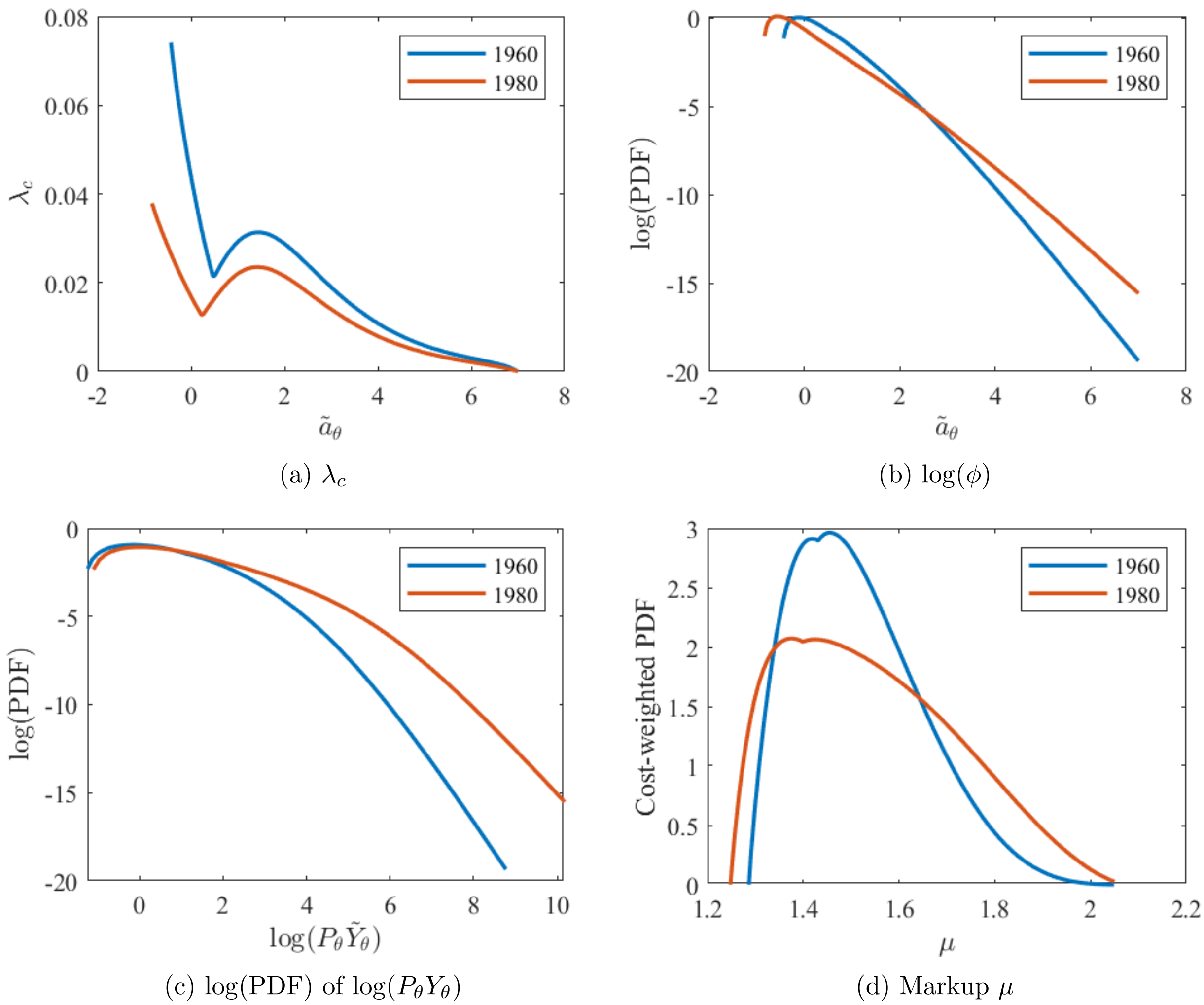
I study the 1960-1980 period to show the effect of economic integration *in conjunction with* harder research. Consider a thought experiment in which there are $N$ identical economies in autarky in 1960. All the economies are on the same travelling wave equilibrium as described by the model. As in the previous section, ideas are harder to find in 1980 than in 1960, and interest rate declines which partially mitigates the harder research (see Section 4 for the calibration). What's more, the $N$ economies integrate into one single economy, in the sense that labor force $L$ and potential number of entrants $M_e$ are multiplied by $N$. [53] 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.

Proposition 6 shows that economic integration promotes growth and mitigates the effect of harder research, as a larger number of customers implies higher profit and hence stronger Schumpeterian incentives. While I estimate the magnitude of harder research using Bloom et al. (2020)'s methodology, the degree of economic integration $N$ is estimated by targeting TFP growth rate of 1.6% around 1960. Concentration, markup and labor share remain non-targeted.

Figure 13 compares the equilibrium in 1960 with the one in 1980. Similar to 11, growth declines for all firms but especially for laggards, 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 mitigating forces of lower interest rate and larger market size. Top 1% firm sales concentration increases by 14.43%, which is in the ballpark of Kwon et al. (2022)'s estimate and slightly more than the change from 1980 to 2020. As growth rates in 1960 (this section) and in 1980 (Section 4.3) have been targeted, the increase in concentration resonates Proposition 5 which links fatter tail with lower growth regardless of market power. Changes in markup and labor share, however, are largely attenuated. Table 7 shows that while market share reallocates towards high markup and low labor share firms as the 1980-2020 period, the reallocation component is largely compensated by the within-firm component due to economic integration. The within-firm component can act as a compensating force against fatter tail for markup and labor share, because the equilibrium log-productivity distribution is overall log-concave. [54] The combined effect of within- and between- components means strongly attenuated markup and labor share changes, despite an even higher increase in concentration. To summarize, the dynamic growth effect of harder research and the static effect of economic integration jointly reconcile fatter Pareto tails with stable markup and labor share. 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.

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

[54] See further discussions in Section 4.4.




(a) $\lambda_c$

(b) $\log(\phi)$

(c) log(PDF) of $\log(P_\theta Y_\theta)$

(d) Markup $\mu$

Figure 13: Comparison between 1960 and 1980. 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

| | Within | Between | Net entry | Total |
|---|---|---|---|---|
| $\Delta$Markup | $-5.29\%$ | $8.74\%$ | $-0.36\%$ | $3.09\%$ |
| $\Delta$Labor share | $2.86\%$ | $-4.32\%$ | $0.14\%$ | $-1.32\%$ |

Table 7: Melitz-Polanec decomposition of markup and labor share changes between 1960 and 1980.

# 7 Conclusion

The present study constructs a continuous-time Schumpeterian growth model which complements the existing literature by emphasizing how growth shapes market structure. The model is reminiscent of endogenous sunk cost models of Sutton (1998) and shares his insights in a growth setting. 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 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 desctruction rates of US in the last four decades. The explanation is compatible with growth decline and increasing concentration in a longer historical horizon. The framework can accommodate fatter Pareto tail with stable markup and labor share in the pre-1980 period by introducing economic integration in addition to harder research.

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* recommended 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 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. [55] Even though the model cannot fully speak to that history, as 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 US, Antolin-Diaz and Surico (2022) shows military spending increases productivity in the long run through public R&D. Gruber and Johnson (2019) interprets the post-war experience of US in a similar fashion, suggesting policy implications centering around public R&D. This framework can hopefully open the door for further fruitful discussions.

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

## References


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

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022). Income and wealth distribution in macroeconomics: A continuous-time approach. *The review of economic studies*, 89(1):45–86.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (forthcoming). A theory of falling growth and rising rents. *Forthcoming in The review of economic studies.*

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.

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.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica: Journal of the Econometric Society*, pages 323–351.

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.

Akcigit, U. and Ates, S. T. (forthcoming). What happened to us business dynamism? *Forthcoming in Journal of Political Economy.*

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

Alvarez, F., Argente, D., Lippi, F., Méndez, E., and Van Patten, D. (2022). Strategic complementarities in a dynamic model of fintech adoption. Technical report, mimeo, New York University.

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

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.

Arkolakis, C. and Morlacco, M. (2017). Variable demand elasticity, markups, and pass-through. *Manuscript, Yale University.*

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

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

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.

Baqaee, D., Farhi, E., and Sangani, K. (2020). The darwinian returns to scale. Technical report, National Bureau of Economic Research.

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

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.

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

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.

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.

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

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly journal of economics*, 121(2):541–585.

Burstein, A., Carvalho, V. M., and Grassi, B. (2020). Bottom-up markup fluctuations. Technical report, National Bureau of Economic Research.

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).*

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

Chen, Z. (2022). Economic growth and the rise of large firms.

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

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

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

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.

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

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.

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.

Edmond, C., Midrigan, V., and Xu, D. Y. (forthcoming). How costly are markups? *Forthcoming in Journal of Politcal Economy.*

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

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

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

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

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

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

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.

Grossman, G. M. and Oberfield, E. (2021). The elusive explanation for the declining labor share. Technical report, National Bureau of Economic Research.

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

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

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

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.

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.

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

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

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

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

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.

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

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

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

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

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.

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

Kortum, S. S. (1997). Research, patenting, and technological change. *Econometrica: Journal of the Econometric Society*, pages 1389–1419.

Kwon, S. Y., Ma, Y., and Zimmermann, K. (2022). 100 years of rising corporate concentration.

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.

Lane, N. (2022). Manufacturing revolutions: Industrial policy and industrialization in south korea. *Available at SSRN 3890311.*

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

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

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

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

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

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

Luttmer, E. G. J. (2020). Bounded learning from incumbent firms. Technical report, Federal Reserve Bank of Minneapolis.

Matsuyama, K. (forthcoming). Non-ces aggregators: A guided tour. *Forthcoming in Annual Review of Economics.*

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.

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

Melitz, M. J. and Ottaviano, G. I. (2008). Market size, trade, and productivity. *The review of economic studies*, 75(1):295–316.

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.

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.

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

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

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

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

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

Philippon, T. (2022). Additive growth. Technical report, National Bureau of Economic Research.

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

Sutton, J. (1998). *Technology and market structure: theory and history*. MIT press.