Economic Growth and the Rise of Large Firms*

Zhang Chen
Princeton University

August 2023

Abstract

Rich and poor countries differ in the size distribution of business firms. In this paper, I document that the right tail of the firm size distribution systematically grows thicker with economic development, both within countries over time and across countries. I develop a simple idea search model with both endogenous growth and an endogenous firm size distribution. The economy features an asymptotic balanced growth path. Along the transition, Gibrat's law holds at each date, and the right tail of the firm size distribution becomes monotonically thicker. The firm size distribution converges to Zipf's distribution. The model also implies that policies favoring large firms can improve welfare due to the externality associated with idea search. Finally, I extend the results obtained in the simple model to a general class of idea search models. Under common functional form assumptions, my model stands out as the only model within that class that is consistent with both Gibrat's law and a thickening right tail.

Keywords: Firm Size Distribution, Right Tail, Idea Diffusion, Growth, Gibrat's Law, Zipf's Law

*I am greatly indebted to my advisors Gene Grossman, Ezra Oberfield, Stephen Redding, and Richard Rogerson for their continued guidance and support. I am grateful for comments from Mark Aguiar, David Argente, Cheng Chen, Pablo Fajgelbaum, Nobuhiro Kiyotaki, Ernest Liu, Yueran Ma, Kiminori Matsuyama, Jonathan Payne, Gianluca Violante, Wei Xiong, various seminar participants, and particularly from Manuel García-Santana. This research receives generous financial support from the International Economics Section (IES) at Princeton University and the STEG Ph.D. Grant. All errors are my own. Email: zhangchen.econ@gmail.com

1 Introduction

Rich and poor countries differ in many respects, one of which is the size distribution of business firms. Whereas giant corporations are sometimes viewed as symbols of economic success, a notable feature of developing economies is the prevalence of small firms. A fundamental question of economic growth is how the firm size distribution varies with the level of economic development. Following the seminal paper by Lucas (1978), existing research has focused on the relationship between economic development and average firm size. This paper instead studies the relationship between economic development and the right tail of the firm size distribution.

I contribute to this issue both empirically and theoretically and derive novel policy implications. Empirically, I document that the right tail of the firm size distribution systematically grows thicker with economic development, both within countries over time and across countries. Theoretically, I propose a novel idea search model that rationalizes this relationship as a generic feature of the growth process. My model has four major properties: 1) the economy features an asymptotic balanced growth path; 2) Gibrat's law holds at each date; 3) the right tail of the firm size distribution becomes thicker along the transition; 4) the firm size distribution converges to Zipf's distribution. On the policy side, the model sheds light on how policies favoring large firms improve social welfare in the presence of an externality associated with idea diffusion.

One challenge when comparing firm size distributions across countries is the potential for missing data on small firms. In my empirical analysis, I construct a measure of the thickness of the right tail using a transformation of the relative employment share between large and not-so-small firms, which excludes small firms. Readily available statistics on firms by employment size bin suffice to compute this measure, making comparable measurement feasible and convenient across a wide variety of countries and periods. Three distinct but complementary datasets are suitable for this task: the OECD Structural Business Statistics (SBS), the World Bank Enterprise Survey (WBES), and the US census Business Dynamics Statistics (BDS). I find in all these data a positive correlation between GDP per capita and the right tail thickness. Importantly, this positive correlation holds across countries and within countries over time, in developing and developed countries, and by major sectors.

The robustness of this positive relationship suggests that a thickening right tail might be an innate feature of the process of economic growth. To pursue this, I build on recent development in endogenous growth theory that study idea flows among heterogeneous firms as a source of growth. I develop a novel idea search model in which both growth and the firm size distribution are endogenous, and the right tail of the firm size distribution thickens along the transition path.

The model economy has a continuum of firms with heterogeneous productivity. Firms increase their productivity by learning from more productive firms via random meetings. Meetings between firms are Poisson events: firms decide on how much to invest in idea search, which determines the arrival rate of meeting opportunities. Given a realized meeting opportunity, firms take a random draw among all firms that are more productive than they are and update their productivity to the level of the firm they encounter. Firm-level productivity growth from learning fuels economic growth, and the collective learning activities of firms continuously reshape the productivity distribution.

My model has the property that if the firm productivity distribution is Pareto at time t, then it will also be Pareto at $t+h$. Assuming that the initial distribution is Pareto, I am able to obtain a complete analytical characterization of the equilibrium path, which exhibits four key properties. First, the economy features an asymptotic balanced growth path. Second. Gibrat's law holds at all times; namely, average firm growth is always independent of firm size. Third, the distribution of firm productivity is always Pareto with a constant scale but a varying shape k, i.e., $F(x) = 1 - x^{-k}$, in which x is a firm's productivity, and $x \geq 1$. Fourth, the shape parameter k decreases monotonically over time and converges to 1. In other words, the right tail thickens over time, and the limiting productivity distribution is Zipf's distribution, a Pareto distribution with shape parameter 1.

A key departure of my model from existing idea search models is the source distribution from which firms draw ideas. Earlier models in the literature assume all firms search from a common source of ideas. This implies that firms with higher productivity benefit less from each search, as fewer of their meetings would yield improvement. It follows that more productive firms have lower expected growth, and Gibrat's law does not hold. My model instead assumes that more productive firms draw ideas from better source distributions. In equilibrium, each firm faces the same source distribution in terms of *relative* productivity, and the expected growth rate is the same across firms, consistent with Gibrat's law. Additionally, the assumption that all firms draw ideas from the same distribution in earlier models implies that all firms have the same probability of adopting state-of-the-art technology. This implication is at odds with empirical work which finds larger firms are more likely to adopt the most advanced technologies.

This departure of my model yields two important insights about the firm size distribution and growth. First, my model presents a novel growth mechanism termed ***tail growth***: growth is generated by the thickening of the right tail. Earlier models in the literature assume a balanced growth path in which the distribution of *relative* firm productivity is stationary. The productivity distribution is scaled up proportionately so that its shape remains unchanged. In my model, the distribution of *relative* firm productivity varies along the equilibrium path. Aggregate productivity improves due to the redistribution of mass from lower productivity to higher productivity in *relative* terms. This redistribution manifests as a thickening of the right tail, capturing the rising share of high productivity firms. Specifically, in the model, output per capita y is the mean of firms' productivity, $k/(k-1)$, in which k is the aforementioned Pareto shape parameter. In other words, my model predicts a tight relationship between GDP per capita and right tail thickness. The model economy features an asymptotic balanced growth path since with firms' optimal search intensity, $k-1$ decreases towards zero at a constant rate, i.e., the firm size distribution converges to Zipf's distribution. My model thus offers an explanation for why advanced countries such as the US have firm size distributions with Pareto tails close to 1: they are further along the development path.

Second, my model offers new insights into the relationship between growth and the rise in concentration. Up to some normalization, a stationary firm size distribution has been a standard component of growth models with heterogeneous firms. Nevertheless, recent research suggests that the rise in market concentration is a secular trend in the US. How do we reconcile constant growth with the rise in concentration? Notice that the right tail thickness is a concentration measure by itself and partly determines other common measures such as the Herfindahl-Hirschman Index. Thus, the model describes an economy on an asymptotic balanced growth path that exhibits continually rising concentration. In my model, output growth $\dot{y}/y$ converges to a constant as $k \quad 1$ decreases to 0 at a constant rate. Especially when k is close to 1, the output growth varies very little while the concentration still grows steadily.

Additionally, the model delivers new policy implications. Idea search by each firm has externality on other firms since it affects the productivity distribution, which determines future search efficiency. While search by large firms thickens the right tail and has positive externality on all firms in the economy, search by small firms has few externality on large firms. Thus, relative to first-best outcomes, large firms under-invest in idea search, and policy should encourage more search by large firms. I consider two policy exercises. In the first exercise, the social planner chooses a productivity threshold and imposes an additional tax on firms below the threshold. As a result, search is conducted only by firms above the threshold. I show that the equilibrium long-run growth rate increases with the level of productivity threshold, and so does welfare. The second exercise solves the social planner's problem. The optimal individual search intensity grows with the level of productivity at an approximate power rate. In this respect, the socially optimal search intensity differs from the equilibrium intensity, which is uniform across all firms. Both exercises indicate that policies favoring large firms better capture the diffusion externality and improve welfare.

In the final analysis, I revisit the results obtained in the simple model and extend them to a general framework that encompasses both my model and other prominent idea search models in the literature. This analysis delivers three noteworthy findings. Firstly, Gibrat's law and a thickening right tail provide disciplines for idea search models. Under common functional form assumptions, I show that Gibrat's law and a thickening right tail identify the parameters of the idea search process, which are exactly what I assumed in my model. My model then stands out as the only model within this class that generates both Gibrat's law and a thickening right tail. Secondly, the dynamics of tail index in my model can be extended to the general case if the idea search process satisfies an internal search condition. It further implies that Gibrat's law holds if and only if the right tail becomes thicker. Therefore, existing idea search models characterized by a stationary firm size distribution do not adhere to Gibrat's law. Lastly, I provide a necessary and sufficient condition for Zipf's law to hold in the limit, beyond the specific case of the simple model. This condition reveals that a wide range of idea search processes can be consistent with Zipf's law, thus bolstering the plausibility of idea diffusion as an explanation for Zipf's law.

Related Literature This paper makes contributions to four strands of literature. First, it is related to the literature on the relationship between economic development and firm size distribution. Many papers have documented and explained the positive correlation between average firm or establishment size and the level of economic development, typically measured by GDP per capita or worker.[1] Misallocation is often posited as a leading explanation for this empirical finding: the smaller size of firms in developing countries stems from under-investment, driven by institutional distortions that disproportionately affect larger firms. Most existing work employ steady-state comparisons to capture cross-country differences, which rely on exogenous cross-country variations in distortions. This paper targets instead the right tail of the firm size distribution and rationalize the relationship as an endogenous feature of the transition path.

Second, there have been growing interests in the rise of market concentration in the US for the past several decades, as evidenced by papers such as Gutiérrez and Philippon (2018), De Loecker et al. (2020), Autor et al. (2020), and Kwon et al. (2022). This paper contributes to this literature by introducing a useful concentration measure based on the thickness of the right tail. In particular, I document and explain the thickening of the right tail of the US firm size distribution over the past 40 years. Oberfield (2018) is one of the few exceptions in the literature which study the right tail thickness of the firm size distribution. He proposes a theory of endogenous production network formation based on input choices, which associates lower labor share with thicker right tails. Kwon et al. (2022) takes a similar perspective to this paper, viewing the rise in concentration as a normal feature of growth. They use historical data for the US to show that the rise in market concentration has been observed for about a century, together with growth. I complement their work with a theoretical framework wherein asymptotic balanced growth aligns with the rise in market concentration. Moreover, I present new empirical evidence that supports the predictions of my theory.

Third, this paper builds heavily upon a burgeoning literature on endogenous growth models with idea flows. Buera and Lucas (2018) provide a comprehensive survey of this literature.[2] This paper differs from the existing literature pioneered by Kortum (1997) in two significant respects. First, existing models focus on balanced growth paths in which the distribution of relative firm productivity is stationary. In contrast, my model features an asymptotic balanced growth path in which the distribution of relative firm productivity varies. Specifically, the right tail of the distribution becomes thicker over time. Second, Gibrat's law does not hold in those idea search models. I further show that in a general class of idea search models, including these earlier models, Gibrat's law holds if and only if the right tail becomes thicker.

Notably, this paper is related to recent work by Jones (2022). Both studies explore alternative approaches to the classical conditions outlined in Kortum (1997) for constant growth, which involve a Pareto source distribution and exponential growth in the idea flow. Jones (2022) relaxes the thick-tailedness assumption but strengthens exponential growth to combinatorial growth. In contrast,

[1] Notable references are Lucas (1978), Tybout (2000), Alfaro et al. (2008), Hsieh and Olken (2014), Hsieh and Klenow (2014), García-Santana and Ramos (2015), Poschke (2018) and Bento and Restuccia (2017, 2021)

[2] An incomplete list goes as follows: Jovanovic and Rob (1989), Kortum (1997), Alvarez et al. (2008), Lucas and Moll (2014), Perla and Tonetti (2014), Sampson (2016), Buera and Oberfield (2020), Perla et al. (2021), Akcigit et al. (2018), Benhabib et al. (2021), König et al. (2016), and König et al. (2022).

my approach allows a constant population (decreasing idea flows) but make a stronger assumption on the source distribution. Instead of searching randomly over the same Pareto distribution, firms in my model search more efficiently over Pareto distributions left-truncated by their productivity.

Fourth, this paper contributes to the study of Zipf's law. Axtell (2001) documented that the US firm size distribution has a remarkable resemblance to Zipf's distribution. Other prominent papers on Zipf's law include Gabaix (1999) and Luttmer (2007, 2012). Both obtain a limiting Pareto distribution from a geometric Brownian motion with barriers. They show that parameters can be chosen such that the Pareto tail is close to one. However, they do not provide an economic rationale for the emergence of Zipf's law. My model explains Zipf's law as a result of idea search. Other attempts to micro-found Zipf's law include Geerolf (2017), which rationalizes this empirical regularity using a static model of endogenous span-of-control.

Roadmap The rest of the paper is organized as follows. Section 2 presents evidence in favor of a positive relationship between the thickness of the right tail and the level of development. Section 3 and 4 propose and analyze the baseline idea search model. Section 5 discusses the policy implications of the model. Section 6 extends the baseline model and presents some general results. Finally, Section 7 concludes.

2 Stylized Facts

In this section, I provide empirical evidence supporting a positive relationship between the right tail thickness of the firm size distribution and the level of economic development. I outline the construction of my thickness measure and explore its implications in Section 2.1. To compute the right tail thickness, I introduce three complementary datasets in Section 2.2. Section 2.3 presents the results, whose robustness is confirmed through multiple validations conducted in various settings, Collectively, these evidence suggest the necessity of a generic theory.

2.1 A Tail Thickness Measure

Measuring the right tail of firm size distributions across multiple countries presents a significant challenge. While the literature has extensively studied estimating the tail index of thick-tailed distributions, sophisticated statistical procedures require a relatively large number of individual observations, e.g., data on individual firms.[3] However, obtaining a large sample of countries with administrative micro-data on firms is difficult as such data is usually confidential in each country. Instead, tables of employment or number of firms by firm size bins are more widely accessible. Therefore, an appropriate measure must capture the tail thickness in a simple manner and be suitable for coarsely tabulated data on firm size distributions. I construct such measure based on the notion of right tail index in statistics.

[3] Resnick (2007) introduces standard estimators of the tail index. Also, see Gabaix (2009) for an introduction of tail estimation with economic applications.

One widely-used right tail index has the following definition. Consider a continuous random variable with cumulative distribution function (CDF) F and probability density function (PDF) f. Furthermore, let $\tilde{F} \equiv 1 - F$ denote the complement CDF. Then, $\tilde{F}$ has tail index k if $\lim_{x \to \infty} \frac{\tilde{F}(tx)}{\tilde{F}(x)} = t^{-k}$ for all $t > 0$.[4] Conceptually, the right tail index reflects how quickly the relative fraction of population above thresholds in fixed ratio changes with the lower threshold. The right tail is thinner if there is faster decay in the relative fractions when moving the threshold to the right. The right tail index motivates a simple way to measure the right tail thickness. Formally, let T_L be the firm size threshold for large firms and T_S for small firms, I construct a right tail thickness measure $\tilde{R}^f$ as the following:

$$\tilde{R}^f = \log \frac{\tilde{F}(T_L)}{\tilde{F}(T_S)} / \log \frac{T_L}{T_S}.$$

In words, $\tilde{R}^f$ is the normalized log relative share of large firms.

There is a close relationship between the right tail thickness measure $\tilde{R}^f$ and the right tail index k. First of all, $\tilde{R}^f$ is a finite analog of $-k$. With $x = T_S$ and $t = T_L/T_S$, $k = -\tilde{R}^f$ satisfies the equation $\frac{\tilde{F}(tx)}{\tilde{F}(x)} = t^{-k}$. Next, that $\tilde{R}^f \approx -k$ holds for large x or T_S. This follows immediately from the definition of tail index k and the previous point. Finally, it is straightforward to show that $\tilde{R}^f = -k$ if and only if the underlying firm size distribution is exactly Pareto, in which k is the shape parameter. This property implies that $\tilde{R}^f$ is less influenced by choices of thresholds if the underlying distribution is approximately Pareto distributed in the right tail. Hence, this measure is particularly suitable in the current setting given widespread evidence that firm size distributions have Pareto right tails.

In addition, the right thickness measure $\tilde{R}^f$ has four implications that are useful in practice. First, this measure only uses two size thresholds and the fractions of firms above them. Specifically, the simple structure works well for cross-country aggregate data, which usually present firm size distributions in tables of number of firms or employment by firm size bins. Second, countries may differ slightly in the division of firm size bins. For example, if measuring firm size by the number of employees, a few countries may use thresholds 5 and 19 and skip 10, while the majority use 10 as a threshold. The normalization term T_L/T_S then takes care of these discrepancies across countries. Third, poor coverage or low quality of the data on the smallest firms in developing countries has been a known challenge in the literature for reasons like informality and self-employment. Data limitations on these firms is of less concern here since the thickness measure targets on the right tail of the firm size distribution. It is clear from the construction that the threshold T_S excludes the left tail. Fourth, the right tail thickness is also a measure of market concentration. It is straightforward to verify that the tail index is a sufficient statistic for the Herfindahl-Hirschman Index (HHI) if the underlying sales distribution is Pareto or Fréchet.

In many occasions, data often provide information on the partial sum of firm sizes rather than

[4] $\tilde{F}$ needs to be regularly varying to have a well-defined tail index. Please refer to section 6.2 and the references therein for relevant mathematical background on regular variation.

the density of firm sizes. For example, tables may report the total employment in firms with a specific number of employees, which represents the sum of employees across firms of that size. Analogously, I construct a right tail thickness measure $\tilde{R}^{emp}$ for this type of data:

$$\tilde{R}^{emp} = \log \frac{\int_{T_L}^{1} x \mathrm{d}F}{\int_{T_S}^{1} x \mathrm{d}F} / \log \frac{T_L}{T_S}.$$

When employment measures the firm size, $\tilde{R}^{emp}$ is the normalized log relative employment share of large firms. Similar arguments can be used to show that $\tilde{R}^{emp}$ and $1 \quad k$ have the same theoretical linkages as those between $\tilde{R}^{f}$ and $\quad k$.

2.2 Data

I make empirical investigations into the correlation between the right tail thickness and the level of economic development. On the one hand, I use GDP per capita as a proxy for the level of economic development and obtain data on real GDP per capita in constant international dollars from the Penn World Table (PWT) version 10.0. On the other hand, I use the number of employees as a proxy for firm size since the majority of countries report firm size distributions by the number of employees. The primary data source on firm size distributions is the OECD database of Structural Business Statistics (SBS) by ISIC Rev 4. The final sample consists of a panel of 33 OECD countries between 2008 and 2017.[5] It classifies firms into five size bins by the following number of employees: 0-9, 10-19, 20-49, 50-249, and 250+. This dataset also reports firm size distributions by major sectors and thus addresses concerns for sectoral composition.

While the OECD data is of high quality, it has two limitations. First, the representativeness may be questionable since it is only about advanced economies. Second, the time span is relatively short and special. It has only 10 years and covers the post-crisis recession period. The following two datasets complement the OECD data to address these issues. One complementary dataset is the World Bank Enterprise Survey (WBES), which is a collection of surveys conducted by the World Bank aiming for a representative portray of a country's business economy. It has surveyed over 130 countries between 2006 and 2019, 113 of which are low, lower-middle and upper-middle income countries.[6] Firms are divided into three size bins: 5-19, 20-99, and 100+. Since most of the countries are surveyed only once or twice, I use it as evidence on the cross-section. The other dataset is the Business Dynamics Statistics (BDS) of the US census. It contains detailed information on the US firm size distribution in the past four decades (1978-2019). The BDS tracks well the path of the varying right tail in a representative growing economy for a protracted period of time. Moreover, US firms are clustered into 10 size bins ranging from 1 to 10,000+ employees.

[5]It covers all OECD countries by the end of 2018 except Chile, Mexico and Korea, which are excluded due to data mismatch or incompleteness. The original database covers these countries from 2005 to 2018. I focus on the period 2008-2017 because only a few countries have data on 2005-2007. Data on 2007 and 2018 are systematic breaks which may due to changes in measurement.

[6]Admittedly, the WBES has the limitation that the observation is at plant rather than firm level. To the best of my knowledge, it is however the most comprehensive dataset on the business structure of developing countries.

This relatively refined classification enables sensitivity checks on the choice of thresholds.

All these datasets offer two types of firm size distributions, namely, the number of firms or total employment by the number of employees. Therefore, I compute both statistics $\tilde{R}^f$ and $\tilde{R}^{emp}$ and use them to measure the right tail thickness. I obtain highly consistent results with both measures. For the sake of brevity, the rest of this section only presents the results with the number-based statistics $\tilde{R}^f_t$. Additional results using the employment-based statistics $\tilde{R}^{emp}_t$ are relegated to Appendix F.

2.3 Results

In this subsection, I discuss the computation of the right tail thickness for each dataset. Then, I detail the corresponding regression specifications and present the estimation results of the correlation between the right tail thickness and the GDP per capita.

2.3.1 the OECD SBS

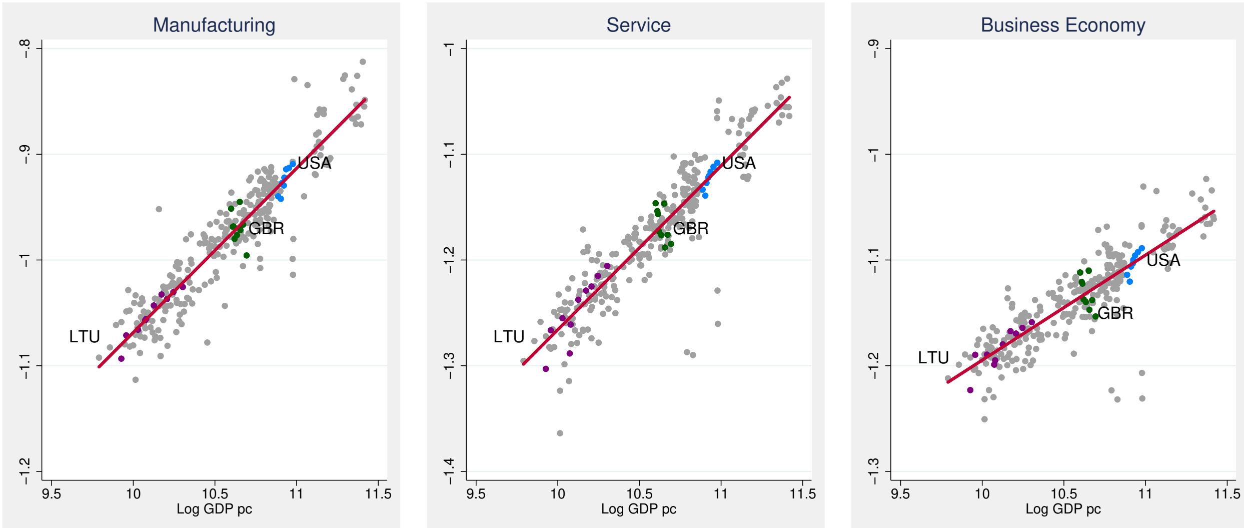
With OECD countries, I construct the right tail thickness measure $\tilde{R}^f_t$ for each country-year pair with a small firm threshold $T_S = 10$ and a large firm threshold $T_L = 250$. This choice of thresholds follows the OECD small and medium enterprise (SME) standard. Notably, the small firm threshold $T_S = 10$ has already captured well the right tail of the distribution: the median fraction of firms with less than 10 employees is 91.6% across all country-year pairs in our sample.[7] To obtain the correlation, I regress the right tail thickness measure on log GDP per capita and country fixed effects. Formally,

$$\tilde{R}^f_{c,t} = \alpha + \beta \text{logGDPpc}_{c,t} + \gamma_c + \varepsilon_{c,t},$$

in which γ_c is the country fixed effect. With $\bar{\gamma}\quad \frac{\sum_{c=1}^{N_c}\gamma_c}{N_c}$, I plot modified data dots $(\text{logGDPpc}_{c,t}, \hat{\tilde{R}}^f_{c,t})$, in which $\hat{\tilde{R}}^f_{c,t} = \tilde{R}^f_{c,t}\quad \gamma_c + \bar{\gamma}$. In this way, the scatter plot in figure 1 presents the correlation between the right tail thickness and log GDP per capita ***after controlling for*** country fixed effects. The estimated correlation, $\hat{\beta}$, then manifests itself as the slope of the linear fit and is significantly positive. Appendix F.1 stores the full regression results. By filtering out country-specific components, figure 1 tracks the changes in the right tail thickness over a wide span of development stages in the synthetic country. Three annotated data series exhibit the trajectories of Lithuania (in purple), the UK (in green) and the US (in blue), which represents countries on different parts of the development spectrum.

This exercise with the OECD countries suggests that there is a positive correlation between the right tail thickness and log GDP per capita. Moreover, this correlation suggests that the right tail tends to become thicker as a country progresses along its growth trajectory. It is important to note that these conclusions are based on within-country and over-time variations, making theories relying on cross-country differences in exogenous factors inadequate for explaining the thickening of the right tail during the transition. In addition to the evidence for the overall economy (see figure 1c),

[7] The average of this fraction over the sample period is below 80% in only three countries: the USA (79.2%), Switzerland (68.1%) and New Zealand (79.5%).



(a) Manufacturing (b) Service (c) Business Economy

Figure 1: Right tail thickness and the level of development in OECD countries

Notes. This figure plots the right tail thickness $\tilde{R}_t^f$ against log GDP per capita for each country-year pair in manufacturing, service and the whole business economy. The scatter dots are readjusted by country fixed effects, and the red lines are the linear fits. Appendix F.1 documents the details of the construction. The right tail thickness $\tilde{R}_t^f$ is calculated using the OECD SBS data and with $T_S = 10$ and $T_L = 250$. Data on GDP per capita are from the PWT 10.0. Three annotated countries are Lithuania (LTU), the UK (GBR) and the USA.

the positive correlation holds true for both manufacturing (see figure 1a) and service (see figure 1b). This indicates that the thickening of the right tail is not merely a result of composition effects driven by unique features of specific sectors. The fact that the positive correlation also holds in the service sector suggests that the underlying mechanism is not solely dependent on international trade, as it extends to less-tradable sectors. Additionally, Appendix F.1 demonstrates the robustness of the positive correlation in the absence of country fixed effects. In other words, there are also cross-sectional evidence supporting the notion that country-year pairs with higher log GDP per capita exhibit thicker right tails. To conclude, this positive correlation appears to be a fundamental characteristic of the growth process, representing a common feature across different countries.

2.3.2 the WBES

In this section, I focus on developing countries to complement with the results presented for the OECD countries. I restrict the WBES sample to the 113 low, lower-middle and upper-middle income countries covered by the survey. I construct the right thickness measure $\tilde{R}_t^f$ for each country-year (survey year) pair based on a small firm threshold $T_S = 5$ and a large firm threshold $T_L = 100$.[8] Figure 2 plots the right tail thickness against log GDP per capita without any fixed

[8]The WBES presents truncated data including only establishments with at least 5 employees, so it is difficult to know the percentile of 5 employees in the true distribution. Bento and Restuccia (2021) publish data on average establishment size by country which they compute from micro establishment data. Using their data, I find that the average size of non-agricultural establishments are below 5 employees in 90% of the countries with log GDP below

effects. The slope of the fitted line in red presents the positive correlation between the right tail thickness and log GDP per capita among the sampled developing countries. In other words, there are cross-sectional evidence from developing countries that the positive relationship holds true.

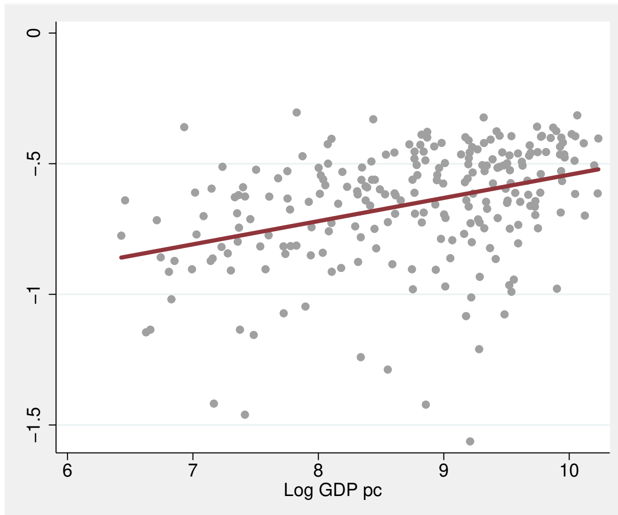


Figure 2: Right tail thickness and the level of development in developing countries

Notes. This figure plots the right tail thickness $\tilde{R}_t^f$ against the log GDP per capita for each country-year pair in the business economy. The red line is the linear fit. The right tail thickness $\tilde{R}_t^f$ is calculated using data on developing countries of the WBES and with $T_S = 5$ and $T_L = 100$. Data on GDP per capita are from the PWT 10.0.

Figure 2 presents new versions of the results documented by García-Santana and Ramos (2015). Also using the WBES (an older version), they find a negative cross-country association between productivity (or GDP per worker) and the share of employment in small plants.[9] Appendix F.2 obtains the full regression results and other robustness checks. It confirms that the positive correlation from the cross-section is robust to the inclusion of high-income countries or year fixed effects.

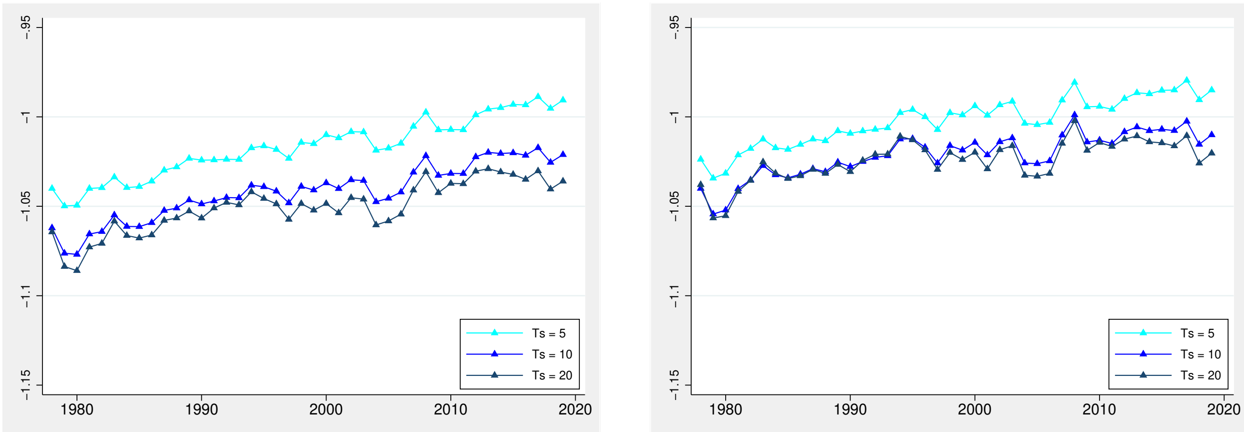
2.3.3 the US BDS

In this section, I study the case of the United States using detailed data on the size distribution of US business firms from the Census. The benchmark choice of the size thresholds are a small firm size threshold of 20 employees and a large firm size threshold of 500 employees. The former captures approximately the top 10% of firms in the US based on employment size, while the latter represents the top 0.5%. Additionally, the richness of the BDS data allows me to examine the sensitivity of the results to the choice of thresholds. For this purpose, I also experiment with alternative thresholds

10.23, the highest log GDP in my baseline WBES sample. Recall that the WBES divides the establishments by 5, 20, or 100 employees. Then, I choose 5 employees as the small firm threshold since a threshold 5 can already capture the right tail in most countries.

[9] Their sample consists of 104 countries surveyed between 2006 and 2010.

$T_S = 5$ or 10 and $T_L = 1000$. In contrast to previous exercises, I plot the right tail thickness against years instead of log GDP per capita since the data is a time-series of a well-known economy.



(a) $T_L = 500$ (b) $T_L = 1000$

Figure 3: Changes in the right tail thickness in the US (1978-2019)

Notes. This figure plots the right tail thickness $\tilde{R}_t^f$ of the size distribution of all US business firms from 1978 to 2019. The right tail thickness $\tilde{R}_t^f$ is calculated using the census BDS data and with various T_S and T_L indicated in the figure.

Figure 3 plots the changes in the right tail thickness the US from 1978 to 2019 and points out positive trends in it. These trends are also robust to the choice of thresholds, with all connected lines closely aligned with each other. These results then indicate that the US firm size distribution has grown thicker right tails in the past forty years.

That the right tail becomes thicker in the US joins a strand of growing literature documenting the rise of market concentration in the US. It is particularly close to the findings in Autor et al. (2020) and Kwon et al. (2022). Autor et al. (2020) find that in various industries, the share of the top 4 or 20 firms in that industry in total sales or employment has been increasing between 1980 and 2010. Likewise, Kwon et al. (2022) confirm that the aggregate employment share of the top 1% and top 0.1% firms has increased in the past forty years. On top of that, they leverage historical data on the financial metrics of firms, such as assets, sales, and net incomes. They find that the rise in concentration in the US may start at a much earlier date, up to a hundred years ago.

Robustness Appendix A provides supplementary evidence regarding the thickening of the right tail in the United States. It addresses three specific concerns raised in relation to the current findings. Firstly, it is questionable whether the 1978-2019 duration is sufficient to establish a long-term trend. Using the data shared publicly by Kwon et al. (2022), I extend the analysis to cover the period from 1918 to 2018. These expanded time series, which measures firm size based on assets, receipts, or net income, confirm that the thickening of the right tail persists over a longer horizon.

Secondly, I discuss potential measurement errors associated with the current estimator, arising from fixed size thresholds and the arbitrary selection of two thresholds. The estimator proposed by Toda and Wang (2021) greatly alleviates these concerns by taking advantages of all top firm shares and corresponding top percentiles. Notably, estimates based on their method also indicate a thickening right tail over the past four decades using the US BDS.

Lastly, there are doubts regarding the representativeness of aggregate data, particularly its ability to accurately capture changes in the tail thickness among very large firms. However, recent studies by Cao et al. (2022) and Chen et al. (2023) support the trend of thicker right tails. These papers respectively estimate the tail thickness among very large US firms using microdata from the US census and Compustat.

2.4 Conclusion

By synthesizing results across multiple data sources, I present compelling empirical evidence that supports a positive relationship between the level of economic development and the right tail thickness of the firm size distribution. This relationship holds true across different contexts. Specifically, the positive correlation remains consistent across countries and for within-country changes over time, in developed and developing countries, as well as across major sectors. The robustness of this relationship suggests that the underlying mechanism driving this phenomenon is generic to the growth process itself.

3 An Idea Search Growth Model

In this section, I propose a simple growth model based on idea search. The key ingredient of the model is that firms enhance their productivity by actively searching for and learning from more productive firms. The process of idea search by firms continuously reshapes the firm size distribution, which in turn affects firms' incentive to search. Consequently, there is a close interplay between changes in firm size distribution, especially the right tail, and the overall growth rate of the economy. In the following, I detail the model setup and discuss key underlying assumptions.

3.1 Model Setup

The economy consists of a continuum of firms, which are heterogeneous in their productivity. Firms produce homogeneous goods and engage in perfect competition in the market. The output good is the numeraire. In order to highlight the role of idea search, I make a number of assumptions to abstract away from other firm's decisions. Specifically, firms make no input decisions, and there are no entry or exit of firms. Each firm is endowed with one unit of capital, which is the sole input of production. The production function is such that the output is linear in firms' productivity and capital, i.e., $y = zk$. Capital does not depreciate, is specific to firms, and cannot be produced, so all firms use one unit of capital for production at all times. Then, a firm with productivity z produces z goods in each instant of time.[10] In the rest of the paper, I use firm z to denote a firm

[10] In this simple model, productivity is isomorphic to firms' size since it is equal to firms' revenue (output), and there is no production labor employment (only researchers). It is easy to incorporate production labor and measure

with productivity z.

Idea Search Absent input and entry/exit decisions, firms can only increase their profits by enhancing their productivity. They actively search for and learn from more productive firms in the economy through random meetings. The details of the meeting and learning process are as follows. At each instant, firms endogenously choose the arrival rate of their meeting opportunities, which is a Poisson process. Search intensity $\eta(z,t)$ is then the arrival rate chosen by firms. Once a meeting takes place, the other firm in the meeting is drawn randomly among all firms that are more productive than the searching firm. Let $F(\cdot,t)$ denote the productivity distribution of firms at time t and $f(\cdot,t)$ the corresponding density. Then, the probability that the meeting firm has productivity y is given by $f(y|y\geq z,t)$. After the meeting, the searching firm upgrades its productivity to the same level as the meeting firm's. A key departure of this learning process from that in standard idea search models[11] is that upon meetings, firms take random draws among firms that are more productive than they are, instead of the universe of firms in the economy. I will later discuss in depth the plausibility and the implications of this assumption.

Idea search is costly. Firms need to hire researchers to conduct the search and upgrade the technology. Moreover, more productive firms need more researchers to achieve the same level of search intensity. This reflects that it takes more efforts to absorb more sophisticated ideas, since more productive firms expect to obtain more sophisticated technology in each search. Besides, a higher search intensity also requires more researchers. Therefore, the cost function is such that firm z has to hire $z\eta$ researchers to achieve an arrival rate of η. It pays a search (gross of adoption) cost $z\eta w(t)$ given researcher's wage $w(t)$.

Firms' Problem Now I present firms' problem. Taking interest rate $r(t)$ and wage $w(t)$ as given, each firm solves the following profit maximization problem:

$$\begin{aligned} v(z,t) &= \max_{\eta(z,s)\geq 0} \mathsf{E}\left[\int_t^{\infty} e^{-\int_t^s r(\tau)\mathrm{d}\tau}\left(z(s) - z\eta(z,s)w(s)\right)\mathrm{d}s\right], \\ &\text{s.t. } \mathrm{d}z = (\tilde{X}(z,s) - z)\mathrm{d}J(\eta(z,s)), \end{aligned}$$

in which $J(\eta)$ is a jump process with rate η, and $\tilde{X}(z,t)$ is a random variable on $[z,\infty)$ and follows the conditional firm productivity distribution

$$F(x|x\geq z,t) = \frac{F(x,t) - F(z,t)}{1 - F(z,t)}.$$

firms' size by employment. For example, one can adopt this production function with labor input: $y = z^{1-\alpha}l^{\alpha}$, and $\alpha \in (0,1)$. Then, $l \propto y \propto R \propto \pi \propto z$. In this sense, productivity is a sufficient statistics for all common firm size measures (employment, output, revenue, profit).

[11] For example, Kortum (1997), Alvarez et al. (2008), Lucas and Moll (2014), and Perla and Tonetti (2014).

The associated Hamilton-Jacobi-Bellman (HJB) equation is then:

$$r(t)v(z,t) = z + \max_{\eta \geq 0}\left\{\eta \int_z^{\infty} [v(x,t) - v(z,t)]\mathrm{d}F(x \mid x \geq z, t) - z\eta w(t)\right\} + \partial_t v(z,t). \tag{1}$$

Equation (1) says that the flow value of the firm (the LHS) is the sum of four terms on the RHS: the flow profit (the first term), the total expected gains from learning (the second term), the total search and adoption cost (the third term), and the option value due to changes in the aggregate state (the fourth term).

Households' Problem There are a continuum of representative households with measure L. Households are infinitely lived, and each of them is endowed with one unit of labor. Each household inelastically supplies one unit of labor due to the absence of disutility from work. There is also no population growth, so the aggregate labor supply is fixed to L. It should be reminded that capital is the only input for production, and all labors are employed as researchers and earn wage $w(t)$. Households are owners of firms and claim their profits. Let $\pi(t)$ denote the average profit of all firms at time t. The flow income of each household is then given by $y(t) = w(t) + \pi(t)/L$. Taking interest rate $r(t)$ and income $y(t)$ as given, households maximize the present value of their CRRA utilities with respect to a lifetime budget constraint:

$$\begin{aligned} &\max_{c(\tau) \geq 0} \int_t^{\infty} e^{-\rho(\tau - t)} \frac{c(\tau)^{1-\theta} - 1}{1-\theta} \mathrm{d}\tau. \\ &\text{s.t.} \int_t^{\infty} e^{-\int_t^s r(\tau)\mathrm{d}\tau} c(s)\mathrm{d}s \leq \int_t^{\infty} e^{-\int_t^s r(\tau)\mathrm{d}\tau} y(s)\mathrm{d}s. \end{aligned} \tag{2}$$

Goods and labor markets With representative households, the total consumption of goods $C(t) = Lc(t)$. Conversely, the total output of goods $Y(t)$ aggregates the output of each individual firm. Then, goods market clears at time t if

$$C(t) = Y(t) \iff Lc(t) = \int_0^{\infty} zf(z,t)\mathrm{d}z. \tag{3}$$

Similarly, labor market clears at time t if

$$\int z\eta(z,t)f(z,t)\mathrm{d}z = L, \tag{4}$$

in which the LHS is the total labor demand by aggregating individual labor demand $z\eta(z,t)$ over firms, and the RHS is the inelastic labor supply L.

3.2 Equilibrium Path

The distribution of firm productivity summarizes the aggregate state of the economy. Most importantly, it determines the return of each search for all firms and evolves with the joint search behaviors across firms. Given firms' decisions on search intensity $\eta(z,t)$, the fraction of firms with productivity below z undergoes the following changes between time t and $t+\mathrm{d}t$:

$$F(z,t+\mathrm{d}t) = \int_0^z \left[1 - \eta(x,t)\mathrm{d}t + \eta(x,t)\mathrm{d}t F(z|z \geq x,t)\right] f(x,t)\mathrm{d}x.$$

In words, firms with productivity below z at $t+\mathrm{d}t$ are those which had productivity below z at time t and did not encounter another firm with productivity greater than z during the interval $[t,t+\mathrm{d}t]$. These firms either failed to meet any other firms or only encountered firms with productivity below z. For a firm with productivity x at time t, the former event occurs with productivity $1 - \eta(x,t)\mathrm{d}t$, while the latter event occurs with productivity $\eta(x,t)\mathrm{d}tF(z|z \geq x,t)$. The bracket term of the integrand is then the probability that a firm x at time t has productivity below z at $t+\mathrm{d}t$. The total fraction of firms with productivity below z at $t+\mathrm{d}t$ is then obtained by integrating the remaining firms with productivity below z at t. By rearranging terms and considering it at the limit as $\mathrm{d}t \to 0$, I derive the following Kolmogorov Forward Equation on the productivity distribution:

$$\frac{\partial F(z,t)}{\partial t} = -\int_0^z \eta(x,t)(1 - F(z|z \geq x,t))\mathrm{d}F(x,t). \tag{5}$$

It is evident that the productivity distribution is stochastically increasing over time since firms' productivity never decrease.

The initial productivity distributions plays a significant role in the transition dynamics. As mentioned in section 2.1, the standard notion of right tail index applies only to thick-tailed distributions. Then, the initial productivity distribution is restricted to the class of thick-tailed distributions in order to obtain meaningful changes in the right tail thickness.[12] Furthermore, I impose the following assumption 1 that the initial productivity distribution is exactly Pareto. With this assumption, I am able to obtain a closed-form characterization of the complete transition dynamics. Section 6 contains further results that apply to more general cases.

Assumption 1. *The initial productivity distribution is Pareto, i.e.,*

$$F(z,0) = 1 - z^{-k_0} \tag{6}$$

for $z \geq 1$ *and* $k_0 > 1$.

This condition implies that the initial distribution is a finite mean, has a lower support 1, and has the right tail index k_0. I am now ready to describe the equilibrium as the following.

Definition (Equilibrium)**.** A competitive equilibrium for this economy consists of wages $w(t)$, interest rates $r(t)$, firms' value functions $v(z,t)$, firms' search intensity $\eta(z,t)$, household's consumption $c(t)$, and the productivity distribution $F(z,t)$ that satisfy the following:

[12] The tail index will be infinity for thin-tailed distribution such as normal distributions. One can hardly conclude anything if the tail index is changing from infinity to infinity. If, instead, the tail index drops to a finite number, it is then without loss to start from there, and the thick-tailedness assumption applies.

(i) Given $\{w(t), r(t), F(z,t)\}$, $v(z,t)$ solves the HJB equation (1), and $\eta(z,t)$ is the associated policy function;

(ii) Given $\{w(t), r(t)\}$, $c(t)$ solves the households' problem (2);

(iii) Both good and labor markets clear, i.e., (3) and (4) are satisfied;

(iv) $F(z,t)$ solves the KFE (5) given $\eta(z,t)$ and satisfies the initial condition (6).

Finally, I introduce the notion of asymptotic balanced growth path (BGP). While an exact BGP assumes that consumption growth is always constant over time, an asymptotic BGP only requires that consumption growth is constant in the limit.

Definition (BGP)**.** An asymptotic balanced growth path (BGP) is an equilibrium in which consumption growth converges to a constant $g > 0$, i.e.,

$$\lim_{t \to \infty} \frac{\dot{c}(t)}{c(t)} = g.$$

3.3 Discussion on the search assumption

A slight but crucial deviation of this model from existing idea search models is the assumption that firms only search among more productive firms. The following discussion delves into the plausibility of this assumption, considering three key aspects. First, this assumption yields implications that align with the widely recognized Gibrat's law, while standard assumptions fall short in this regard. Second, there are supportive empirical evidence from the technology adoption literature. Third, public information and selection provide intuitions behind this assumption.

Gibrat's law Gibrat's law is an important empirical regularity about firms' growth which states that the growth rate of firms on average are independent of their size. Therefore, I consider it a useful criterion to evaluate idea search models with different search assumptions. Notably, models in the earlier literature such as Lucas and Moll (2014) and Perla and Tonetti (2014) assume that conditional on a realized meeting, firms take a random draw among all firms in the economy. The probability of meeting a particular firm will be the same irrespective of the productivity of the firm that does the searching. More productive firms then find idea search less rewarding since they are less likely to meet firms with higher productivity. Then, they search less intensively and grow less from each search. In short, Gibrat's law does not hold in these models.

In contrast, Gibrat's law holds in my model. Once a meeting realizes, firms only draw from firms that are more productive than they are. Hence, the probability of meeting another firm depends on the productivity of the searcher, or more precisely, the productivity gap between the searching and meeting firms. I will show later that in equilibrium, all firms face the same source distributions in terms of relative productivity and search at the same intensity. The expected growth rate is then the same across all firms.

A more subtle and technical point about Gibrat's law is that it disciplines a wide class of idea search models. I elaborate on this point in section 6 by showing that among those models, Gibrat's law holds if and only if the right tail of the firm size distribution becomes thicker. With additional functional form assumptions, my model turns out to be the only one that is consistent with Gibrat's law and a thickening right tail.

Empirical Support Technology adoption is perhaps the most relevant economic activity that idea search models intend to capture. Data on firms' adoption behaviors then provide direct empirical evidence to validate model predictions under various search assumptions. Following the discussion in the last paragraph, the search assumption in existing models tends to imply that all firms have the same probability of adopting state-of-the-art technology. Low-productivity firms can leapfrog to adopt the most advanced technology without getting them upgraded to medium-level technology first. Whereas, my assumption implies that technology adoption is a much more gradual process. Medium-productivity firms are closer to top-productivity firms and are more likely to adopt the most advanced technology than low-productivity firms.

Comprehensive data on firms' adoption behaviors is fairly uncommon. In a recent paper, Cirera et al. (2022) study detailed technology adoption decisions using a novel Firm-level Adoption of Technology (FAT) survey. The survey has data on the sophistication of technologies used at the business function level for firms in 11 mostly developing countries, making it suitable for studying within-industry cross-firm heterogeneity in technology adoption. Consistent with the implications of my assumption, they find that technology leapfrogging is rare. Specifically, they estimate the probability of firms of different sizes adopting digital and frontier technologies, including both general-purpose technologies and sector-specific business function technologies. As a result, the probability of adopting frontier technologies increases with firms' size. In addition, larger firms already use more sophisticated technologies. Therefore, they conclude that firms gradually upgrade their technology instead of jumping into the state-of-the-art.

Potential Mechanisms The reduced-form assumption that firms exclusively search among more productive firms can be intuitively understood through various mechanisms. Firstly, the availability of abundant public information allows firms to easily discern which firms are more productive than others. This information can include indicators such as market capitalization, employment size, number of patents, third-party rankings, and more. In a broader sense, it is common for individuals or organizations to focus their efforts on learning from peers who are known to be superior in the profession. An analogous example in the realm of research is the organization of seminars. Researchers often invite more established speakers from higher-ranked institutions.

Secondly, the process of selection and sorting connects the matching probability with the characteristics of the parties to be matched. This observation aligns with my assumption, which implies that firms are most likely to learn from firms with slightly higher productivity than their own. Managers of firms of similar size tend to share similar educational backgrounds and working experiences (Kodama and Li, 2018), fostering increased interactions among themselves. Additionally, the spatial literature presents extensive evidence on the spatial sorting of firms. (Combes et al., 2012; Gaubert, 2018). Consequently, knowledge diffusion tends to be stronger among firms that are similar in size due to their closer geographical distance. This mechanism can also be applied to the earlier example of seminars. Academic visits between low-ranked and top-ranked institutions are relatively more limited, as researchers in these places often have distinct past experiences and attend different conferences.

4 Transition Dynamics

This section provides a complete analytical characterization of the equilibrium path. In Section 4.1, I derive the equilibrium trajectory of the firm size distribution and discuss its implications on firm growth and output growth. In Section 4.2, I present similar transition dynamics results on equilibrium prices. In Section 4.3, I complement the baseline results with discussions on the selection of equilibrium. Lastly, I conduct numerical exercises in Section 4.4 to confront the model predictions with data. All proofs are relegated to Appendix B.

4.1 The dynamics of firm size distribution

Notice that the aggregate state variable—the productivity distribution—is an infinite dimensional object. Hence, I introduce the following two results and show that the right tail index is a sufficient statistic of the aggregate state.

Lemma 1. *In equilibrium, the value function is linear in productivity, i.e.,* $v(z,t) = v(t)z$*.*

This first result comes from the linearity of the HJB equation (1), which implies the gains from each search must be non-positive for all firms. Otherwise, firms will demand an infinite amount of labor. In addition, it is independent of the initial conditions on the productivity distribution. Nevertheless, assuming a Pareto initial distribution yields significant gains in tractability. As suggested by the proposition below, the greatest convenience of having a Pareto initial distribution is that the productivity distribution remains Pareto along the equilibrium path. Consequently, the Pareto parameters of the distribution are sufficient to characterize the aggregate state.

Proposition 1. *With assumption 1, there exists an equilibrium with the following properties.*

(i) The search intensity is invariant with regard to productivity, i.e., $\eta(z,t) = \eta(t)$*;*

(ii) The equilibrium productivity distribution at time t *is*

$$F(z,t) = 1 - z^{-k(t)} \quad \text{for} \quad z \geq 1, \tag{7}$$

in which $k(t)$ *satisfies*

$$\frac{\dot{k}(t)}{k(t)} = -\eta(t), \tag{8}$$

and $k(0) = k_0$*.*

This proposition states that despite heterogeneity in productivity, all firms search for ideas at the same intensity. The resulting productivity distribution is always a Pareto distribution, in which the scale remains constant, while the shape varies over time. Furthermore, the growth rate of the shape parameter is governed solely by the uniform search intensity. Proposition 1 delivers two important implications on the right tail thickness and firms' individual growth, respectively.

A thickening right tail In line with previous empirical findings, the model predicts a positive relationship between the right tail thickness of the firm size distribution and output per capita. With Pareto distributions, the shape parameter k represents the right tail index.[13] Based on Equation (8), it follows that with positive $\eta(t)$, the right tail index $k(t)$ exhibits a strict decrease at the given moment, resulting in a thicker right tail. That all firms search at a positive intensity ($\eta(t) > 0$) is also guaranteed in equilibrium as the labor market clears. Additionally, in the absence of population growth, the growth in output per capita is simply the growth in aggregate output, which is the sum of all firms' output. Therefore, the aggregate output can be expressed as the mean productivity of firms, given a unit measure of firms and the production function. Due to the Pareto nature of the productivity distribution, the mean productivity is given by $\frac{k(t)}{k(t)-1}$, which decreases with respect to $k(t)$ and consequently grows over time. Collectively, the model suggests that the process of idea search can lead to a thickening right tail along the growth path since it generates growth by increasing the proportion of relatively more productive firms.

Gibrat's law The model implies that firms always have the same expected output growth, i.e., Gibrat's law holds. While Gibrat's law is a well-known empirical regularity of firm growth, existing idea search models, in general, are incompatible with it. I leave a detailed discussion on why existing models fail to obtain Gibrat's law to Section 6.3. Instead, I discuss in below why Gibrat's law is consistent with this model.

In the current setting, the output growth of firms is equivalent to their productivity growth, which is jointly determined by the number and quality of firms' idea searches. On one hand, Lemma 1 shows that gains from each search are proportionate to firm productivity when the productivity distribution follows a Pareto distribution. This result leads to the first part of Proposition 1: all firms are indifferent to any level of search intensity in equilibrium and can agree upon the same search intensity. On the other hand, firms engage in random draws during realized meetings, and the distribution of productivity growth from each draw remains the same for all firms. This property stems from the fact that Pareto distributions are preserved under left truncation. Let $\tilde{x}(z,t)$ represent the productivity growth of firm z in each meeting, i.e., $\tilde{x}(z,t) = \frac{\tilde{X}(z,t)}{z}$. It can be further demonstrated that $\text{Prob}(\tilde{x}(z,t) \leq y) = F(y,t)$. This idea search process has a similar implication to the traditional assumption in quality ladder models, where the size of quality improvement is independent of the current quality. In summary, the expected productivity growth of firm z is the product of the search intensity and the expected productivity growth per search. Since both

[13] This is easy to check with the previous definition.

components are independent of firm productivity, the expected productivity growth is the same across firms and can be expressed as:

$$\lambda(z,t) \quad \frac{\mathsf{E}[\mathrm{d}z]}{z} = \eta\left(\mathsf{E}[\tilde{x}(z,t)] \quad 1\right) = \quad \frac{\dot{k}(t)}{k(t)\left(k(t) \quad 1\right)}, \tag{9}$$

where the last equality is obtained using equation (8).

With Proposition 1, it is sufficient to characterize the equilibrium path by the trajectory of the Pareto shape parameter. Equation (8) outlines a differential equation on $k(t)$ given the trajectory of $\eta(t)$. Besides, the labor market clearing condition (4) pins down the search intensity $\eta(t)$ as a function of the aggregate state. Integrating out the LHS of (4) using equation (7), I obtain that

$$\eta(t)\frac{k(t)}{k(t) \quad 1} = L.$$

Plugging it into the law of motion on $k(t)$, the following differential equation describes the dynamics of $k(t)$:

$$\dot{k}(t) = \quad L\left(k(t) \quad 1\right).$$

This is a first-order linear ordinary differential equation. With initial condition $k(0) = k_0 > 1$, its solution has the following simple expression:

$$k(t) = 1 + (k_0 \quad 1)e^{\quad Lt}. \tag{10}$$

Based on the dynamics of the aggregate state given by equation (10), it is straightforward to derive the equilibrium trajectories of other variables. Figure 4 illustrates the evolution of the firm size distribution, $F(\ ,t)$, and the average search intensity, $\eta(t)$. Furthermore, I obtain the following two asymptotic results. They demonstrate that the model economy converges to a state in which firms' productivity follows Zipf's distribution, and the output per capita grows at a constant rate.

Zipf's law An immediate consequence of equation (10) is that $k(t)$ converges to 1. In other words, the productivity distribution converges to Zipf's distribution, which is a Pareto distribution with shape parameter 1. The so-called Zipf's law is the form taken by a remarkable number of regularities in economics and finance.[14] If X denotes the variable of interest, Zipf's law states that $P(X > x) \quad 1/x$. In the context of firms, Zipf's law states that advanced economies such as the US have firm size distributions very close to Zipf's distribution. This model then provides a growth-based theory of Zipf's law on firm size distributions. The process of idea search generates aggregate productivity growth by thickening the right tail of the productivity distribution. Rich countries are further displaced along the growth path, so their distributions are closer to the limit, which is Zipf's distribution. Additionally, it is worth noting that there are barely any restrictions

[14] Outside economics, Zipf's law is best known for its applicability in the word frequencies in natural languages.

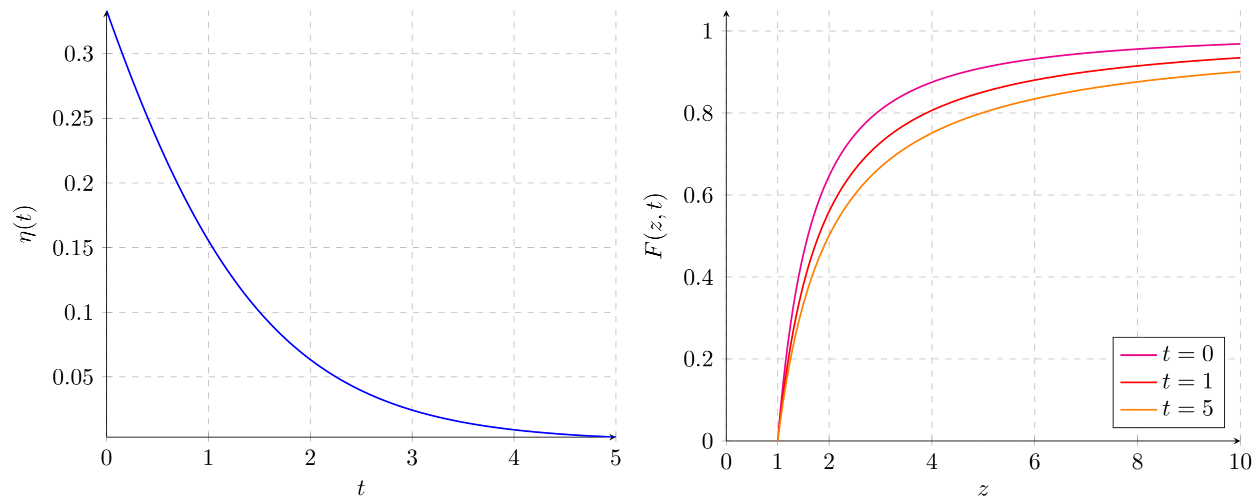


Figure 4: An illustration of $\eta(t)$ and $F(z,t)$

Notes. These figures illustrate the dynamics in the search intensity and productivity distribution with initial shape parameter $k_0 = 1.5$ and population size $L = 1$. The left figure plots the average search intensity, $\eta(t)$, over time. The right figure displays the selected CDFs of productivity, $F(z,t)$, at time 0, 1 and 5.

on the model parameters for the convergence to Zipf's law. With an initial Pareto distribution, the condition $k_0 > 1$ is necessary to ensure that the economy starts with finite output.

Asymptotic balanced growth path As discussed before, the output per capita is equal to the mean firm productivity divided by the measure of households, i.e., $y(t) = \frac{k(t)}{k(t)-1}\frac{1}{L}$. Differentiating it with respect to t, the growth rate of output per capita is as follows:

$$g(t) \equiv \frac{\dot{y}(t)}{y(t)} = \underbrace{\frac{1}{k(t)}}_{\to 1}\underbrace{\left(\frac{-\dot{k}(t)}{k(t)-1}\right)}_{=L} = \frac{L}{k(t)} \Longrightarrow g(t) \to g^* \equiv L, \tag{11}$$

i.e., $g(t)$ converges to the long-run growth rate g^* as $k(t) \to 1$. Therefore, the equilibrium path is an asymptotic balanced growth path, and the long-run growth rate is the average employment of researchers across firms.[15]

Two notable features of this asymptotic balanced growth path are worth emphasizing. First, the product market concentration rises steadily in an economy with nearly constant growth. Specifically, we may measure the product market concentration by a transformed right tail thickness measure, $k-1$. The equilibrium trajectory of $k(t)$ implies that $k-1$ decreases to zero at a constant

[15]It may seem like a scale effect in growth rate since the long-run growth rate is given by L, the population size. However, since the total measure of firms is normalized to one, the total employment of researchers coincides with the average employment per firm. The case in Section 5.1 makes it clear that the average employment of researchers per firm determines the growth rate. In principle, one can extend the model with both population growth and firm entry and still achieve asymptotic balanced growth. This extension is in the same spirit as a class of endogenous growth models without scale effect (Young, 1998), where extensive margin growth in the measure of firms offsets the population growth.

rate. In addition, commonly-used concentration measures such as HHI decrease on k, given that the underlying distribution of firms' sales is Pareto. This aspect of the model is particularly noteworthy in that it reconciles the tension between rising market concentration and constant growth. In standard growth models, the firm size distribution remains stationary along an exactly balanced growth path. However, recent studies present mounting evidence that the market concentration is increasing in the US—an economy marked with constant growth over a long history—even in the long run.

Second, equation (11) reveals a growth mechanism that distinguishes this model from existing idea search growth models in the literature. In existing models, idea search processes enable the existence of balanced growth paths (BGPs) with stationary productivity distributions. Therefore, growth is sustained without affecting the productivity distribution. Namely, the relative weight between low-productivity and high-productivity firms remains constant. In contrast, the idea search process here increases aggregate output by relocating mass from relatively low-productivity firms to relatively high-productivity firms. This process, referred to as ***tail growth***, sustains growth by thickening the right tail of the productivity distribution. In this context, Zipf's distribution emerges as a natural limit, as it represents the minimum level of thickness required for the mean productivity to approach infinity. Conversely, constant growth implies that the distribution must converge to Zipf's distribution; otherwise, growth will eventually cease. However, it should be noted that constant growth is only a sufficient condition for the convergence to Zipf's distribution. I leave more discussion on this point to Section 6.3.

I conclude the equilibrium analysis in this section with the following proposition, which summarizes the four major properties discussed above.

Proposition 2. *With assumption 1 and $\rho > L(1-\theta)$, there exists an equilibrium that satisfies the following:*

(i) The right tail of the productivity distribution thickens along the transition path;

(ii) The expected productivity growth is the same across all firms, i.e., Gibrat's law holds;

(iii) The productivity distribution converges to Zipf's distribution;

(iv) The equilibrium is an asymptotic balanced growth path.

The parametric condition $\rho > L(1-\theta)$ ensures that the representative households' utility is finite and the relevant transversality condition is satisfied.

4.2 The dynamics of equilibrium prices

I complete the analysis of the transition dynamics with a discussion on the dynamics of prices. Two prices, the wage of researchers and the interest rate, clear the goods and labor markets. I first characterize the dynamics of the interest rate. The Euler equation implies that $r(t) = \theta\frac{\dot{c}(t)}{c(t)} + \rho$.

Since the goods market clears and output per capita growth satisfies equation (11), the equilibrium interest rate can be written as follows as a function of the aggregate state $k(t)$, i.e.,

$$r(t) = \theta \frac{L}{k(t)} + \rho.$$

Then, $r(t)$ increases to $r \;\; = \theta L + \rho$ as the right tail index $k(t)$ decreases to 1.

Next, I discuss the dynamics of the wage of researchers. Proposition 1 implies that the marginal cost of firms' idea search equals its marginal return. Therefore, the productivity-adjusted expected gains from each search determine the wage of researchers:

$$w(t) = \frac{1}{z} \int_z^{1} [v(x,t) \quad v(z,t)] \mathrm{d}F(x j x \quad z,t) = \frac{v(t)}{k(t) \quad 1},$$

in which the second equality is obtained by simplifying the integral using Lemma 1 and Proposition 1. Also notice that firms' value of unit productivity, $v(t)$, is the present value of a dividend flow of unit output, i.e.,

$$v(t) = \int_t^{1} e^{\quad \int_t^x r(s) \mathrm{d}s} \mathrm{d}x.$$

This equation is intuitive since each unit of productivity yields one unit of output at each instant. The value of an additional unit of productivity is then the discounted sum of its output flow, or equivalently, the inverse of average interest rate $\tilde{r}(t)$.[16] As $r(t)$ converges to $r \;\;$, the value of unit productivity $v(t)$ converges to $1/r \;\;$. The growth rate of the researchers' wage then converges to a constant, as the transformed right tail thickness $k(t) \quad 1$ decreases at a constant rate to zero. Therefore, I consider instead a normalized wage $\tilde{w}(t)$, which is the ratio between the wage of researchers and households' income. Formally,

$$\tilde{w}(t) = \frac{w(t)}{y(t)} = \frac{w(t)}{\frac{k(t)}{k(t) \quad 1} \frac{1}{L}} = \frac{v(t)L}{k(t)} = \frac{g(t)}{\tilde{r}(t)},$$

which can be further written as the ratio between the output per capita growth and the average interest rate. The following lemma characterizes the dynamics of the normalized wage.

Lemma 2. *The normalized wage $\tilde{w}(t)$ increases to $\tilde{w} \;\; = 1/(\rho/L + \theta)$.*

It is evident to see that the long-run normalized wage is the ratio of the long-run growth rate (L) and long-run interest rate ($r \;\;$). The monotonicity, however, is less evident since both the growth rate and average interest rate are increasing. Note that the instantaneous interest rate $r(t)$ is the sum of the discount rate (ρ) and a linear term of the instantaneous growth rate ($\theta g(t)$). Then, $r(t)$ grows less than $g(t)$, i.e., $g(t)/r(t)$ increases over time. When t is sufficiently large, the average interest rate $\tilde{r}(t)$ and the instantaneous interest rate $r(t)$ are very close to each other. The above

[16] See the proof of Lemma 1 for the detailed derivation. The average interest rate $\tilde{r}(t)$ is a uniform interest rate that discounts the dividend flows to match the same present value, i.e., $v(t) = \int_0^{1} e^{\quad \tilde{r}(t)s} \mathrm{d}s = 1/\tilde{r}(t)$.

intuition carries over to see that the normalized wage increases over time. In Lemma 2, I prove that the same intuition holds with average interest rate $\tilde{r}(t)$. As the economy grows over time, the value of unit productivity is more discounted with higher interest rates, but the size of productivity improvement from each meeting increases with a thicker right tail. Together, the return of each idea search increases over time at a growth rate faster than the output per capita. In words, idea search becomes more valuable with thicker right tails of the firm size distribution.

Lastly, this property has implications for R&D intensity. Since the payment to researchers is the only expenditure to increase productivity in this model, the normalized wage is equivalent to the share of R&D expenditure to GDP. The model then predicts a positive relationship between R&D intensity and the level of economic development.

4.3 A discussion on equilibrium selection

The basis of the previous analysis is Proposition 1, which constructs an equilibrium of a particular form and proves its existence. However, it does not rule out other types of equilibrium. With a linear cost function of idea search, firms are indifferent between any level of search intensity in the baseline equilibrium. That is the reason why Gibrat's law holds, and the productivity distribution remains Pareto. Conversely, an asymmetric equilibrium arises if not all firms agree on the same search intensity. Some firms could search strictly more intensively than other firms, despite that all of them face zero net gain of search in that time, and the labor market still clears. The productivity distribution will lose its Pareto shape afterwards but still evolve according to the KFE (5).

I argue that the baseline symmetric equilibrium should be preferred to those equilibria. While this set of alternative equilibria is essentially the product of linear search cost, the following shows that the baseline equilibrium is the limiting case of a class of equilibria with non-linear search cost. To see this, I consider the case that an additional adjustment cost is required for firms to complete the same amount of searches. Particularly, a firm z now has to hire $z(\eta + g(\eta))$ researchers to achieve a search intensity of η. g is assumed to be a strictly increasing and strictly convex function and satisfies $g(0) = g'(0) = 0$. Immediately, Lemma 1 does not always hold with adjustment cost: a linear value function is not a feature of all equilibria. Nevertheless, Proposition 3 presents that there still exists an equilibrium satisfying all properties stated in Lemma 1 and Proposition 1.

Proposition 3. *With assumption 1 and adjustment cost $g(\eta)$, there exists an equilibrium with the following properties:*

(i) the value function is linear in productivity, i.e., $v(z,t) = v(t)z$;

(ii) The search intensity is invariant of productivity, i.e., $\eta(z,t) = \eta(t)$;

(iii) The equilibrium productivity distribution at time t is

$$F(z,t) = 1 - z^{-k(t)} \quad \textit{for} \quad z \geq 1,$$

in which $k(t)$ satisfies

$$\frac{\dot{k}(t)}{k(t)} = -\eta(t),$$

and $k(0) = k_0$.

The proof of proposition 3 follows very closely the arguments for proving proposition 1. In below, I verify that this equilibrium with adjustment cost inherits all properties stated in Proposition 2 and discuss its connections with the baseline equilibrium. First, Gibrat's law holds immediately in this equilibrium. With a linear value function, the first order condition on η implies a unique optimal search intensity, i.e.,

$$\frac{v(t)}{k(t)-1} = (1 + g'(\eta(t)))w(t).$$

In contrast to the baseline, the adjustment cost creates a wedge between the gains from each search and the researchers' wage, which pins down the search intensity and resolves indeterminacy.

Next, there is still convergence to Zipf's distribution. The labor market clearing condition implies that

$$(\eta(t) + g(\eta(t)))\,\frac{k(t)}{k(t)-1} = L.$$

Taking this equation into the law of motion on k, I obtain that

$$k(t) = 1 + (k_0 - 1)\exp\left(-\int_0^t \tilde{L}(s)\mathrm{d}s\right), \tag{12}$$

in which $\tilde{L}(t) = \frac{L}{1+g(\eta(t))/\eta(t)}$. This is similar to the baseline equilibrium trajectory of $k(t)$ except that the transformed thickness measure $k-1$ now decreases to zero at a varying rate $\tilde{L}(t)$. Since k is decreasing, the monotonicity of g implies that η also decreases over time. $g(\eta)/\eta$ then increases on η due to the convexity of g. Therefore, $\tilde{L}(t)$ increases over time, and $\tilde{L}(t) \geq \tilde{L}(0) > 0$. As t goes to infinity, the integral on $\tilde{L}(t)$ goes to infinity, and $k(t)$ goes to one. Then, the limiting productivity distribution is still Zipf's distribution. Technically, one can still solve for $k(t)$ by substituting η as a function of k into the law of motion, although there is generally no explicit solution. In comparison, the baseline equilibrium with linear search cost has the merits of admitting an explicit solution.

Lastly, the output growth still converges to a constant given by L. Recall that

$$\frac{\dot{y}(t)}{y(t)} = -\frac{\dot{k}(t)}{k(t)\,(k(t)-1)} = \frac{\tilde{L}(t)}{k(t)},$$

where the second equality is obtained using equation (12). With $k(t)$ decreasing to one, $\eta(t)$ is decreasing to zero to equalize the supply and demand of labor. $g(\eta(t))/\eta(t)$ also decreases to zero since $g'(0) = 0$. As a result, $\tilde{L}(t)$ converges to L. It is also worth noting that the wedge between the gains from each search and the researchers' wage vanishes as η goes to zero. Following the arguments in section 4.2, both wages and interest rates converge to those in the baseline equilibrium.

In conclusion, the baseline equilibrium is preferred over alternative equilibria with linear search costs as it emerges as the limiting case of a broader class of equilibria with generalized search costs. On one hand, for any equilibrium with the specified adjustment cost, its equilibrium path aligns with that of the baseline equilibrium in the long run, specifically for sufficiently large t. On the other hand, the baseline equilibrium can be understood as the equilibrium of a limiting economy, where the adjustment cost approaches zero.[17]

4.4 Numerical Exercises

Despite its parsimony, the model in section 3 makes theoretical predictions that are empirically testable. Specifically, the simple model makes two strong predictions: 1) output per capita y is proportional to $k/(k \quad 1)$, and 2) $k \quad 1$ decreases to 0 at a constant rate which is equal to the long-run growth rate.

In appendix G, I confront both predictions with changes in the US GDP per capita using estimates for k. Both predictions are consistent with the data. For the US from 1978 to 2019, the actual growth in output per capita aligns closely with changes in $k/(k \quad 1)$. Moreover, regressing $\ln(k \quad 1)$ on time, the estimated coefficient perfectly hits the well-known US growth rate of 2%. These results provide compelling evidence supporting the empirical relevance of the parsimonious model. Additionally, as both exercises link GDP per capita growth with changes in the right tail thickness, these results indicate that an important source of economic growth could be the aforementioned tail growth mechanism.

5 Policy Implications

In this section, I show that the model in section 3 delivers a novel perspective to evaluate size-dependent industrial policies. The idea search conducted by each firm in the economy generates an externality on other firms, as it impacts the productivity distribution, which, in turn, determines future search efficiency. Specifically, searches carried out by large firms contribute to the thickening of the right tail and have positive externalities on all firms within the economy. Conversely, searches by small firms exert fewer externalities on large firms. Thus, relative to first-best outcomes, large firms under-invest in idea search, and policy should encourage more searches by large firms.

Two policy exercises shed light on this policy implication. In the first exercise, I consider a scenario where a social planner selects a productivity threshold and implements an additional tax on firms below that threshold. Essentially, small firms are required to pay an extra cost

[17] Readers should be alerted that even with adjustment cost, the equilibrium in Proposition 3 is not guaranteed to be unique. This is because my proof only shows the uniqueness of equilibrium with value functions linear in productivity. There might be equilibria with non-linear value functions and non-Pareto productivity distributions. To prove whether this equilibrium is unique *or not* is beyond the scope of this paper. It is identified as an open mathematical challenge in Achdou et al. (2014) to show the uniqueness of a solution of coupled PDE systems (the HJB and the KFE) in idea search models.

for each search they undertake. As a result, only firms above the threshold conduct search in the equilibrium. I prove that the equilibrium long-run growth rate increases with the level of productivity threshold, and so does social welfare. In the second exercise, I solve the social planner's problem and show that the optimal individual search intensity grows with individual productivity, following an approximate power rate function. The socially optimal search intensity then differs from the equilibrium intensity, which is uniform across all firms. Both exercises demonstrate that policies favoring large firms better capture the diffusion externality and improve welfare.

5.1 A tax on the small firms

In the first exercise, the policymaker chooses a threshold z^* and imposes a positive searching tax τ on firms below the threshold. That is, if a firm has productivity less than z^*, its unit search cost becomes $z(1+\tau)w(t)$ rather than the original $zw(t)$. On the contrary, firms with productivity above that threshold retain the original unit search cost $zw(t)$. The tax revenue is further rebated to the households as a lump sum transfer. The following proposition characterizes the equilibrium with tax τ and threshold z^*.

Proposition 4. *For any threshold productivity $z^* \geq 1$, there exists an equilibrium such that the equilibrium productivity distribution satisfies*

$$F(z,t) = \begin{cases} 1 - z^{-k_0} & \text{if } z \leq z^*, \\ 1 - (z^*)^{k(t)-k_0} z^{-k(t)} & \text{if } z > z^*, \end{cases} \tag{13}$$

in which $k(t) = 1 + (k_0 - 1)\exp\left(-L(z^*)^{k_0-1}t\right)$. In addition, the output per capita growth converges to $L(z^*)^{k_0-1}$.

In the resulting equilibrium, firms below the threshold do not search, so that part of the productivity distribution stays constant. The other part of the distribution evolves just as before, as firms above the threshold still search at the same intensity. The only difference is that with fewer firms searching, each searching firm gets to hire more researchers and increases the average search intensity. A higher average employment of researchers then accelerates output growth. Taxing small firms raises the effective labor endowment per firm by discouraging small firms' inefficient use of labor. Moreover, the long-run growth rate $L(z^*)^{k_0-1}$ can be arbitrarily large since the policymaker can always choose arbitrarily large thresholds z^*. Given sufficiently large long-run growth, social welfare will be infinite regardless of the discount factor.

On the technical side, Proposition 4 also illustrates the multiplicity of equilibrium due to the linear cost assumption. As shown in the proof, an equilibrium with distribution (13) is supported in the tax-free ($\tau = 0$) or baseline economy for any threshold. That is, there is a continuum of threshold equilibria in the baseline economy. Since all firms are indifferent between any level of search with the initial Pareto distribution, any subset of firms can possibly be inactive at time 0. However, once the search process begins, the resulting piece-wise Pareto distribution creates a strictly lower return for firms below the threshold, leading them to stay inactive indefinitely. Among all threshold equilibria, I focus on the one with the lowest threshold for reasons discussed in Section 4.3. The equilibrium in the tax-free economy is then the threshold equilibrium with minimum level $\underline{z} = 0$. With a positive tax, there is no threshold equilibrium with a threshold below the chosen level z^* since firms below z^* are discouraged from searching even at time 0. Then, the minimum threshold binds at the chosen threshold, i.e., $\underline{z} = z^*$, and distribution (13) describes the resulting equilibrium.

5.2 The planner's problem

In the second policy exercise, a utilitarian social planner solves the following maximization problem on the present value of the social welfare given an initial distribution f:

$$\begin{aligned} W(f) = & \max_{\{c(\omega,t),\eta(z,t)\}} \int_0^{\infty} e^{-\rho t} \int_{\Omega} u(c(\omega,t)) \mathrm{d}\omega \mathrm{d}t \\ \text{s.t.} \quad & \int_{\Omega} c(\omega,t)\mathrm{d}\omega \le \int_0^{\infty} z f(z,t)\mathrm{d}z, \\ & \int_0^{\infty} z\eta(z,t) f(z,t)\mathrm{d}z \le L, \\ & \frac{\partial f(z,t)}{\partial t} = f(z,t)\left[\int_0^z \eta(x,t)\frac{f(x,t)}{1-F(x,t)}\mathrm{d}x - \eta(z,t)\right], \\ & f(z,0) = f(z). \end{aligned} \tag{14}$$

The first two constraints are the respective goods and labor market clearing conditions. The third equation is the law of motion on the density of the productivity distribution, and the last one is the initial condition. The social planner chooses the full paths of consumption $c(\omega,t)$ and search intensity $\eta(z,t)$ for each household and firm.

The optimal control problem (14) is challenging because the state variable is an infinite dimensional object–a distribution. Lucas and Moll (2014) and Nuño and Moll (2018) study similar planner's problem in heterogeneous agent models. I borrow techniques developed there to characterize the solution to the planner's problem. The general idea is to transform the above problem into a system of finite-dimensional partial differential equations and solve the policy function from there. In particular, I work with $w(f,z)$, which is the Gateaux derivative of $W(f)$ at point z with a Dirac delta function as increment.[18] $w(f,z)$ is then the marginal social value of a firm z with aggregate state f. Moreover, let $w(z,t) \equiv w(f(\cdot,t),z)$. $w(z,t)$ is then the marginal social value along the trajectory of the productivity distribution $f(z,t)$, which results from the optimal policies.

[18] Formally,

$$w(f,z) \equiv \frac{\delta W(f)}{\delta f(z)} = \lim_{\alpha \to 0} \frac{W(f+\alpha\delta_z) - W(f)}{\alpha} = \frac{d}{d\alpha} W(f+\alpha\delta_z)|_{\alpha=0},$$

in which α is a real scalar, and $\delta_z(x) = \delta(x-z)$ with δ the Dirac delta function.

I show in Appendix C that $w(z,t)$ satisfies the following partial differential equation:

$$
\begin{aligned}
\rho w(z,t) = \hat{\lambda} z + \frac{\partial w(z,t)}{\partial t} + \max_{\eta} \left\{ \eta \int_z^{\infty} [w(y,t) - w(z,t)] \frac{f(y,t)}{1 - F(z,t)} \mathrm{d}y - \hat{\mu} z \eta \right\} \\
+ \int_0^{\infty} \left\{ w(y,t) \int_0^{\max\{y,z\}} \eta^*(x,t) \frac{\varphi(x,t)}{1 - F(x,t)} \mathrm{d}x + w(z,t) \int_0^{z} \eta^*(x,t) \varphi(x,t) \mathrm{d}x \right\} f(y,t) \mathrm{d}y,
\end{aligned}
\tag{15}
$$

in which $\hat{\lambda}$ and $\hat{\mu}$ are the respective Lagrangian multipliers of the goods and labor market clearing conditions.

Equation (15) is the counterpart of HJB equation (1) for the social planner. The LHS is the flow social value of firm z based on the discounting factor ρ rather than the interest rate r_t. Three items on the RHS compose this flow value. The first item is the value of static output with $\hat{\lambda}$ the shadow price of the output in utility. The second item is the option value which is the sum of an incremental change, $\partial w(z,t)/\partial t$, and the net return of idea searches. The third and additional item that only shows up in the social planner's HJB equation is the last term in (15), capturing the externality of other firms' learning. The conditional source distribution makes the externality term very complicated since any firm can affect others in various ways. To see this, consider an increase in the density $f(z)$ and a firm y with $y < z$. The density of the source distribution facing by firm y is $\frac{f(x)}{1-F(y)}$ for $x \geq y$. An increase in the portion of firm z increases the likelihood for firm y to meet a firm z. Yet it also decreases the likelihood of meeting other firms x with $x \neq z$, which could also benefit firm y. The following proposition characterizes the optimal search policy associated with the dynamic programming problem (15).[19]

Proposition 5. *Given that $F(z,t)$ has tail index $k(t)$, the optimal search intensity is regularly varying with exponent $(k(t) - 1)/2$, i.e.,*

$$
\eta^*(z,t) = z^{\frac{k(t)-1}{2}} L(z,t),
$$

in which $L(z,t)$ is a slow varying function. Moreover, $L(z,t)$ is constant over z if $f(z,t)$ is exactly Pareto.

Proposition 5 indicates that the optimal search strategy allocates more search intensity to more productive firms at an approximate power rate. The optimal allocation echoes with the prior intuition that large firms search too little in the market equilibrium. To illustrate this further, figure 5 visualizes the comparison between search intensities in the competitive equilibrium and the planner's problem at time 0. With assumption 1, Proposition 5 implies that the optimal search intensity $\eta^*(z,0) = Cz^{\frac{k_0-1}{2}}$ for some constant C. The labor market clearing condition pins down this constant to be $\frac{(k_0-1)L}{2k_0}$. Contrastingly, all firms search at the same intensity $\eta(z,0) = \frac{k_0-1}{k_0}L$ in the competitive equilibrium, as shown in Section 4. It is obvious that relative to the market outcome, a social planner would reallocate researchers from low-productivity firms to high-productivity firms.

[19] See Section 6.2 for the definitions of tail index and regular and slow variations.

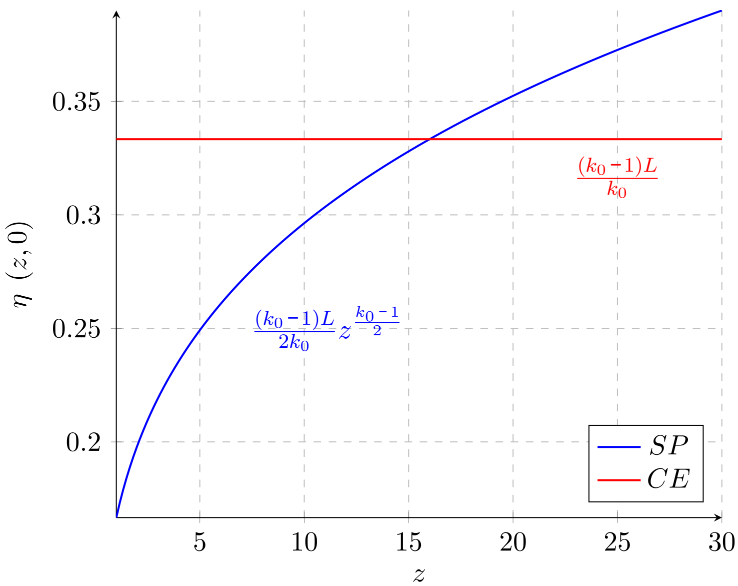


Figure 5: Comparing search intensities ($k_0 = 1.5$, $L = 1$)

Notes. The blue line (SP) plots the optimal search intensity $\eta\ (z, 0)$, and the red line (CE) the equilibrium search intensity $\eta(z, 0)$. Each is given by the respective formula in label with $k_0 = 1.5$ and $L = 1$.

Another implication of the optimal search policy is that it induces jumps in the tail index of the productivity distribution. For a distribution $F(z,t)$ with tail index $k(t)$, its tail index drops instantly from $k(t)$ to $\frac{k(t)+1}{2}$, once the optimal search policy is in place. Considering an infinitesimal time interval $[t, t+h]$, the law of motion of the productivity distribution implies that

$$\tilde{F}(z, t+h) = \tilde{F}(z,t)\left[1 + h\int_0^z \frac{\eta\ (x,t)}{x}\frac{xf(x,t)}{\tilde{F}(x,t)}\mathrm{d}x\right] = \tilde{F}(z,t)\left[1 + h\int_0^z x^{\frac{k(t)\ \ 1}{2}\ \ 1}\tilde{L}(x,t)\mathrm{d}x\right],$$

in which $\tilde{L}(z,t) = L(z,t)k(z,t)$ and is slow varying.[20] When $k(t) > 1$, the Karamata's theorem applies to show that the integral is regularly varying with exponent $\frac{k(t)\ \ 1}{2}$.[21] $\tilde{F}(z, t+h)$ is then a regularly varying function with exponent $\ \ k(t) + \frac{k(t)\ \ 1}{2} = \ \frac{k(t)\ \ 1}{2}$ for arbitrary h. Therefore, there is a jump in the tail index from $k(t)$ to $\frac{k(t)+1}{2}$. If $k(t) = 1$, there will be no changes in the tail index since the integral term is also slowly varying.

Now it is clear what the optimal search policy does to the whole economy. Given that the initial distribution is Pareto with shape $k_0 > 1$, a sequence $\left\{1 + \frac{k_0\ \ 1}{2^n}\right\}$ characterizes the dynamics in the tail index. That is, the tail index declines to one in countably many steps. In continuous time, the tail index immediately becomes one, and Zipf's law holds. Figure A.3 illustrates this process in the appendix. Furthermore, the output must become infinite instantly. Otherwise, one can choose a common $\hat{\eta} > 0$ for all firms that respect the labor market clearing condition. Such a choice then decreases the tail index to below one and violates the optimality.[22] Finally, although both policy

[20] $k(z,t) = zf(z,t)/\tilde{F}(z,t)$. Since $f(z,t)$ is regularly varying, $k(z,t)$ is slowly varying.

[21] For the Karamata's theorem, see, for example, theorem 1.5.11 of Bingham et al. (1987).

[22] A tail index below one must imply infinite output and raise welfare. Then, there is no interval of time in which

exercises imply an infinite present value of social welfare, the solution to the planner's problem is still stronger in the following sense. In any interval of time, any tax policy in the first exercise only generates finite social welfare, whereas the social welfare under the optimal search policy is always infinite. Albeit an unusual result, this exercise demonstrates the potential of the tail growth mechanism.

6 General Results

Preceding sections have shown that a simple idea search model generates a number of desirable empirical regularities, namely, Gibrat's law, the thickening of the right tail, and Zipf's law. In this section, I revisit these results in a more general framework and examine key assumptions behind them. In Section 6.1, I describe a general idea search model which nests my simple model and major idea search models in the literature. In Section 6.2, I introduce technical definitions and regularity conditions that are fundamental to the results with the general setup. In Section 6.3, I show that Gibrat's law and a thickening right tail discipline the choice of idea search processes. Furthermore, the dynamics of the tail index in the simple model extends to the general case. In Section 6.4, I establish that Zipf's law can hold under weaker conditions than those assumed in the simple model. Results in this section suggest that assumptions in the simple model are not arbitrary, and insights from there carry over to a more general environment.

6.1 A general idea search model

I begin with a description of a general idea search model, which encompasses the simple model in Section 3 and most models in the existing literature. All the notations follow very closely to those in the simple model. As before, there is a continuum of firms in the economy with heterogeneous productivity. The distribution of firm productivity at time t is captured by a cumulative distribution function (CDF) $F(\cdot, t)$ and has a well-defined density $f(\cdot, t)$.

Firms actively or passively obtain ideas with heterogeneous productivity. An idea has productivity z if adopting it enables firms to have productivity z. Therefore, I use productivity to index both firms and ideas, and a firm x adopts an idea z if and only if $z \geq x$. The arrival of ideas is a Poisson event that depends on idea productivity z, firm productivity x, and time t. A learning function $m(z, x, t)$ characterizes the process of idea arrivals by specifying the arrival rate of ideas with productivity at least z for firm x at time t. Furthermore, let $n(z, x, t)$ be the arrival rate of idea z for firm x at time t. I assume that $m(z, x, t)$ is differentiable with respect to z, and $m(z, x, t) = \int_z^{\infty} n(y, x, t)\mathrm{d}y$. There are no other productivity shocks to firms.

the output is finite, or the current strategy is suboptimal.

I focus on the case in which $m(z,x,t)$ is separable on $z \geq x$, i.e.,[23]

$$m(z,x,t) = \mu(x,t)\tilde{H}(z,t), \quad \forall z \geq x.$$

The arrival rate function is then the product of a firm-specific component $\mu(x,t)$ and an idea-specific component $\tilde{H}(z,t)$. Intuitively, $\mu(x,t)$ can be viewed as the search intensity of an individual firm. This definition further implies that for each x, $m(z,x,t)$ decreases on z, $\lim_{z\to\infty} m(z,x,t) = 0$, and $m(0,x,t) < \infty$. It follows that $\tilde{H}(z,t)$ also decreases on z, and $\lim_{z\to\infty} \tilde{H}(z,t) = 0$. Thus, $\tilde{H}$ is a valid complement CDF (under normalization) whenever it is bounded.[24] It is also useful to define $h = -\partial \tilde{H}(z,t)/\partial z$, or equivalently, $\tilde{H}(z,t) = \int_z^\infty h(z,t)\mathrm{d}z$. Even though it is not always a distribution function, I abusively refer to $\tilde{H}$ as the source distribution and h as the associated density.

I leave an extensive discussion on the generality of the learning function $m(z,x,t)$ to appendix D.2. Importantly, even with the separability assumption, this function remains compatible with complex learning processes and encompasses multiple forms of learning heterogeneity. In particular, the learning function accommodates scenarios in which firms can search for ideas at different intensities, draw ideas from different sources, and absorb ideas to different degrees. I offer a further demonstration of how the learning function summarizes the simple model in Section 3 and various idea search models in the literature.

The evolution of productivity distribution is a straightforward generalization of that in the simple model. The same intuition applies: newly added firms with productivity above z are those which previously had productivity below z but adopted ideas with productivity above z in the past interval. Formally, the following partial differential equation depicts how the productivity distribution evolves at time t with learning function $m(z,x,t)$:

$$\frac{\partial \tilde{F}(z,t)}{\partial t} = \int_0^z m(z,x,t)\mathrm{d}F(x,t) = \tilde{H}(z,t)\int_0^z \mu(x,t)f(x,t)\mathrm{d}x, \tag{16}$$

where the tilde notation denotes the complement CDF. With an initial distribution $F_0(z)$, the solution to the PDE (16) characterizes the transition of the productivity distribution.

Similar to the simple model, the productivity process in PDE (16) avoids explicit modeling of firms' entry and exit. I discuss in appendix D.4 that such omission is with little loss of generality. In fact, all the results in this section are unaltered with firms' entry and exit under two conditions: (1) the productivity distribution of entrants has a thinner right tail than the productivity distribution F, and (2) firms' exit rates are non-increasing on their productivity. Both conditions are common features of firm dynamics models.

In the rest of section 6, I explore the equilibrium relationship between productivity distribution

[23] In the appendix D.1, I provide an equivalent condition for m to be separable. That is, the relative arrival rate of any two ideas above firms' productivity has to be independent of firms. This assumption also generalizes Assumption 1 in Buera and Oberfield (2020).

[24] An example when $\tilde{H}$ is not bounded is as follows. The domain of firm productivity is $(0,+\infty)$, and an unbounded source distribution $\tilde{H}(z) = z^{-k}$ on $(0,+\infty)$. Let $m(z,x,t) = (\max\{x,z\})^{-k}$, which satisfies separability on $z \geq x$.

F and learning function m in idea search models of this kind. Before reaching there, it is useful to present technical definitions and regularity conditions that are crucial to the equilibrium analysis.

6.2 Useful definitions and regularity conditions

This section is a technical preliminary to the characterization results in the next section. I first introduce the notion of regular variation, which applies to a general class of distributions. Then, I provide regularity conditions to ensure the validity of the characterization results. Lastly, I present theoretical versions of Gibrat's law and Zipf's law.

Regular variation and tail index Regularly varying functions can be understood as a generalization of power functions. They are first studied by Karamata (1930). I follow the modern treatment in Bingham et al. (1987) to give the below definition.

Definition (Regular Variation)**.** Let f be a positive measurable function, defined on some neighborhood $[x_0, \infty)$, and satisfying

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

for all $t > 0$ and some $\alpha \in \mathbb{R}$; then f is said to be regularly varying (at infinity) with index α. If $\alpha = 0$, f is said to be slowly varying (at infinity).

Illustrative examples of slowly varying functions are constant functions and the log family, such as $\log(x)$ and $\log\log(x)$. Regular variation also has wide applications in probability theory. The tail index defined below generalizes the shape parameter of Pareto distributions and broadly measures the thickness of the right tail of a distribution. Technically, a distribution is thick-tailed if its CDF is a regularly varying function with a finite exponent.

Definition (Tail Index)**.** A non-negative random variable and its distribution are said to have tail indices (or "tail") $k \geq 0$ if the density function is regularly varying with index $-1-k$.

Regularity conditions The following assumption is a smoothness condition on the evolution of a distribution. It ensures that the evolving distribution has a well-defined tail index at all times, and the trajectory of the tail index does not have jumps or kinks. I will impose this assumption on the equilibrium productivity distribution F so that the dynamics of its tail index are smooth enough to be characterized by a differential equation.

Assumption 2. *A distribution $W(z,t)$ satisfies the following conditions:*

(i) $W(z,t)$ has a well defined tail index $k(t)$;

(ii) For all t, $\frac{\partial}{\partial t}\frac{\ln \tilde{W}(z,t)}{\ln z}$, if exists, converges uniformly (possibly to infinity) in its neighborhood as $z \to \infty$.

Similarly, assumption 3 ensures that the search intensity and source distribution are regularly varying. Since $\tilde{H}(x,t)$ decreases on x, it is bounded on $[x_0, \infty)$ for any $x_0 > 0$. As only the right tail is of concern, it is then without loss to treat $\tilde{H}(z,t)$ as a distribution function and consider its tail index.

Assumption 3. *$\mu(x,t)$ and $H(z,t)$ satisfy the following conditions:*

(i) $\mu(x,t)$ is regularly varying with exponent $m(t) \in \mathbb{R}$;

(ii) $H(z,t)$ has tail index $h(t) \in (1, \infty)$.

Parameters $k(t)$, $m(t)$, and $h(t)$ have clear economic interpretations. The exponent of regular variation $m(t)$ captures the elasticity of firms' search intensity with respect to their productivity. Whenever positive, a larger $m(t)$ implies that more productive firms search much more intensively. Similarly, tail indices $k(t)$ and $h(t)$ capture respectively the supply elasticity of idea quality in the population and at the source of learning. Given that idea supply decreases on its quality, a larger tail index indicates that high-productivity ideas are relatively more scarce.

Gibrat's law and Zipf's law Gibrat's law and Zipf's law are empirical regularities that describe the growth of firms and the distribution of firm sizes, respectively. The following definitions outline their theoretical counterparts, with a focus on large firms in the right tail. Let $\lambda(z,t)$ represent the expected productivity growth of a firm z at time t, i.e., $\lambda(z,t) = \frac{\mathbb{E}[\mathrm{d}z]}{z}$. Similarly, let $g^r(t)$ denote the growth rate of large firms, defined as $g^r(t) = \lim_{z \to \infty} \lambda(z,t)$. Then, Gibrat's law holds for large firms if the growth rate of firms converges to a positive constant as the firms' productivity approaches infinity. On the other hand, Zipf's law holds for large firms if the complement CDF of the productivity distribution $\tilde{F}(z)$ is inversely proportional to productivity z as z goes to infinity.

Definition (Gibrat's law for large firms)**.** Gibrat's law holds if $g^r(t) \in (0, \infty)$.

Definition (Zipf's law for large firms)**.** Zipf's law holds if $\lim_{z \to \infty} \frac{\ln \tilde{F}(z)}{\ln z} = -1$.

It is worth pointing out that these modified definitions deviate meaningfully from the standard understanding of Gibrat's law and Zipf's law in two key aspects. Firstly, the definitions are based on firms' productivity rather than size. This deviation is made for theoretical convenience, as most firm dynamics models in the literature express firm size measures such as employment or sales as monotonic functions of productivity. Secondly, the definitions focus on the limit at infinitely large productivity, whereas the standard versions of Gibrat's law and Zipf's law are stated as global properties. The standard Gibrat's law suggests that firms' growth rates are, on average, independent of their size, while the standard Zipf's law describes the distribution of firm sizes as approximately following a Pareto distribution with an exponent close to 1. Therefore, the definitions provided represent weaker versions of Gibrat's law and Zipf's law, which are implied by the standard versions. Not only are the weaker versions sufficient for the main results, but they also offer more accurate descriptions. Empirical evidence has consistently shown violations of Gibrat's law and Zipf's law for small firms. (Haltiwanger et al., 2013; Kondo et al., 2021).

6.3 Gibrat's law and the thickening of the right tail

I now present results pertaining to Gibrat's law and the thickening of the right tail in the general idea search model. To begin with, I show that only specific forms of learning functions are consistent with Gibrat's law and a thickening right tail. Given the empirical challenge of having direct measures of how firms learn ideas, it remains difficult in the literature to discipline the idea search process. These results contribute to this problem by suggesting two indirect moments from firm dynamics that are useful for identification. Indeed, both firms' growth rates (Gibrat's law) and firm size distributions (right tails) are relatively easy to measure.

In the general idea search model established earlier in this section, the first result states that only three types of learning functions can be consistent with an equilibrium in which the right tail of the productivity distribution thickens over time. Furthermore, these learning functions predict distinct trends in firms' growth and thus can be identified from these trends.

Proposition 6. *Consider $F(z,t)$ that solves the initial value problem in (16) and satisfies assumption 2. Suppose assumption 3 holds as well. Then, the learning function and productivity growth satisfy one of the three relationships if $k(t)$ decreases strictly at time t:*

(i) $m(t) = h(t) > k(t)$, $\quad \lambda(z,t) \sim C \ln z$;

(ii) $m(t) < h(t) = k(t)$, $\quad \lambda(z,t) = o(1)$;

(iii) $m(t) = h(t) = k(t)$, $\quad \lambda(z,t) = o(\ln z)$.

It follows immediately from Proposition 6 that only the third type can satisfy Gibrat's law, i.e., $\lim_{z \to \infty} \lambda(z,t) \in (0, \infty)$. Hence, to be consistent with both Gibrat's law and a thickening right tail, an idea search model must have all three elasticities $m(t)$, $h(t)$, and $k(t)$ equal to each other. As a moment of the productivity distribution, $k(t)$ can be estimated from the data. Once it is known, one can instantly recover $m(t)$ and $h(t)$. These arguments lead to the second result below, which assumes the learning function is a power function of z and x.

Corollary 1. *Consider $F(z,t)$ that solves the initial value problem in (16) and satisfies assumption 2. Suppose the learning function takes the form $m(z,x,t) \propto x^{m_t} z^{-h_t}$ for $z \geq x$. Then, $m_t = h_t = k_t$ if at time t, Gibrat's law for large firms holds, and $k(t)$ strictly decreases.*

Recall that the simple idea search model in section 3 has learning function $m(z,x,t) = \eta_t x^{k_t} z^{-k_t}$ for all $z \geq x$ and t. Based on Corollary 1, it is the only idea search model with this functional form assumption that is in line with both Gibrat's law and a thickening right tail. Moreover, that the learning function takes the form of power functions is a widely adopted function form assumption among idea search models. Table A.1 of appendix D.2 validates its widespread use in the existing literature. This result offers a theoretical foundation for the deviation from standard idea search assumptions in the simple model. Such deviations are necessary to construct idea search models that are compatible with both Gibrat's law and the thickening of the right tail in firm size distributions.

Next, I proceed to the third result that the dynamics of the tail index, as described by equation (11) in the simple model, extends to general idea search models with internal search. Gibrat's law and thickening right tail must appear simultaneously among this subclass of models.

Definition (Internal Search)**.** There is an internal search if the source distribution and the productivity distribution are asymptotically equivalent at the infinity, i.e., $\lim_{z \to \infty} \frac{\tilde{H}(z,t)}{\tilde{F}(z,t)} \in (0, \infty)$.

This version of internal search generalizes the usual concept of learning from internal sources, which assumes that all firms engaging in production are the objects of learning. Consequently, it assumes that $H(z,t)$ is equal to $F(z,t)$. In contrast, the internal search here only requires the right tail of source distribution to be asymptotically equivalent to that of the productivity distribution. Therefore, it covers major idea search models in the literature, including those with learning from external or mixed sources. For example, learning from mixed source distributions in Buera and Oberfield (2020) also belongs to the class of internal search. The following proposition characterizes the equilibrium dynamics of the tail index with internal search.

Proposition 7. *Consider $F(z,t)$ that solves the initial value problem in (16) and satisfies the regularity conditions in assumption 2. Suppose assumption 3 also holds, and there is an internal search. Then, the tail index evolves as follows:*

$$\frac{\dot{k}(t)}{k(t)(k(t)-1)} = g^r(t). \tag{17}$$

As a result, Gibrat's law for large firms holds at time t if and only if $k(t)$ decreases strictly.

Equation (17) is the same as equation (11) in the simple model, in which all firms have the same expected growth rate, and $g^r(t)$ is equal to output per capita growth. It confirms that the characterization of tail dynamics in the simple model is, in fact, generic to a much more general class of idea search models. Meanwhile, this equation reveals the dual facets of tail growth. While manifesting itself as gains from thicker right tails at the macro level, it is associated with the growth of large firms at the micro level. In this regard, $g^r(t)$ is a legitimate measure of tail growth.

Notably, equation (17) serves as a valuable benchmark for comparing the simple model with existing models in the literature. It establishes that $g^r(t) = 0$ if and only if $\dot{k}(t) = 0$. When the productivity distribution is stationary with a constant tail index, Gibrat's law is violated as large firms cease to grow. Conversely, a learning function must ensure $g^r(t) = 0$ in order to generate a stationary equilibrium. In the literature, major idea search models achieve a stationary equilibrium by assuming learning functions with $m(t) < h(t)$, resulting in $g^r(t) = 0$ (as demonstrated in Table A.1 and Lemma A.3 in the appendix).

In addition, it is not a mere coincidence that the simple model generates both Gibrat's law and a thickening right tail; these two characteristics must coexist. Intuitively, for firms to jump into the right tail, they need to be close to it. Then, the right tail ceases to evolve if large firms no longer jump into (or grow within) the tail. That is why the dynamics of the right tail are contingent upon the growth dynamics of large firms.

Lastly, I comment on the differences between Corollary 1 and Proposition 7. With more restrictive assumptions on the functional form, the former identifies the model parameters with observed moments, i.e., Gibrat's law and thickening of the right tail. Whereas, the latter characterizes the equilibrium relationship between Gibrat's law and the thickening of the right tail without relying on any observed moments (though with the knowledge that there is an internal search).

6.4 Revisit Zipf's law

In the simple model, tail growth drives economic development and thickens the right tail of the firm size distribution. As economies progress along the growth path, they accumulate more tail growth, resulting in firm size distributions that are closer to the limiting Zipf's distribution. Thus, Zipf's law emerges as a consequence of the cumulative effect of tail growth.

However, it is important to note that the magnitude of tail growth matters for whether Zipf's law holds in the limit. With the absence of tail growth, the stationary distribution in existing idea search models must exhibit a tail index greater than one. In contrast, the convergence to Zipf's law in the simple model relies on a tail growth process that not only thickens the right tail but also sustains economic growth at a constant rate.

Proposition 7 makes it possible to quantify the tail index of the limiting productivity distribution of a given internal search process. Noticing that equation (17) is a logistic differential equation, it then admits a logistic function as the solution. With $k(0) > 1$,

$$k(t) = \frac{1}{1 \quad \left(1 \quad \frac{1}{k(0)}\right) \exp\left(\quad \int_0^t g^r(\tau) \mathrm{d}\tau\right)}.$$

Instantly, $\lim_{t \to \infty} k(t) = 1$ if and only if $\int_0^\infty g^r(\tau)\mathrm{d}\tau = \infty$. Proposition 8 summarizes this finding.

Proposition 8. *Consider $F(z,t)$ that solves the initial value problem in (16) and satisfies the regularity conditions in assumption 2. Suppose assumption 3 also holds, and there is internal search. Then, Zipf's law holds in the limit if and only if the cumulative tail growth is unbounded.*

This result highlights that constant tail growth, as seen in the simple model, is not a necessary condition for the convergence to Zipf's law. Instead, unbounded cumulative tail growth is a weaker necessary condition. This means that Zipf's law can still hold in the limit even if the idea search process exhibits a decreasing tail growth rate over time, such as $g^r(t) \quad 1/t$. The inclusion of such cases allows scenarios where imitation is the main driver of growth. It is reasonable to conjecture that the importance of imitation as a source of growth diminishes as a country progresses towards the technological frontier, where innovation becomes the dominant driver. Proposition 8 is valuable in that it indicates a wide range of tail growth processes that are consistent with Zipf's law. This robustness reinforces the position of tail growth as a plausible explanation for the prevalence of Zipf's law.

7 Conclusions

I present new evidence on the dynamics of firm size distributions in the context of economic development. Empirically, I document a positive correlation between the thickness of the right tail in the firm size distribution and the level of economic development. Theoretically, I develop a growth model based on idea search to provide a coherent explanation for this relationship as a feature of an asymptotic balanced growth path. Within this model, I demonstrate that the rise in market concentration could be a secular trend over development. I also establish the idea search process as a key driver behind the emergence of Zipf's law in firm size distributions. Furthermore, the model renders new implications for size-dependent industrial policies. My findings suggest that policies favoring large firms can capitalize on the diffusion externality resulting from idea search, leading to improved welfare.

A crucial aspect of idea search models lies in the learning function, which governs how firms search for and acquire new ideas. I introduce a novel idea search process termed *tail growth*, which reshapes the firm size distribution by shifting mass from relatively low-productivity firms to relatively high-productivity firms. This process generates growth by thickening the right tail. My model exemplifies how slight changes in the idea search assumption could result into equilibria with very different properties. Hence, estimating the learning function is essential for a thorough understanding of the diffusion of ideas across firms. Although I show that indirect firm dynamics moments are informative about the unobserved learning function, exploring new theories and data of firms' exchange of ideas is an interesting area for future research.

Finally, the central analysis of this paper is carried out in a starkly simple context in order to elucidate the economic forces and the nature of their interactions. I hope the insights and mathematical structure built in here are useful preliminary steps towards a realistic and comprehensive framework, which can quantify various channels of idea production and is suitable for counterfactual analysis.

References

Achdou, Y., F. J. Buera, J.-M. Lasry, P.-L. Lions, and B. Moll (2014): "Partial differential equation models in macroeconomics," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372, 20130397.

Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva, and V. Sterzi (2018): "Dancing with the Stars: Innovation Through Interactions," Working Paper 24466, National Bureau of Economic Research.

Alfaro, L., A. Charlton, and F. Kanczuk (2008): "Plant Size Distribution and Cross-Country Income Differences," *NBER International Seminar on Macroeconomics*, 5, 243–272.

Alvarez, F. E., F. J. Buera, and J. Lucas, Robert E. (2008): "Models of Idea Flows," Working Paper 14135, National Bureau of Economic Research.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The Fall of the Labor Share and the Rise of Superstar Firms," *The Quarterly Journal of Economics*, 135, 645–709.

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–1820.

Benhabib, J., J. Perla, and C. Tonetti (2021): "Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier," *Econometrica*, 89, 2261–2301.

Bento, P. and D. Restuccia (2017): "Misallocation, Establishment Size, and Productivity," *American Economic Journal: Macroeconomics*, 9, 267–303.

——— (2021): "On average establishment size across sectors and countries," *Journal of Monetary Economics*, 117, 220–242.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987): *Regular Variation*, Encyclopedia of Mathematics and its Applications, Cambridge University Press.

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

Buera, F. J. and E. Oberfield (2020): "The Global Diffusion of Ideas," *Econometrica*, 88, 83–114.

Cao, D., H. R. Hyatt, T. Mukoyama, and E. Sager (2022): "Firm Growth Through New Establishments," *Georegetown University Working Paper*.

Chen, Y.-F., W.-T. Hsu, and S.-K. Peng (2023): "Innovation, firm size distribution, and gains from trade," *Theoretical Economics*, 18, 341–380, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/TE4152.

Cirera, X., D. Comin, and M. Cruz (2022): "Bridging the Technological Divide: Technology Adoption by Firms in Developing Countries," Tech. rep., World Bank, Washington, DC.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection," *Econometrica*, 80, 2543–2594.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 135, 561–644.

Gabaix, X. (1999): "Zipf's Law and the Growth of Cities," *The American Economic Review*, 89, 129–132.

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

García-Santana, M. and R. Ramos (2015): "Distortions and the size distribution of plants: evidence from cross-country data," *SERIEs*, 6, 279–312.

Gaubert, C. (2018): "Firm Sorting and Agglomeration," *American Economic Review*, 108, 3117–3153.

Geerolf, F. (2017): "A Theory of Pareto Distributions," Mimeo, UCLA.

Gutiérrez, G. and T. Philippon (2018): "Ownership, Concentration, and Investment," *AEA Papers and Proceedings*, 108, 432–437.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013): "Who Creates Jobs? Small versus Large versus Young," *Review of Economics and Statistics*, 95, 347–361.

Hsieh, C.-T. and P. J. Klenow (2014): "The Life Cycle of Plants in India and Mexico *," *The Quarterly Journal of Economics*, 129, 1035–1084.

Hsieh, C.-T. and B. A. Olken (2014): "The Missing "Missing Middle"," *Journal of Economic Perspectives*, 28, 89–108.

Jones, C. I. (2022): "Recipes and Economic Growth: A Combinatorial March Down an Exponential Tail," *Journal of Political Economy*, Forthcoming.

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

Karamata, J. (1930): "Sur un mode de croissance régulière des fonctions," *Mathematica (Cluj.)*, 38–53.

Kodama, N. and H. Li (2018): "Manager Characteristics and Firm Performance," Discussion Paper 18-E-060, RIETI.

Kondo, I. O., L. T. Lewis, and A. Stella (2021): "Heavy Tailed, but not Zipf: Firm and Establishment Size in the U.S," Tech. Rep. 21-15, Center for Economic Studies, U.S. Census Bureau.

Kortum, S. S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65, 1389–1419.

Kwon, S. Y., Y. Ma, and K. Zimmermann (2022): "100 Years of Rising Corporate Concentration," Working Paper 3936799, SSRN.

König, M., K. Storesletten, Z. Song, and F. Zilibotti (2022): "From Imitation to Innovation: Where Is All That Chinese R&D Going?" *Econometrica*, 90, 1615–1654.

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

Lucas, R. E. (1978): "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9, 508–523.

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

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

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

Nuño, G. and B. Moll (2018): "Social optima in economies with heterogeneous agents," *Review of Economic Dynamics*, 28, 150–180.

Oberfield, E. (2018): "A Theory of Input–Output Architecture," *Econometrica*, 86, 559–589.

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

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

Poschke, M. (2018): "The Firm Size Distribution across Countries and Skill-Biased Change in Entrepreneurial Technology," *American Economic Journal: Macroeconomics*, 10, 1–41.

Resnick, S. I. (2007): *Heavy-tail phenomena: probabilistic and statistical modeling*, Springer series in operations research and financial engineering, New York, N.Y: Springer.

Sampson, T. (2016): "Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth," *The Quarterly Journal of Economics*, 131, 315–380.

Toda, A. A. and Y. Wang (2021): "Efficient minimum distance estimation of Pareto exponent from top income shares," *Journal of Applied Econometrics*, 36, 228–243.

Tybout, J. R. (2000): "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature*, 38, 11–44.

Young, A. (1998): "Growth without Scale Effects," *Journal of Political Economy*, 106, 41–63.

Online Appendix

Appendix A Additional evidence on the thickening right tails in the US

It is well documented in the literature that the US firm size distribution has a Pareto right tail, i.e., $P(\text{size} > x) \propto x^{-k}$. In this section, I report additional estimates of the Pareto exponent k using alternative data sources, firm size measures, and estimation techniques. Recalling that the tail thickness $\tilde{R}^f = -k$ with Pareto right tails, one should then expect decreasing Pareto exponents in the US. These results corroborate the robustness of the findings in section 2.3.3.

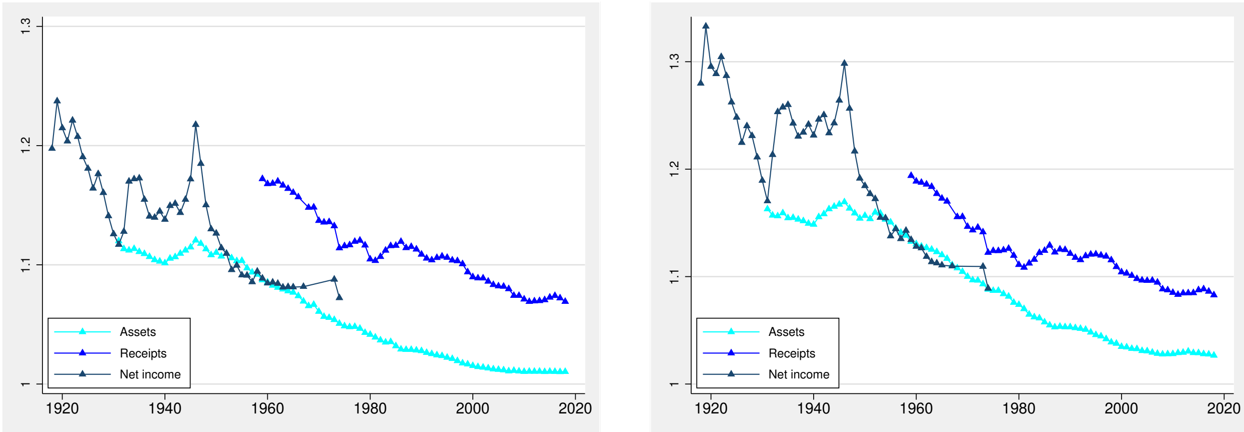
A longer time series (1918-2018) I perform the same tail estimation exercises on long-term time series data published by Kwon et al. (2022). The main data source they use is the Statistics of Income (SOI) and the associated Corporation Source Book published annually by the IRS. The SOI originated from the Revenue Act of 1916, which requires the IRS to report statistics based on the tax returns filed each year. Kwon et al. (2022) digitalized historical SOI publications and published the cleaned data series online at `http://businessconcentration.com`. In particular, they provide data on the business share (by assets, receipts, or net income) held by the top 50%, 10%, 1%, and 0.1% firms. The data series on assets has the best time coverage (1931-2018), followed by receipts (1959-2018) and net income (1918-1974).

Let q_x denote the share of top x% firms. With a Pareto right tail, it is straightforward to show that

$$\frac{\ln\left(\frac{q_{x_1}}{q_{x_2}}\right)}{\ln\left(\frac{x_1}{x_2}\right)} = 1 - \frac{1}{k}, \tag{A.1}$$

for percentiles x_1 and x_2. Since the LHS is data, this equation yields an estimate for k. Consistent with the threshold choices in section 2.3.3, I plot in figure A.1 the time series of the estimated Pareto exponent k for $x_1 = 1$ or 0.1 and $x_2 = 10$. The negative trend in the estimated k verifies figure 2 in Kwon et al. (2022), which reports increasing relative shares over time, i.e., a positive trend in $\frac{q_{x_1}}{q_{x_2}}$.

An alternative tail estimator The main tail estimator in section 2.1 may suffer from two caveats. First, it is based on size thresholds instead of percentiles. Since size thresholds are fixed, their corresponding percentiles are varying over time, which may become a source of measurement error. To see this, consider a horizontal shift of the size distribution. While the right tail remains unaffected, it changes the percentiles of the size thresholds. The variations in the resulting estimates then come solely from the fact that each year's estimator captures different parts of the same size distribution. Second, the current estimator uses only two size thresholds. While a general



(a) Top 1%

(b) Top 0.1%

Figure A.1: Changes in the Pareto exponents in the US (1918-2018)

Notes. This figure plots the estimated Pareto exponent k based on equation (A.1). In the left figure, $x_1 = 1, x_2 = 10$; in the right figure, $x_1 = 0.1, x_2 = 10$. The computation is based on top business share series used by Kwon et al. (2022).

compromise to cross-country comparisons, the US BDS data provides 7 firm size bins even within the top 10%. Utilizing data on all these size bins can plausibly improve the efficiency of the estimator and reduces the arbitrariness of choosing the thresholds.

Toda and Wang (2021) develop an efficient minimum distance estimator of Pareto exponent using top share data. Given a dataset with top shares $q_1, q_2, \ldots, q_{K+1}$ for the top $x_1, x_2, \ldots, x_{K+1}$ percentile, their estimator finds the Pareto exponent k that minimizes the distance between top shares implied by the Pareto distribution (the model) and the data. Therefore, I make the following attempts to alleviate concerns about the above caveats. For the US BDS, I first compute the corresponding percentiles of the size thresholds for each year.[25] Then, I apply the estimator in Toda and Wang (2021) to top 10% firms (employees 20) using data on both top firm shares and top percentiles. The following figure A.2 confirms that the re-estimated Pareto exponent also decreases in the past forty years.

Micro data and very large firms Using micro data on US firms, two papers also find that the right tail of the US firm size distribution has become thicker. Cao et al. (2022) use the Employer Characteristics File maintained by the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Program. This dataset is supposed to be representative of the universe of the US business firms. They estimate the employment size-rank slope (k) of the upper percentiles of the firm size distribution and find that for firm at or above the 99 percentile, the slope changed from -1.17 to -0.99 between 1995 and 2014. Chen et al. (2023) present similar findings using Compustat for publicly-listed firms in North America. They estimate the Pareto exponent for each year during

[25] Toda-Wang estimator requires at least 4 size bins for firms in the right tails and the Pareto exponent must be larger than 1. The cross-country datasets (OECD and WBES) do not meet these requirements.

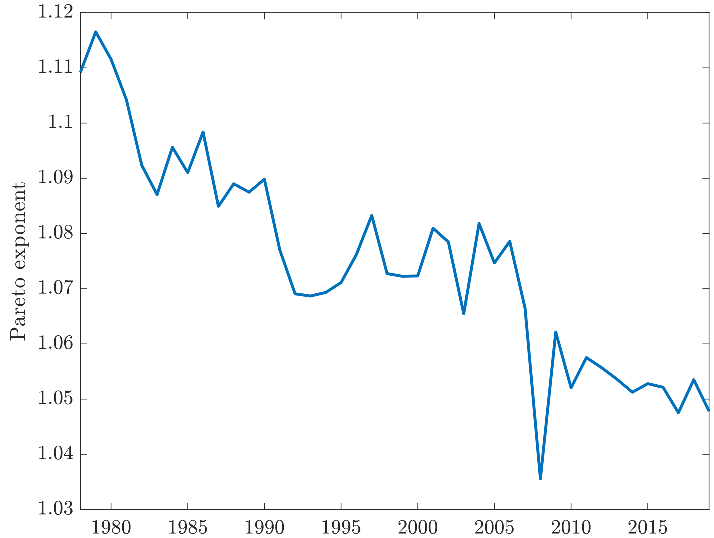


Figure A.2: Pareto exponents (Toda-Wang) in the US (1978-2019)

Notes. This figure plots the estimated Pareto exponent k based on Toda and Wang (2021). The estimator is applied to firms with at least 20 employees (top 10% of the total firms) and uses 7 firm size bins. Data source: US census BDS (1978-2019).

1970-2019 and with firm size measured by real revenue. Similar to my previous results, figure 1 in their paper exhibits declining Pareto exponents during the period for super large firms: firms with real revenues above 2500, 5000, and 7500 million dollars, as well as the top 500, 750, and 1000 firms.

Appendix B Omitted Proofs in Section 4

B.1 Proof of Lemma 1

Proof. From the HJB equation (1), the gains from learning is linear in η. Hence, the gains from learning per unit arrival rate must be non-positive, otherwise firms will choose infinitely large η, i.e.,

$$\int_z^\infty [v(x,t) - v(z,t)]\mathrm{d}F(x|x \geq z,t) - zw(t) \leq 0.$$

Then, the total gains from learning must be zero as firms can always choose $\eta = 0$. In this way,

$$r(t)v(z,t) = z + \partial_t v(z,t) \implies r(t)\frac{v(z,t)}{z} = 1 + \frac{\partial}{\partial t}\frac{v(z,t)}{z}$$

by dividing both sides with z. Then $v(z,t)/z$ is a constant $v(t)$, which satisfies $r(t)v(t) = 1 + v'(t)$. Integrating this ordinary differential equation forward with the transversality condition,

$$v(t) = \int_t^{\infty} e^{-\int_t^x r(s)\mathrm{d}s}\mathrm{d}x. \tag{A.2}$$

That is, firm's unit value, $v(t)$, is the present value of a dividend flow of unit output.

B.2 Proof of Proposition 1

Proof. As in the proof of lemma 1, let $v(z,t) = v(t)z$, in which $v(t)$ is given by (A.2), and $w(t) = \frac{v(t)}{k(t)-1}$. I verify that $v(z,t)$ and $F(z,t)$ solve respectively firm's problem (1) and the law of motion on the productivity distribution (5). First, with $F(z,t)$ given by (7), the gains from learning per unit arrival rate for firm z satisfies

$$\int_z^{\infty} [v(x,t) - v(z,t)]\mathrm{d}F(x|x \geq z,t) - zw(t) = v(t)z\left\{E_{F(\cdot,t)}[x|x \geq z] - \frac{k(t)z}{k(t)-1}\right\} = 0.$$

The second equality uses that the conditional distribution $F(x|x \geq z)$ is also Pareto with scale z and shape $k(t)$. Therefore, each firm is indifferent with any level of η. $\eta(z,t) = \eta(t)$ then qualifies for an optimal choice. $v(z,t)$ satisfies the HJB equation (1) following the proof of lemma 1.

Second, with $\eta(z,t) = \eta(t)$, $F(z,t)$ solves the KFE (5). To see this, rewrite (5) as follows for any $z \geq 1$,

$$\frac{\partial \ln \tilde{F}(z,t)}{\partial t} = \eta(t)\int_1^z \frac{f(x,t)}{\tilde{F}(x,t)}\mathrm{d}x.$$

Inserting $\tilde{F}(z,t) \equiv 1 - F(z,t) = z^{-k(t)}$, the above PDE is reduced to the following ODE:

$$\dot{k}(t)\ln z = \eta(t)k(t)\ln z \implies \frac{\dot{k}(t)}{k(t)} = -\eta(t),$$

which is precisely (8).

Lastly, the labor market clearing condition pins down $\eta(t)$ such that

$$\eta(t)\int_1^{\infty} zf(z,t)\mathrm{d}z = L \implies \eta(t)\frac{k(t)}{k(t)-1} = L.$$

With $k(0) = k_0$, (8) and the above equation determine $k(t)$ and $F(z,t)$ at each date. Consequently, I obtain output per capita $y(t) = \mathsf{E}_{F(\cdot,t)}[z]/L$. The goods market condition and the Euler equation of the households' problem give the interest rate in a standard way, i.e.,

$$r(t) = \theta\frac{\dot{c}(t)}{c(t)} + \rho = \theta\frac{\dot{y}(t)}{y(t)} + \rho.$$

The proof is then complete.

B.3 Proof of Lemma 2

Proof. From section 4.2, the normalized wage satisfies that

$$\tilde{w}(t) = \frac{v(t)L}{k(t)} = \frac{L}{k(t)} \int_t^{\infty} e^{-\int_t^x r(s)\mathrm{d}s} \mathrm{d}x.$$

Noticing that $r(t) = \theta \frac{L}{k(t)} + \rho$ from the Euler equation, and $\frac{\dot{k}}{k-1} = -L$,

$$\int_t^x r(s)\mathrm{d}s = \int_t^x \theta \left(\frac{\dot{k}(s)}{k(s)} + L \right) + \rho \mathrm{d}s = \theta \ln \frac{k(x)}{k(t)} + (\rho + \theta L)(x - t).$$

Therefore,

$$\frac{v(t)L}{k(t)} = \frac{L}{k(t)} \int_t^{\infty} e^{-\theta \ln \frac{k(x)}{k(t)} - (\rho+\theta L)(x-t)} \mathrm{d}x = \frac{L}{k(t)} \int_t^{\infty} \left(\frac{k(x)}{k(t)} \right)^{-\theta} e^{-(\rho+\theta L)(x-t)} \mathrm{d}x.$$

With $k(x+s) = 1 + (k(x) - 1)e^{-Ls}$,

$$\begin{aligned}
\int_t^{\infty} k(x)^{-\theta} e^{-(\rho+\theta L)(x-t)} \mathrm{d}x &= \int_0^{\infty} \left(1 + (k(t)-1)e^{-Ls}\right)^{-\theta} e^{-(\rho+\theta L)s} \mathrm{d}s, \qquad \text{Sub. } (s = x - t) \\
&= \frac{1}{L} \int_0^1 (1 + (k(t)-1)q)^{-\theta} q^{\frac{\rho}{L}+\theta-1} \mathrm{d}q, \qquad \text{Sub. } (q = e^{-Ls}) \\
&= \frac{1}{L\,(k(t)-1)^{\frac{\rho}{L}+\theta}} \int_1^{k(t)} p^{-\theta} (p-1)^{\frac{\rho}{L}+\theta-1} \mathrm{d}p, \qquad \text{Sub. } (p = 1 + (k(t)-1)q) \\
&= \frac{1}{L\,(k(t)-1)^{\frac{\rho}{L}+\theta}} \int_0^{1-\frac{1}{k(t)}} y^{\frac{\rho}{L}+\theta-1} (1-y)^{-\frac{\rho}{L}-1} \mathrm{d}y. \qquad \text{Sub. } (y = 1 - \frac{1}{p})
\end{aligned}$$

Plugging it back and suppressing time variable t, the normalized wage is a function of k:

$$\tilde{w}(k) = \frac{v(t)L}{k(t)} = \frac{k^{\theta-1}}{(k-1)^{\frac{\rho}{L}+\theta}} \int_0^{1-\frac{1}{k}} y^{\frac{\rho}{L}+\theta-1} (1-y)^{-\frac{\rho}{L}-1} \mathrm{d}y = \frac{k^{\nu}}{(k-1)^{\nu+\alpha}} \int_0^{1-\frac{1}{k}} y^{\nu+\alpha-1} (1-y)^{-\alpha} \mathrm{d}y,$$

in which $\nu = \theta - 1 > -1$ and $\alpha = \frac{\rho}{L} + 1 > 1$. The parametric condition $\rho > L(1 - \theta)$ implies that $\nu + \alpha > 1$. To see that the research share increases over time, it suffices to show that $\tilde{w}(k)$ decreases on k. Differentiating it with respect to k,

$$\tilde{w}'(k) = \frac{k^{\nu}}{(k-1)^{\nu+\alpha}} \left\{ \left[\frac{\nu}{k} - \frac{\nu+\alpha}{k-1} \right] \int_0^{1-\frac{1}{k}} y^{\nu+\alpha-1} (1-y)^{-\alpha} \mathrm{d}y + \left(1 - \frac{1}{k}\right)^{\nu+\alpha-1} \left(\frac{1}{k}\right)^{2-\alpha} \right\}$$

Noting that the integral term is an incomplete Beta function $B_{1\quad\frac{1}{k}}(\nu+\alpha, 1\quad\alpha)$, it has the following hypergeometric representation. [26]

$$\int_0^{1\quad\frac{1}{k}} y^{\nu+\alpha\quad 1}(1\quad y)^{\quad\alpha}\mathrm{d}y = \frac{(1\quad\frac{1}{k})^{\nu+\alpha}(\frac{1}{k})^{1\quad\alpha}}{\nu+\alpha}F(\nu+1,1;\nu+\alpha+1;1\quad\frac{1}{k}),$$

in which F is a hypergeometric function such that

$$F(\nu+1,1;\nu+\alpha+1;1\quad\frac{1}{k}) = \frac{\Gamma(\nu+\alpha+1)}{\Gamma(\nu+1)}\sum_{s=0}^{1}\frac{\Gamma(\nu+1+s)}{\Gamma(\nu+\alpha+1+s)}\left(1\quad\frac{1}{k}\right)^s.$$

Then,

$$\begin{aligned}
&\left[\frac{\nu}{k}\quad\frac{\nu+\alpha}{k\quad 1}\right]\int_0^{1\quad\frac{1}{k}} y^{\nu+\alpha\quad 1}(1\quad y)^{\quad\alpha}\mathrm{d}y + \left(1\quad\frac{1}{k}\right)^{\nu+\alpha\quad 1}\left(\frac{1}{k}\right)^{2\quad\alpha} \\
=&\left(1\quad\frac{1}{k}\right)^{\nu+\alpha}\left(\frac{1}{k}\right)^{1\quad\alpha}\left[\left(\frac{\nu}{\nu+\alpha}\frac{1}{k}\quad\frac{1}{k\quad 1}\right]F + \left(1\quad\frac{1}{k}\right)^{\quad 1}\frac{1}{k}\right] \\
=&\left(1\quad\frac{1}{k}\right)^{\nu+\alpha}\left(\frac{1}{k}\right)^{1\quad\alpha}\frac{1}{k\quad 1}\left(1+\frac{\nu}{\nu+\alpha}\left(1\quad\frac{1}{k}\right)F\quad F\right)
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&1+\frac{\nu}{\nu+\alpha}\left(1\quad\frac{1}{k}\right)F(\nu+1,1;\nu+\alpha+1;1\quad\frac{1}{k}) \\
=&1+\frac{\nu}{\nu+\alpha}\left(1\quad\frac{1}{k}\right)\frac{\Gamma(\nu+\alpha+1)}{\Gamma(\nu+1)}\sum_{s=0}^{1}\frac{\Gamma(\nu+1+s)}{\Gamma(\nu+\alpha+1+s)}\left(1\quad\frac{1}{k}\right)^s, \\
=&1+\frac{\Gamma(\nu+\alpha)}{\Gamma(\nu)}\sum_{s=0}^{1}\frac{\Gamma(\nu+1+s)}{\Gamma(\nu+\alpha+1+s)}\left(1\quad\frac{1}{k}\right)^{s+1}, \\
=&\frac{\Gamma(\nu+\alpha)}{\Gamma(\nu)}\sum_{s=0}^{1}\frac{\Gamma(\nu+s)}{\Gamma(\nu+\alpha+s)}\left(1\quad\frac{1}{k}\right)^s = F(\nu,1;\nu+\alpha;1\quad\frac{1}{k}).
\end{aligned}$$

[26] See, for example, equation (11.34) in chapter 11 of Temme (1996).

Then, given that $\alpha > 0$,

$$
\begin{aligned}
&F(\nu, 1; \nu+\alpha; 1-\frac{1}{k}) - F(\nu+1, 1; \nu+\alpha+1; 1-\frac{1}{k}) \\
=&\sum_{s=0}^{\infty}\left(\frac{\Gamma(\nu+\alpha)}{\Gamma(\nu)}\frac{\Gamma(\nu+s)}{\Gamma(\nu+\alpha+s)} - \frac{\Gamma(\nu+\alpha+1)}{\Gamma(\nu+1)}\frac{\Gamma(\nu+1+s)}{\Gamma(\nu+\alpha+1+s)}\right)\left(1-\frac{1}{k}\right)^{s}, \\
=&\sum_{s=1}^{\infty}\frac{\Gamma(\nu+\alpha+1)}{\Gamma(\nu+1)}\frac{\Gamma(\nu+s)}{\Gamma(\nu+\alpha+s)}\left(\frac{v}{v+\alpha} - \frac{v+s}{v+\alpha+s}\right)\left(1-\frac{1}{k}\right)^{s}, \\
=&\sum_{s=1}^{\infty}\frac{\Gamma(\nu+\alpha+1)}{\Gamma(\nu+1)}\frac{\Gamma(\nu+s)}{\Gamma(\nu+\alpha+s)}\frac{-\alpha s}{(v+\alpha)(v+\alpha+s)}\left(1-\frac{1}{k}\right)^{s} < 0.
\end{aligned}
$$

In sum,

$$
\begin{aligned}
\tilde{w}'(k) =&\frac{k^{\nu}}{(k-1)^{\nu+\alpha}}\left(1-\frac{1}{k}\right)^{\nu+\alpha}\left(\frac{1}{k}\right)^{1-\alpha}\frac{1}{k-1}\left(1+\frac{\nu}{\nu+\alpha}\left(1-\frac{1}{k}\right)F - F\right), \\
=&\frac{1}{k(k-1)}\left(F(\nu, 1; \nu+\alpha; 1-\frac{1}{k}) - F(\nu+1, 1; \nu+\alpha+1; 1-\frac{1}{k})\right) < 0.
\end{aligned}
$$

Thus, $\tilde{w}$ increases over time as $k(t)$ is decreasing to one. Using the same representation results,[27]

$$
\begin{aligned}
\tilde{w}(k) =&\frac{k^{\nu}}{(k-1)^{\nu+\alpha}}\frac{(1-\frac{1}{k})^{\nu+\alpha}(\frac{1}{k})^{1-\alpha}}{\nu+\alpha}F(\nu+1, 1; \nu+\alpha+1; 1-\frac{1}{k}), \\
=&\frac{1}{\nu+\alpha}\frac{1}{k}F(\nu+1, 1; \nu+\alpha+1; 1-\frac{1}{k}).
\end{aligned}
$$

As $k \to 1$, $F(\nu+1, 1; \nu+\alpha+1; 1-\frac{1}{k}) \to 1$. Then,

$$
\tilde{w}(t) \to \tilde{w}^{*} = \frac{1}{\rho/L+\theta} < 1.
$$

Besides, note that both $F(\nu+1, 1; \nu+\alpha+1; 1-\frac{1}{k})$ and $1/(\nu+\alpha)$ decrease on α. Then, $\tilde{w}(k; \alpha)$ decreases on α for all $k > 1$. An increase in L lowers α and increases $\tilde{w}(k; \alpha)$.

B.4 Proof of Proposition 3

Proof. I verify that there is an equilibrium with productivity distributions $F(z,t) = 1 - z^{-k(t)}$ for $z \geq 1$. Assume that the productivity distribution $F(z,t) = 1 - z^{-k(t)}$ for $z \geq 1$ and $v(z,t) = v(t)z$. Then,

$$
\int_{z}^{\infty}[v(x,t) - v(z,t)]\mathrm{d}F(x|x \geq z, t) = \frac{v(t)z}{k(t)-1},
$$

[27] One can also obtain the same $s'_R(k)$ from here, but the computation will be much harder.

Given wage $w(t)$, the first order condition implies that for each firm,

$$\frac{v(t)z}{k(t)-1} = z(1+g'(\eta))w(t).$$

Then, all firms search at the same intensity. The solution must be interior otherwise the labor demand will be zero, and the labor market will not clear. Given that $\eta(z,t) = \eta(t) > 0$, the law of motion on the productivity distribution (5) implies that it is consistent to have Pareto $F(z,t)$ at all times. To see this, rewrite (5) as follows for any $z \geq 1$,

$$\frac{\partial \ln \tilde{F}(z,t)}{\partial t} = \eta(t) \int_1^z \frac{f(x,t)}{\tilde{F}(x,t)} \mathrm{d}x \implies -\dot{k}(t) \ln z = \eta(t) k(t) \ln z,$$

in which I insert both sides with $\tilde{F}(z,t) \equiv 1 - F(z,t) = z^{-k(t)}$. With initial Pareto distribution, the above equation verified that $F(z,t)$ remains Pareto with shape parameter $k(t)$ satisfying $\dot{k}(t) = -\eta(t)k(t)$.

Next, I verify that $v(z,t) = v(t)z$ given $\eta(z,t) = \eta(t)$. Note that the total return of learning with optimal search intensity is given by

$$\eta(t) \int_z^{\infty} [v(x,t) - v(z,t)] \mathrm{d}F(x|x \geq z,t) - z(\eta(t) + g(\eta(t)))w(t) = \frac{\eta g'(\eta) - g(\eta)}{1+g'(\eta)} \frac{v(t)z}{k(t)-1},$$

which is positive (strictly convex g) and linear in z. Hence, $v(z,t) = v(t)z$ satisfies the HJB equation of each firm at all times if

$$r(t)v(t) = 1 + \frac{\eta g'(\eta) - g(\eta)}{1+g'(\eta)} \frac{v(t)}{k(t)-1} + v'(t)$$

with $\eta = \eta(t)$.

In the end, the following two equations characterize the equilibrium path on $\{k(t), \eta(t)\}$ given $k(t) = k(0)$:

$$(\eta(t) + g(\eta(t))) \frac{k(t)}{k(t)-1} = L, \quad \text{and} \quad \frac{\dot{k}(t)}{k(t)} = -\eta(t).$$

The goods market clear trivially. I can solve for $r(t)$, $v(t)$ and $w(t)$ in the same way as before. Thus, all equilibrium conditions are satisfied.

Appendix C Omitted Proofs in Section 5

C.1 Proof of proposition 4

Proof. I verify that the described equilibrium distributions can be supported by the following search strategy. While firms with productivity below z^* do not search, those with with productivity above

z search at the same intensity $\eta(t)$, i.e.,

$$\eta(z,t) = \begin{cases} 0, & \text{if } z \quad z \;, \\ \eta(t), & \text{if } z > z \;. \end{cases} \tag{A.3}$$

I show in the following that this search strategy is consistent with $\tau = 0$. Therefore, it holds trivially with $\tau > 0$ in which firms below the threshold are more discouraged to search. Proposition 4 then presents a continuum of equilibria of the simple model, making use of the multiplicity due to the linear cost assumption.

To begin with, it is straightforward that the part of distribution that is below z does not change since more productive firms do not drop from above. Hence, $F(z,t) = F(z,0)$ for all $z \quad z$ and t. For the part that is above the threshold, the evolution is identical to the baseline case if we normalize the threshold. Consider an infinitesimal time break h, for $z > z$,

$$\tilde{F}(z,t+h) = \tilde{F}(z,t) + \int_{z}^{z} \eta(x,t) h \frac{\tilde{F}(z,t)}{\tilde{F}(x,t)} f(x,t) \mathrm{d}x$$

$$) \; \frac{\partial \tilde{F}(z,t)}{\partial t} = \tilde{F}(z,t) \int_{z}^{z} \eta(x,t) \frac{f(x,t)}{\tilde{F}(x,t)} \mathrm{d}x. \tag{A.4}$$

This is almost the same as the baseline except that the integral now begins with the threshold z instead of the minimum. Lastly, the total measure of firms are constant, so we always have $\tilde{F}(z\;,t) = \tilde{F}(z\;,0)$.

To see that (13) satisfies all three conditions, it suffices to verify that it satisfies the law of motion (A.4). Note that for $z > z$,

$$\frac{\partial \tilde{F}(z,t)}{\partial t} \frac{1}{\tilde{F}(z,t)} = \frac{\partial \ln \tilde{F}(z,t)}{\partial t} = \quad \dot{k}(t) \left(\ln z \quad \ln z\;\right),$$

$$\int_{z}^{z} \eta(x,t) \frac{f(x,t)}{\tilde{F}(x,t)} \mathrm{d}x = \eta(t) \int_{z}^{z} \frac{k(t)}{x} \mathrm{d}x = \eta(t) k(t) \left(\ln z \quad \ln z\;\right).$$

Hence, we only need to solve for $k(t)$ given $\dot{k}(t)/k(t) = \quad \eta(t)$. The labor market clearing condition implies that

$$\eta(t) \int_{z}^{1} y f(y,t) \mathrm{d}y = L,$$

since only firms above z demand labor for search. Solving this equation gives us

$$\eta(t) \frac{k(t)}{k(t) \quad 1} = L(z\;)^{k_0 \quad 1}.$$

When $z \; = 1$, we are back to the baseline equilibrium. It is intuitive that with fewer firms doing search, the available labor per firm goes up. This alternative search strategy is essentially an increase in effective labor endowment. As before, this equation completes the equilibrium path of the tail index $k(t)$, as described in proposition 4. Along this equilibrium,

$$y(t) = \int_1^\infty z \mathrm{d}F(z,t) = \frac{k_0}{k_0 - 1}\left[1 - (z^*)^{1-k_0}\right] + \frac{k(t)}{k(t)-1}(z^*)^{1-k_0}.$$

It is straightforward that $\dot{y}/y \to L(z^*)^{k_0-1}$. This is a strategy which trades off short run output for long run growth.

In the end, we show that this strategy is optimal for firms as well. Recall the HJB equation,

$$r(t)v(z,t) = z + \max_\eta \eta \left\{ \int_z^\infty \left[v(x,t) - v(z,t)\right] \mathrm{d}F(x|x \geq z, t) - zw(t) \right\} + \partial_t v(z,t).$$

In equilibrium, we must have

$$\int_z^\infty \left[v(x,t) - v(z,t)\right] \mathrm{d}F(x|x \geq z, t) - zw(t) \leq 0.$$

Therefore, the above HJB equation becomes $r(t)v(z,t) = z + \partial_t v(z,t)$. $v(z,t)/z$ is then independent of z. Let $v(t) = v(z,t)/z$. Solving $r(t)v(t) = 1 + v'(t)$ forward, we obtain

$$v(t) = \int_t^\infty e^{-\int_t^x r(s)ds} \mathrm{d}x,$$

in which $r(t)$ is determined by the equilibrium output growth. Given $w(t) = v(t)/(k(t)-1)$ and the equilibrium distributions, we show that the expected gain from search is negative for $z \leq z^*$ and zero for $z > z^*$. Noticing that

$$\int_z^\infty \left[v(x,t) - v(z,t)\right] \mathrm{d}F(x|x \geq z, t) - zw(t) = v(t)z \left\{ \int_z^\infty \frac{x}{z} \mathrm{d}F(x|x \geq z, t) - \frac{k(t)}{k(t)-1} \right\},$$

it then suffices to compute the compare the value of the integral with $k(t)/(k(t)-1)$. For $z \leq z^*$,

$$\begin{aligned}
\int_z^\infty \frac{x}{z} \mathrm{d}F(x|x \geq z, t) &= \frac{1}{z[1-F(z,t)]} \int_z^\infty x \mathrm{d}F(x,t), \\
&= \frac{1}{z^{1-k_0}} \left\{ \int_z^{z^*} k_0 x^{-k_0} \mathrm{d}x + (z^*)^{k(t)-k_0} \int_{z^*}^\infty k(t) x^{-k(t)} \mathrm{d}x \right\}, \\
&= \frac{1}{z^{1-k_0}} \left\{ \frac{k_0}{k_0-1}\left[z^{1-k_0} - (z^*)^{1-k_0}\right] + (z^*)^{k(t)-k_0} \frac{k(t)}{k(t)-1} (z^*)^{1-k(t)} \right\}, \\
&= \frac{k_0}{k_0-1}\left[1 - \left(\frac{z^*}{z}\right)^{1-k_0}\right] + \frac{k(t)}{k(t)-1}\left(\frac{z^*}{z}\right)^{1-k_0}, \\
&< \frac{k(t)}{k(t)-1}.
\end{aligned}$$

The last equality uses that $(z^*/z)^{1-k_0} < 1$ and $k(t) < k_0$. For $z > z^*$, the truncated distribution

$F(x|x \geq z, t)$ is exactly Pareto with shape parameter $k(t)$ and then has mean $zk(t)/(k(t)-1)$. Then, that all firms above the threshold search at the same intensity satisfies the optimality of firm's problem.

At this point, it is straightforward to see what a positive tax does. Let $\hat{z}$ be the threshold of a threshold equilibrium (13). With $\tau > 0$, the policy maker can choose the threshold z^* to eliminate threshold equilibria with $\hat{z} < z^*$.

C.2 Derivation of equation (15) on $w(z,t)$

Following Lucas and Moll (2014), the corresponding HJB equation of problem (14) is

$$\begin{aligned} \rho W(f) = & \max_{\{c(\omega),\eta(y)\}} \int_\Omega u(c(\omega))\mathrm{d}\omega + \int_0^\infty \frac{\delta W(f)}{\delta f(y)} f(y) \left[\int_0^y \eta(x)\varphi(x)\mathrm{d}x - \eta(y)\right] \mathrm{d}y \\ \text{s.t.} & \int_\Omega c(\omega)\mathrm{d}\omega \leq \int_0^\infty y f(y)\mathrm{d}y, \quad \int_0^\infty y\eta(y)f(y)\mathrm{d}y \leq L, \end{aligned} \tag{A.5}$$

in which $\varphi(x) = f(x)/(1-F(x))$. $\hat{\lambda}$ and $\hat{\mu}$ are the respective Lagrangian multipliers on the goods and labor market clearing conditions. Let $w(f,z) = \delta W(f)/\delta f(z)$. The first order condition on consumption gives us

$$u'(c(\omega)) = \hat{\lambda}. \tag{A.6}$$

The first order condition on the search intensity $\eta(y)$ implies that

$$\begin{aligned} & \int_z^\infty w(f,y)f(y)\varphi(z)\mathrm{d}y - w(f,z)f(z) - \hat{\mu} z f(z) = 0 \\ \Longrightarrow & \int_z^\infty \left[w(f,y) - w(f,z)\right] \frac{f(y)}{1-F(z)}\mathrm{d}y = \hat{\mu} z \end{aligned} \tag{A.7}$$

Differentiating both sides of the HJB equation (A.5) with respect to $f(z)$,

$$\begin{aligned} \rho w(f,z) = & \int_0^\infty \frac{\delta w(f,z)}{\delta f(y)} f(y) \left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right] \mathrm{d}y \\ & + \int_0^\infty w(f,y) \frac{\delta}{\delta f(z)} \left\{ f(y) \left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right]\right\} \mathrm{d}y + \hat{\lambda} z - \hat{\mu} z \eta^*(z). \end{aligned} \tag{A.8}$$

Let $w(z,t) \equiv w(f(\cdot,t),z)$, then

$$\frac{\partial w(z,t)}{\partial t} = \int_0^\infty \frac{\partial w(z,f(\cdot,t))}{\partial f(y,t)} \frac{\partial f(y,t)}{\partial t}\mathrm{d}y = \int_0^\infty \frac{\delta w(f,z)}{\delta f(y)} f(y) \left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right] \mathrm{d}y$$

with $f(\ ,t) = f$. Hence, the first term on the RHS of (A.8) is simply $\partial w(z,t)/\partial t$. To calculate the second term, note that

$$\frac{\delta \varphi(y)}{\delta f(z)} = \begin{cases} \frac{f(y)}{[1 \quad F(y)]^2} & \text{if } y < z, \\ \frac{1}{1 \quad F(y)} \quad \frac{f(y)}{[1 \quad F(y)]^2} & \text{if } y = z, \\ 0 & \text{if } y > z. \end{cases}$$

Therefore,

$$\frac{\delta}{\delta f(z)} f(y) \left[\int_0^y \eta \ (x)\varphi(x)\mathrm{d}x \quad \eta \ (y) \right] = \begin{cases} f(y) \int_0^y \eta \ (x) \frac{\varphi(x)}{1 \quad F(x)} \mathrm{d}x & \text{if } y < z, \\ \int_0^z \eta \ (x)\varphi(x)\mathrm{d}x \quad \eta \ (z) & \\ +f(z) \left[\quad \int_0^z \eta \ (x) \frac{\varphi(x)}{1 \quad F(x)} \mathrm{d}x + \frac{\eta \ (z)}{1 \quad F(z)} \right] & \text{if } y = z, \\ f(y) \left[\quad \int_0^z \eta \ (x) \frac{\varphi(x)}{1 \quad F(x)} \mathrm{d}x + \frac{\eta \ (z)}{1 \quad F(z)} \right] & \text{if } y > z. \end{cases}$$

Then,

$$\begin{aligned} & \int_0^1 w(f,y) \frac{\delta}{\delta f(z)} \left\{ f(y) \left[\int_0^y \eta \ (x)\varphi(x)\mathrm{d}x \quad \eta \ (y) \right] \right\} \mathrm{d}y \\ = & \int_0^z w(f,y)f(y) \int_0^y \eta \ (x) \frac{\varphi(x)}{1 \quad F(x)} \mathrm{d}x\mathrm{d}y + \int_z^1 w(f,y)f(y) \left[\int_0^z \eta \ (x) \frac{\varphi(x)}{1 \quad F(x)} \mathrm{d}x + \frac{\eta \ (z)}{1 \quad F(z)} \right] \mathrm{d}y \\ & + w(f,z) \left[\int_0^z \eta \ (x)\varphi(x)\mathrm{d}x \quad \eta \ (z) \right] \\ = & \int_0^1 \left\{ w(f,y) \int_0^{\max\{y,z\}} \eta \ (x) \frac{\varphi(x)}{1 \quad F(x)} \mathrm{d}x + w(f,z) \int_0^z \eta \ (x)\varphi(x)\mathrm{d}x \right\} f(y)\mathrm{d}y \\ & + \eta \ (z) \int_z^1 [w(f,y) \quad w(f,z)] \frac{f(y)}{1 \quad F(z)} \mathrm{d}y \end{aligned}$$

Finally, rewriting (A.8) gives equation (15):

$$\begin{aligned} \rho w(z,t) = & \frac{\partial w(z,t)}{\partial t} + \hat{\lambda} z + \max_{\eta} \left\{ \eta \int_z^1 [w(y,t) \quad w(z,t)] \frac{f(y,t)}{1 \quad F(z,t)} \mathrm{d}y \quad \hat{\mu} z \eta \right\} \\ & + \int_0^1 \left\{ w(y,t) \int_0^{\max\{y,z\}} \eta \ (x,t) \frac{\varphi(x,t)}{1 \quad F(x,t)} \mathrm{d}x + w(z,t) \int_0^z \eta \ (x,t)\varphi(x,t)\mathrm{d}x \right\} f(y,t)\mathrm{d}y, \end{aligned}$$

in which the max operator comes from first order condition (A.7).

C.3 Proof of proposition 5

Proof. To solve for the optimal policy, I first differentiate equation (15) with respect to z.

$$\begin{aligned}
\rho w_z(z,t) =& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} + \eta\ (z,t)\frac{\partial}{\partial z}\left\{\int_z^1 [w(y,t) - w(z,t)]\frac{f(y,t)}{1-F(z,t)}\mathrm{d}y - \hat{\mu}z\right\} \\
&+ w(z,t)f(z,t)\int_0^z \eta\ (x,t)\frac{\varphi(x,t)}{1-F(x,t)}\mathrm{d}x - w(z,t)f(z,t)\int_0^z \eta\ (x,t)\frac{\varphi(x,t)}{1-F(x,t)}\mathrm{d}x \\
&+ \int_z^1 w(y,t)f(y,t)\mathrm{d}y\eta\ (z,t)\frac{\varphi(z,t)}{1-F(z,t)} + w_z(z,t)\int_0^z \eta\ (x,t)\varphi(x,t)\mathrm{d}x + w(z,t)\eta\ (z,t)\varphi(z,t) \\
=& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} - \eta\ (z,t)\varphi(z,t)\int_z^1 (w(y,t) - w(z,t))\frac{f(y,t)}{1-F(z,t)}\mathrm{d}y + w_z(z,t)\int_0^z \eta\ (x,t)\varphi(x,t)\mathrm{d}x \\
=& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} - \hat{\mu}\eta\ (z,t)\varphi(z,t)z + w_z(z,t)\int_0^z \eta\ (x,t)\varphi(x,t)\mathrm{d}x
\end{aligned} \tag{A.9}$$

The first equality is the result of an envelope theorem, and the second and the third use the first order condition (A.7), which holds for all z at any time t. Let $k(z,t) = zf(z,t)/(1-F(z,t))$, I obtain $w_z(z,t)$ and $\partial w_z(z,t)/\partial t$ by differentiating (A.7):

$$\begin{aligned}
w_z(z,t) &= \hat{\mu}\left(k(z,t) - 1\right), \\
\frac{\partial w_z(z,t)}{\partial t} &= \dot{\hat{\mu}}(k(z,t) - 1) + \hat{\mu}\frac{\partial k(z,t)}{\partial t} = \dot{\hat{\mu}}(k(z,t) - 1) - \hat{\mu}\eta\ (z,t)\varphi(z,t)z,
\end{aligned}$$

in which the last equality comes from the law of motion on $\varphi(z,t)$. Formally, the original law of motion (5) implies that by differentiating both sides on z,

$$\frac{\partial \ln(1-F(z,t))}{\partial t} = -\int_0^z \eta(x,t)\frac{f(x,t)}{1-F(x,t)}\mathrm{d}x \Rightarrow \frac{\partial \varphi(z,t)}{\partial t} = -\eta(z,t)\varphi(z,t).$$

Inserting them back to (A.9), the following differential equation characterizes the optimal policy.

$$\begin{aligned}
&\rho\hat{\mu}\left(k(z,t) - 1\right) = \dot{\hat{\mu}}(k(z,t) - 1) + \hat{\lambda} - 2\hat{\mu}\eta\ (z,t)\varphi(z,t)z + \hat{\mu}\left(k(z,t) - 1\right)\int_0^z \eta\ (x,t)\varphi(x,t)\mathrm{d}x \\
\Rightarrow &\left(\rho - \frac{\dot{\hat{\mu}}}{\hat{\mu}}\right)(1 - k(z,t)) + \frac{\hat{\lambda}}{\hat{\mu}} = (1 - k(z,t))\int_0^z \eta\ (x,t)\varphi(x,t)\mathrm{d}x + 2\hat{\mu}\eta\ (z,t)\varphi(z,t)z.
\end{aligned}$$

Fixing time t, this is a first order linear ordinary differential equation in $\int_0^z \eta\ (y,t)\varphi(y,t)\mathrm{d}y$, so it can be solved analytically. It admits the following solution:

$$\begin{aligned}
\int_0^z \eta\ (y,t)\varphi(y,t)\mathrm{d}y &= e^{-\int_0^z \frac{1-k(x,t)}{2x}dx}\int_0^z \frac{1}{2y}e^{\int_0^y \frac{1-k(x,t)}{2x}dx}\left\{\left(\rho - \frac{\dot{\hat{\mu}}}{\hat{\mu}}\right)[1 - k(y,t)] + \frac{\hat{\lambda}}{\hat{\mu}}\right\}\mathrm{d}y \\
&= \left[e^{\int_0^z \frac{k(x,t)-1}{2x}dx} - 1\right]\left(\frac{\dot{\hat{\mu}}}{\hat{\mu}} - \rho\right) + \frac{\hat{\lambda}}{\hat{\mu}}e^{\int_0^z \frac{k(x,t)-1}{2x}dx}\int_0^z \frac{1}{2y}e^{\int_0^y \frac{1-k(x,t)}{2x}dx}\mathrm{d}y
\end{aligned}$$

Given this solution, the above ODE gives $\eta(z,t)$:

$$
\begin{aligned}
\eta(z,t) &= \frac{k(z,t)-1}{2k(z,t)}\left\{\int_0^z \eta(y,t)\varphi(y,t)\mathrm{d}y + \frac{\dot{\hat{\mu}}}{\hat{\mu}} - \rho\right\} + \frac{\hat{\lambda}}{\hat{\mu}}\frac{1}{2k(z,t)} \\
&= \frac{k(z,t)-1}{2k(z,t)} e^{\int_0^z \frac{k(x,t)-1}{2x}dx}\left\{\left(\frac{\dot{\hat{\mu}}}{\hat{\mu}} - \rho\right) + \frac{\hat{\lambda}}{\hat{\mu}}\int_0^z \frac{1}{2y} e^{\int_0^y \frac{1-k(x,t)}{2x}dx}\mathrm{d}y\right\} + \frac{\hat{\lambda}}{\hat{\mu}}\frac{1}{2k(z,t)} \\
&= \frac{k(z,t)-1}{2k(z,t)} e^{\int_0^z \frac{k(x,t)-1}{2x}dx}\Bigg\{\left(\frac{\dot{\hat{\mu}}}{\hat{\mu}} - \rho\right) + \frac{\hat{\lambda}}{\hat{\mu}}\Bigg[\frac{e^{\int_0^z \frac{1-k(x,t)}{2x}dx}}{1-k(z,t)} - \frac{1}{1-k(0,t)} \\
&\quad - \int_0^z \frac{k_y(y,t)}{(1-k(y,t))^2} e^{\int_0^y \frac{1-k(x,t)}{2x}dx}\mathrm{d}y\Bigg]\Bigg\} + \frac{\hat{\lambda}}{\hat{\mu}}\frac{1}{2k(z,t)} \\
&= \frac{k(z,t)-1}{2k(z,t)} e^{\int_0^z \frac{k(x,t)-1}{2x}dx}\Bigg\{\left(\frac{\dot{\hat{\mu}}}{\hat{\mu}} - \rho\right) + \frac{\hat{\lambda}}{\hat{\mu}}\frac{1}{(k(0,t)-1)} - \frac{\hat{\lambda}}{\hat{\mu}}\int_0^z \frac{k_y(y,t)}{(1-k(y,t))^2} e^{\int_0^y \frac{1-k(x,t)}{2x}dx}\mathrm{d}y\Bigg]\Bigg\}
\end{aligned}
$$

The third equality comes from integration by parts. If $F(z,t)$ has tail index $k(t)$, $\lim_{z\to\infty} k(z,t) = k(t)$, and the search intensity is a regularly varying function with exponent $(k(t)-1)/2$, i.e.,

$$
\eta(z,t) = z^{\frac{k(t)-1}{2}} L(z,t)
$$

in which $L(z,t)$ is a slow varying function. This is a direct application of the Karamata's representation theorem. With a Pareto initial distribution, $k_z(z,0) = 0$, and

$$
\eta(z,0) = \left[\left(\frac{\dot{\hat{\mu}}}{\hat{\mu}} - \rho\right) + \frac{\hat{\lambda}}{\hat{\mu}}\frac{1}{k_0 - 1}\right]\frac{k_0 - 1}{2k_0} z^{\frac{k_0-1}{2}},
$$

a power function. The proof is then complete.

C.4 Illustration of the tail dynamics with optimal search policy

See Figure A.3.

Appendix D Additional Results in Section 6

D.1 An equivalent condition of a separable arrival rate function

Recall that for each x, $m(z,x,t)$ decreases on z, $\lim_{z\to\infty} m(z,x,t) = 0$, and $m(0,x,t) < \infty$. The following lemma says the arrival rate function is separable on $z - x$ if and only if the relative arrival rate of any two ideas above firms' productivity is independent of firms.

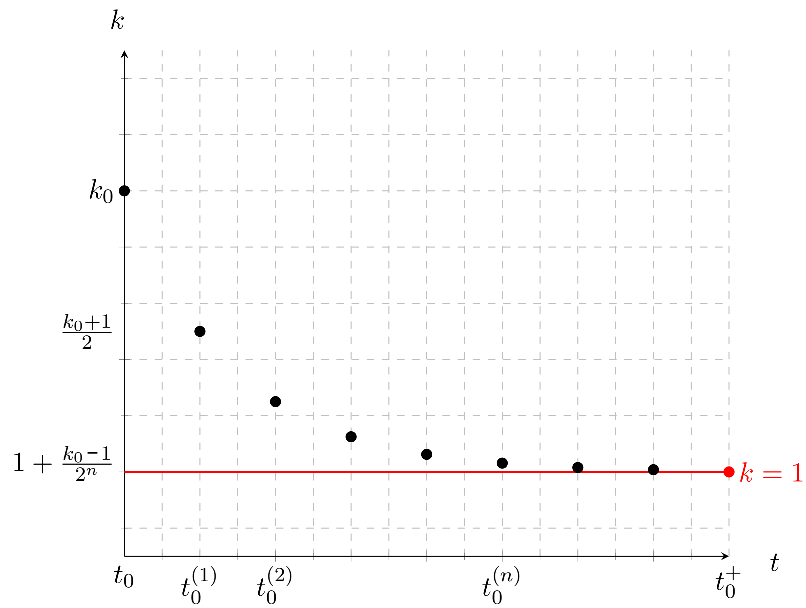


Figure A.3: Illustration of jumps in tail indices

Notes. Given an initial tail index k_0, the tail index is $1 + \frac{k_0-1}{2}$ after n jumps. It converges to one in countably many steps, which take zero measure of time.

Lemma A.1. *$m(z, x, t)$ is separable on $z \geq x$ if and only if for any $z', z \geq \max\{x, x'\}$,*

$$\frac{m(z', x, t)}{m(z, x, t)} = \frac{m(z', x', t)}{m(z, x', t)}. \tag{A.10}$$

Proof. The "only if" part is obvious. To see the "if" part, it suffices to show that given condition (A.10), there exists $\mu(x, t)$ and $\tilde{H}(z, t)$ such that $m(z, x, t) = \mu(x, t)\tilde{H}(z, t)$ for $z \geq x$. For each x and t, let $\bar{z}(x, t)$ denote the upper support of $m(z, x, t)$ on z, i.e., $\bar{z}(x, t) = \inf\{z' : m(z', x, t) = 0\}$. I focus on the case that there exists x_0 such that $\bar{z}(x_0, t) > x_0$. Lemma A.1 will be a tautology if no such point exists. Then, I define $\tilde{H}(z, t)$ as follows,

$$\tilde{H}(z,t) = \begin{cases} \frac{m(z,x_0,t)}{m(x_0,x_0,t)} & \text{if } z \geq x_0, \\ \frac{m(z,z,t)}{m(x_0,z,t)} & \text{if } z < x_0. \end{cases} \tag{A.11}$$

It is evident that $\tilde{H}(z, t)$ has upper support $\bar{z}(x_0, t)$. Also, let $\mu(x, t) = \frac{m(x,x,t)}{\tilde{H}(x,t)}$ for $x \leq \bar{z}(x_0, t)$ and be well-defined otherwise.

First, I show that $\tilde{H}(z, t)$ is well-defined. The part on $z \geq x_0$ is trivial. For $z < x_0$, it suffices to show $m(x_0, z, t) > 0$ holds. Suppose not, consider $x' \in (x_0, \bar{z}(x_0, t))$. Then,

$$\frac{m(x', z, t)}{m(x_0, z, t)} = \frac{0}{0},$$

which is undefined. On the contrary, $\frac{m(x',x_0,t)}{m(x_0,x_0,t)} > 0$ is well-defined, violating condition (A.10). Next, I show that $m(\max\{x, \bar{z}(x_0,t)\}, x, t) = 0$. This is straightforward for $x < x_0$ since (A.10) implies

$$\frac{m(\bar{z}(x_0,t), x, t)}{m(x_0, x, t)} = \frac{m(\bar{z}(x_0,t), x_0, t)}{m(x_0, x_0, t)} = 0.$$

Then, $m(\bar{z}(x_0,t), x, t) = 0$.[28] For $x_0 < x < \bar{z}(x_0,t)$,

$$\frac{m(\bar{z}(x_0,t), x, t)}{m(x, x, t)} = \frac{m(\bar{z}(x_0,t), x_0, t)}{m(x, x_0, t)} = 0,$$

so $m(\bar{z}(x_0,t), x, t) = 0$ and $m(x,x,t) > 0$. For $x > \bar{z}(x_0,t)$, I must have $m(x,x,t) = 0$. In other words, $\bar{z}(x,t) = \bar{z}(x_0,t)$ for $x < \bar{z}(x_0,t)$. To see this, suppose there exists x' such that $m(x', x', t) > 0$. Picking $x'' > x'$ such that $m(x'', x, t) > 0$, $\frac{m(x'', x', t)}{m(x', x', t)} > 0$. Whereas, $\frac{m(x'', x_0, t)}{m(x', x_0, t)}$ is undefined. Thus, $m(z,x,t) = \mu(x,t)\tilde{H}(z,t)$ whenever $z \geq x$ and $m(z,x,t) = 0$.

It remains to show that $\tilde{H}(z,t)$ is decreasing on z. Since $m(z,x,t)$ decreases on z for any x, then $\tilde{H}(z,t) \leq 1$ for $z \geq x_0$ and ≥ 1 for $z < x_0$. It is trivial that $\tilde{H}(z,t)$ decreases on z for $z \geq x_0$. Now suppose that $z < z' < x_0$,

$$\tilde{H}(z',t) = \frac{m(z', z', t)}{m(x_0, z', t)} = \frac{m(z', z, t)}{m(x_0, z, t)} \leq \frac{m(z, z, t)}{m(x_0, z, t)} = \tilde{H}(z,t),$$

in which the second equality uses condition (A.10) and the inequality uses that $m(z,x,t)$ decreases on z. Thus, it also decreases on z for $z < x_0$. Besides, It is straightforward that $\lim_{z \to \infty} \tilde{H}(z,t) = 0$ as $\lim_{z \to \infty} m(z, x_0, t) = 0$.

Finally, I verify that $m(z,x,t) = \mu(x,t)\tilde{H}(z,t)$ for $z \geq x$ and $m(z,x,t) > 0$ in three cases.

1. $\underline{z > x > x_0}$. In this case,

$$m(z,x,t) = \frac{m(z,x,t)}{m(x,x,t)} m(x,x,t) = \underbrace{\frac{m(z,x_0,t)}{m(x_0,x_0,t)}}_{\tilde{H}(z,t)} \underbrace{\frac{m(x_0,x_0,t)}{m(x,x_0,t)}}_{\frac{1}{\tilde{H}(x,t)}} m(x,x,t) = \tilde{H}(z,t)\mu(x,t).$$

   The second equality uses condition (A.10) on the first term. In addition, $\tilde{H}(z,t)$ and $\tilde{H}(x,t)$ follow the definition (A.11) with $z > x_0$ and $x > x_0$ respectively.

2. $\underline{z > x_0 > x}$. In this case,

$$m(z,x,t) = \frac{m(z,x,t)}{m(x_0,x,t)} m(x_0,x,t) = \underbrace{\frac{m(z,x_0,t)}{m(x_0,x_0,t)}}_{\tilde{H}(z,t)} \underbrace{\frac{m(x_0,x,t)}{m(x,x,t)}}_{\frac{1}{\tilde{H}(x,t)}} m(x,x,t) = \tilde{H}(z,t)\mu(x,t).$$

   The second equality uses condition (A.10) on the first term. In addition, $\tilde{H}(z,t)$ and $\tilde{H}(x,t)$ follow the definition (A.11) with $z > x_0$ and $x < x_0$ respectively.

[28] More rigorously, it holds with $\bar{z}(x_0,t) + \epsilon$ for any $\epsilon > 0$.

3. $\underline{x_0 > z > x}$. In this case,

$$m(z,x,t) = \frac{m(z,x,t)}{m(x_0,x,t)} m(x_0,x,t) = \underbrace{\frac{m(z,z,t)}{m(x_0,z,t)}}_{\tilde{H}(z,t)} \underbrace{\frac{m(x_0,x,t)}{m(x,x,t)}}_{\frac{1}{\tilde{H}(x,t)}} m(x,x,t) = \tilde{H}(z,t)\mu(x,t).$$

The second equality uses condition (A.10) on the first term. In addition, $\tilde{H}(z,t)$ and $\tilde{H}(x,t)$ follow the definition (A.11) with $z < x_0$ and $x < x_0$ respectively.

The proof is then complete.

D.2 A discussion on $m(z, x, t)$

The arrival rate function $m(z,x,t)$ is essentially the transition probability of firm x to next period's productivity of at least z. It then summarizes the description of any idea search process. To see this, I consider in the following a detailed process of how firms learn ideas, in which each step is a common element in the literature. The first step of firms' learning is to obtain learning opportunities, which arrives at rate $\mu(x,t)$ for each firm x. Upon its arrival, the probability that a firm x recognizes an idea with original productivity z is $h(z,x,t)$. However, it might not always be easy to absorb new ideas. As mentioned before, there are abundant empirical evidence that technology gap hinders the adoption of technology for firms in developing countries. For example, engineers in low-productivity firms might find it difficult to understand the state-of-the-art technology due to lack of college education. Alternatively, they might only understand a portion of the new materials and upgrade their technology partially rather than fully. To capture this, a firm x has probability $p(x,z)$ to absorb an idea of productivity z. Conditional on successful absorption, firms develop new technologies based on new ideas, and $q(x,z)$ represents the productivity of the resulting technology from firms' adaption of new ideas. That is to say, $q(x,z)$ is the final realized productivity that a firm x can attain from an idea with original productivity z. The equation below gives the total arrival rate of an option technology with productivity at least z:

$$m(z,x,t) = \mu(x,t) \int_{\Omega(z,x)} p(x,y)h(y,x,t)\mathrm{d}y, \tag{A.12}$$

in which $\Omega(z,x) = \{y : q(x,y) \geq z\}$ and is the set of qualified ideas before adaptation. In words, it is the product of search intensity and the total probability that a firm obtains and absorbs qualified ideas. $m(z,x,t)$ then precisely measures the potential change in firms' productivity due to learning.

Furthermore, it is with little loss of generality to assume separability in that separable arrival rate functions still cover a plethora of heterogeneity. Consider again the learning process described above. All four elements–search intensity, source distribution, absorption probability and idea adaptation–could be heterogeneous across firms. With commonly used functional forms, table A.1 shows that separable arrival rate functions are able to cover all four types of heterogeneity. For better illustration, each row focuses on one type of heterogeneity and assumes homogeneity in the

Search Intensity $\mu(x,t)$	Source Distribution $\tilde{H}(z,x,t)$	Absorption Probability $p(x,z)$	Idea Adaptation $q(x,z)$	Arrival Rate Function $m(z,x,t)$
$\mu(x,t)$	$\tilde{H}(z,t)$	1	z	$\mu(x,t)\tilde{H}(z,t)$
μ_t	$\left(\frac{z}{x}\right)^{-k_t}$, $z \geq x$	1	z	$\mu_t x^{k_t} z^{-k_t}$, $z \geq x$
μ_t	$z^{-\theta}$	$\max\{1, \left(\frac{z}{x}\right)^{-\gamma}\}$	z	$\frac{\mu_t \theta}{\gamma+\theta} x^{\gamma} z^{-(\gamma+\theta)}$, $z \geq x$
μ_t	$z^{-\theta}$	1	$\left(\frac{z}{x}\right)^{-\beta} z$	$\mu_t x^{\frac{\beta}{1-\beta}\theta} z^{-\frac{\theta}{1-\beta}}$

Table A.1: Examples of separable $m(z,x,t)$

Notes: This table presents four separable arrival rate functions with respective heterogeneity in search intensity, source distribution, adoption probability and modification function. In particular, $\gamma > 0$, and $\beta \in (0,1)$. $F(z,t)$ is the population productivity distribution. $m(z,x,t)$ is calculated based on (A.12). $m(z,x,t)$ without labeling $z \geq x$ is separable all over its domain.

other three.

The first row has heterogeneous search intensity with a common source distribution. Lucas and Moll (2014), Perla and Tonetti (2014) and Sampson (2016) fall into this category. Both Perla and Tonetti (2014) and Sampson (2016) assumes that only the entrants adopt ideas from existing firms, therefore $\mu(x,t) = 0$ for any firms that are above the productivity threshold. In Lucas and Moll (2014), $\mu(x,t)$ is a decreasing function on firm productivity and converges to zero. It also includes the more basic case in which search intensities are uniform over firms, as in Kortum (1997), Alvarez et al. (2008) and Buera and Oberfield (2020). Additionally, the source distribution H in these models are Pareto distributions. Hence, the elasticity of search intensity $m(t) = 0$, while the supply elasticity of idea quality at the source $H(t) > 0$.

The second row summarizes the simple model in section 3. Firms search at the same intensity but draw ideas from the population productivity distribution left truncated at their own productivity. Both the third and fourth rows capture the heterogeneity in firms' learning capacity. While one models it as differences in absorption probability and the other in idea adaptation, they are isomorphic in terms of overall learning performance. The third row says that firms are less likely to absorb more advanced ideas. Namely, the absorption probability declines with the productivity gap, i.e., $p(x,z) = \max\{1, \left(\frac{z}{x}\right)^{-\gamma}\}$, which is similar to that in Van Patten (2020). An alternative formulation is that more advanced ideas are less useful to unproductive firms, even if they still benefit these firms. In the fourth row, $q(x,z) = z^{1-\beta} x^{\beta} \in (\min\{x,z\}, \max\{x,z\})$ and increases with x. An idea z increases the productivity of any firm $x < z$ with certainty but are more useful to more productive firms. Note that with $\gamma = \frac{\beta}{1-\beta}\theta$, $m(z,x,t)$ in the third and fourth rows are the same over $z \geq x$, up to a constant. Besides, that $m(t) < h(t)$ holds with both learning functions in the third and fourth row.

D.3 Useful derivations and lemmas

Derivation of $\lambda(x,t)$ The expected productivity at $t+\mathrm{d}t$ satisfies that

$$\mathsf{E}[x(t+\mathrm{d}t)] = \left(1 - \int_0^\infty n(y,x,t)\mathrm{d}y\mathrm{d}t\right)x + \int_0^\infty n(y,x,t)\mathrm{d}t \max\{y,x\}\mathrm{d}y.$$

The first term captures the probability of no idea arrivals, and the second term is the weighted sum of realized productivity next instant. In particular, the first term is well-defined with $m(0,x,t) < \infty$. Let $\lambda(x,t)$ be the expected productivity growth of a firm with productivity x, then

$$\lambda(x,t) = \lim_{h \to 0} \frac{\mathsf{E}[x(t+h)] - x}{xh} = \frac{\mu(x,t)}{x} \int_x^\infty (y - x)h(y,t)\mathrm{d}y. \tag{A.13}$$

The equation holds since $n(z,x,t) = \mu(x,t)h(z,t)$ for $z \geq x$.

Lemma A.2. *Consider $F(z,t)$ that solves the initial value problem of (16) and satisfies assumption 2. Then, $k(t)$ is decreasing. Whenever it is finite, $k(t)$ is differentiable and has the following finite derivative:*

$$\dot{k}(t) = -\lim_{z \to \infty} \frac{1}{\ln z} \frac{\tilde{H}(z,t)}{\tilde{F}(z,t)} \int_0^z \mu(x,t) f(x,t)\mathrm{d}x > -\infty. \tag{A.14}$$

Proof. See appendix E.

If the equilibrium productivity F satisfies the regularity conditions (are sufficiently smooth), lemma A.2 states that the time derivative of the tail index $k(t)$ can be written as the above expression of F, H, and μ.

Lemma A.3. *Suppose condition (ii) of assumption 3 holds. Then, the productivity growth $\lambda(x,t)$ has the following asymptotic equivalence as $x \to \infty$:*

$$\lambda(x,t) \sim \frac{1}{h-1}\mu(x,t)\tilde{H}(x,t) \tag{A.15}$$

Proof. See appendix E.

With regular varying functions, lemma A.3 shows that for large firms, the expected productivity growth is proportional to the arrival rate of having ideas better than their own productivity, i.e., $\lambda(x,t) \propto m(x,x,t)$. Consequently, $g^r(t) = 0$ if $m(t) < h(t)$.

D.4 A discussion on entry and exit

Similar analysis extends to idea search models with entry and exit. Let $\hat{F}(z,t)$ be the measure of firms with productivity at least z at time t, which generalizes the previous complement CDF $\tilde{F}(z,t)$. Similarly, $\hat{f}(z,t) \equiv -\partial\hat{F}(z,t)/\partial z$ is the relevant density. I use $E(z,t)$ and $\delta(z,t)$ to capture entry and exit respectively. At time t, $E(z,t)$ is the measure of entrants with productivity at least

z, and $\delta(z,t)$ the exit probability of a firm z. The evolution of firm size distribution is revised as follows:

$$\frac{\partial \hat{F}(z,t)}{\partial t} = \tilde{H}(z,t)\int_0^z \mu(x,t)\hat{f}(x,t)\mathrm{d}x + E(z,t) - \int_z^\infty \delta(x,t)\hat{f}(x,t)\mathrm{d}x.$$

Like $\tilde{H}(z,t)$, similar arguments justify the tail index of $\hat{F}$. Following the proof of lemma A.2, the change in tail index of $\hat{F}$ satisfies that

$$\dot{k}(t) = \lim_{z\to\infty}\frac{1}{\ln z}\frac{\tilde{H}(z,t)}{\hat{F}(z,t)}\int_0^z \mu(x,t)\hat{f}(x,t)\mathrm{d}x + \lim_{z\to\infty}\frac{E(z,t)}{\hat{F}(z,t)\ln z} - \lim_{z\to\infty}\frac{\int_z^\infty \delta(x,t)\hat{f}(x,t)\mathrm{d}x}{\hat{F}(z,t)\ln z}. \tag{A.16}$$

With regularly varying $E(z,t)$ and $\delta(z,t)$, same arguments can immediately extend proposition 6 with entry and exit. It is important to note two observations. First, both entry and exit can make direct impact on the right tail. Entry thickens the right tail, whereas exit makes the tail thinner. Second, entry and exit affect the tail index only if either the entrant distribution has a sufficiently thick right tail, or larger firms are more likely to exit. More commonly, $E(z,t)$ has a thinner right tail than $\hat{F}(z,t)$, and $\delta(z,t)$ is non-increasing on z, Then, equation (A.16) still reduces to equation (A.2), and lemma A.2 and proposition 6 apply without any modifications. For this reason, the general mechanism in this section excludes explicit entry and exit.

Appendix E Omitted Proofs in Section 6

E.1 Proof of lemma A.2

Proof. $k(t)$ decreases over time since (16) implies that $F(z,t)$ increases stochastically in t. By the Karamata's theorem (cf. Bingham et al. (1987), BGT, Proposition 1.5.10), that $f(z,t)$ is regularly varying with index $-(1+k(t))$ implies that $\lim_{x\to\infty}\frac{xf(x,t)}{\tilde{F}(x,t)} = k(t)$ if $k(t) > 0$. Then, $\tilde{F}(x,t)$ is regularly varying with exponent $-k(t)$. With $k(t) = 0$ and $\lim_{z\to\infty}\tilde{F}(z,t) = 0$, Proposition 1.5.9b of BGT applies to show that $\lim_{x\to\infty}\frac{xf(x,t)}{\tilde{F}(x,t)} = 0$, and $\tilde{F}(x,t)$ is slowly varying. With the representation theorem (BGT, Theorem 1.3.1), $\tilde{F}(x,t)$ can be written as follows:

$$\tilde{F}(x,t) = x^{-k(t)}c(x,t)\exp\left\{\int_{a_t}^x \frac{\varepsilon(u,t)}{u}\mathrm{d}u\right\} \quad (x \geq a_t)$$

for some $a_t > 0$, where $c(x,t) \to c_t \in (0,\infty)$, $\varepsilon(x,t) \to 0$ as $x \to \infty$. Therefore,

$$\lim_{x\to\infty}\frac{\ln \tilde{F}(x,t)}{\ln x} = -k(t) + \lim_{x\to\infty}\frac{\ln c(x,t)}{\ln x} + \lim_{x\to\infty}\frac{1}{x}\int_{a_t}^x \frac{\varepsilon(u,t)}{u}\mathrm{d}u = -k(t).$$

Let $G(z,t) = \frac{\ln \tilde{F}(z,t)}{\ln z}$, then $G(z,t)$ converges to $k(t)$ for all t. Since $F(z,t)$ is the solution, we obtain the following by rewriting (16):

$$\frac{\partial}{\partial t}\frac{\ln \tilde{F}(z,t)}{\ln z} = \frac{1}{\ln z}\frac{\tilde{H}(z,t)}{\tilde{F}(z,t)}\int_0^z \mu(x,t)f(x,t)\mathrm{d}x. \tag{A.17}$$

Then, $\partial G(z,t)/\partial t$ exists for all z and t. In sum, $G(z,\cdot)$ is differentiable for all z, and $G(z,t)$ is convergent for all t as $z \to \infty$. Note that $\partial G(z,t)/\partial t$ is positive, condition (ii) of assumption 2 then implies that its limit exists and is positive. We first consider the case when the limit is finite. One can find $\delta > 0$ such that for any $\varepsilon > 0$, there exists $M > 0$ such that $|\partial G(z', t')/\partial t - \partial G(z'', t')/\partial t| < \varepsilon$ for all $t' \in [t-\delta, t+\delta]$ and $z', z'' > M$. Hence, $G(z,t)$ satisfy all conditions of Theorem 7.17 in Rudin (1976), which states as a result that limit and derivative can be interchanged. Consequently, $k(t)$ is differentiable and satisfies

$$\dot{k}(t) \equiv \frac{\mathrm{d}}{\mathrm{d}t}\lim_{z\to\infty} G(z,t) = \lim_{z\to\infty}\frac{\partial G(z,t)}{\partial t} = \lim_{z\to\infty}\frac{1}{\ln z}\frac{\tilde{H}(z,t)}{\tilde{F}(z,t)}\int_0^z \mu(x,t)f(x,t)\mathrm{d}x < \infty.$$

When $\lim_{z\to\infty}\partial G(z,t)/\partial t = \infty$, integrating both sides of (A.17) gives us

$$\begin{aligned}
\frac{\ln \tilde{F}(z,t+\delta)}{\ln z} - \frac{\ln \tilde{F}(z,t)}{\ln z} &= \int_t^{t+\delta}\frac{1}{\ln z}\frac{\tilde{H}(z,\tau)}{\tilde{F}(z,\tau)}\int_0^z \mu(x,\tau)f(x,\tau)\mathrm{d}x\mathrm{d}\tau \\
\Rightarrow \quad -k(t) - k(t+\delta) &\geq \liminf_{z\to\infty}\int_t^{t+\delta}\frac{1}{\ln z}\frac{\tilde{H}(z,\tau)}{\tilde{F}(z,\tau)}\int_0^z \mu(x,\tau)f(x,\tau)\mathrm{d}x\mathrm{d}\tau \\
&\geq \int_t^{t+\delta}\lim_{z\to\infty}\frac{1}{\ln z}\frac{\tilde{H}(z,\tau)}{\tilde{F}(z,\tau)}\int_0^z \mu(x,\tau)f(x,\tau)\mathrm{d}x\mathrm{d}\tau = \infty
\end{aligned}$$

The last inequality comes from Fatou's lemma and the last equality from the uniform convergence of $\partial G(z,t)/\partial t$ on $[t, t+\delta]$. Since $k(t+\delta) \geq 0$, we must have $k(t) = -\infty$. In other words, we have proved that when $k(t) < \infty$, $\lim_{z\to\infty}\partial G(z,t)/\partial t < \infty$. This justifies the last part of lemma A.2 and completes the proof.

E.2 Proof of Lemma A.3

Proof. Note that $\tilde{H}(z,t)$ is regularly varying with exponent $-h(t) < -1$. So $\lim_{z\to\infty} z\tilde{H}(z,t) = 0$. Then, equation (A.13) implies that

$$\lambda(x,t) = \frac{\mu(x,t)}{x}\int_x^\infty (y - x)h(y,t)\mathrm{d}y = \frac{\mu(x,t)}{x}\int_x^\infty \tilde{H}(y,t)\mathrm{d}y,$$

in which the last equality is obtained from integration by parts. By the Karamata's theorem,

$$\int_x^\infty \tilde{H}(y,t)\mathrm{d}y \sim \frac{1}{h-1}x\tilde{H}(x,t).$$

Taking it back into the above equation, we obtain equation (A.15).

E.3 Proof of Proposition 6

Proof. Since $k(t)$ decreases strictly, lemma A.2 and its remarks imply that $k(t) < \infty$ for all t, and

$$\dot{k}(t) = -\lim_{z\to\infty} \frac{1}{\ln z}\frac{\tilde{H}(z,t)}{\tilde{F}(z,t)}\int_0^z \mu(x,t)f(x,t)\mathrm{d}x > -\infty.$$

Then, we must have $h(t) \geq k(t)$. Otherwise, $\tilde{H}(z,t)/\tilde{F}(z,t)$ will be regularly varying with index $-h(t) + k(t) > 0$. The limit must be a contradictory minus infinity. That $k(t)$ decreases strictly also implies $k(t) > 0$. Otherwise, it violates that $k(t) \geq 0$ for all t. Note that

$$\eta(t) \equiv -\frac{\dot{k}(t)}{k(t)} = \lim_{z\to\infty} \frac{\mu(z,t)\tilde{H}(z,t)}{\ln z}\frac{\int_0^z \mu(x,t)f(x,t)\mathrm{d}x}{\mu(z,t)zf(z,t)} \in (0, \infty), \tag{A.18}$$

in which we use that $\lim_z \frac{zf(z,t)}{\tilde{F}(z,t)} = k(t)$. We discuss below all three cases on the relationship between $m(t)$ and $k(t)$.

Case 1: $m(t) > k(t)$. In this case, $\mu(x,t)f(x,t)$ is regularly varying with index $m(t) - k(t) - 1 > -1$. The Karamata's theorem implies that

$$\lim_{z\to\infty} \frac{\int_0^z \mu(x,t)f(x,t)\mathrm{d}x}{\mu(z,t)zf(z,t)} = \frac{1}{m(t)-k(t)}.$$

Therefore, $\eta(t)$ can be rewritten as

$$\begin{aligned}\eta(t) &= \lim_{z\to\infty} \frac{\int_0^z \mu(x,t)f(x,t)\mathrm{d}x}{\mu(z,t)zf(z,t)} \lim_{z\to\infty} \frac{\mu(z,t)\tilde{H}(z,t)}{\ln z} \\ &= \frac{h(t)-1}{m(t)-k(t)} \lim_{z\to\infty} \frac{\lambda(z,t)}{\ln z},\end{aligned}$$

in which the second equality follows from lemma A.3. Then $\lambda(z,t) \sim C\ln z$ for $C = \frac{\eta(t)(m(t)-k(t))}{h(t)-1}$. Since $\ln z$ is slow varying, $\mu(z,t)\tilde{H}(z,t)$ is also slow varying. Thus, we have $m(t) = h(t)$.

Case 2: $m(t) < k(t)$. In this case, $\mu(x,t)f(x,t)$ is regularly varying with index $m(t) - k(t) - 1 < -1$. The dominated convergence theorem implies that $\lim_{z\to\infty}\int_0^z \mu(x,t)f(x,t)\mathrm{d}x$ exists and is finite. Denoting this limit as $A(t)$, we have

$$\eta(t) = A(t)\lim_{z\to\infty}\frac{\tilde{H}(z,t)}{zf(z,t)\ln z} \Rightarrow \tilde{H}(z,t) \sim \frac{\eta(t)}{A(t)}zf(z,t)\ln z.$$

Since $zf(z,t)\ln z$ is regularly varying with index $-k(t)$, then we have $h(t) = k(t)$. Thus, $\mu(z,t)\tilde{H}(z,t)$ is regularly varying with index $m(t) - k(t) < 0$ and then converges to 0. Lemma A.3 implies that so does $\lambda(z,t)$.

Case 3: $m(t) = k(t)$**.** In this case, $\mu(x,t)f(x,t)$ is regularly varying with index -1. Proposition 1.5.9a of BGT shows that

$$\lim_{z\to\infty} \frac{\int_0^z \mu(x,t)f(x,t)\mathrm{d}x}{\mu(z,t)zf(z,t)} = \infty$$

and $\int_0^z \mu(x,t)f(x,t)\mathrm{d}x$ is slow varying. Then, (A.18) implies that $\lim_{z\to\infty} \frac{\mu(z,t)\tilde{H}(z,t)}{\ln z} = 0$, so do we have $\lambda(z,t) = o(\ln z)$ and $m(t) \geq h(t)$. Suppose we have $m(t) < h(t)$. Then there exists $\varepsilon > 0$, such that $\lim_{z\to\infty} \frac{\mu(z,t)\tilde{H}(z,t)}{z^{-\varepsilon}\ln z} = 0$. On the other hand,

$$\lim_{z\to\infty} z^{-\varepsilon}\frac{\int_0^z \mu(x,t)f(x,t)\mathrm{d}x}{\mu(z,t)zf(z,t)} = 0$$

since both numerator and denominator of the ratio are slow varying. Thus,

$$\eta(t) = \lim_{z\to\infty} \frac{\mu(z,t)\tilde{H}(z,t)}{z^{-\varepsilon}\ln z} \lim_{z\to\infty} \frac{z^{-\varepsilon}\int_0^z \mu(x,t)f(x,t)\mathrm{d}x}{\mu(z,t)zf(z,t)} = 0,$$

contradicting (A.18). Then we must have $m(t) = h(t)$ and complete the proof.

E.4 Proof of Proposition 7

Proof. I first show the "if" part that a thickening tail implies Gibrat's law. Suppose we have $\lim_{z\to\infty} \frac{\tilde{H}(z,t)}{\tilde{F}(z,t)} = B(t) \in (0, \infty)$, lemma A.2 implies

$$\dot{k}(t) = B(t) \lim_{z\to\infty} \frac{\int_0^z \mu(x,t)f(x,t)\mathrm{d}x}{\ln z}. \tag{A.19}$$

Since $g^r(t)$ exists, lemma A.3 and assumption 3 imply that $\lim_{z\to\infty} \mu(z,t)\tilde{H}(z,t)$ exists. With internal search and a regularly varying $\tilde{F}$, $\lim_{z\to\infty} \mu(z,t)zf(z,t)$ exists. The L'Hospital rule implies that

$$\lim_{z\to\infty} \mu(z,t)zf(z,t) = \frac{\dot{k}(t)}{B(t)} \in (0, \infty).$$

Internal search further implies that $h(t) = k(t)$. Therefore,

$$
\begin{aligned}
\lim_{z \to \infty} \lambda(z,t) &= \frac{1}{h(t) - 1} \lim_{z \to \infty} \mu(z,t)\tilde{H}(z,t) \\
&= \frac{1}{k(t) - 1} \lim_{z \to \infty} \mu(z,t) z f(z,t) \frac{\tilde{F}(z,t)}{z f(z,t)} \frac{\tilde{H}(z,t)}{\tilde{F}(z,t)} \\
&= \frac{1}{k(t) - 1} \lim_{z \to \infty} \mu(z,t) z f(z,t) \lim_{z \to \infty} \frac{\tilde{F}(z,t)}{z f(z,t)} \lim_{z \to \infty} \frac{\tilde{H}(z,t)}{\tilde{F}(z,t)} \\
&= \frac{\dot{k}(t)}{k(t)(k(t) - 1)}.
\end{aligned}
$$

Next, I show that the "only if" part also stands. From (A.19), it suffices to show that if $g^r(t) \in (0, \infty)$,

$$
\frac{\int_0^z \mu(x,t) f(x,t) \mathrm{d}x}{\ln z} \in (0, \infty).
$$

With $g^r(t) \in (0, \infty)$, lemma A.3 implies $\mu(z,t)\tilde{H}(z,t) = (k(t) - 1)g^r(t) + o(1)$. Therefore,

$$
\mu(x,t) f(x,t) = \mu(x,t)\tilde{H}(x,t) \frac{x f(x,t)}{\tilde{F}(x,t)} \frac{\tilde{F}(x,t)}{\tilde{H}(x,t)} \frac{1}{x} = \frac{c + o(1)}{x},
$$

since $\mu(x,t)\tilde{H}(x,t)$, $\frac{x f(x,t)}{\tilde{F}(x,t)}$ and $\frac{\tilde{F}(x,t)}{\tilde{H}(x,t)}$ all converge to positive and finite constants. Therefore, it is straightforward to verify that $\lim_{z \to \infty} \frac{\int_0^z \mu(x,t) f(x,t) \mathrm{d}x}{\ln z}$ is positive and finite, and obtain (17)

$$
\dot{k}(t) = k(t)(k(t) - 1)g^r(t).
$$

Finally, it is straightforward that when $k(t) \in (1, \infty)$, $g^r(t)$ is ∞ (or 0) if $\dot{k} = \infty$ (or 0).

Supplementary References

Rudin, W. (1976): *Principles of mathematical analysis*, New York: McGraw-Hill, 3 ed.

Temme, N. M. (1996): *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*, John Wiley & Sons.

Van Patten, D. (2020): "International Diffusion of Technology:," Mimeo, Yale SOM.

Additional Materials (Not For Publication)

Appendix F   Additional Results of Section 2

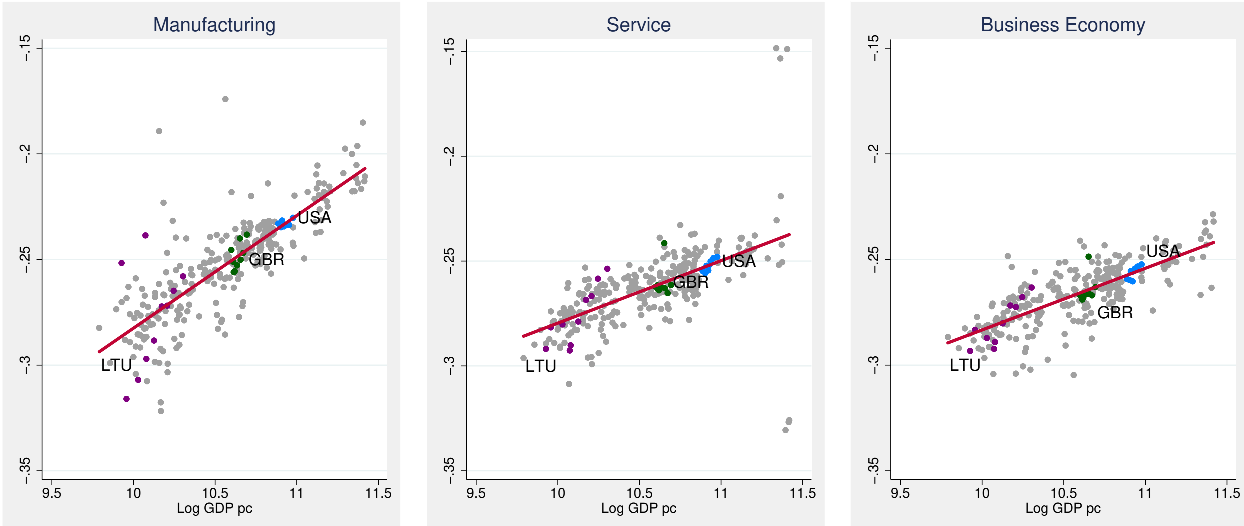
F.1   Additional Results with OECD data

Table A.2: Table: Regression results with OECD countries

	Tot	Tot	Mft	Mft	Sev	Sev
Number of firms						
logGDPpc	0.05543^a	0.09936^a	0.1007^a	0.1553^a	0.1047^a	0.1550^a
	(0.017)	(0.027)	(0.020)	(0.030)	(0.016)	(0.028)
_cons	-1.7249^a	-2.1894^a	-2.0420^a	-2.6186^a	-2.2882^a	-2.8205^a
	(0.187)	(0.283)	(0.215)	(0.316)	(0.172)	(0.299)
Country FE	No	Yes	No	Yes	No	Yes
Obs.	299	299	301	301	300	300
R-sq	0.04718	0.9441	0.09608	0.9648	0.1713	0.9188
Employment						
logGDPpc	0.05431^a	0.02918^a	0.09221^a	0.05327^a	0.02541^c	0.02977^a
	(0.010)	(0.009)	(0.011)	(0.017)	(0.014)	(0.011)
_cons	-0.8418^a	-0.5762^a	-1.2254^a	-0.8137^a	-0.5338^a	-0.5798^a
	(0.107)	(0.096)	(0.117)	(0.184)	(0.149)	(0.119)
Country FE	No	Yes	No	Yes	No	Yes
Obs.	297	297	302	302	298	298
R-sq	0.1134	0.9772	0.1900	0.9677	0.02226	0.9408

Robust standard errors in parentheses. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$.

Notes: This tables reports the correlation between the right tail thickness and log GDP per capita among OECD countres Each observation is a country-year pair. The right tail thickness is measured by either $\tilde{R}_t^f$ or $\tilde{R}_t^{emp}$. GDP per capita is in constant international dollar. Sources: the OECD SBS and PWT 10.0



(a) Manufacturing (b) Service (c) Business Economy

Figure A.4: Right tail thickness $\tilde{R}_t^{emp}$ and the level of development in OECD countries

Notes. This figure plots the right tail thickness $\tilde{R}_t^{emp}$ against the log GDP per capita for each country-year pair in manufacturing, service and the whole business economy. The scatter dots are readjusted by country fixed effects, and the red lines are the linear fits. Appendix F.1 documents the details of the construction. The right tail thickness $\tilde{R}_t^{emp}$ is calculated using the OECD SBS data and with $T_S = 10$ and $T_L = 250$. Data on GDP per capita are from the PWT 10.0. Three annotated countries are Lithuania (LTU), the UK (GBR) and the USA.

F.2 Additional Results with WBES data

Countries in the sample High income countries (17): Bahamas, Barbados, Cyprus, Czech Republic, Estonia, Greece, Hungary, Israel, Italy, Latvia, Malta, Poland, Portugal, Slovak Republic, Slovenia, Sweden, Trinidad and Tobago.

Upper-middle income countries (41): Antigua and Barbuda, Argentina, Azerbaijan, Belarus, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Chile, China, Costa Rica, Croatia, Dominica, Dominican Republic, Fiji, Gabon, Grenada, Jamaica, Jordan, Kazakhstan, Lebanon, Lithuania, Malaysia, Mauritius, Mexico, Montenegro, North Macedonia, Panama, Romania, Russian Federation, Serbia, South Africa, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines,Suriname, Thailand, Tunisia, Turkey, Uruguay, Venezuela.

Lower-middle income countries (39): Albania, Angola, Armenia, Belize, Bhutan, Bolivia, Cabo Verde, Cameroon, Colombia, Congo, Rep., Côte d'Ivoire, Djibouti, Ecuador, Egypt, El Salvador, Eswatini, Georgia, Guatemala, Guyana, Honduras, India, Indonesia, Iraq,Lesotho, Moldova, Mongolia, Morocco, Myanmar, Namibia, Nicaragua, Paraguay, Peru, Philippines, Sri Lanka, Sudan, Ukraine, Vietnam, West Bank and Gaza, Yemen.

Low income countries (36): Bangladesh, Benin, Burkina Faso, Burundi, Cambodia, Central African Republic, Chad, Congo, Dem. Rep., Ethiopia, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Kyrgyz Republic, Lao PDR, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Nepal, Niger, Nigeria, Pakistan, Rwanda, Senegal, Sierra Leone,Tajikistan, Tanzania, Togo, Uganda, Uzbekistan, Zambia ,Zimbabwe.

Table A.3: Table: Regression results with developing countries

	Num	Num	Num	Emp	Emp	Emp
logGDPpc	0.08878^a	0.08359^a	0.07770^a	0.03461^a	0.02567^b	0.03794^a
	(0.014)	(0.011)	(0.015)	(0.013)	(0.010)	(0.013)
_cons	-1.4300^a	-1.3866^a	-1.3332^a	-0.5877^a	-0.5140^a	-0.6169^a
	(0.127)	(0.106)	(0.132)	(0.113)	(0.096)	(0.119)
HI countries	No	Yes	No	No	Yes	No
Year FE	No	No	Yes	No	No	Yes
Obs.	236	274	236	232	270	232
R-sq	0.1352	0.1526	0.2542	0.03568	0.02582	0.09235

Robust standard errors in parentheses. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Notes: This table reports the correlation between the right tail thickness and log GDP per capita among countries in the WBES. Each observation is a country-year pair among countries. The right tail thickness is measured by either $\tilde{R}_t^f$ (Num) or $\tilde{R}_t^{emp}$ (Emp). GDP per capita is in constant international dollar. Sources: the OECD SBS and PWT 10.0

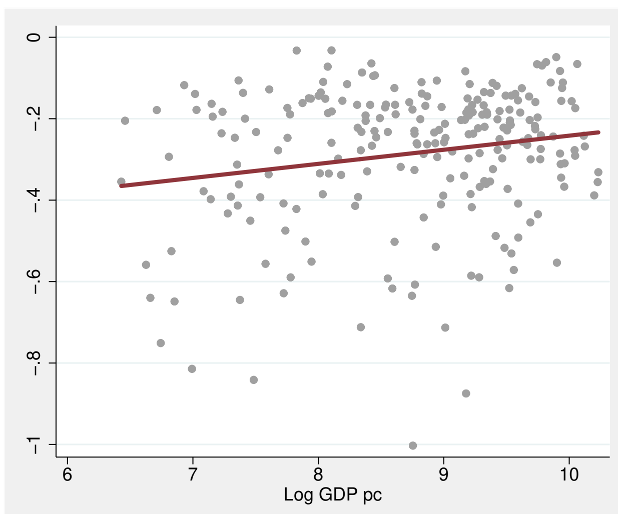
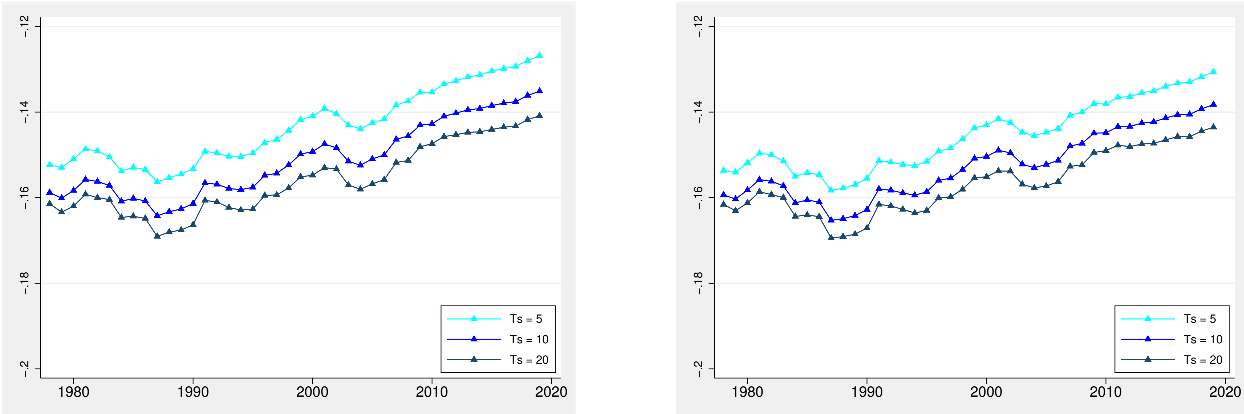


Figure A.5: Right tail thickness $\tilde{R}_t^{emp}$ and the level of development in developing countries

Notes. This figure plots the right tail thickness $\tilde{R}_t^{emp}$ against the log GDP per capita for each country-year pair in the business economy. The red line is the linear fit. The right tail thickness $\tilde{R}_t^{emp}$ is calculated using data on developing countries of the WBES and with $T_S = 5$ and $T_L = 100$. Data on GDP per capita are from the PWT 10.0.

F.3 Additional Results with the US BDS data



(a) $T_L = 500$ (b) $T_L = 1000$

Figure A.6: Changes in the right tail thickness $\tilde{R}_t^{emp}$ in the US (1978-2019)

Notes. This figure plots the right tail thickness $\tilde{R}_t^{emp}$ of the size distribution of all US business firms from 1978 to 2019. The right tail thickness $\tilde{R}_t^{emp}$ is calculated using the census BDS data and with various T_S and T_L indicated in the figure.

Appendix G Numerical Exercises in Section 4.4

Despite its parsimony, the model in section 3 makes theoretical predictions that are consistent with data. Specifically, the simple model makes two strong predictions: 1) output per capita y is proportional to $k/(k \quad 1)$, and 2) $k \quad 1$ decreases to 0 at a constant rate which equals to the long-run growth rate. I confront both predictions with data in the first two exercises. In the last one, I calibrate model parameters and test model predictions on untargeted moments.

G.1 y vs. $k/(k \quad 1)$

The model yields a simple relationship between output per capita and the right tail index of the firm size distribution, i.e.,

$$y(t) \ / \ \frac{k(t)}{k(t) \quad 1}, \quad \text{or} \quad \frac{y(t)}{y(t_0)} = \frac{k(t)}{k(t) \quad 1} \Big/ \frac{k(t_0)}{k(t_0) \quad 1},$$

in which $k(t)$ is the Pareto shape parameter, and t_0 is a chosen date. As discussed in section 2, the number-based thickness measure $\tilde{R}^f$ is equal to $\quad k$ with Pareto firm size distributions, enabling $\hat{k} = \quad \tilde{R}^f$ an estimator of k.

I test this relationship using data from the United States during 1978-2019. First, I use my preferred thickness measure from the US BDS ($\tilde{R}^f$ with $T_S = 20$ and $T_L = 500$) and compute a data series of $\hat{k}/(\hat{k} \quad 1)$ for each year between 1978 and 2019. Next, I obtain data on the US GDP per capita in the same period from the Federal Reserve Economic Data (FRED). I further

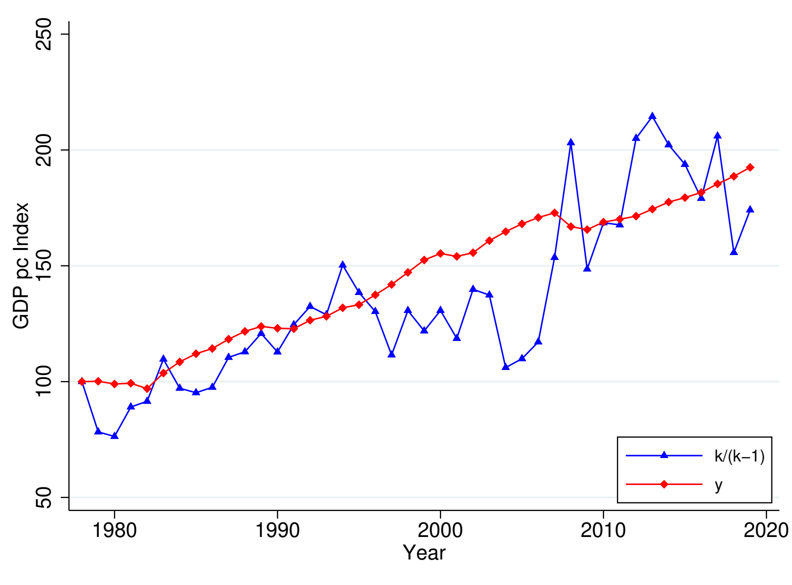


Figure A.7: y vs. $k/(k \quad 1)$

Notes. This figure compares the relationship between y and $k/(k \quad 1)$ for the US between 1978 and 2019. The estimates $\hat{k} = \quad \tilde{R}^f$, in which $\tilde{R}^f$ is from section 2.3.3 with $T_S = 20$ and $T_L = 500$. The data on y, GDP per capita, comes from FRED. Both data series are in level and normalized with respect to their 1978 level.

normalize both data series relative to their respective 1978 levels, both of which are assigned an index value of 100. It is evident from figure A.7 that the two variables exhibit a strikingly close alignment over the 1978-2019 period in the United States.

G.2 Changes in $k \quad 1$

The second testable prediction is about the changes in $k \quad 1$. Equation (11) makes it transparent that $k \quad 1$ decreases to zero at a constant rate, which is equal to the long-run output per capita growth g . That is, if regressing $\ln(k_t \quad 1)$ over time, i.e.,

$$\ln(k_t \quad 1) = \alpha \quad \beta t, \tag{A.20}$$

the model will predict that $\beta = g$.

As before, I obtain the dependent variable, $\ln(k_t \quad 1)$, using the number-of-firms-based thickness measure with $T_S = 20$ and $T_L = 500$. Regressing it on time, I obtain the following results:

$$\ln(k_t \quad 1) = \underset{(0.044)}{2.61} \quad \underset{(0.002)}{0.02}\, t.$$

The estimated $\hat{\beta}$ is 0.02, with a standard deviation of 0.002. It hits strikingly well with the well-known US long-run growth rate of 2%.

G.3 Calibrating the model parameters

Finally, the equilibrium analysis in section 4.1 shows that the initial shape parameter k_0 and the population size L completely determine the equilibrium allocation. In this exercise, I target the right tail thickness of US in 1978-2019 to calibrate the model and compare the model predictions on GDP per capita with that in the data.

Note that equation (10) implies that

$$\ln(k_t \quad 1) = \ln(k_0 \quad 1) \quad Lt.$$

Hence, $k_0 = e^{\alpha} + 1$ and $L = \beta$, in which α and β are in the regression (A.20). Using estimates $\hat{\alpha}$ and $\hat{\beta}$ from the previous exercise, the estimated initial shape $\hat{k}_0$ is 1.073, and the estimated population size $\hat{L}$ is 0.02. Note that the initial shape corresponds to the 1978 level. Based on these two parameters and equation (10), the model generates a sequence of $\hat{k}_t$ for each year t. Notably, the model predicted shape parameter in 1997 is 1.050, which is very close to the 1.059 reported by Axtell (2001). The following figure A.8 shows that the model fits targeted right tail thickness very well.

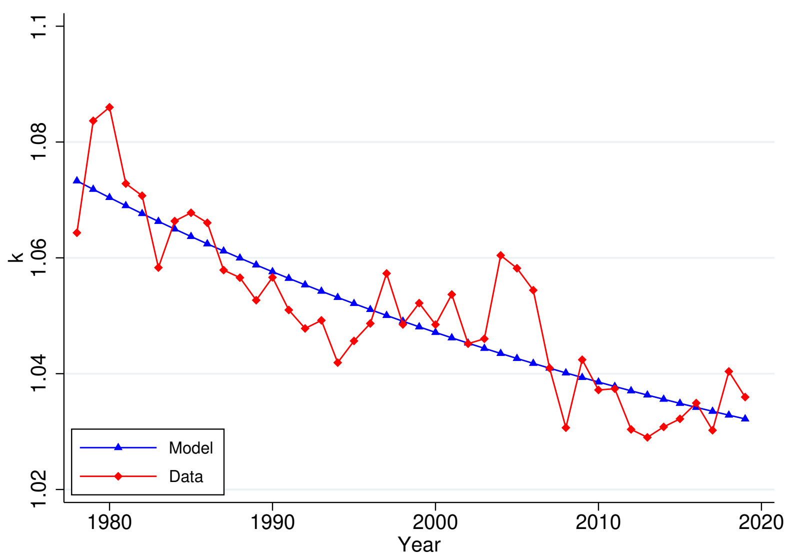


Figure A.8: Targeted: The right tail shape k

Notes. This figure compares the right tail shape predicted by the model with that measured by the data. The model predicted k is obtained using equation (10) with $k_0 = 1.073$ and $L = 0.02$. The data measured k is recovered from the number-of-firms-based thickness measure with $T_S = 20$ and $T_L = 500$.

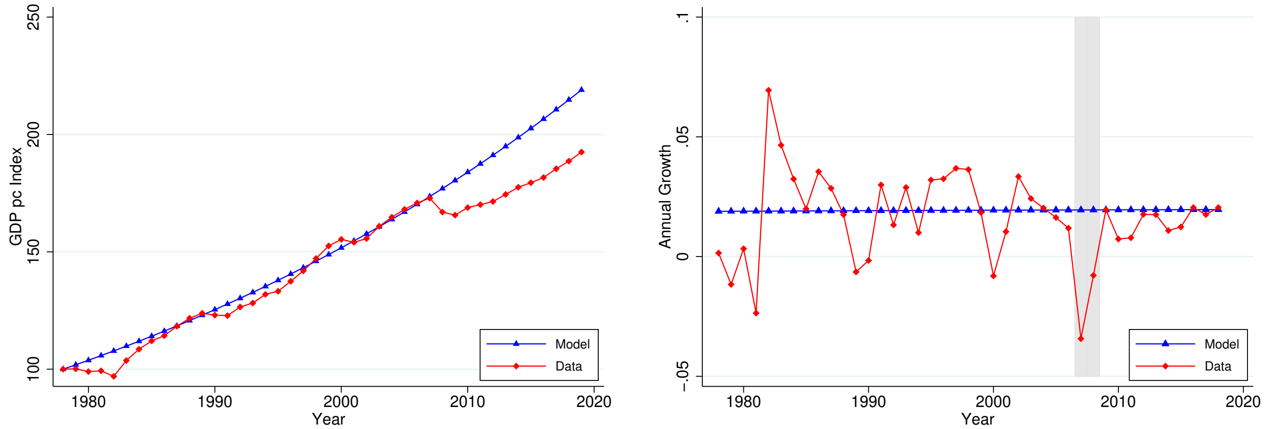


Figure A.9: Untargeted: GDP per capita growth

Notes. This figure compares the 1978-2019 US GDP per capita growth predicted by the model with that in the data. The left figure plots the US GDP per capita in index with base year 1978. The right figure shows the US GDP per capita in annual growth rate, and the shaded bar is the crisis period 2008-2009. The model implied GDP per capita index is obtained using parameters $\hat{k}_0 = 1.073$ and $\hat{L} = 0.02$. The data on the GDP per capita index is from FRED.

Using the sequence of $\hat{k}_t$ and $\hat{L}$, I obtain a sequence of model predicted GDP per capita $\hat{y}_t$. The calibrated match closely matches the untargeted normalized US GDP per capita index. In figure A.9, I compare the fitted value generated by the model with data over the period 1978-2019. The left figure plots the normalized GDP per capita index, and the right figure shows the annual growth rate. Both figures show remarkable alignment between the data and the model. The model can match the data almost perfectly before the financial crisis. Both model and data have an average growth rate of 1.92% from 1978 to 2007. The deviation between the model and data in the crisis period 2008-2009 is unsurprising given that this is a growth model without any negative shocks. Nevertheless, it is obvious from the right figure that the annual growth rate in the data is reversing to the model level, or the long-run level, after the crisis.