Economic Growth and the Rise of Large Firms*

Zhang Chen

Princeton University

Job Market Paper

November 15, 2022

Click here for the lastest version

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 diffusion 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. Despite its parsimony, the model provides a good quantitative fit to the US GDP per capita growth. I prove that, in a general class of idea diffusion models, Gibrat's law holds if and only if the right tail of the firm size distribution grows thicker. The simple model is the only one consistent with Gibrat's law and a thickening tail under common functional form assumptions. Finally, I show that policies favoring large firms can improve welfare due to the externality associated with idea diffusion.

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, Treb Allen, David Argente, Cheng Chen, Pablo Fajgelbaum, Nobuhiro Kiyotaki, Ernest Liu, Yueran Ma, Kiminori Matsuyama, Jonathan Payne, Esteban Rossi-Hansberg, 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. All errors are my own.

Email: zhangc@princeton.edu

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 diffusion 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 developments in endogenous growth theory that study idea diffusion among heterogeneous firms as a source of growth. I develop a novel idea diffusion 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 heterogenous productivities. 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 in $t+h$. 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 diffusion 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 that 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: growth is generated by a 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. That is, 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 square 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 $k-1$ 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.

Despite its parsimony, the model makes theoretical predictions consistent with the data. Notably, the simple model makes two strong predictions: 1) output per capita $y = k/(k-1)$, and 2) $k-1$ decreases to 0 at a constant rate equal to the long-run growth rate. Both predictions can be readily tested using my estimates for k. For the US from 1978 to 2019, the actual growth in output per capita aligns closely with changes in $k/(k-1)$. Moreover, regressing $\ln(k-1)$ on time, the estimated coefficient perfectly hits the well-known US growth rate of 2%.

In idea diffusion models, the learning function, which describes how ideas arrive, determines the equilibrium properties. As discussed earlier, my model's departure in the learning function relative to existing models results in equilibria with very different properties. Moreover, relatively little is known in the literature about how to discipline the learning function using data, partly because agents' learning processes are hard to measure directly. To address this issue, I ask whether indirect empirical moments can provide discipline for the learning function. Specifically, I consider a general class of idea diffusion models that nests my model, most other existing models, as well as other learning heterogeneities. I show that, in this class of models, Gibrat's law holds if and only if the right tail becomes thicker. That is why Gibrat's law does not hold in existing idea diffusion models that are characterized by a stationary firm size distribution. Furthermore, I show that Gibrat's law and a thickening tail discipline the learning function. Under common functional form assumptions on the learning function, I show that Gibrat's law and a thickening tail identify the learning function, which is exactly the one I assumed in my model. In this sense, my model is the only idea diffusion model that generates both Gibrat's law and a thickening right tail.

Finally, the model delivers new policy implications. Idea search by each firm has externalities 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 externalities on all firms in the economy, search by small firms has few externalities 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 approximately a 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.

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 a leading explanation for this finding: the fact that firms are small in developing countries is the outcome of under-investment because larger firms are more exposed to institutional distortions. Most existing work uses steady-state comparisons to capture cross-country differences, which rely on exogenous cross-country variation in distortions. This paper targets instead the right tail of the firm size distributions and generates the relationship as a feature of the transition path.

Second, there have been recent interests in the rise of market concentration in the US for the past several decades, for example, Gutiérrez and Philippon (2018), De Loecker et al. (2020), Autor et al. (2020), and Kwon et al. (2022). This paper contributes to this literature with a novel 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. Most papers do not study the right tail of the firm size distribution except for Oberfield (2018), which proposes a theory of the endogenous formation of a production network based on input choices. His model implies that a decline in the labor share results in a thicker right tail. 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 theory in which balanced growth is compatible with the rise in concentration, and I present new empirical evidence consistent with the predictions of my theory.

Third, this paper builds heavily upon burgeoning literature on endogenous growth models with idea diffusion. Buera and Lucas (2018) provide a comprehensive survey of this literature.[2] This paper differs 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 diffusion models. I further show that in a general class of idea diffusion models, which includes these earlier models, Gibrat's law holds if and only if the right tail becomes thicker.

Fourth, this paper contributes to the study of Zipf's law. Axtell (2001) documented that the

[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).

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 directly as the result of tail growth. 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 proposes and analyzes the simple model, and section 4 validates it with quantitative exercises. Section 5 further extends the simple model and presents some general results. Section 6 discusses the policy implications. Finally, section 7 concludes.

2 Stylized Facts

In this section, I present empirical evidence of a positive relationship between the right tail thickness of the firm size distribution and the level of economic development. In section 2.1, I describe the construction of my thickness measure and discuss its implications. I introduce three complementary datasets in section 2.2 to compute the right tail thickness. In section 2.3, I show the positive correlation between the right tail thickness and log GDP per capita at the country-year level. The robustness of this correlation receives multiple validations in various settings and suggests the necessity of a generic theory.

2.1 A Tail Thickness Measure

It remains challenging to measure the right tail of firm size distributions in multiple countries. Despite a large literature on estimating the tail index of thick-tailed distributions, sophisticated statistical procedures rely on a relatively large number of individual observations.[3] Meanwhile, it is very difficult to obtain a large sample of countries with adminstrative micro data on firms, which are usually confidential in each country. Hence, a suitable measure should capture the tail thickness in a simple manner and work with coarsely tabulated data on firm size distributions.

I construct a tail thickness measure using the share of large and small firms. Let $F(x,t)$ be the CDF of the underlying distribution of firm employment size, and f the density function. Then, $\tilde{F}(x,t) \equiv 1 - F(x,t)$ denotes the fraction of firms with size greater than x, and $\tilde{F}^{emp}(x,t) \equiv \int_x^\infty y \mathrm{d}F(y,t)$ the total employment in firm with size greater than x. In addition, let T_L be the employment size threshold for large firms and T_S for small firms. I use the the normalized log

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

relative share of large firms $\tilde{R}_t^f$ to measure the right tail thickness. Formally,

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

Conceptually, the right tail index reflects how quickly the fraction of population above the threshold drop with the threshold. A tail is thinner if there is faster decay in the fraction. In the same spirit, $\tilde{R}_t^f$ captures the drop from the fraction of above-small firms to large firms if raising the threshold from small to large firms. A larger $\tilde{R}_t^f$ in absoulte value then implies a thinner tail. More generally, it is an empirical counterpart of the tail index if the underlying distribution is thick-tailed.[4] A special case is that $\tilde{R}_t^f = -k_t$ when $F(x,t)$ is a Pareto distribution with shape parameter k_t. On the other hand, a similar construction is to use the employment share instead of the number share. Let $\tilde{R}_t^{emp}$ be the normalized log relative employment share of large firms, i.e.,

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

Similarly, it is easy to show that $\tilde{R}_t^{emp} = 1 - k_t$ if the underlying firm size distribution is Pareto with shape parameter k_t. For expositional brevity , I use the number-based statistics $\tilde{R}_t^f$ as the measure of the right tail thickness for the rest of this section. Additional results using the employment-based statistics $\tilde{R}_t^{emp}$ can be found in the appendix. I obtain consistent results using both measures.

It is worth noting that this thickness measure has the following implications. First, aggregate data on the number of enterprises or employment by firm size bins are sufficient to compute this measure. It also works with coarse tabulations such as a simple division of small, medium and large firms. The simple structure is particularly useful for cross-country datasets, which are mostly tabulated aggregate statistics. Second, countries may differ in firm size bins. For example, measuring firm size by the number of employees, a few countries may use thresholds such as 5 and 19, while the majority use 10 as the small firm threshold. The normalization term makes the thickness measure adjustable for slight discrepancies in the thresholds 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 less of a challenge 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.

[4] A distribution function has tail index k if $\lim_{x\to\infty} \frac{\tilde{F}(tx)}{\tilde{F}(x)} = t^{-k}$ for all $t > 0$. Note that the LHS of the equation is the ratio between fractions above different thresholds. Then, $\tilde{R}^f$ is obtained with a particular choice of thresholds. With a Pareto distribution, the above equation holds for any t and x.

2.2 Data

The goal of the empirical exercise is to investigate the correlation between the right tail thickness and the level of economic development. A common proxy for the level of economic development is GDP per capita. I obtain data on real GDP per capita in constant international dollars from the Penn World Table (PWT) version 10.0. For better cross-country comparisons, I use total employment to measure a firm's size. The primary data source on firm size distributions is the OECD database of Structural Business Statistics (SBS) by ISIC Rev 4. The final sample is a panel of 33 OECD countries between 2008 and 2017.[5] It classifies firms into five size bins by the number of employees: 0-9, 10-19, 20-49, 50-249, and 250+. This dataset also includes sectoral information on firm size distributions, which are useful to test the correlation at the sector level.

There are two caveats with the OECD data. 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. The World Bank Enterprise Survey (WBES) 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+. Most of the countries are surveyed only once or twice, so it seems most plausible to view the WBES as cross-sectional evidence. 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). This long time series is adequate to present how the right tail changes in a representative growing economy. Specifically, US firms are clustered into 10 size bins ranging from 1 to 10,000+ employees. This level of precision enables sensitivity checks on the choice of thresholds.

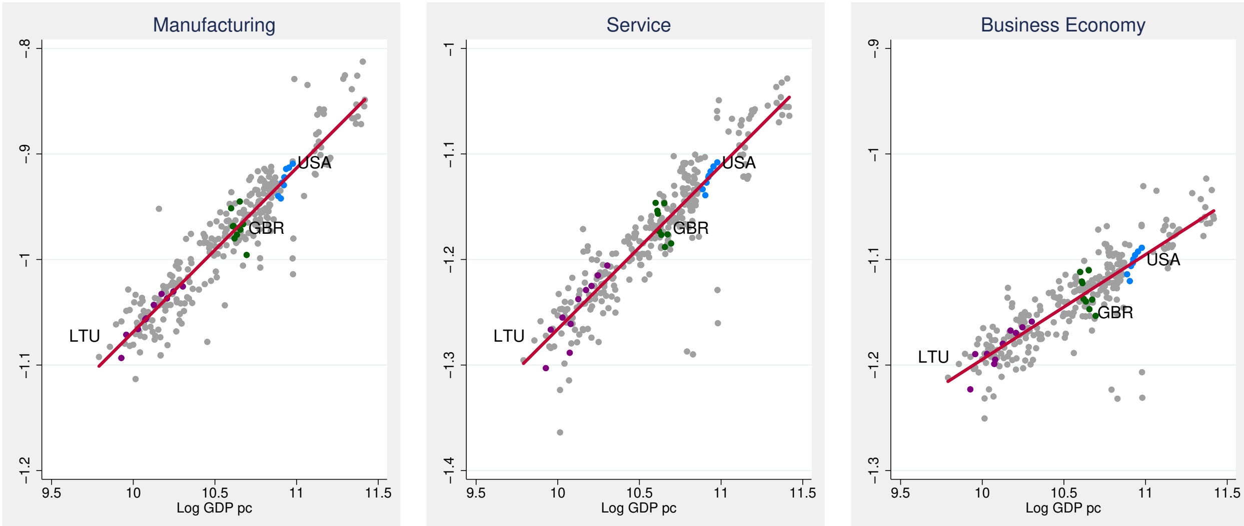
2.3 Results

This part describes the estimation of the correlation between the right tail thickness and the level of development in each dataset and presents the results in figures.

OECD countries With OECD countries, I construct the right thickness measure $\tilde{R}_t^f$ for each country-year pair based on small firm threshold $T_S = 10$ and large firm threshold $T_L = 250$. The choice of thresholds follows the OECD small and medium enterprise (SME) standard. To obtain the correlation, I regress the right tail thickness measure on the log GDP per capita and control for country fixed effect. In figure 1 , the correlation manifests itself as the slope of the linear fit

[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.



(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 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 A.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.

line in red. The dots in figure 1 are obtained by evaluating the residual thickness measure at the unweighted average of country fixed effects. Filtering out country-specific components, I construct a synthetic country with a protracted span of development stages out of all the countries. Figure 1 then plots the trajectory of the right tail thickness in this synthetic country in dots and the trend in lines. Three annotated data series exhibit the trajectories of three countries–Lithuania, the UK and the US. Together, they suggest that the positive correlation seems to hold across countries with distinct levels and growth rates of economic development.

In a nutshell, the main message from the OECD data is that the right tail becomes thicker as the economy grows. Removing country fixed effect is pivotal to this statement since the correlation is identified using only within-country over-time variations. Therefore, theories relying on cross-country differences in exogenous factors are unlikely to explain this correlation. Besides, not only does the positive correlation hold in the overall business economy (panel 1c), it also holds separately in manufacturing (panel 1a) and service (panel 1b). That is to say, a thickening right tail is not merely a composition effect driven by specific sectors. Nor is it largely a result of international trade, given that the non-tradable sector (service) witnesses similar changes as well. In sum, this positive relationship seems plausibly a generic feature of the growth process.

Appendix A.1 stores additional results of the estimation using the OECD data. It provides details on the construction of figure 1. Furthermore, full regression results, with and without country fixed effect, can also be found there.

Developing countries To complement with the above finding, I restrict the WBES sample to developing countries, i.e., the 113 low, lower-middle and upper-middle income countries covered in the survey. I construct the right thickness measure $\tilde{R}_t^f$ for each country-survey year pair based on small firm threshold $T_S = 5$ and large firm threshold $T_L = 100$. Figure 2 plots the right tail thickness against log GDP per capita without controling for any fixed effects. The slope of the red linear fit line visualizes the positive correlation between the right tail thickness and the level of development among the sampled developing countries. In words, figure 2 suggests that richer countries tend to have thicker tails.

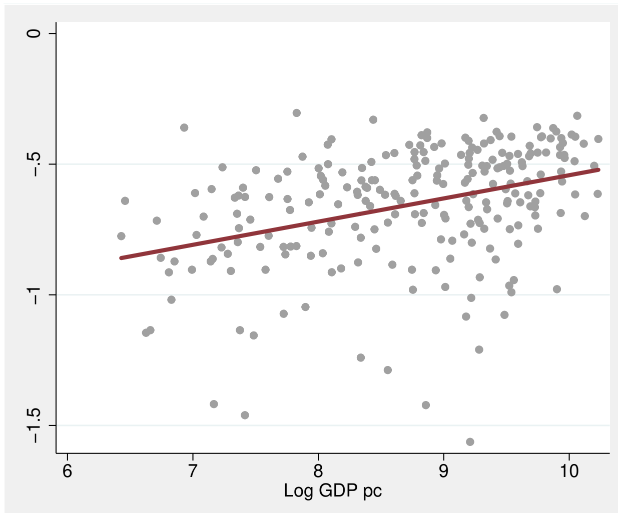


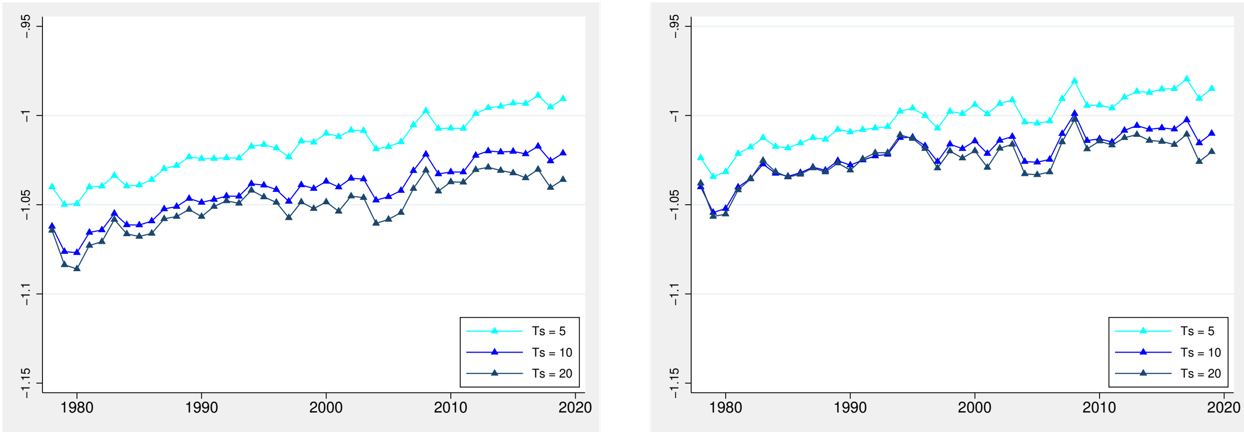
Figure 2: Right tail thickness and the level of development in developing countries

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.

This finding is consistent with figure 3 of García-Santana and Ramos (2015). They use an older version of the WBES and document a negative cross-country association between productivity (or GDP per worker) and the share of employment in small plants.[7] Appendix A.2 presents the regression results and additional checks. It validates that the positive correlation holds with and without high-income countries and using both thickness measures.

the US The choice of thresholds is more diverse in the case of the United States. The US Small Business Administration has a table of size thresholds for firms of different industries to qualify for federal government small business programs. The thresholds range from 100 to 1500 across industries and are usually much larger than the 250 large firm threshold by the OECD SME standard. I choose two large firm thresholds $T_L = 500$ and $T_L = 1000$ in order to be more comparable with the previous results. Similarly, I experiment with three small firm thresholds

[7] Their sample consists of 104 countries surveyed betwen 2006 and 2010.



(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.

$T_S = 5, 10, 20$, around the OECD threshold 10. Combinations of small and large firm thresholds also test the sensitivity of the results to the choice of thresholds. It is clear from figure 3 that there is a positive trend (or a negative trend in absolute value) on the right tail thickness of the US firm size distribution in the past forty years. In addition, the trend is very stable regardless of the threshold choices.

That the right tail becomes thicker in the US joins a growing empirical literature on 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) documents that in various industries, the share of the top 4 or 20 firms in that industry in total sales or employment has been increasing from 1980 to 2010. Kwon et al. (2022) also confirms 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.

In a nutshell, the empirical evidence in this section suggests a widespread positive relationship between the level of economic development and the thickness of the right tail. It holds across countries and for within-country changes over time, in developed and developing countries, and by major sectors. The robustness of this relationship signals a generic underlying mechanism.

3 A Simple Idea Diffusion Model

This section presents a simple idea diffusion model with endogenous growth and firm size distribution. In the model equilibrium, the right tail of the firm size distribution becomes thicker along the transition path, which converges to a balanced growth path. Hence, that the right tails are thicker with higher development can be understood as an inherent feature of the growth path. The model also rationalizes well-known stylized facts about firm growth, namely Gibrat's and Zipf's law. The following begins with a description of the model, proceeds with an equilibrium analysis, and ends with a discussion on specific assumptions.

3.1 Model Description

Production Consider an economy with a continuum of firms, which are indexed by j. There are no entry and exit of firms, so it is without loss to normalize the measure of total firms to one. The structure of production is very similar to that in an endowment economy. The only production factor is machine. Each firm owns one machine and uses it to produce homogeneous goods. Machines are specific to firms, so firms have no incentive to sell their machines to other firms (for zero price). They also do not depreciate and cannot be replicated. Therefore, firms make no production decisions and receive output from their machines. Machines are heterogeneous in their productivities. A machine with productivity z produces z goods at each instant. Since productivity is the only source of firm heterogeneity, firms are equivalently indexed by the productivity of their machines, z_j. The market is perfectly competitive, and the output good is the numeraire. Taking the price of the output as given, a firm with productivity z, or a firm z, sells z goods and makes z profits in each instant. The only way that a firm can increase its profit is by upgrading its machine.

Learning Firms enhance the productivity of their machines by learning from more productive firms in meetings. The details of the meeting and learning process are as follows. The meeting opportunity facing a firm z follows a non-homogeneous Possion process, in which firm chooses the arrival rate $\eta(z,t)$ at each instant. 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. Contrasting to other idea diffusion models[8], firms' search for more advanced technology are directed in the sense that they target on more productive firms. Whereas conditional on the pool of more productive firms, the realization of the other firm in the meeting is still random as usual.

On the other hand, firms hire researchers to conduct the search and upgrade the technology. The required research efforts rise with the complexity of the expected transferring technology. That is, with the same arrival rate, more productive firms will hire more researchers to complete adoption since the expected new technology will be more sophisticated. Firms also need to hire more researchers with a higher arrival rate because the likelihood of technology upgrade increases and so does the workload. Specifically, firm z has to hire $z\eta$ researchers to achieve an arrival rate

[8] For example, Kortum (1997), Alvarez, Buera and Lucas (2008), Perla and Tonetti (2014), and Lucas and Moll (2014).

of η. It pays a search (or adoption) cost $z\eta w(t)$ given researcher's wage $w(t)$.

Given interest rate $r(t)$ and wage $w(t)$, each firm solves the following profit maximization problem:

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

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

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

The associated HJB equation is then:

$$r(t)v(z,t) = z + \max_{\eta \geq 0}\left\{\eta \int_z^\infty [v(x,t) - v(z,t)]\mathrm{d}F(x|x \geq z, t) - z\eta w(t)\right\} + \partial_t v(z,t). \tag{3.1}$$

Equation (3.1) says that the flow value of the firm (the LHS term) is the sum of 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 aggreagte state (the fourth term).

Consumption In this economy, there live a continuum of representative households with measure L. Households are infinitely lived and indexed by $i \in [0, L]$. Each of them is endowed with one unit of time. Since there is no opportunity cost of work, households spend all their time working for firms as researchers, i.e., the labor supply is inelastic. Hence, a unit of labor means to hire a labor for all her time. There is no population growth, so the total labor supply is fixed to L. Households are owners of all firms in the economy 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 $y(t) = w(t) + \pi(t)/L$. Households have CRRA utility and maximize the present value of their utilities subject to their income flow at all times:

$$\max_{c(\tau)\geq 0} \quad \int_t^\infty e^{-\rho(\tau - t)}\frac{c(\tau)^{1-\theta} - 1}{1-\theta}\mathrm{d}\tau. \tag{3.2}$$

Also, households may borrow and lend in the financial market at interest rate $r(t)$.

Market Clearing There are markets for goods and labor. With representative households, total consumption of output goods $C(t) = Lc(t)$. Goods market clears such that

$$C(t) = Y(t) \equiv \int_0^\infty zf(z,t)\mathrm{d}z. \tag{3.3}$$

Similarly, I have

$$\int z\eta(z,t)f(z,t)\mathrm{d}z = L \tag{3.4}$$

as the labor market clearing condition. The LHS is the total labor demand from aggregating over individual labor demand $z\eta(z,t)$, and the RHS is the inelastic labor supply.

Aggregate Dynamics The aggregate state variable of this economy is the distribution of firm productivity. Firms' learning behaviors continuously shape the firm size distribution. Given the productivity distribution at time t, $F(\cdot,t)$, and firms' choices on the arrival rate $\eta(z,t)$, the productivity distribution at time $t+\mathrm{d}t$ satisfies

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

In words, firms with productivity no greater than z at $t+\mathrm{d}t$ are those which have productivity no greater than z at time t and have not met another firm with productivity greater than z during the interval $[t,t+\mathrm{d}t]$. These firms either do not have any meeting opportunities or only meet firms with productivity lower than z. For a firm with productivity x at time t, the former event happens with productivity $1-\eta(x,t)\mathrm{d}t$, and the latter event with productivity $\eta(x,t)\mathrm{d}tF(z|z\geq x,t)$. The bracket term of the integrand is then the fraction of firm x that stay in the region where productivity is lower than z. The total fraction at $t+\mathrm{d}t$ is the sum of all remainders from the least productive to those with productivity z. Rearranging terms and considering it at the limit with $\mathrm{d}t\to 0$, I obtain the Kolmogorov forward equation on the productivity distribution:

$$\frac{\partial F(z,t)}{\partial t} = -\int_0^z \eta(x,t)(1-F(z|z\geq x,t))\mathrm{d}F(x,t). \tag{3.5}$$

Indeed, the productivity distribution is stochastically increasing since firms only increase but never decrease their productivities. The change in the fraction equals to the rate of escaping from that productivity region.

Description of the initial productivity distribution is necessary to complement the law of motion. To gain tractability, I assume the initial productivity distribution is Pareto, in which the productivity of the least productive firm is normalized to unity.

Assumption 3.1. *The initial productivity distribution is Pareto, i.e., $F(z,0)=1-z^{-k_0}$ for $z\geq 1$ and $k_0>1$.*

Equilibrium Concept This section ends with a discussion on the equilibrium concept used for further analysis. I consider a perfect foresight competitive equilibrium.

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

(i) Given $\{w(t),r(t),F(z,t)\}$, $v(z,t)$ solves the HJB equation (3.1), and $\eta(z,t)$ is the associated policy function;

(ii) Given $\{w(t), r(t)\}$, $c(t)$ solves the households' problem (3.2);

(iii) Both good and labor markets clear, i.e., (3.3) and (3.4) are satisfied;

(iv) $F(z,t)$ solves the KFE (3.5) given $\eta(z,t)$ and satisfies the initial condition.

It is also useful to define the notion of balanced growth path. Comparing to the standard definition, which assumes constant consumption growth at all times, the following definition only requires that growth is asymptotically constant.

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

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

3.2 Equilibrium Characterization

In this section, I discuss important properties of the equilibrium and gives a complete analytical characterization of the transition dynamics. The following is a very useful lemma. Omitted proofs can be found in the appendix.

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

This result is straightforward from the HJB equation (3.1) that the gains from learning per search must be non-positve for all firms. Otherwise, that firm will demand infinite amount of labor. The equilibrium search strategy $\eta(z,t)$ ensures that the total gains from learning are zero. Then, I show that there exists an equilibrium in which the productivity distribution remains Pareto at all times but with varying shapes.

Proposition 1. *With assumption 3.1, there exists an equilibrium with the following properties.*

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

(ii) The equilibrium productivity distribution at time t is

$$F(z,t) = 1 - z^{-k(t)} \quad \textit{for} \quad z \geq 1, \tag{3.6}$$

in which $k(t)$ satisfies

$$\frac{\dot{k}(t)}{k(t)} = -\eta(t), \tag{3.7}$$

and $k(0) = k_0$.

This proposition gives a sharp characterization on an equilibrium path. It says that all firms search at the same intensity, and the productivity distribution retains its Pareto shape at all times. In addition, the growth rate of the shape parameter is determined by the average search intensity.

There are two major implications. First, Gibrat's law holds. Lemma 1 implies that the gains from each meeting are determined by firm's productivity growth there, so firms' learning decisions are based on productivity growth. Let $\tilde{x}(z,t)$ be the productivity growth of firm z in each meeting, i.e., $\tilde{x}(z,t) = \tilde{X}(z,t)/z$. Since the productivity distribution is Pareto, left truncation does not change the shape of the conditional distribution. It is straightforward that the random variable $\tilde{x}(z,t)$ is invariant of firm's productivity, and follows the same Pareto distribution as the productivity distribution, i.e., $\text{Prob}(\tilde{x}(z,t) \le x) = F(x,t)$. Firms have the same distribution of productivity growth in each meeting and then choose the same arrival rate of meetings. Consequently, they grow at the same rate. The expected output growth of each firm, or equivalently the expected productivity growth, is given by

$$\lambda(z,t) \equiv \frac{\mathbb{E}[\mathrm{d}z]}{z} = \eta\left(\mathbb{E}[\tilde{x}(z,t)] - 1\right) = -\frac{\dot{k}(t)}{k(t)\,(k(t)-1)}. \tag{3.8}$$

Note that the growth rate of the total output, $Y(t)$, is the average growth rate of firm's output weighted by their output shares. With a constant population, the output growth is equal to the output per capita growth. Note that the household income $y(t)$ is equal to the output per capita, i.e., $y(t) = Y(t)/L$. Then, $\dot{y}/y$ denotes the output per capita growth and equals to the average firm growth $\lambda(t) \equiv \lambda(z,t)$.

Second, the right tail of the firm size distribution becomes thicker as the economy grows. It is intuitive to see that the right tail thickens over time. We learn from the KFE (3.5) that the productivity distribution is stochastically increasing, so the right tail cannot become thinner. The reason why it becomes strictly thicker is that even the largest firms grow, or "jump" in the model context, at a positive rate. If these firms remain inactive, the rightest part of the distribution will be unaltered in that firms on the left side can hardly reach there. I leave a more in-depth discussion on the connection between Gibrat's law and thicker tails to section 5. Since the productivity distribution is always Pareto, the thickness of its right tail is exactly the Pareto shape parameter $k(t)$. Last paragraph shows that the economic growth is equal to $\lambda(t)$. Then, equation (3.8) displays a transparent negative relationship between the output per capita growth, $\dot{y}/y$, and the growth of right tail thickness, $\dot{k}/k$, provided that $k(t) > 1$ at all times. This condition will be verified instantly.

With proposition 1, the aggregate state variable is reduced from the entire distribution $F(\cdot,t)$ to a single parameter $k(t)$. It is sufficient to characterize the equilibrium path by the trajectory of the Pareto shape parameter. The labor market clearing condition (3.4) implies that

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

which is obtained using both properties in proposition 1. Plugging it into the law of motion on

$k(t)$, the following differential equation describes the dynamics of $k(t)$:

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

This is a first order linear ordinary differential equation. With initial condition $k(0) = k_0$, it admits a simple solution,

$$k(t) = 1 + (k_0 - 1)e^{-Lt}. \tag{3.9}$$

Therefore, $k(t) > 1$ for all t, and $k(t)$ strictly decreases to one. Zipf's distribution emerges as the limiting distribution of productivity. Using equation (3.8), the output per capita growth satisfies

$$g(t) \equiv \frac{\dot{y}(t)}{y(t)} = \frac{L}{k(t)}, \tag{3.10}$$

which converges to L as $k(t) \to 1$. In this way, I show the equilibrium path is an asymptotic balanced growth path.

In this model, it is the increasing share of high productivity firms that improves the average productivity and generates output growth. Like all endogenous growth models, there is a source of increasing return to scale that sustains the long run growth. Considering again equation (3.8), doubling the search intensity doubles the growth rate if the distribution is held constant. That a thicker right tail increases firms' growth per search is analogous to that in the Romer model, an increase in the number of varieties increases output per capita. Yet the difference is that the increasing return to scale in this model is on the *growth rate* rather than the *level*. Therefore, a constant growth rate is achieved with increasing growth rates per search and decling search intensities. The latter is the consequence of a constant population: the number of searches goes down when it takes more labor to complete a search for an average firm in the economy. Figure 4 illustrates the evolution of the firm size distribution, $F(\cdot, t)$, and the average search intensity, $\eta(t)$.

The model also offers an explanation on Zipf's law. The thickening tail, which captures the aggregate effect of firms' learning, is the only source of growth in the model economy. If the productivity distribution converges to one with a tail thinner than the Zipf's distribution's, it has a finite mean. Then, there exists an upper bound on the level of output per capita, and eventually there will be no output growth. That being said, that we live in a world with balanced growth and Zipf's law in turn suggests that firms' learning is an important driver of the growth. It pushes continously the firm size distribution to Zipf's distribution. Section 5 continues this discussion and extends the intuition to more general cases.

I summarize the above equilibrium analysis in the following proposition. The parametric condition ensures that the representative households' utility is finite and relevant transversality condition is satisfied.

Proposition 2. *With assumption 3.1 and $\rho > L(1-\theta)$, there exists an equilibrium that satisfies the following:*

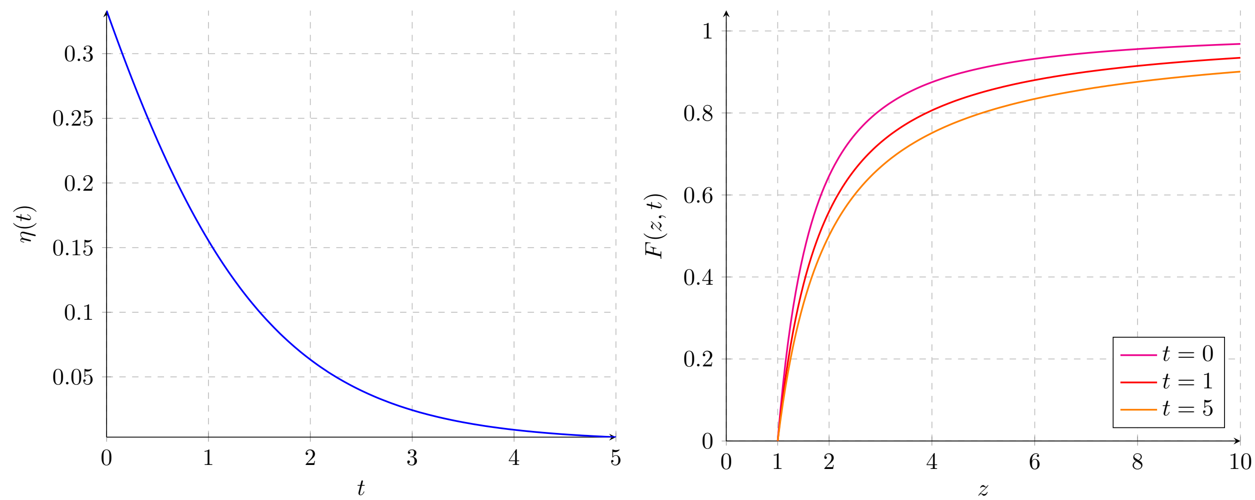


Figure 4: An illustration of $\eta(t)$ and $F(z,t)$

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

(i) *The equilibrium is an asymptotic BGP with limit output per capita growth L;*

(ii) *Firms always grow at the same rate, so Gibrat's law holds;*

(iii) *The productivity distribution has a thicker tail over time and converges to a Zipf's distribution in the limit.*

3.3 Prices and Comparative Statics

I complete the equilibrium analysis starting from the last section with a discussion on the price dynamics and comparative statics. Two prices, the wage of researcher and the interest rate, clear the goods and labor markets. It is a standard result that the Euler equation on household's consumption prices the interest rate, i.e., $r(t) = \theta\dot{c}(t)/c(t) + \rho$. Given that goods market clears and output per capita growth satisfies equation (3.10), the equilibrium interest rate can be written as a function of the state variable $k(t)$, namely,

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

Then, $r(t)$ increases to $r^* = \theta L + \rho$ as the growth rate increases to L.

On the other hand, the marginal cost of firms' learning must be equal to its marginal return. Therefore, the expected gains from learning normalized by the productivity determine researcher's wage:

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

in which lemma 1 and proposition 1 imply the second equality. Note that the unit value of productivity, $v(t)$, is the present value of a dividend flow of unit output, i.e.,

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

This is intuitive from the pricing of researcher's wage that the total return of frms' learning is zero. The value of a machine with unit productivity is then simply the discounted sum of its output, or the inverse of average interest rate $\tilde{r}(t)$.[9] As $r(t)$ converges to r^*, the value of unit productivity becomes $1/r^*$ in the limit. Then, wage grows unboundedly in the equilibrium, which is expected in endogeneous growth models. Following the standard treatment in these models, I consider instead the normalized wage by households' income and denote it by $\tilde{w}$. Thus,

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

which is the ratio between the growth rate and average interest rate. The following lemma characterizes the dynamics of the normalized wage.

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

The long-run normalized wage is the ratio of the long-run growth rate (L) and interest rate (r^*). The monotonicity, however, is less evident since both the growth rate and average interest rate are increasing. I show above that the instantaneous interest rate $r(t)$ is the sum of the discount factor and a linear term in output growth. Then, it grows less than the growth rate, i.e., $g(t)/r(t)$ increases over time. Lemma 2 shows that the same trend holds with average interest rate $\tilde{r}(t)$. While the value of unit productivity drops with larger discounting, the size of productivity improvement per meeting increases with a thicker tail, even in relative terms to the average productivity. This result suggests that the latter dominates in the aggregate, and learning becomes more valuable over time. Additionally, the normalized wage is equivalent to the share of research expenditure to GDP. The model can also predict that research intensity grows with economic development. Having said that, that both prices converge to some constant in the equilibrium further validates the convergence to a balanced growth path.

The above anatomy of the model makes the comparative statics fairly straightforward. Evidently, the qualitative properties of the model do not depend on the configurations of primitives. There are four model parameters: $\{k_0, L, \rho, \theta\}$. As shown in the last section, the shape parameter of the initial productivity distribution k_0 and the population size L determine the equilibrium productivity distributions and hence the equilibrium allocation. Whereas the two preference parameters, discount factor ρ and risk aversion θ, only affect the prices. Intuitively, a higher interest rate is required if households are less patient (a larger ρ), or the intertemporal subsitution is less

[9] See the proof of lemma 1 for the detailed derivation. An average interest rate is the equivalent interest rate with which the output flow has the same present value, i.e., $v(t) = 1/\tilde{r}(t)$.

elastic (a larger θ). The normalized wage goes down in the mean time for the average interest rate is higher but the growth rate remains unaffected.

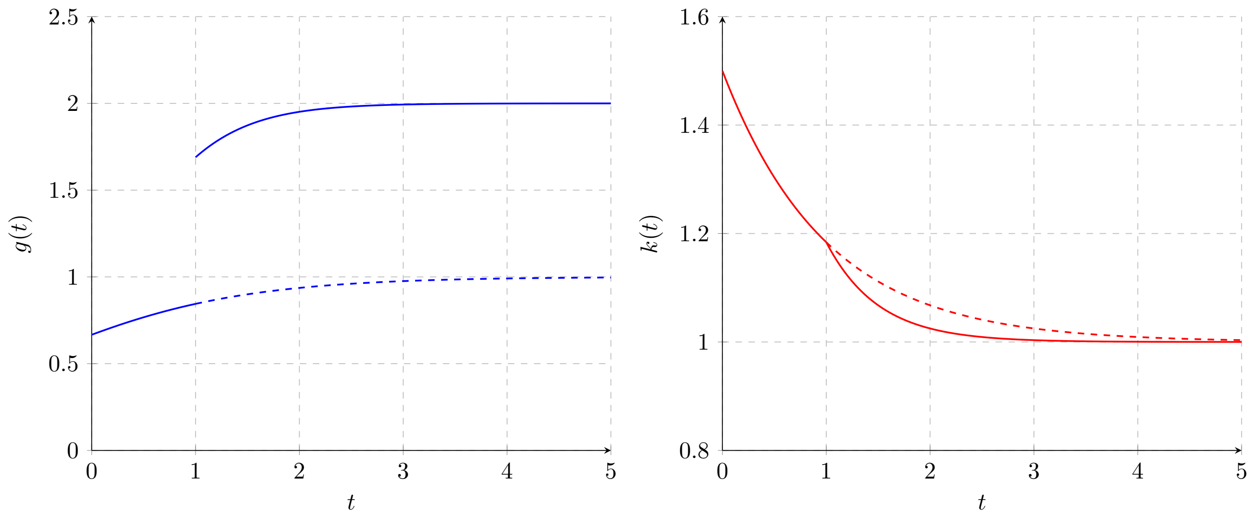


Figure 5: An illustration of population expansion on $g(t)$ and $k(t)$

Notes. This figure illustrates the dynamics in the output per capita growth and the shape of productivity distribution before and after an expansion in population. $k_0 = 1.5$. Population L doubles from 1 to 2 at $t = 1$. The left subfigure plots the growth rate, $g(t)$, over time. The right subfigure depicts the evolution of the shape parameter $k(t)$. The dashed lines are trajectories without the population expansion. At $t = 1$, there is a jump in the growth rate and a kink in the tracjectory of the shape.

Changes in the initial shape do not shift the equilibrium path but rather move along the equilibrium path. Economies with different shape parameters are just the past or the future of the others. It accurately delivers the model interpretation of the cross-country relationship between the right tail thickness and economic development: developing countries have thinner right tails because they are at early stages of development. Contrastingly, changes in population size shift the equilibrium path. Equation (3.10) makes it clear that there is a scale effect on the growth rate of output per capita. This is reminiscent of the endogenous growth tradition that R&D employment governs the productivity growth. (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992). In these models, knowledge spillover is so strong that the innovation step is independent of the level of productivity. Directed search provides a micro-foundation on such knowledge spillover: that firms target on better firms implies that learning efficiency does not have to decline with productivity. Therefore, population expansion increases the number of searches (the arrival rate) of each firm and consequently the overall growth rate. Figure 5 illustrates the impact of a population expansion on the output per capita growth and the shape of productivity distribution. When population is doubled, the growth rate jumps to the doubled level and converges at a faster steep. A kink emerges on the trajectory of $k(t)$ since $k(t)$ decreases at a faster speed to one with larger population. On the price side, higher growth rate raises the interest rate. The normalized wage also increases as the direct effect of population expansion on the growth rate outweighs its indirect effect on the interest rate.[10]

To conclude, table 1 below summarizes the discussion on the comparative statics.

	ρ	θ	L	k_0
r	+	+	+	0
$\tilde{w}$	−	−	+	0
g	0	0	+	0

Table 1: Comparative Statics

Notes: This table summarizes the comparative statics on equilibrium interest rate, normalized wage and growth rate. Let x be the row variable and y the column parameter. Each entry is the sign of partial derivative of the row variable with respect to the column parameter holding state variable k, i.e, $\text{sign}(x_y(k; y))$.

3.4 Discussion

In this part, I discuss the empirical relevance of the assumptions on the learning function. On the other hand, proposition 1 only states the existence of equilibrium. Therefore, I also discuss issues with multiple equilibria.

Directed Search The key deviation of this model from standard models in the idea diffusion literature is the assumption that firms only search among more productive peers. In contrast, search for ideas in those models are completely random. The probability of meeting a firm with certain productivity is the same irrespective of the searching firms' identities. More productive firms then have little incentives to search since there is larger probability to meet firms with productivity lower than them. They are less likely to search, and growth driven by idea diffusion declines with firm productivity. Therefore, Gibrat's law does not hold in these models. In this model, the probability of meeting another firm depends on the productivity of the searcher and is determined by the relative productivity. All firms face the same source distributions in terms of relative productivity and have same incentives to search. Gibrat's law then holds.

At the macro level, there are empirical evidence that productivity gap determines the rate of technology adoption. In a very recent World Bank report, Cirera et al. (2022) study technology adoption by firms in developing countries 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 cross-firm within-industry heterogeneity in technology adoption. One finding is that leapfrogging a technology in a business function is rare. That is, low productivity firms upgrate their technology gradually instead of jumping into the state-of-the-art. They present the estimated probability of firms using digital and frontier technologies, which include both general-purpose technologies and sector-specific business function technologies. Among all these technologies, the probability of adoption increases

[10] See the proof of lemma 2 for the details.

with firms' size. Besides, they also find that larger firms use more sophistcated technology (figure 2.6). Hence, firms upgrade their technologies at relatively similar paces. Similar observations have been documented on international knowledge diffusion. Among country-industry pairs, Van Patten (2020) finds that trading with technologically more sophiscated country-industry does not lead to more productivity growth. Rather, sectoral productivity growth decreases on the gap between its productivity and that of its trading partners. She then suggests that technology gap reduces international knowledge spillover by lowering the likelihood of technology adoption.

An alternative interpretation of this assumption at the micro level is that firms are more likely to learn from peers which are similar to them. It is not necessary that small firms benefit more from the knowledge of their peers. Three plausible mechanisms help to understand this interpretation. First, selection makes firms of similar size or productivity more likely to interact with each other. Managers of firms similar in size are likely to be invited to the same conference or social event at their level and have more meeting opportunities. It would be far less common for the owner of a local coffee shop to attend award ceremonies organized by the Fortune magazine. Sorting is another plausible reason of the homophily among firms. Using a comprehensive dataset on Japanese firms, Kodama and Li (2018) find that managers of larger firms tend to be more educated and from a prefecture different from the firms' locations. It is also well known in the urban literature that there is a positive relationship between firm productivity and city size. (Combes et al., 2012; Gaubert, 2018). As large firms are geographically more concentrated (in large cities), knowledge diffusion are stronger among them.

Second, larger firms are more likely to do R&D cooperation, which promotes knowledge diffusion. One view is the absorptive capacity argument that it takes internal knowledge base to absorb external knowledge. Larger firms are more equipped with R&D experience and can utilize outside knowledge. A strand of literature in technology management investigates the effect of firm size on R&D cooperation. Despite the lack of consensus, a large number of studies find that large firms cooperate more often in many countries, e.g., Cassiman and Veugelers (2002) in Belgium, Becker and Dietz (2004) in Germany, Negassi (2004) in France, and Badillo et al. (2017) in Spain.

Third, it might be more difficult for small firms to exchange useful information among themselves due to search frictions. Evidence from randomized control trials suggest that lack of interfirm information exchange is an important barrier to the growth of SMEs in developing countries. Cai and Szeidl (2018) Cai and Szeidl (2018) organize meetings between owner-managers of young Chinese firms and find that regular meetings increase firms' performance compared to the control group. They argue that managers did not organize meeting for themselves may be due to search cost and trust barriers. Similar randomized experiment also show that group-based consulting is very effective in elevating the performance of SMEs partly because it facilitates the spreading of localized and specific knowledge. (Iacovone et al., 2022; Brooks et al., 2018).

Multiplicity of Equilibrium Proposition 1 is a statement on the existence of an equilibrium. With a linear cost function $z\eta$, Lemma 1 and the initial Pareto productivity distribution imply that the initial researcher wage has to be equal to the normalized expected gains from learning. Therefore, firms are indifferent between search intensities of all levels at time 0. Any initial assignment of labor that respects the labor market clearing condition will lead to an equilbrium in which productivity distribution evloves by the KFE (3.5). Hence, subsequent productivity distribution is not necessarily Pareto, and Gibart's law does not have to hold. It then raises concerns on the relevance of the above equilibrium analysis.

I argue that this type of multiplicity is an artifact of the linear cost assumption rather than a fundamental caveat of the learning protocol. Perhaps the most transperant way to see it is by constructing a similar equilibrium with a non-linear cost function. Suppose that it now takes an additional adjustment cost $g(\eta)$ to complete search and adoption tasks with arrival rate η. In other words, a firm z will have to hire $z(\eta + g(\eta))$ researchers to achieve an arrival rate of η. Moreover, g is a strictly increasing and strictly convex function such that $g(0) = g'(0) = 0$. Note that lemma 1 does not necessarily hold with adjustment cost: a linear value function is not a feature of all equilibria. In constrast, I show that there exists an equilibrium that still satisfies properties in lemma 1 and proposition 1.

Proposition 1'. *With assumption 3.1 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 $k(0) = k_0$.

It is quite straightforward to verify that proposition 2 holds with adjustment cost. First, Gibrat's law holds immediately in this equilibrium using the same arguments. With a linear value function, the first order condition on η implies a unique optimal search intensity, i.e.,

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

Comparing to the baseline, the adjustment cost creates a wedge between the normalized expected gains from learning and wage, which pins down the search intensity and resolves indeterminacy.

Next, it is not hard to see the 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,

$$k(t)=1+(k_0-1)\exp\left(-\int_0^t \tilde{L}(s)\mathrm{d}s\right), \tag{3.11}$$

in which $\tilde{L}(t)=\frac{L}{1+g(\eta(t))/\eta(t)}$. This is similar to the baseline trajectory on $k(t)$ except that $\tilde{L}(t)$ now deviates from the constant labor per firm L. Since k is decreasing, the monotonicity of g implies that η also decreases over time. In addition, $g(\eta)/\eta$ increases on η due to the convexity of g. Then, $\tilde{L}(t)$ 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. Technically, one can still solve for $k(t)$ by substituting η as a function of k into the law of motion. Whereas, there is in general no explicit solution in the presence of the adjustment cost. That is one reason why linear cost is favored.

Lastly, the output growth still converges to a constant given by the labor per firm. Recall that

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

where I obtain the second equality using equation (3.11). 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 $g'(0)=0$. Then, $\tilde{L}(t)$ converges to L. Going back to an earlier point, the wedge between normalized expected gains from learning and wage vanishes as η goes to zero. All the prices, namely, wage and interest rate, converge to those in the baseline equilibrium. Hence, the baseline equilibrium is preferred to other equilibria in the benchmark setting in two ways: (a) it is indistinguishable from any equilibrium with the above adjustment cost in the long run, and (b) it can be viewed as the equilibrium in an limiting economy in which the adjustment cost approaches to zero.[11]

4 Quantitative Exercises

The model in section 3 is made intentionally parsimonious so as to highlight the growth mechanism by varying tail. For this purpose, other sources of growth such as capital accumulation, selection or R&D are muted. In addition, a common feature of existing idea diffusion models is that growth is generated by scaling the firm size distributon. From the perspective of growth implications,

[11] A caveat is that even with adjustment cost, the equilibrium in proposition 1' is not guaranteed to be unique. This is because my proof only shows the uniqueness of equilibrium with a linear value function. There might be equilibria with non-linear value functions and, accordingly, non-Pareto productivity distributions. Proving the more general uniqueness of the equilibrium 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 a coupled PDE systems (the HJB and the KFE) in idea diffusion models.

scaling is isomorphic to a non-rival TFP shock that augments the productivity of all firms to the same extent. It is apparent that scaling is absent from the model since all firms have a positive probability to retain their initial productivities forever. That is also why in the model, the output $Y(t)$ is given by $k(t)/(k(t)-1)$ instead of $A(t)k(t)/(k(t)-1)$, in which $A(t)$ represents potential scaling factors. Therefore, output growth comes exclusively from changes in the shape of firm size distribution. I refer to this growth mechanism as "tail growth". To further our understanding on the novel tail growth, it seems natural to investigate its relevance to the real world. In other words, how far is our world away from a world driven solely by tail growth? The following two quantitative exercises explore this question by testing the model implications.

4.1 A First Look

The model makes predictions that there is a tight connection between output per capita and thickness of the right tail. In particular,

$$y(t) = \frac{1}{L}\frac{k(t)}{k(t)-1}.$$

Given a constant population, one can obtain the model implied output per capita growth once there are estimates for the right tail thickness. That the equilibrium size distribution in the model is Pareto makes this task considerably simpler.[12] Recalling from section 2, the thickness measure based on the number of firms equals to $-k$ with a Pareto firm size distribution. It is then the emprical counterpart of the model moment $k(t)$.

Figure 6 compares the model implied US GDP per capita growth with the real data from 1978 to 2019. Both series are indexed relative to the 1978 year level, which is indexed by 100. For estimates of the US right tail thickness, I rely on the series of thickness measures in figure 3 with $T_S = 20$ and $T_L = 500$. The model implied $\hat{y}$ is given by $\hat{k}/(\hat{k}-1)$, in which $\hat{k}$ is the additive inverse of the thickness measure. Data on the US GDP per capita growth, on the other hand, are taken directly in index scale from the Federal Reserve Economic Data (FRED). Although more volatile, the model implied GDP per capita growth tracks the real data very closely. With all the simplifying assumptions, the model predictions are so strong that calibration is unnecessary. Hence, the alignment between the two series is unexpected a priori and is indicative of the importance of the tail growth.

4.2 A Simple Calibration

The last exercise tests directly on the relationship between the thickeness of the right tail and GDP per capita using externally estimated shape parameters. However, the model itself also has predictions on changes in the thickness of the right tail, which could be tested as well. In other words, the model predicts simultaneously on both changes in the thickness of the right tail and

[12] Note that in the model, the employment size of each firm is its number of researchers. Since firms search at the same intensity, employment sizes are proportional to productivities.

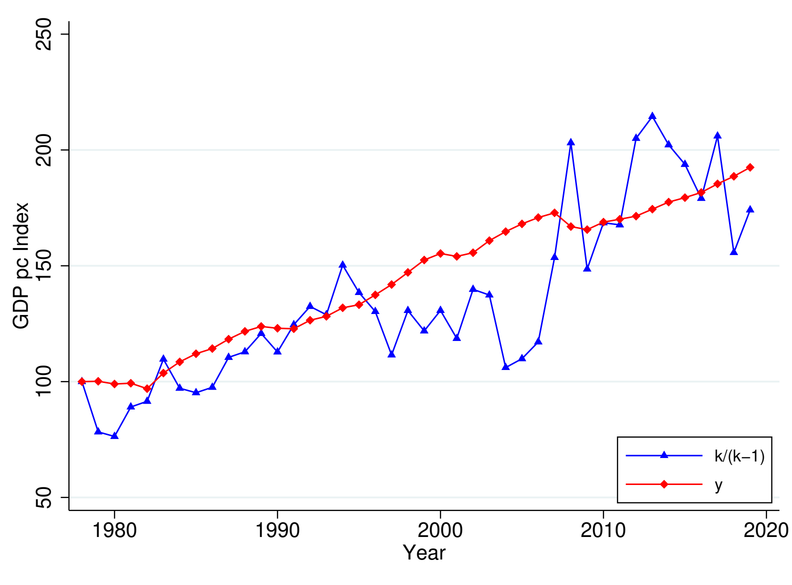


Figure 6: Model vs. Data: US GDP per capita growth

Notes. This figure compares the 1978-2019 US GDP per capita growth predicted by the model with that in the data. The model implied GDP per capita index is obtained using $\hat{y} = \hat{k}/(\hat{k}-1)$, in which $\hat{k}$ is minus the number-of-firms-based thickness measure with $T_S = 20$ and $T_L = 500$. The data on the GDP per capita index is from FRED.

output per capita holding. One can then calibrate the model using one moment and test its performance on the other moment.

In this exercise, I target the right tail thickness of US in 1978-2019 to calibrate the model and test the model with the US GDP per capita growth of that period. The equilibrium analysis in section 3.2 shows that the initial shape parameter k_0 and the population size L completely determine the equilibrium allocation. It then suffices to calibrate these two parameters. From equation (3.9), the equilibrium shape parameter satisfies that

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

As before, I obtain the dependent variable, $\ln(k_t - 1)$, using the number-of-firms-based thickness measure with $T_S = 20$ and $T_L = 500$. Regressing it on time, the slope and constant coefficients respectively identify k_0 and L. 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 (3.9), the model generates a sequence of $\hat{k}_t$ for each year t. Notably, the model predicted shape paramter in 1997 is 1.050, which is very close to the classic 1.059 obtained in Axtell (2001). The following figure 7 shows that the model fits targeted right tail thickness very well.

Using the sequence of $\hat{k}_t$ and $\hat{L}$, I obtain a sequence of model predicted GDP per capita $\hat{y}_t$. The model performs very well in fitting the untargeted US GDP per capita growth. First, the model

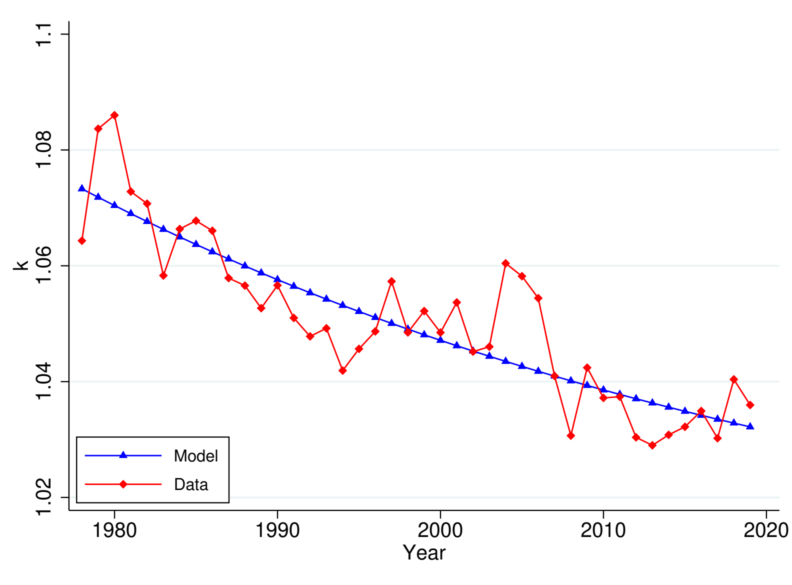


Figure 7: Targeted: The right tail shape k

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 (3.9) 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 $T_L = 500$.

implies that L is the long-run output per capita growth rate, as shown in equation (3.10). The estimated $\hat{L}$ perfectly hits the well-known 2% growth rate. Next, figure 8 compares the time series on GDP per capita growth generated by the model with that in the data. The left figure plots the indexed GDP per capita, while the right shows the annual growth rate. Both figures show close alignment between the data and model predictions. 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.

In sum, both exercises suggest that a model with only tail growth can capture the growth in the real world surprisingly well. Given the extremely simplified model structure, these quantitative results are best viewed as supporting evidence that tail growth is a significant source of growth, among others. It needs more sophisticated work to quantify the role of tail growth together with other growth mechanisms.

5 General Results

The simple idea diffusion model generates economic growth from firm growth with marked consistency to empirical regularities. At the firm level, Gibrat's law holds at all times on the equilibrium, and Zipf's law emerges as the limiting firm size distribution. The economy, at the same time, grows

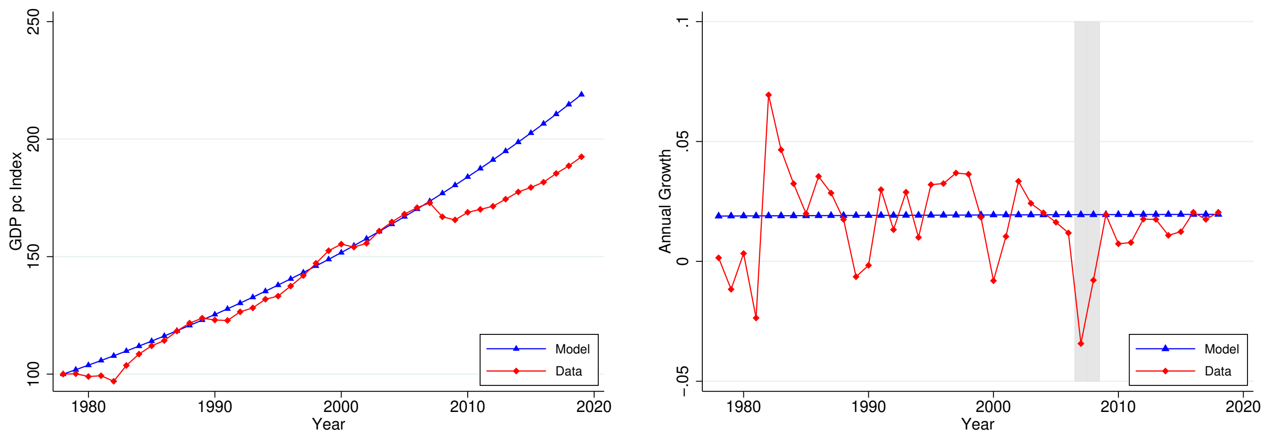


Figure 8: Untargeted: GDP per capita growth

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.

asymptotically at a constant rate with a thickening right tail. It is only natural to ask whether the joint appearance of these four stylized facts–two micro facts and two macro facts–is merely an artifact of the model or hints deeper connections between firm and economic growth. The answer is likely to be the latter since firms are the basic units of the economy. I show that idea diffusion is a good mechanism to think about the linkage between firm and economic growth. Extra model assumptions or other details might muddle the effect of idea diffusion and have to be removed. Therefore, I propose the following general idea diffusion mechanism to focus on the bare bones of idea diffusion models. This section describes the general idea diffusion mechanism, establishes basic properties and then discusses the close relationship between Gibrat's law and thicker tails and between Zipf's law and long-run growth.

5.1 A General Idea Diffusion Mechanism

There are a continuum of firms in the economy with heterogeneous productivities. Productivity is positive. The distribution of firm productivity at time t is captured by a cumulative distribution function (CDF) $F(\cdot, t)$ and has well-defined density $f(\cdot, t)$. Firms can upgrade their technologies by adopting new ideas, whose arrival is a Possion event. For a firm with productivity x, the arrival rate of ideas with productivity z is $n(z, x, t)$. Then, $m(z, x, t) \equiv \int_z^\infty n(y, x, t)\mathrm{d}y$ gives the arrival rate of ideas with productivity at least z. If an idea has a productivity higher than its current level, the firm adopts the idea and increase its productivity to that level. Otherwise it retains the current productivity. Namely, the new firm productivity $x' = \max\{x, z\}$, where z is idea's productivity. There are no aggregate productivity shocks, so growth in the aggregate productivity is a weighted average of all firms' productivity growth. In the rest of section 5, I focus on the case in which

$m(z,x,t)$ is separable on $z \geq x$, i.e.,[13]

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

That is, the arrival rate function is made of two components: a firm-specific $\mu(x,t)$ and an idea-specific $\tilde{H}(z,t)$. Intuitively, $\mu(x,t)$ can be viewed as the search intensity of an individual firm. Note that the definition of $m(z,x,t)$ 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 is then straightforward that $\tilde{H}(z,t)$ also decreases on z, and $\lim_{z\to\infty} \tilde{H}(z,t) = 0$. Whenever $\tilde{H}$ is bounded, it will be a valid complement CDF with normalization.[14] It is also useful to define $h = -\partial H(z,t)/\partial z$, 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 the density.

The productivity distribution then evolves as follows.

$$\begin{aligned} \tilde{F}(z,t+\mathrm{d}t) &= \tilde{F}(z,t) + \int_0^z m(z,x,t)\mathrm{d}t\mathrm{d}F(x,t), \\ \Rightarrow \quad \frac{\partial \tilde{F}(z,t)}{\partial t} &= \tilde{H}(z,t)\int_0^z \mu(x,t)f(x,t)\mathrm{d}x, \end{aligned} \tag{5.1}$$

The tilde notation denotes the complement CDF, for instance, $\tilde{F}(z,t) = \int_z^\infty f(y,t)\mathrm{d}y$. The first equation is intuitive: the fraction of firms with productivity at least z at $t+\mathrm{d}t$ is given by the fraction of firms with productivity at least z at t and the fraction of firms with productivity below z at t but adopt ideas with productivity above z between t and $t+\mathrm{d}t$. It implies the second equation with $\mathrm{d}t \to 0$ since integrand satisfies $z \geq x$. With an initial distribution $F_0(z)$, solution to the partial differential equation (5.1) characterizes the entire transition of the productivity distribution.

On the other hand, the expected productivity at $t + \mathrm{d}t$ satisfies that

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

The first term captues 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

$$\lambda(x,t) \equiv \lim_{h\to 0} \frac{\mathbb{E}[x(t+h)] - x}{xh} = \frac{\mu(x,t)}{x}\int_x^\infty (y-x)h(y,t)\mathrm{d}y. \tag{5.2}$$

The equation holds since $n(z,x,t) = \mu(x,t)h(z,t)$ for $z \geq x$. The task of this section is to explore the relationship between firm growth $\lambda(z,t)$ and the evolution of firm size distribution $F(z,t)$.

[13] In the appendix C.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).

[14] 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 $z \geq x$.

Discussion The arrival rate function $m(z,x,t)$ summarizes the overall learning efficiency of a firm x. More broadly, the idea productivity z is the final productivity, probably at the end of a lengthy learning process, which the firm can attain if it choose to adopt the idea. Accordingly, $\max\{x,z\}$ captures the firm's final decision on whether to continue with the existing technology or replace it with the outcome technology of learning. To see this, the following considers a detailed learning process, 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. In sum, a firm x has probability $p(x,z)$ to absorb an idea of productivity z. Conditional on successful absorptions, 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:

$$m(z,x,t) = \mu(x,t)\int_{\Omega(z,x)} p(x,y)h(y,x,t)\mathrm{d}y, \tag{5.3}$$

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.

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 2: 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 (5.3). $m(z,x,t)$ without labeling $z \geq x$ is sparable all over its domain.

Furthermore, it is with little loss of generality to assume separability in that separable arrival rate functions also cover a plethora of heterogeneities. 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 2 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 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. 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). 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 productivities. Both the thrid and fourth rows capture the heterogenity 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.

5.2 Useful Preliminaries

This part introduces technical definitions, assumptions and lemmas that are essential to the general results. First, I introduce the notion of regular variations, which can be understood as a generalization of power law. Regularly varying functions are first studied by Karamata (1930). I follow closely Bingham et al. (1987), which is a modern encyclopedia on regular variations.

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 slow 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.

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 $-1-k$.

Next, I present regularity conditions under which the dynamics in the tail index is characterized by the following lemma.

Assumption 5.1. *A distribution $W(z,t)$ satisfies the following regularity 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 $z \to \infty$.

Lemma 3. *Consider $F(z,t)$ that solves the initial value problem of (5.1) and satisfies assumption 5.1. Then, $k(t)$ is decreasing. Whenever it is finite, $k(t)$ is differentiable and has the following finite derivative:*

$$\dot{k}(t) = -\lim_{z\to\infty}\frac{1}{\ln z}\frac{\tilde{H}(z,t)}{\tilde{F}(z,t)}\int_0^z \mu(x,t)f(x,t)\mathrm{d}x > -\infty. \tag{5.4}$$

I center on equilibrium in which the productivity distributions satisfy the smoothness conditions in assumption 5.1. They are sufficiently smooth such that the tail dynamics in those equilibria has no jumps or kinks. Thin-tailed productivity distributions are excluded from the analysis in that they are less nontrivial. A thin-tailed distribution has tail index $k = \infty$. Then, k must be decreasing since it either stays thin-tailed or jumps into a fat-tailed distribution with $k < \infty$. Besides, a thin-tailed productivity distribution will immediately become thick-tailed if the source distribution is thick-tailed. This is straightforward using the law of motion in (5.1). For an infinitesimal time interval $[t, t+\delta]$,

$$\tilde{F}(z,t+h) = \tilde{F}(z,t) + \delta\tilde{H}(z,t)\int_0^z \mu(x,t)f(x,t)\mathrm{d}x.$$

If $\tilde{H}(z,t)$ has tail index $h < \infty$, then $\lim_{z\to\infty} z^k\tilde{F}(z,t+h) = \infty$ for any $k > h$. The productivity distribution is then thick-tailed with a right tail no thinner than the source distribution’s.

I consider search intensity and source distribution that are regularly varying. Note that $\tilde{H}(x,t)$ is bounded on $[x_0, \infty)$ for any $x_0 > 0$. As only the right tail is of consideration, it is without loss to treat $\tilde{H}(z,t)$ as a distribution function and define a tail index on it. The following formally states the assumptions.

Assumption 5.2. *$\mu(x,t)$ and $H(z,t)$ satisfy the following conditions:*

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

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

Lemma 4. *Suppose condition (ii) of assumption 5.2 holds. Then, the productivity growth $\lambda(x,t)$ has the following asymptotic equivalence as $x \to \infty$:*

$$\lambda(x,t) \sim \frac{1}{h-1}\mu(x,t)\tilde{H}(x,t) \tag{5.5}$$

With regular varying functions, lemma 4 shows that for large firms, the expected productivity growth is proportional to the arrival rate of having ideas better than their own productivities, i.e., $\lambda(x,t) \propto m(x,x,t)$. Then, it becomes clear on the relationship between firm productivity growth and changes in the tails, since both are determined by the arrival rate function. The following proposition explores this linkage.

Proposition 3. *Consider $F(z,t)$ that solves the initial value problem in (5.1) and satisfies assumption 5.1. Suppose assumption 5.2 holds as well. Then, the arrival rate 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) $m(t) = h(t) = k(t), \quad \lambda(z,t) = o(\ln z)$.

Smooth changes in the tail indices are very informative about the arrival of ideas and the heterogeneity of productivity growth. Proposition 3 points out that there are only three types of equilibrium consistent with a varying tail. Among all three types, the source distribution has a right tail no thicker than the productivity distribution's, i.e., $h(t) \geq k(t)$. Otherwise, the productivity distribution will instantly have a right tail no thinner than the source distribution's, as shown before. Besides, $\ln z$ in the differential equation (5.4) puts an upper bound on the speed of the productivity growth over firms' productivity. $\lambda(z,t)$ can grow at the same or a lesser rate as $\ln z$, which includes zero growth for large firms.

Each type of equilibrium corresponds to a distinct interpretation. The first type implies that firms search over a source distribution with a tail thinner than the productivity distribution's. Representative examples of this type are search-theoretic R&D models such as Kortum (1997), in which research ideas flow from a fixed and exogenous pool of undiscovered ideas. The source distribution stays constant while the productivity distribution changes over time. In order to have a thickening right tail, large firms search sufficiently more times to compensate the decline in learning efficiency in each search. Furthermore, they grow unboundedly, which is clearly counterfactual. In the second type, the source distribution also varies over time as the productivity distribution varies. This type of equilibrium reminisces traditional diffusion models in which diffusion is exogenous and undirected. For example, Eaton and Kortum (1999) study technology diffusion across countries. In such models, changes in the productivity distribution (of the domestic country) are driven solely by changes in the source distribution, which might be the productivity distribution of a technology-leader country. However, if the thickening of the right tail of the domestic country is due to that of the technology-leader country, what is an internally consistent explanation for the thickening of the right tail there? These models have limited abilities to offer self-contained solutions since they rely heavily on exogenous variations to vary the productivity distribution. Besides, inconsistent with Gibrat's law, large firms do not grow.

The simple model in section 3 falls into the third type of equilibrium, which is the only type where Gibrat's law can hold. This proposition makes it clear that Gibrat's law and thicker tails, both of which are observable from the data, are useful moments to identify the unobserved learning function. In particular, search intensity, source distribution and productivity distribution all grow or diminish at the same rate on productivity. The arrival rate function in the simple model, $m(z,x,t) = \eta_t x^{k_t} z^{-k_t}$, is exemplary of this relationship. To conclude, above discussions on the first two types of equilibrium clarifies that a thickening tail is unlikely the consequence of learning from external sources such as the pool of future ideas or frontier countries. More plausibly, it indicates learning from existing ideas within the economy, which are activities like imitation or technology adoption. Hence, the internal idea diffusion in the simple model is not an arbitrary modeling choice: thicker tails hint the importance of internal diffusion.

Discussion Similar analysis extends to idea diffusion 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 3, the change in tail index of $\hat{F}$ satisfies that

$$-\dot{k}(t) = \lim_{z\to\infty}\frac{1}{\ln z}\frac{\tilde{H}(z,t)}{\hat{F}(z,t)}\int_0^z \mu(x,t)\hat{f}(x,t)\mathrm{d}x + \lim_{z\to\infty}\frac{E(z,t)}{\hat{F}(z,t)\ln z} - \lim_{z\to\infty}\frac{\int_z^\infty \delta(x,t)\hat{f}(x,t)\mathrm{d}x}{\hat{F}(z,t)\ln z}. \quad (5.6)$$

With regularly varying $E(z,t)$ and $\delta(z,t)$, same arguments can immediately extend proposition 3 with entry and exit. Thus, I omit a complete analysis but point out 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 $\delta(z,t)$ decreases on z, Then, equation (5.6) still reduces to equation (3), and lemma 3 and proposition 3 apply without any modifications. For this reason, the general mechanism in this section excludes explicit entry and exit.

5.3 Understanding Gibrat's and Zipf's Law

I present the main results of section 5 that there are close relationships between Gibrat's law and thicker tails and betwen Zipf's law and long-run growth. Before that, proper definitions on Gibrat's law and Zipf's law seem useful. Let $g^r(t)$ be the growth rate of large firms, that is, $g^r(t) = \lim_{z\to\infty} \lambda(z,t)$. Later, I will show that $g^r(t)$ is also the rate of tail growth. Throughout this section, $g^r(t)$ is well defined and possibly infinite. The following presents a theoretical formulation of Gibrat's law and Zipf's law.

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 $\lim_{z\to\infty} \frac{\ln \tilde{F}(z)}{\ln z} = -1$.

In most firm dynamics models, productivity is isomorphic to firm size measures such as employment or sales. That is why I define both laws over productivity, which is used abusively as a proxy of firm size. These definitions are weaker but more accurate than the usual versions. Standard Gibrat's law states that firm growth is independent of size. Having said that, abundant evidence show that the departures from Gibrat's law are primarily for small and young firms. (Haltiwanger et al., 2013). This version of Gibrat's law then focuses on large firms and avoid these departures. Similarly, the standard Zipf's law is that firm size distributions in developed economies are well approximated by a power law with an exponent close to 1. (Gabaix, 2009). It remains debatable whether a log-normal or Pareto distribution fits better the US firm size distribution. (Kondo et al., 2021) Notwithstanding, a right tail with tail index 1 captures the essence of Zipf's law given that both log-normal and Pareto are thick-tailed distributions. I restrict to idea diffusion models with internal search for a sharper characterization.

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

The above definition of internal search generalizes the usual concept of learning from internal sources, which typically assumes that $H(z,t) = F(z,t)$. That is, firms make random draws among all firms producing in the economy. In contrast, the generalized internal search only requires the source distribution to have a similar right tail as the productivity distribution. It then covers nearly all existing idea diffusion models regardless of whether the source is internal. For example, the class of source distributions in Buera and Oberfield (2020) qualify for internal search. With internal search, the following result says that there is an equivalence between Gibrat's law and thicker tails.

Proposition 4. *Consider $F(z,t)$ that solves the initial value problem in (5.1) and satisfies the regularity conditions in assumption 5.1. Suppose assumption 5.2 also holds, and there is internal search. Then, Gibrat's law holds at time t if and only if $k(t)$ decreases strictly. In addition, the tail index evolves as follows:*

$$-\frac{\dot{k}(t)}{k(t)(k(t)-1)} = g^r(t). \tag{5.7}$$

The above results significantly refine the predictions of proposition 3. With internal search, the source distribution has the same tail index as the productivity distribution, i.e., $h(t) = k(t)$. Therefore, a direct application of proposition 3 only eliminates the first type of equilibrium under internal search. Proposition 4 improves this prediction by showing that the only possibility is a special case of the third type, in which Gibrat's law holds. It also shows that the growth rate of large firms completely pins down the law of motion on the tail index of the productivity distribution, so $g^r(t)$ is a legitimate measure of tail growth. Indeed, the evolution of the Pareto shape in the simple model is a perfect illustration of the general tail dynamics. Since tail growth is the only source of growth there, the tail growth rate $g^r(t)$ equals to the aggregate growth in output per capita, and equation (3.8) coincides with (5.7). That is to say, the tail dynamics in the simple model is in fact a general feature of idea diffusion models with internal search.

With a stationary firm size distribution, there must be no change in the right tail, or $\dot{k}(t) = 0$. Then, $g^r(t) = 0$, large firms do not grow, and Gibrat's law does not hold. This is why Gibrat's law does not hold on the balanced growth paths of existing idea diffusion models which assume stationary firm size distributions. Moreover, their learning functions determine that neither does Gibrat's law hold in the transition. Lemma 4 implies that $\lambda(z,t) \sim \mu(z,t)\tilde{H}(z,t) = m(z,z,t)$. To obtain Gibrat's law, the growth of $\mu(z,t)$ in productivity has to offset the decline in $\tilde{H}(z,t)$. In words, large firms has to be sufficiently more efficient in searching so as to compensate their low return in each search. The required advantage of large firms is stronger than what is assumed in typical existing models. To see this, consider again table 2 and a class of arrival rate functions in power form, i.e., $m(z,x,t) \propto x^{m_t} z^{-h_t}$. Gibrat's law requires that $m_t = h_t$, so it certainly fails in models of pure random search (the first row), namely with $\mu(x,t) = \mu_t$. With specifications in the third and fourth rows, large firms have advantages in learning since they can better absorb or adapt high quality ideas. However, these advantages are not strong enough for Gibrat's law to hold given that $m_t < h_t$ in those cases. At this stage, it becomes transparent that the only arrival rate function that takes a power form, respects Gibrat's law and generates a thickening right tail is the one in the simple model (the second row), i.e., $m(z,x,t) = \eta_t x^{k_t} z^{-k_t}$.[15] In other words, under functional form assumptions that the arrival rate function is in power form, the tail index of productivity distribution uniquely identifies the arrival rate function, up to a constant. In this sense, the simple model presents not just one possibility to obtain Gibrat's law and thicker tails but the only possibility.

The differential equation (5.7) is a logistic differential equation, so it admits a logistic function as the solution. With $k(0) > 1$,

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

Instantly, $\lim_{t\to\infty} k(t) = 1$ if and only if $\int_0^\infty g^r(\tau)\mathrm{d}\tau = \infty$. Putting them in words, Zipf's law holds

[15]Note that internal search is not necessary here. Proposition 3 and the functional form assumptions are sufficient to obtain this identification result.

in the limit if and only if there is unbounded cumulative tail growth. Intuitively, tail growth, or growth of large firms, is the only force to change the right tail. The result is more interpretable if tail growth is systematically related to the aggregate productivity growth. Let $g(t)$ be the productivity growth of the whole economy, then

$$g(t) = \frac{\int_0^\infty \lambda(z,t) z f(z,t) \mathrm{d}z}{\int_0^\infty z f(z,t) \mathrm{d}z},$$

which is the average of firm growth weighted by their productivity. There is an equivalence between unbounded tail and aggregate productivity growth if their ratio is uniformly bounded over time. That is, there exist $\underline{M}, \bar{M} > 0$ such that $g(t)/g^r(t) \in (\underline{M}, \bar{M})$ for all z and t. In this way, Zipf's law holds if and only if there is unbounded aggregate growth. The uniform boundness condition has a clear economic interpretation that tail growth plays a nontrivial role in the overall economic growth. In idea diffusion models with stationary firm size distributions, there is no tail growth, so this condition is obviously violated with $g^r(t) = 0$. Then, idea diffusion provides a simple explanation of the Zipf's law in the firm size distribution. Firm size distributions in highly developed economies resemble Zipf's distribution because a significant portion of their economic development is contributed by tail growth, or idea diffusion, which thickens the right tail considerably and brings it toward the limit level. Finally, I summarize the discussion on Zipf's law into the following proposition.

Proposition 5. *Consider $F(z,t)$ that solves the initial value problem in (5.1) and satisfies the regularity conditions in assumption 5.1. Suppose assumption 5.2 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.*

6 Policy Implications

This section returns to the simple model in section 3 to explore the normative implications of the tail growth mechanism. In the model, thicker tails increase the efficiency of each search, and searching for ideas in turn fuel the thickening of the right tail. The key part inside this feedback loop–the thickening right tail–works only if large firms are also searching and growing. Searches by large firms thicken the right tail and improve the learning efficiecny of all firms in the economy. In comparison, searches by small firms have little impact on large firms since the latter do not learn from the former. Hence, the market equilibrium in which all firms search at the same intesnity seems inefficient: small firms might search too much while large firms might search too little. The asymmetric externality of large and small firms' search marks room for policy interventions. Particularly, policies favoring large firms should improve social welfare since they take better advantages of the diffusion externality from large firms. The rest of this section presents two such policy exercises. The first considers a tax on small firms' search, and the second solves the planner's problem. Both exercises show that having large firms search more significantly improves social welfare.

6.1 A tax on the small firms

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

Proposition 6. *For any threshold productivity $z^* \geq 1$, there exists an equilibrium such that the equilibrium productivity distribution satisfies*

$$F(z,t) = \begin{cases} 1 - z^{-k_0} & \text{if } z \leq z^*, \\ 1 - (z^*)^{k(t)-k_0} z^{-k(t)} & \text{if } z > z^*, \end{cases} \tag{6.1}$$

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 firm gets to hire more scientists and increases the average arrival rate of ideas. As discussed in section 3, a higher labor endowment (scientists) per firm accelerates output growth. Taxing the small firms raises the effective labor endowment 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 policy maker can always choose sufficiently large threshold z^*. The social welfare will be infinite regardless of the discount factor, given sufficiently large long-run growth.

On the technical side, proposition 6 also illustrates the multiplicity of equilibrium due to the linear cost assumption. As shown in the proof, an equilibrium with distribution (6.1) is supported in the tax-free ($\tau = 0$) or baseline economy for any threshold. That is, there are 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 set of firms can possibility be inactive at time 0. But henceforth, the resulting piece-wise Pareto distribution makes the return of search strictly less for firms below the threshold, so they stay inactive forever. Among all threshold equilibria, I focus on the one with the lowest threshold for reasons discussed in section 3.4. 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 to search even at time 0. Then, minimum threshold binds at the chosen threshold, i.e., $\underline{z} = z^*$, and distribution (6.1) describes the resulting equilibrium.

6.2 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:

$$\begin{aligned} W(f) = & \max_{\{c(\omega,t),\eta(z,t)\}} \quad \int_0^\infty e^{-\rho t} \int_\Omega u(c(\omega,t)) \mathrm{d}\omega \mathrm{d}t \\ \text{s.t.} \quad & \int_\Omega c(\omega,t)\mathrm{d}\omega \leq \int_0^\infty z f(z,t)\mathrm{d}z, \\ & \int_0^\infty z\eta(z,t)f(z,t)\mathrm{d}z \leq L, \\ & \frac{\partial f(z,t)}{\partial t} = f(z,t)\left[\int_0^z \eta(x,t)\frac{f(x,t)}{1-F(x,t)}\mathrm{d}x - \eta(z,t)\right], \\ & f(z,0) = f(z). \end{aligned} \tag{6.2}$$

The first two constraints are the respective goods and labor market clearing conditions. The third equation is the law of motion on the denisty 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 (6.2) is challenging because the state variable is an infinite dimensional object–a distribution. I borrow techniques developed in Lucas and Moll (2014) and Nuño and Moll (2018) which study optimal policies in heterogeneous agent models. 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.[16] $w(f, z)$ is then the marginal social value of a firm z. Moreover, let $w(z, t) \equiv w(f(\cdot, t), z)$. That is, $w(z, t)$ is the marginal social value along the trajectory of the productivity distribution $f(z, t)$, which results from the optimal policies. I show in the appendix that $w(z, t)$ satisfies the following partial differential equation:

$$\begin{aligned} \rho w(z,t) = & \hat{\lambda} z + \frac{\partial w(z,t)}{\partial t} + \max_\eta \left\{ \eta \int_z^\infty [w(y,t) - w(z,t)] \frac{f(y,t)}{1-F(z,t)}\mathrm{d}y - \hat{\mu} z \eta \right\} \\ & + \int_0^\infty \left\{ w(y,t) \int_0^{\max\{y,z\}} \eta^*(x,t)\frac{-\varphi(x,t)}{1-F(x,t)}\mathrm{d}x + w(z,t)\int_0^z \eta^*(x,t)\varphi(x,t)\mathrm{d}x \right\} f(y,t)\mathrm{d}y, \end{aligned} \tag{6.3}$$

in which $\hat{\lambda}$ and $\hat{\mu}$ are the respective Lagrangian multipliers of the goods and labor market clearing conditions. Equation (6.3) is the counterpart of HJB equation (3.1) for the social planner. The left hand side is the flow social welfare, which is based on the discounting factor ρ rather than the interest rate r_t. Three items on the right hand side compose this flow value of a firm z. The first

[16] 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.

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 learning. The third and additional item that only shows up in the social planner's HJB equation is the last term in (6.3), capturing the externality of other firms' learning. The conditional source distribution has made the externality term very complicated since any firm can affect any 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 to meet 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 programing problem (6.3).

Proposition 7. *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 7 indicates that the optimal search strategy allocates more search intensity to more productive firms at an approximately power rate. The optimal allocation echoes with the prior intuition that large firms search too little in the market equilibrium. To make a clear illustration, figure 9 visualizes the comparison between search intensities in the competitive equilibrium and planner's problem at time 0. With assumption 3.1, proposition 7 implies that the optimal search intensity $\eta^*(z,0) = Cz^{\frac{k_0-1}{2}}$ for some constant C. The labor market clearing condition further 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 3.2. It is obvious that relative to the market outcome, a social planner would reallocate scientists from low-productivity firms to high-productivity firms.

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. Specifically, consider an infinitesimal time interval $[t, t+h]$. As before, the law of motion of the productivity distribution implies that

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

in which $\tilde{L}(z,t) = L(z,t)k(z,t)$ and is slow varying. When $k(t) > 1$, the Karamata's theorem applies to show that the integral is regularly varying with exponent $\frac{k(t)-1}{2}$.[17] $\tilde{F}(z,t+h)$ 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

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

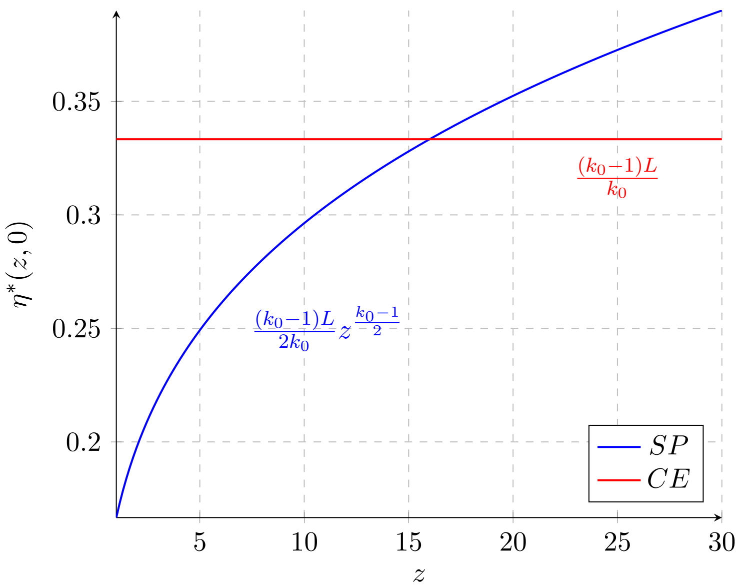


Figure 9: Comparing search intensities ($k_0 = 1.5$, $L = 1$)

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 $L = 1$.

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. With continuous time, the tail index immediately becomes one, and Zipf's law holds. Figure A.4 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 respects the labor market clearing condition. With a tail index one, such choice decreases the tail index below one and then violates the optimality.[18] Even though both policy exercises give 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.

7 Conclusions

In this paper, I establish a positive relationship between the right tail thickness of the firm size distribution and the level of economic development. I develop a growth model based on idea diffusion to rationalize this relationship as a feature of an asymptotic balanced growth path. Specifically,

[18] A tail index below one must imply infinite output and raise welfare. Then, there is no interval of time in which the output is finite, or the current strategy is suboptimal.

firms' learning from more productive firms thickens the right tail of the firm size distribution. A thicker tail further increases learning efficiency and sustains long-run growth. Gibrat's law also holds at all times, and Zipf's distribution emerges as the limiting firm size distribution. The idea diffusion mechanism then sheds light on the close relationship between Gibrat's law and thicker tails and between Zipf's law and long-run growth. The heterogeneous diffusion externality across firms unveils new insights into industrial policies. Policies favoring large firms take advantage of the diffusion externality and improve welfare.

That being said, the model can be enriched in several ways to quantify the importance of tail growth. First of all, the simple model needs more production decisions. The only factor endowment, labor, is used for search and adoption, not production. Second, the model abstracts from the entry and exit of firms. Therefore, the right tail cannot be thinner, even though it is not guaranteed to become thicker. Selection has long been known as an important source of growth, and it might also change the right tail, as in Luttmer (2007). Third, technology adoption in developing countries is often in the form of international knowledge diffusion. An open economy margin is pivotal to studying industrial policies in these countries. Lastly, a perfectly competitive market assumes away important questions on monopoly power. A sophisticated policymaker should take into account these aspects.

A key component of idea diffusion models is the learning function–how firms obtain and adopt ideas. This model shows that a slight deviation from standard assumptions has an enormous impact on firms and the whole economy. In particular, the optimal search policy of the social planner's problem reveals a huge potential for diffusion growth. It makes us rethink the growth accounting of innovation and imitation: how much growth is driven by genuinely original ideas, and how much is just transition growth from technology adoption? The idea diffusion model in this paper could be a useful framework to address this problem.

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.

Aghion, P. and P. Howitt (1992): "A Model of Growth Through Creative Destruction," *Econometrica*, 60, 323–351.

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.

Badillo, E. R., F. L. Galera, and R. Moreno Serrano (2017): "Cooperation in R&D, firm size and type of partnership: Evidence for the Spanish automotive industry," *European Journal of Management and Business Economics*, 26, 123–143.

Becker, W. and J. Dietz (2004): "R&D cooperation and innovation activities of firms—evidence for the German manufacturing industry," *Research Policy*, 33, 209–223.

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.

Brooks, W., K. Donovan, and T. R. Johnson (2018): "Mentors or Teachers? Microenterprise Training in Kenya," *American Economic Journal: Applied Economics*, 10, 196–221.

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.

Cai, J. and A. Szeidl (2018): "Interfirm Relationships and Business Performance*," *The Quarterly Journal of Economics*, 133, 1229–1282.

Cassiman, B. and R. Veugelers (2002): "R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium," *American Economic Review*, 92, 1169–1184.

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.

Eaton, J. and S. Kortum (1999): "International Technology Diffusion: Theory and Measurement," *International Economic Review*, 40, 537–570.

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.

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

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.

Iacovone, L., W. Maloney, and D. McKenzie (2022): "Improving Management with Individual and Group-Based Consulting: Results from a Randomized Experiment in Colombia," *The Review of Economic Studies*, 89, 346–371.

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.

Negassi, S. (2004): "R&D co-operation and innovation a microeconometric study on French firms," *Research Policy*, 33, 365–384.

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.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98, S71–S102.

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

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

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

Appendix A Additional Results of Section 2

A.1 Additonal Results with OECD data

Construction of figure 1 To obtain figure 1, I run the following regression

$$y_{c,t} = \alpha + \beta \text{logGDPpc}_{c,t} + \gamma_c + \varepsilon_{c,t},$$

in which $y_{c,t}$ is the relevant tail thickness metric and γ_c the country fixed effect. Let $\bar{\gamma} = \frac{\sum_{c=1}^{N_c} \gamma_c}{N_c}$. Then, each dot in the graph $(\text{logGDPpc}_{c,t}, \hat{y}_{c,t})$ is given by

$$\hat{y}_{c,t} = y_{c,t} - \gamma_c + \bar{\gamma},$$

i.e., $\hat{y}_{c,t}$ is the demeaned $y_{c,t}$ evaluated at the average country FE.

Table A.1: Table: Regression results with OECD countries

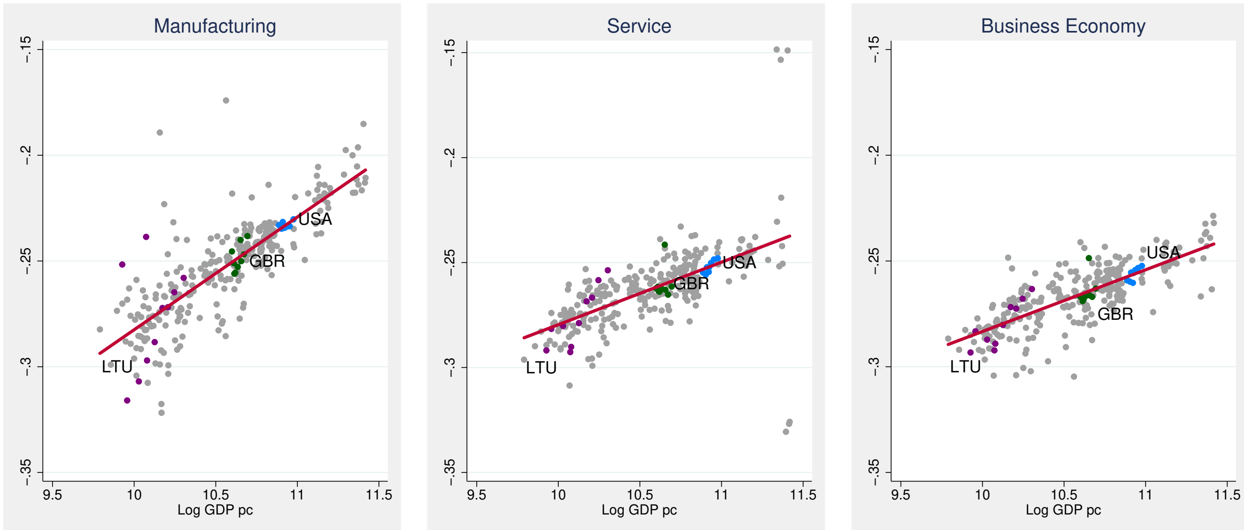
	Tot	Tot	Mft	Mft	Sev	Sev
Number of firms						
logGDPpc	0.05543^a	0.09936^a	0.1007^a	0.1553^a	0.1047^a	0.1550^a
	(0.017)	(0.027)	(0.020)	(0.030)	(0.016)	(0.028)
_cons	-1.7249^a	-2.1894^a	-2.0420^a	-2.6186^a	-2.2882^a	-2.8205^a
	(0.187)	(0.283)	(0.215)	(0.316)	(0.172)	(0.299)
Country FE	No	Yes	No	Yes	No	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.2 Additonal 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,



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

Figure A.1: 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 A.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.

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.2: 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 countres. The right tail thickness is measured by either $\tilde{R}^f_t$ (Num) or $\tilde{R}^{emp}_t$ (Emp). GDP per capita is in constant international dollar. Sources: the OECD SBS and PWT 10.0

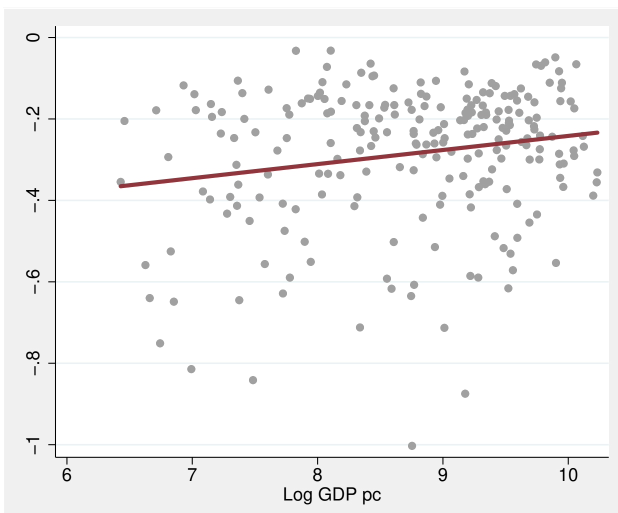
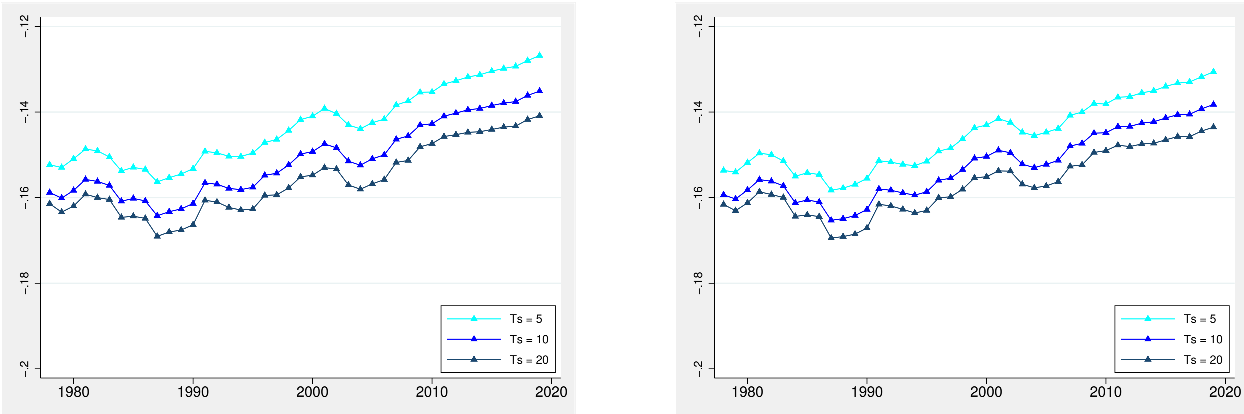


Figure A.2: Right tail thickness $\tilde{R}^{emp}_t$ and the level of development in developing countries

Notes. This figure plots the right tail thickness $\tilde{R}^{emp}_t$ 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 $\tilde{R}^{emp}_t$ 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.

A.3 Additonal Results with the US BDS data



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

Figure A.3: 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 B Omitted Proofs in Section 3

B.1 Proof of Lemma 1

Proof. From the HJB equation (3.1), the gains from learning is linear in η. Hence, the gains from learning per unit arrival rate must be non-positive, otherwsie 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,

$$v(t) = \int_t^\infty e^{-\int_t^x r(s)\mathrm{d}s}\mathrm{d}x. \tag{B.1}$$

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

B.2 Proof of Propostion 1

Proof. As in the proof of lemma 1, let $v(z,t) = v(t)z$, in which $v(t)$ is given by (B.1), and $w(t) = \frac{v(t)}{k(t)-1}$. I verify that $v(z,t)$ and $F(z,t)$ solve respectively firm's problem (3.1) and the law of motion on the productivity distribution (3.5). First, with $F(z,t)$ given by (3.6), 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 (3.1) following the proof of lemma 1.

Second, with $\eta(z,t) = \eta(t)$, $F(z,t)$ solves the KFE (3.5). To see this, rewrite (3.5) 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 (3.7).

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 $k(0) = k_0$, (3.7) and the above equation determine $k(t)$ and $F(z,t)$ at each date. Consequently, I obtain output per capita $y(t) = \mathbb{E}_{F(\cdot,t)}[z]/L$. 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 3.3, 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 $B_{1-\frac{1}{k}}(\nu+\alpha, 1-\alpha)$, it has the following hypergeometric representation. [19]

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

in which F is a hypergeometric function such that

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

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

Then,

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

Furthermore,

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

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,[20]

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

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 1'

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,

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

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 (3.5) implies that it is consistent to have Pareto $F(z,t)$ at all times. To see this, rewrite (3.5) 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

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

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 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 $z\geq x$ if and only if the relative arrival rate of any two ideas above firms' productivity is independent of firms.

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

$$\frac{m(z',x,t)}{m(z,x,t)}=\frac{m(z',x',t)}{m(z,x',t)}. \tag{C.1}$$

Proof. The "only if" part is obvious. To see the "if" part, it suffices to show that given condition (C.1), 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,

$$\tilde{H}(z,t)=\begin{cases}\frac{m(z,x_0,t)}{m(x_0,x_0,t)} & \text{if } z\geq x_0,\\ \frac{m(z,z,t)}{m(x_0,z,t)} & \text{if } z<x_0.\end{cases} \tag{C.2}$$

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 (C.1). Next, I show that $m(\max\{x, \bar{z}(x_0,t)\}, x, t) = 0$. This is straightforward for $x < x_0$ since (C.1) 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, $m(\bar{z}(x_0,t),x,t) = 0$.[21] 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 (C.1) 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 condtion (C.1) on the first term. In addition, $\tilde{H}(z,t)$ and $\tilde{H}(x,t)$ follow the definition (C.2) with $z > x_0$ and $x > x_0$ respectively.

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

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 condtion (C.1) on the first term. In addition, $\tilde{H}(z,t)$ and $\tilde{H}(x,t)$ follow the definition (C.2) with $z > x_0$ and $x < x_0$ respectively.

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 condtion (C.1) on the first term. In addition, $\tilde{H}(z,t)$ and $\tilde{H}(x,t)$ follow the definition (C.2) with $z < x_0$ and $x < x_0$ respectively.

The proof is then complete. ■

C.2 Proof of lemma 3

Proof. $k(t)$ decreases over time since (5.1) 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 (5.1):

$$\frac{\partial}{\partial t} \frac{\ln \tilde{F}(z,t)}{\ln z} = \frac{1}{\ln z} \frac{\tilde{H}(z,t)}{\tilde{F}(z,t)} \int_0^z \mu(x,t) f(x,t) \mathrm{d}x. \tag{C.3}$$

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 5.1 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 (C.3) 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 3 and completes the proof.

■

C.3 Proof of Lemma 4

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 (5.2) 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 (5.5).

■

C.4 Proof of Proposition 3

Proof. Since $k(t)$ decreases strictly, lemma 3 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

$$\eta(t) \equiv -\frac{\dot{k}(t)}{k(t)} = \lim_{z\to\infty} \frac{\mu(z,t)\tilde{H}(z,t)}{\ln z} \frac{\int_0^z \mu(x,t) f(x,t)\mathrm{d}x}{\mu(z,t) z f(z,t)} \in (0,\infty), \tag{C.4}$$

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

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

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

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

in which the second equality follows from lemma 4. 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

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

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 4 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, (C.4) 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 $m(t) \le h(t)$. 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 (C.4). Then we must have $m(t) = h(t)$ and complete the proof.

■

C.5 Proof of Proposition 4

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 3 implies

$$-\dot{k}(t) = B(t) \lim_{z\to\infty} \frac{\int_0^z \mu(x,t)f(x,t)\mathrm{d}x}{\ln z}. \tag{C.5}$$

Since $g^r(t)$ exists, lemma 4 and assumption 5.2 imply that $\lim_{z\to\infty} \mu(z,t)\tilde{H}(z,t)$ exists. With internal search and a reguar 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,

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

Next, I show that the "only if" part also stands. From (C.5), 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 4 implies $\mu(z,t)\tilde{H}(z,t) = (k(t) - 1)g^r(t) + o(1)$. Therefore,

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

since $\mu(x,t)\tilde{H}(x,t)$, $\frac{xf(x,t)}{\tilde{F}(x,t)}$ 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 (5.7)

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

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

■

Appendix D Omitted Proofs in Section 6

D.1 Proof of proposition 6

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.,

$$\eta(z,t) = \begin{cases} 0, & \text{if } z \leq z^*, \\ \eta(t), & \text{if } z > z^*. \end{cases} \tag{D.1}$$

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 6 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 $z \leq z^*$ 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 $z > z^*$,

$$\begin{aligned} \tilde{F}(z,t+h) &= \tilde{F}(z,t) + \int_{z^*}^{z} \eta(x,t)h\frac{\tilde{F}(z,t)}{\tilde{F}(x,t)}f(x,t)\mathrm{d}x \\ \Rightarrow \frac{\partial \tilde{F}(z,t)}{\partial t} &= \tilde{F}(z,t)\int_{z^*}^{z} \eta(x,t)\frac{f(x,t)}{\tilde{F}(x,t)}\mathrm{d}x. \end{aligned} \tag{D.2}$$

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 (6.1) satisfies all three conditions, it suffices to verify that it satisfies the law of motion (D.2). Note that for $z > z^*$,

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

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

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

$$\eta(t)\int_{z^*}^{\infty} yf(y,t)\mathrm{d}y = L,$$

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

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

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 6. 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 $z \le z^*$ and zero for $z > z^*$. Noticing that

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

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

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

The last equality uses that $(z^*/z)^{1-k_0} < 1$ and $k(t) < k_0$. For $z > z^*$, the truncated distribution $F(x|x \ge z, t)$ is exactly Pareto with shape paramter $k(t)$ 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 (6.1). With $\tau > 0$, the policy maker can choose the threshold z^* to eliminate threshold equilibria with $\hat{z} < z^*$.

■

D.2 Derivation of equation (6.3) on $w(z,t)$

Following Lucas and Moll (2014), the corresponding HJB equation of problem (6.2) is

$$\begin{aligned}
\rho W(f) = &\max_{\{c(\omega), \eta(y)\}} \int_\Omega u(c(\omega)) \mathrm{d}\omega + \int_0^\infty \frac{\delta W(f)}{\delta f(y)} f(y) \left[\int_0^y \eta(x) \varphi(x) \mathrm{d}x - \eta(y) \right] \mathrm{d}y \\
&\text{s.t.} \quad \int_\Omega c(\omega) \mathrm{d}\omega \le \int_0^\infty y f(y) \mathrm{d}y, \quad \int_0^\infty y \eta(y) f(y) \mathrm{d}y \le L,
\end{aligned} \tag{D.3}$$

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

$$u'(c(\omega)) = \hat{\lambda}. \tag{D.4}$$

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

$$\begin{aligned}
&\int_z^\infty w(f,y)f(y)\varphi(z)\mathrm{d}y - w(f,z)f(z) - \hat{\mu}zf(z) = 0 \\
\Longrightarrow &\int_z^\infty [w(f,y) - w(f,z)]\,\frac{f(y)}{1-F(z)}\mathrm{d}y = \hat{\mu}z
\end{aligned} \tag{D.5}$$

Differentiating both sides of the HJB equation (D.3) with respect to $f(z)$,

$$\begin{aligned}
\rho w(f,z) = &\int_0^\infty \frac{\delta w(f,z)}{\delta f(y)} f(y)\left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right]\mathrm{d}y \\
&+ \int_0^\infty w(f,y)\frac{\delta}{\delta f(z)}\left\{f(y)\left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right]\right\}\mathrm{d}y + \hat{\lambda}z - \hat{\mu}z\eta^*(z).
\end{aligned} \tag{D.6}$$

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 $f(\cdot,t) = f$. Hence, the first term on the RHS of (D.6) is simply $\partial w(z,t)/\partial t$. To calculate the second term, note that

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

Therefore,

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

Then,

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

Finally, rewriting (D.6) gives equation (6.3):

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

in which the max operator comes from first order condition (D.5).

D.3 Proof of propostion 7

Proof. To solve for the optimal policy, I first differentiate equation (6.3) with respect to z.

$$
\begin{aligned}
\rho w_z(z,t)=&\frac{\partial w_z(z,t)}{\partial t}+\hat{\lambda}+\eta^*(z,t)\frac{\partial}{\partial z}\left\{\int_z^\infty\left[w(y,t)-w(z,t)\right]\frac{f(y,t)}{1-F(z,t)}\mathrm{d}y-\hat{\mu}z\right\}\\
&+w(z,t)f(z,t)\int_0^z\eta^*(x,t)\frac{-\varphi(x,t)}{1-F(x,t)}\mathrm{d}x-w(z,t)f(z,t)\int_0^z\eta^*(x,t)\frac{-\varphi(x,t)}{1-F(x,t)}\mathrm{d}x\\
&+\int_z^\infty w(y,t)f(y,t)\mathrm{d}y\eta^*(z,t)\frac{-\varphi(z,t)}{1-F(z,t)}+w_z(z,t)\int_0^z\eta^*(x,t)\varphi(x,t)\mathrm{d}x+w(z,t)\eta^*(z,t)\varphi(z,t)\\
=&\frac{\partial w_z(z,t)}{\partial t}+\hat{\lambda}-\eta^*(z,t)\varphi(z,t)\int_z^\infty\left(w(y,t)-w(z,t)\right)\frac{f(y,t)}{1-F(z,t)}\mathrm{d}y+w_z(z,t)\int_0^z\eta^*(x,t)\varphi(x,t)\mathrm{d}x\\
=&\frac{\partial w_z(z,t)}{\partial t}+\hat{\lambda}-\hat{\mu}\eta^*(z,t)\varphi(z,t)z+w_z(z,t)\int_0^z\eta^*(x,t)\varphi(x,t)\mathrm{d}x
\end{aligned}
\tag{D.7}
$$

The first equality is the result of an evenlope theorem, and the second and the third use the first order condition (D.5), 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 (D.5):

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

in which the last equality comes from the law of motion on $\varphi(z,t)$. Formally, the original law of motion (3.5) implies that by differentiating both sides on z,

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

Inserting them back to (D.7), the following differential equation characterizes the optimal policy.

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

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:

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

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

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

The thrid equality comes from integraton 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.

∎

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

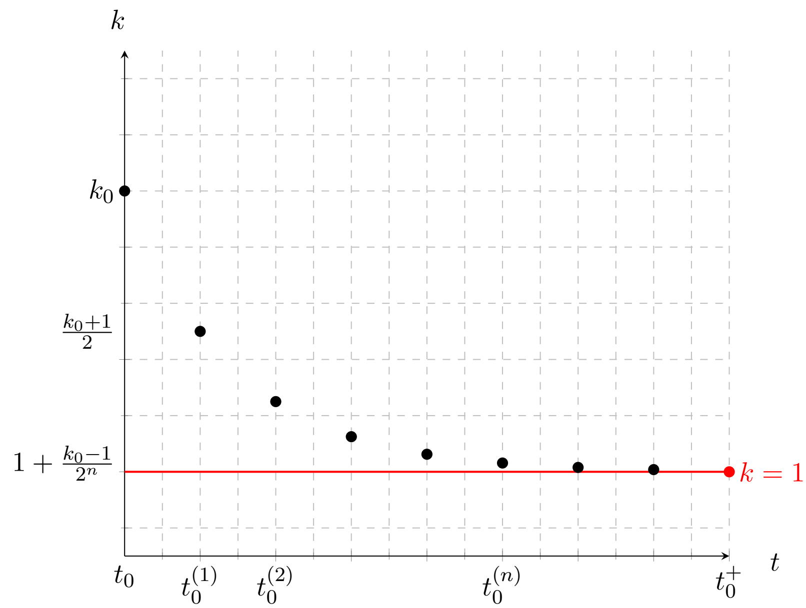


Figure A.4: Illustration of jumps in tail indices

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.

Supplementary References

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

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

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.