Economic Growth and the Rise of Large Firms

Zhang Chen*

March 2025

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. I develop a simple idea search model with both endogenous growth and an endogenous firm size distribution. The economy features an asymptotic balanced growth path. Along the transition, Gibrat's law holds at each date, and the right tail of the firm size distribution becomes monotonically thicker. The firm size distribution converges to Zipf's distribution. The model also implies that policies favoring large firms can improve welfare due to the externality associated with idea search.

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

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

1 Introduction

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

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

It is widely known that power laws characterize the right tails of firm size distributions in advanced economies. The tail index (or Pareto exponent) naturally serves as a measure of tail thickness. Meanwhile, national statistical offices often report firm size distributions in tabulated form, summarizing the number of firms and employees by firm size bins. To accommodate this data structure, I employ two complementary tail estimators specifically designed for tabulated firm size distributions and cross-validate all empirical results using both estimators.

My main empirical analysis focuses on OECD countries and the United States. I estimate the tail index for each country-year pair using data from the OECD Structural Business Statistics (SBS) and the U.S. Census Business Dynamics Statistics (BDS). For both the OECD and U.S. samples, I document a negative relationship between the tail index and log GDP per capita within countries over time. In other words, as an economy grows, the right tail of the firm size distribution becomes thicker. Additional robustness checks confirm that this relationship holds across major sectors, across countries, over longer time horizons, and within the subset of large firms.

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

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

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

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

state-of-the-art technology. This implication is at odds with empirical work which finds larger firms are more likely to adopt the most advanced technologies.

This departure of my model yields two important insights about the firm size distribution and growth. First, my model presents a novel growth mechanism termed *tail growth*: growth is generated by the thickening of the right tail. Earlier models in the literature assume a balanced growth path in which the distribution of *relative* firm productivity is stationary. The productivity distribution is scaled up proportionately so that its shape remains unchanged. In my model, the distribution of *relative* firm productivity varies along the equilibrium path. Aggregate productivity improves due to the redistribution of mass from lower productivity to higher productivity in *relative* terms. This redistribution manifests as a thickening of the right tail, capturing the rising share of high productivity firms. Specifically, in the model, output per capita y is the mean of firms' productivity, $k/(k-1)$, in which k is the aforementioned Pareto shape parameter. In other words, my model predicts a tight relationship between GDP per capita and right tail thickness. The model economy features an asymptotic balanced growth path since with firms' optimal search intensity, $k-1$ decreases towards zero at a constant rate, i.e., the firm size distribution converges to Zipf's distribution. My model thus offers an explanation for why advanced countries such as the U.S. 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 U.S.. How do we reconcile constant growth with the rise in concentration? Notice that the right tail thickness is a concentration measure by itself and partly determines other common measures such as the Herfindahl-Hirschman Index. Thus, the model describes an economy on an asymptotic balanced growth path that exhibits continually rising concentration. In my model, output growth $\dot{y}/y$ converges to a constant as $k-1$ 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 is proportional to $k/(k-1)$, and 2) $k-1$ decreases to 0 at a constant rate. Both predictions can be readily tested using my tail estimates. For the U.S. from

1978 to 2019, the observed growth in output per capita closely tracks changes in $k/(k-1)$. Plotting $\ln(k-1)$ against time also reveals a clear linear trend with a correlation coefficient of -0.88.

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

Finally, I explore several extensions that relax the assumptions of the baseline model. Despite incorporating a richer set of model elements, including population growth and endogenous firm entry, I demonstrate that the extended model remains as analytically tractable as the baseline model while preserving all its desirable theoretical properties. Moreover, the extended model explicitly confirms that scale effects are not fundamental to the tail growth mechanism and further generalizes the key insight from the baseline model—that tail growth represents a persistent source of reallocation gains.

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. Numerous studies 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] Most existing work employ steady-state comparisons to capture cross-

[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)

country differences, which are often attributed to exogenous cross-country variations in distortions. Instead, this paper targets the right tail of the firm size distribution and rationalizes the relationship as an endogenous feature of the transition path.

Second, there has been growing interests in the rise of market concentration in the United States over the past several decades, as evidenced by studies such as Gutiérrez and Philippon (2018), De Loecker et al. (2020), and Autor et al. (2020). Closely related to this paper, Cao et al. (2022) and Chen et al. (2023) provide both evidence and explanations for the thickening of the right tail in firm size distributions in the U.S. or North America. They characterize the tail index of the stationary firm size distribution within models that incorporate innovation choices and argue that a reduction in innovation cost leads to a thicker right tail.[2]

Comparing to their work, this paper focuses on the firm size distribution during the transition period rather than in the steady state. Instead of relying on exogenous technological shocks, I demonstrate that a thickening right tail can naturally emerge along an asymptotically balanced growth path. Kwon et al. (2024) share this perspective, viewing the rise in concentration as a normal feature of economic growth. They utilize historical data from the U.S. to show that the rise in market concentration has been observed for approximately a century, concomitant with balanced growth.

Third, this paper builds heavily upon a burgeoning literature on endogenous growth models with idea flows. Buera and Lucas (2018) provide a comprehensive survey of this literature.[3] This paper differs from the existing literature pioneered by Kortum (1997) in two significant respects. First, existing models focus on balanced growth paths in which the distribution of relative firm productivity is stationary. In contrast, my model features an asymptotic balanced growth path in which the distribution of relative firm productivity varies. Specifically, the right tail of the distribution becomes thicker over time. Second, Gibrat's law does not

[2]Relevant to this discussion, Oberfield (2018) and Gouin-Bonenfant (2022) develop tractable models of firm dynamics and derive explicit formulas linking a lower labor share to thicker right tails. Oberfield (2018) examines an endogenous process of buyer-seller network formation and derives an equilibrium employment distribution across firms, where the tail index is endogenous and expressed as a function of an exogenous firm-level labor share. In contrast, Gouin-Bonenfant (2022) incorporates search frictions and wage posting in the labor market to endogenize both the labor share and the distribution of firm employment. Given an exogenous Pareto distribution of productivity, he derives the aggregate labor share as a function of its tail index.

[3]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).

hold in those idea search models.[4]

Notably, this paper is related to recent work by Jones (2023). Both studies explore alternative approaches to the classical conditions outlined in Kortum (1997) for constant growth, which involve a Pareto source distribution and exponential growth in the idea flow. Jones (2023) relaxes the thick-tailedness assumption but strengthens exponential growth to combinatorial growth. In contrast, my approach allows a constant population (decreasing idea flows) but make a stronger assumption on the source distribution. Instead of searching randomly over the same Pareto distribution, firms in my model search more efficiently over Pareto distributions left-truncated by their productivity.

Fourth, this paper is related to the study of Zipf's law. Axtell (2001) documented that the US firm size distribution has a remarkable resemblance to Zipf's distribution. Gabaix (1999) and Luttmer (2007, 2012) develop the dominant approach in the literature to explain Zipf's law through random growth theory. In particular, they derive stationary distributions with Pareto tails by modeling firm productivity as geometric Brownian motions with barriers and establish parametric conditions under which the tail index converges to one. They show that parameters can be chosen such that the tail index is close to one. More recently, Toda (2017) introduces a framework that relaxes these parametric restrictions by incorporating general equilibrium forces. Through calibration exercises, he finds that the model-implied tail index can approach one across a broad range of parameter values.

The contribution of this paper is to provide an explicit economic rationale for the emergence of Zipf's law. In my model, Zipf's law in the firm size distribution arises as a necessary consequence of an idea search process that progressively thickens the right tail and generates asymptotically balanced growth. Furthermore, I present empirical evidence supporting the thickening of the right tail along the growth path.

2 Stylized Facts

In this section, I provide empirical evidence supporting a positive relationship between the right tail thickness of the firm size distribution and the level of economic development. In Section 2.1, I describe my measure of the right tail thickness —the

[4]In the working paper version, Chen (2023), I further show that in a general class of idea search models, including these earlier models, Gibrat's law holds if and only if the right tail becomes thicker.

tail index —and the method for estimating it. In Section 2.2, I present the main results on this relationship based on a panel of OECD countries and U.S. data. In Section 2.3, I provide further evidence showing that the positive relationship remains robust across sectors, countries, time periods, and among large firms. Collectively, these evidence suggest that a thickening right tail of the firm size distribution is an inherent feature along the path of economic growth.

2.1 Measuring the right tail thickness

The literature has established that the right tail of the firm size distribution in advanced economies follows a power law (Gabaix, 2009). Let $\tilde{F}(x)$ denote the complementary cumulative distribution function (CCDF) of firm size in the economy. Then, $\tilde{F}$ is well approximated by the following form for x in the right tail:

$$\Pr(\text{Size} > x) \equiv \tilde{F}(x) = cx^{-k}, \tag{1}$$

where c and k are constants. Here, k is commonly referred to as the (right) tail index or the Pareto exponent. For a fixed x, a lower k implies a larger value of $\tilde{F}(x)$, indicating greater concentration in the right tail. In other words, a smaller k corresponds to a thicker right tail.

While the literature extensively addresses methods for estimating the tail index of thick-tailed distributions, these statistical procedures generally require access to individual observations.[5] In this context, estimating the right tail of firm size distributions requires firm-level data. However, census data on firms are often less accessible due to confidentiality constraints. In contrast, aggregate statistics on firm size distributions are usually public and widely available, offering a practical advantage in cross-country studies where access to administrative firm-level data across multiple countries is challenging.

Table 1 provides a typical example of how statistical offices report firm size distributions at an aggregate level. The table summarizes both the number of firms and employment by firm size, which is measured by the number of employees. In the first row, firms are grouped into size bins based on a sequence of increasing thresholds, $\{T_i\}_{i=0}^{n}$. The second and third rows respectively display the number of firms and total employment within each size bin. I define the top firm share at threshold T, $N(T)$, as the proportion of firms with more than T employees. Similarly, the top employment share at threshold T, $E(T)$, represents the propor-

[5] Resnick (2007) introduces standard estimators for the tail index. See also Gabaix (2009) for an overview of tail estimation techniques and their applications in economics.

tion of total employment in firms exceeding T employees. Using the values in the second and third rows, I calculate the respective top firm and employment shares.

Table 1: Number of firms/employment by firm size

SIZE	$[T_0, T_1)$	$[T_1, T_2)$	$\cdots$	$[T_{n-1}, T_n)$	$[T_n, +\infty)$
NUM	N_0	N_1	$\cdots$	N_{n-1}	N_n
EMP	E_0	E_1	$\cdots$	E_{n-1}	E_n

I use two approaches to estimate the tail index k based on tabulated firm size distributions in the form of Table 1. The first approach, which I refer to as the "simple" estimator, calculates k directly by inverting the top firm shares. Given two size thresholds, T_L and T_S, where $T_L > T_S$, I take the logarithm of both sides of equation (1) and rearrange terms to obtain:

$$k^f = -\ln \frac{N(T_L)}{N(T_S)} / \ln \frac{T_L}{T_S}.$$

In practice, T_L and T_S are selected from the set of thresholds $\{T_i\}_{i=0}^n$ with $L > S$. Then the ratio of the top firm shares at thresholds T_L and T_S, $N(T_L)/N(T_S)$, is calculated as $\sum_{i \geq L} N_i / \sum_{i \geq S} N_i$.

The second approach closely follows the estimation method developed by Toda and Wang (2021) for studying the tail index of income distributions. They introduce an efficient minimum distance estimator that uses only top income shares and percentiles to estimate the tail index k. Specifically, this estimator identifies the tail index k that minimizes the distance between normalized top income shares predicted by a Pareto distribution with shape parameter k (the model) and those observed in the data.[6] Toda and Wang (2021) further prove that their estimator

[6] Given a sequence of top income percentiles $0 < p_1 < \cdots < p_K < p_{K+1}$ and the corresponding top income shares $S_1 < \cdots < S_K < S_{K+1}$, Toda and Wang (2021) consider the following vector of self-normalized non-overlapping top income shares $\bar{s}$:

$$\bar{s} \equiv (\bar{s}_1, \ldots, \bar{s}_{K-1})^T \equiv \left(\frac{S_2 - S_1}{S_{K+1} - S_K}, \ldots, \frac{S_K - S_{K-1}}{S_{K+1} - S_K} \right)^T.$$

In a Pareto distribution with tail index k, $S_k = p_k^{1-\frac{1}{k}}$. Using $\xi = 1/k$, these shares take the following form with a Pareto distribution:

$$r(\xi) \equiv (r_1(\xi), \ldots, r_{K-1}(\xi))^T \equiv \left(\frac{p_2^{1-\xi} - p_1^{1-\xi}}{p_{K+1}^{1-\xi} - p_K^{1-\xi}}, \ldots, \frac{p_K^{1-\xi} - p_{K-1}^{1-\xi}}{p_{K+1}^{1-\xi} - p_K^{1-\xi}} \right)^T.$$

Toda and Wang (2021) derive estimator ξ^{TW} by solving the distance minimization problem $\min_{\xi \in (0,1)} (r(\xi) - \bar{s})^T \tilde{\Omega}(\xi)^{-1} (r(\xi) - \bar{s})$, where $\tilde{\Omega}$ is the weighting matrix that achieves efficiency.

is the most efficient among a class of classical minimum distance estimators.

For firm size distributions, the top firm shares are analogous to top income percentiles (firms' ranking by employment size) and the top employment shares analogous to top income shares (shares of employment among the largest firms). Given a threshold T at the right tail, their estimator can be directly applied to these top firm and employment shares. I denote this Toda-Wang estimator of the tail index as k^{TW}.

Discussion Before presenting the results based on the two tail estimators, it is instructive to compare them. The Toda-Wang estimator k^{TW} leverages all the information about top firm and employment shares in the right tail and is proven to be efficient. However, it requires that $k > 1$.[7] In practice, their algorithm imposes a lower bound of $k = 1.001$ when minimizing the distance of moments over k. This constraint can cause estimates to mechanically cluster around 1.001, introducing attenuation bias into the regression results. To mitigate this issue, I exclude observations where k^{TW} is between 1.001 and 1.0012 when applying this estimator. As a result, the sample size is significantly reduced due to the well-known Zipf's law in firm size distributions—that is, k is close to 1.

In contrast, the simple estimator k^f does not impose the restriction that $k > 1$, rendering it more suitable for analyzing firm size distributions. However, k^f relies solely on two size thresholds and fails to utilize the information contained in employment shares. The simple estimator then tends to be less efficient. The discrepancies between the Toda-Wang estimator and the simple estimator highlight a trade-off between sample size and efficiency. To address both of these aspects, I present the empirical results obtained using both estimators.[8]

[7] This is because a Pareto distribution with $k \leq 1$ has no finite first moment; consequently, the top employment share cannot be defined.

[8] Analogous to k^f, it is straightforward to construct a simple estimator k^{emp} using only top employment shares. Specifically, given thresholds T_L and T_S with $T_L > T_S$,

$$k^{emp} = 1 - \ln \frac{E(T_L)}{E(T_S)} / \ln \frac{T_L}{T_S},$$

where $E(T_L)/E(T_S)$ represents the ratio of top employment shares at thresholds T_L and T_S. In Supplemental Appendix E.1, I present all the results using k^{emp} for robustness and confirm that they remain consistent.

This arrangement reflects the fact that k^{emp} is as restrictive as k^{TW} but less efficient. Like k^f, k^{emp} is a simple estimator that relies on only two size thresholds, making it inherently less efficient than k^{TW}. Moreover, not only does k^{emp} impose the same requirement as k^{TW} that $k > 1$, it also ensures that the tail estimate remains strictly above one. This property makes it at least as restrictive as k^{TW}, if not more so. In Supplemental Appendix E.2, I provide further comparisons on the performance of the three estimators in firm size distributions.

2.2 Main results

I apply the above tail estimators to firm size distributions in OECD countries and the United States. In both cases, the firm size distributions are structured as shown in Table 1, which reports both the number of firms and employment by firm size (measured by the number of employees). Among OECD countries, I identify a negative relationship between the tail index and the log of GDP per capita. In the U.S., I observe an downward trend in the tail indices over time. The remainder of this section provides a detailed account of the data sources, empirical specifications, and results in each exercise.

2.2.1 The OECD countries

The OECD Structural Business Statistics (SBS) database provides detailed, harmonized statistics on the industrial structure of OECD economies. Most importantly, it offers a breakdown of the number of firms and employment by sector and size class for each economy. My final sample comprises a panel of 33 OECD countries from 2008 to 2017.[9] Firms are categorized into four size bins based on the number of employees: 1-9, 10-49, 50-249, and 250+.[10]

For OECD countries, I focus on firms with at least 10 employees, setting the right tail threshold T at 10. This threshold is selected to approximately capture the top 10% of firms in a typical OECD economy.[11] Additionally, the Toda-Wang estimator requires a minimum of three size thresholds. With $T = 10$, this minimum requirement is met using top firm and employment shares at thresholds 10, 50, and 250. For the simple estimator, I set $T_S = T = 10$ and $T_L = 250$, aligning with the OECD's standard for large firms.

I proceed by regressing the tail index k on the log of GDP per capita to examine their relationship. Notably, I include country fixed effects to ensure that the coefficient on log GDP per capita is identified solely from within-country

[9] This sample includes all OECD countries as of the end of 2018, except for Chile, Mexico, and Korea, which are excluded due to data inconsistencies or incompleteness. Although the original database spans 2005 to 2018, I focus on 2008-2017 because few countries have data for 2005-2007, and 2007 and 2018 exhibit systematic breaks likely due to changes in measurement practices.

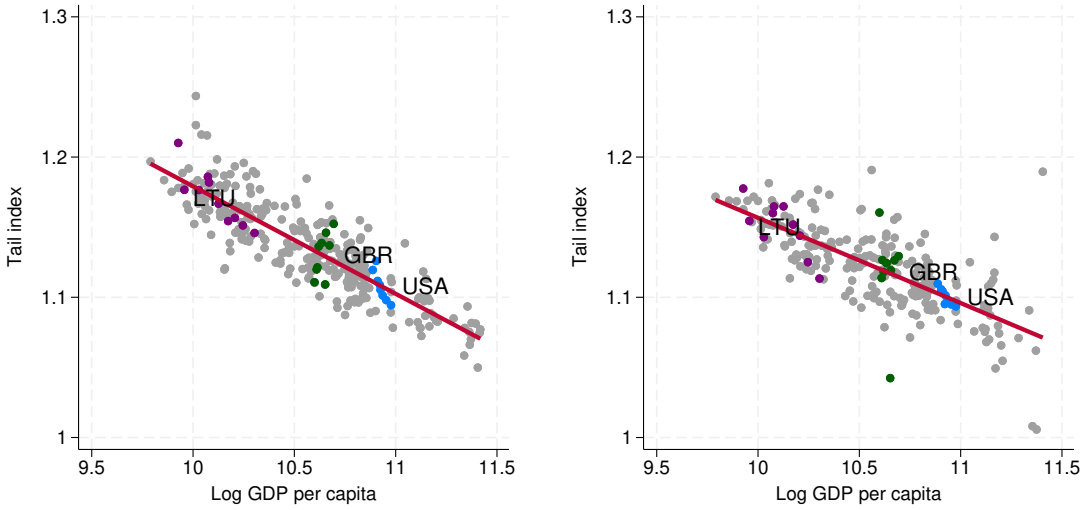
[10] The "number of employees" refers to the total number of persons working in or for the unit during the reference year, including both full-time and part-time workers. The original dataset contains various size classes—1-249, 1-9, 10-19, 10-49, 20-49, 250+, 50-249, and Total. I standardized these into the four size classes above to enhance cross-country comparability.

[11] Across all country-year pairs in my OECD sample, the median proportion of firms with fewer than 10 employees is 91.6%. Only three countries have fewer than 80% of firms with fewer than 10 employees: the U.S. (79.2%), Switzerland (68.1%), and New Zealand (79.5%).

variations over time. Formally,

$$k_{c,t} = \alpha + \beta \text{logGDPpc}_{c,t} + \gamma_c + \varepsilon_{c,t}, \tag{2}$$

in which γ_c represents the country fixed effect, and $k_{c,t}$ denotes either the simple tail index $k^f_{c,t}$ or the Toda-Wang tail index $k^{TW}_{c,t}$. The variable $\text{logGDPpc}_{c,t}$ is real GDP per capita, measured in constant international dollars and sourced from the Penn World Table (PWT) version 10.0.



(a) k^f (b) k^{TW}

Figure 1: Tail index and log GDP per capita in OECD countries

Notes. This figure plots the tail index k against log GDP per capita among a panel of OECD countries. The scatter dots are readjusted by country fixed effects, and the red lines are the linear fits. The tail index in the left figure is k^f with $T_S = 10$ and $T_L = 250$, and that in the right figure is k^{TW} with $T = 10$. The log GDP per capita is in constant international dollars. Three annotated countries are Lithuania (LTU, purple), the UK (GBR, green) and the USA (blue). Data sources: OECD SBS and PWT 10.0.

Figure 1 presents a summary and visualization of the key regression results in three layers. First, each dot represents a pair $(\text{logGDPpc}_{c,t}, \hat{k}_{c,t})$, in which $\hat{k}_{c,t} = k_{c,t} - \gamma_c + \bar{\gamma}$, with $\bar{\gamma}$ as the mean of the country fixed effects, i.e., $\bar{\gamma} \equiv \frac{\sum_{c=1}^{N_c} \gamma_c}{N_c}$. These gray dots thus illustrate the relationship between the tail index and log GDP per capita, *after controlling for* cross-country differences. Second, the slope of the fitted line represents the estimated coefficient $\hat{\beta}$, which is negative. Third, I annotate specific points to illustrate the trajectories of the tail index across development stages for selected countries. The annotated series for Lithuania, the U.K., and the U.S. highlight examples of countries at different stages of economic development. Full regression results can be found in Appendix A.1.

Figure 1 indicates that lower tail indices k are associated with higher levels of log GDP per capita. Equivalently, it suggests a positive relationship between the right tail thickness and the level of development. It is important to note that this observation is based solely on within-country over time variations. That is, the right tail tends to become thicker as a country advances along its growth trajectory. This observation also implies that theories relying on cross-country differences in exogenous factors are inadequate for explaining the thickening of the right tail during the transition.

2.2.2 The United States

The Business Dynamics Statistics (BDS) from the U.S. Census Bureau provides detailed data on the distribution of firm sizes in the United States over the past four decades (1978-2019). It categorizes all U.S. business firms into 10 size classes based on the number of employees, offering information on both the number of firms and employment within each size class.[12]

Consistent with the OECD analysis, I focus on the top 10% firms in the U.S. by the number of employees, corresponding to a right tail threshold T of 20 employees. I apply the Toda-Wang estimator to the top firm and employment shares at all seven thresholds equal to or exceeding 20. For the simple estimator, I set $T_S = T = 20$ and $T_L = 500$, maintaining the same T_L/T_S ratio as in the OECD case. Firms with more than 500 employees represent approximately the top 0.5% of firms in the U.S. business economy. Furthermore, the richness of the BDS data enables an examination of the sensitivity of the simple estimator to different threshold choices. To this end, I also test alternative thresholds with $T_S = 5, 10$ and $T_L = 1000$. The time series of these k^f estimates are available in Appendix A.1.[13]

Contrary to the OECD analysis, I plot the tail indices against years instead of log GDP per capita to examine the temporal evolution within an archetypal growing economy. Figure 2 illustrates a declining tail index in the U.S. firm size distribution from 1978 to 2019. Specifically, the right tail of the U.S. firm size distribution has been continuously thickening over the past forty years.

The thickening of the right tail in the US aligns with recent evidence of rising

[12]The 10 size classes are: 1-4, 5-9, 10-19, 20-99, 100-499, 500-999, 1000-2499, 2500-4999, 5000-9999, and 10000+. The employee count includes both full-time and part-time workers.

[13]The time series of k^f with $T_S = 20$ and $T_L = 1000$ is noteworthy. The median fraction of firms with more than 250 employees in the OECD sample is 0.002, and in the U.S., the median fraction of firms with more than 1000 employees is also 0.002. This specification aligns with the OECD simple estimator in terms of top firm shares at the thresholds, despite doubling the T_L/T_S ratio.

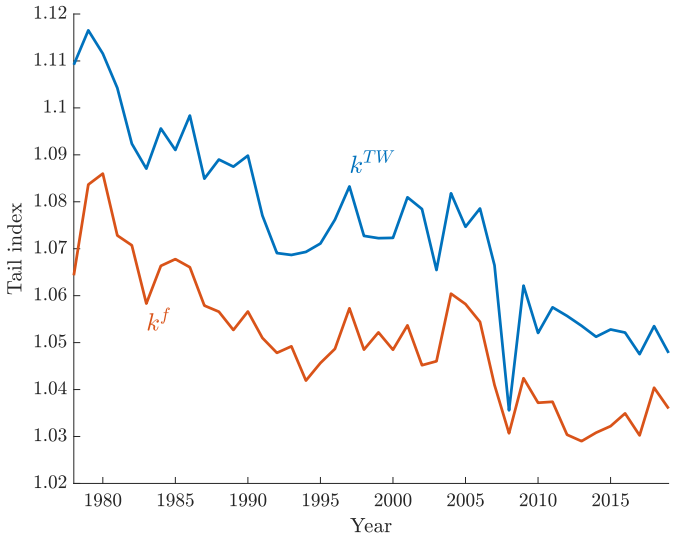


Figure 2: Tail index in the US (1978-2019)

Notes. This figure plots the tail index k of the size distribution of all US business firms from 1978 to 2019. The tail indices are obtained by either the Toda-Wang estimator k^{TW} with $T = 20$ (in blue) or the simple estimator k^{f} with $T_S = 20$ and $T_L = 500$ (in red). Data source: the U.S. Census BDS.

market concentration. This trend echoes the findings by Autor et al. (2020) and Kwon et al. (2024). Autor et al. (2020) document that, across various industries, the market shares of the top 4 or 20 firms in terms of total sales or employment have steadily increased from 1980 to 2010. Similarly, Kwon et al. (2024) show that the aggregate employment shares of the top 1% and top 0.1% of firms have risen over the past forty years. Furthermore, by utilizing historical data on firms' financial metrics—such as assets, sales, and net incomes—they suggest that this trend in U.S. market concentration may date back a century.

2.3 Additional evidence

Evidence from OECD countries and the United States suggests that the right tail of the firm size distribution grows thicker with economic development. The thickening right tail is then a plausibly general characteristic of the growth process. In this section, I provide further evidence supporting this interpretation. All results in this section are delegated to Appendix A.2.

Across sectors The tail index is used as an aggregate statistic in the main analyses, estimated from the firm size distribution across all business firms within the economy. At the same time, economic development is known to drive structural transformation. This raises the possibility that economic growth may not directly

affect the tail indices of firm size distributions; rather, the thickening of the right tail could primarily be a compositional effect resulting from shifts across sectors due to structural transformation.

In Table A.2, I confirm that as the economy grows, the right tail of the firm size distribution thickens within both the manufacturing and service sectors, regardless of tail index measures. Additionally, the observation that the right tail also thickens within the service sector implies that the underlying mechanism is not solely tied to international trade, as it affects even less-tradable sectors.

Across countries In Table A.1, I demonstrate that the positive relationship observed in the OECD sample remains robust in specifications without country fixed effects. These cross-sectional evidence supports the notion that country-year pairs with higher log GDP per capita exhibit thicker right tails, suggesting that differences in tail indices may reflect variations in countries' positions along a common trajectory of development.

Beyond the OECD countries, it is natural to question whether a similar cross-sectional relationship exists in developing countries. The World Bank Enterprise Survey (WBES) is one of the few datasets that provide information on firm size distributions across a wide range of developing countries. Using an early WBES sample (2006–2010), both Ayyagari et al. (2014) and García-Santana and Ramos (2015) find that the employment share of large firms (those with more than 100 employees) is higher in high-income countries than in low-income countries. Assuming a Pareto distribution for firm size, a higher share of employment in large firms is isomorphic to a thicker right tail. In Figure A.2 and Table A.3, I replicate their analyses in terms of tail indices and confirm that, in a more recent WBES sample (2006–2019), higher-income countries also tend to exhibit thicker right tails.

However, I suggest exercising caution in concluding that the positive relationship between right-tail thickness and economic development extends to developing countries. First, the WBES is an establishment-level survey that covers only formal firms, while informal firms—prevalent and generally smaller in developing countries—are excluded. This likely results in downward bias in the estimated tail indices. Second, there exists a longstanding but debated hypothesis that firm size distributions in developing countries are bimodal, with a relative scarcity of mid-sized firms compared to developed economies. This potential "missing middle" casts doubt on whether the Pareto distribution is still a reasonable approxima-

tion for the right tail.[14] Third, the literature highlights the role of institutional distortions in shaping firm size distributions in developing countries, which is contrastingly low in developed countries. Since economic development itself may impact the level of such distortions, the dynamics of the tail index in developing countries might diverge from those documented in developed countries.[15]

Across time The longest time series used in the main analysis is the one of U.S. firm size distributions from 1978 to 2019. It is still worth considering whether data from the past four decades are sufficient to establish a long-term trend. Kwon et al. (2024) address this by examining U.S. firm size distributions using historical corporate income data from the IRS. They compile a dataset of annual U.S. firm size distributions from 1918 to 2018, with firm size measured by assets, receipts, and net income. In Figure A.3, I present the historical tail index of U.S. business firms using their dataset and confirm that the trend of a thickening right tail persists over this extended horizon (1918–2018).

Large firms There are potential concerns about whether aggregate statistics accurately capture changes in the tail index among the largest firms. However, findings in Cao et al. (2022) and Chen et al. (2023) based on micro-level firm data suggest that the thickening of the right tail persists even among firms in the highest percentiles. Specifically, Cao et al. (2022) estimate the tail index of the firm size distribution for the top 1% of firms (by employment) using U.S. Census Longitudinal Employer-Household Dynamics (LEHD) data, finding that the index is 1.17 in 1995 but only 0.99 in 2014. Similarly, Chen et al. (2023) analyze the tail index of the largest public firms in North America using Compustat, documenting a declining trend in the tail index from 1970 to 2018.[16]

In sum, these additional evidence indicate that the positive relationship between the right tail thickness and level of development remains consistent across

[14] Tybout (2000, 2014) and Hsieh and Olken (2014) provide opposing views on the existence of a "missing middle" among mid-sized firms in developing countries. Notably, Tybout (2014) shows that mid-sized firms represent a much smaller share of employment than expected if the underlying distribution were Pareto. Abreha et al. (2022) provide more recent evidence supporting the "missing middle" hypothesis using establishment-level censuses covering both formal and informal firms in Sub-Saharan Africa.

[15] One reason for focusing on OECD countries in the main empirical analysis is that my theoretical framework assumes away institutional distortions. Investigating the dynamic relationship between tail indices and economic development in developing countries, however, may present a fruitful direction for future research.

[16] They estimate the tail index annually from 1970 to 2019 using firm size by real revenue and show that this decline holds for firms with real revenues above 2500, 5000, and 7500 million dollars, as well as the top 500, 750, and 1000 firms.

sectors, countries, time periods, and among large firms. The robustness of this relationship suggests that the mechanism driving this phenomenon is generic to the growth process itself.

3 An Idea Search Growth Model

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

3.1 Model Setup

The economy consists of a continuum of firms, which are heterogeneous in their productivity. Firms produce homogeneous goods and engage in perfect competition in the market. The output good is the numeraire. In order to highlight the role of idea search, I make a number of assumptions to abstract away from other firms' decisions. Specifically, firms make no input decisions, and there are no entry or exit of firms. Each firm is endowed with one unit of capital, which is the sole input of production. The production function is such that the output is linear in firms' productivity and capital, i.e., $y = zk$. Capital does not depreciate, is specific to firms, and cannot be produced, so all firms use one unit of capital for production at all times. Then, a firm with productivity z produces z goods in each instant of time.[17] The output good is the numeraire and then always has a unit price. In the rest of the paper, I use productivity z to index firms.

Idea Search Absent input and entry/exit decisions, firms can only increase their profits by enhancing their productivity. They actively search for and learn from more productive firms in the economy through random meetings. The details of

[17] In this baseline model, productivity is isomorphic to firms' size since it is equal to firms' revenue (output), and there is no production labor employment (only researchers). In Section 7, I incorporate production labor to this model and show that the linearity between firms' revenue and employment can still be preserved.

the meeting and learning process are as follows. At each instant, firms endogenously choose the arrival rate of their meeting opportunities, which is a Poisson process. Search intensity $\eta(z,t)$ is then the arrival rate chosen by firms. Once a meeting takes place, the other firm in the meeting is drawn randomly among all firms that are more productive than the searching firm. Let $F(\cdot,t)$ denote the productivity distribution of firms at time t and $f(\cdot,t)$ the corresponding density. Then, the probability that the meeting firm has productivity y is given by $f(y|y \geq z,t)$. After the meeting, the searching firm upgrades its productivity to the same level as the meeting firm's. A key departure of this learning process from that in standard idea search models is that upon meetings, firms take random draws among firms that are more productive than they are, instead of the universe of firms in the economy. I then call this way of idea search as *truncated sampling*, as opposed to the *random sampling* in the literature.[18] I will later discuss in depth the plausibility and the implications of this assumption.

Idea search is costly. Firms need to hire researchers to conduct the search and upgrade the technology. Moreover, more productive firms need more researchers to achieve the same level of search intensity. This reflects that it takes more efforts to absorb more sophisticated ideas, since more productive firms expect to obtain more sophisticated technology in each search. Besides, a higher search intensity also requires more researchers. Therefore, the cost function is such that firm z has to hire $z\eta$ researchers to achieve an arrival rate of η. It pays a search (gross of adoption) cost $z\eta w(t)$ given researcher's wage $w(t)$.[19]

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

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

[18] I thank an anonymous referee for suggesting these terminologies.

[19] The assumption that the unit search cost scales with firm size follows the tradition in firm-based growth models, such as Klette and Kortum (2004), Atkeson and Burstein (2010), and Peters (2020). In these models, innovation costs are assumed to increase proportionally with firm size, effectively neutralizing the impact of size on the innovation process. This condition ensures that R&D intensity is independent of firm size, thereby preserving Gibrat's law. Similarly, the analogous assumption on search costs in my model ensures that search intensity is independent of firm size, and that Gibrat's law holds. In the context of idea search, I also provide a possible micro-foundation in Supplemental Appendix F to motivate this assumption.

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

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

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

$$r(t)v(z,t) = z + \max_{\eta \geq 0} \left\{ \eta \int_z^\infty [v(x,t) - v(z,t)] \mathrm{d}F(x|x \geq z, t) - z\eta w(t) \right\} + \partial_t v(z,t). \quad (3)$$

Equation (3) says that the flow value of the firm (the LHS) is the sum of four terms on the RHS: the flow profit (the first term), the total expected gains from learning (the second term), the total search and adoption cost (the third term), and the option value due to changes in the aggregate state (the fourth term).

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

$$\begin{aligned} &\max_{c(\tau) \geq 0} \quad \int_t^\infty e^{-\rho(\tau - t)} \frac{c(\tau)^{1-\theta} - 1}{1 - \theta} \mathrm{d}\tau. \\ &\text{s.t.} \quad \int_t^\infty e^{-\int_t^s r(\tau)\mathrm{d}\tau} c(s)\mathrm{d}s \leq \int_t^\infty e^{-\int_t^s r(\tau)\mathrm{d}\tau} y(s)\mathrm{d}s. \end{aligned} \quad (4)$$

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

$$C(t) = Y(t) \iff Lc(t) = \int_0^\infty z f(z,t) \mathrm{d}z. \quad (5)$$

Similarly, labor market clears at time t if

$$\int z\eta(z,t)f(z,t)\mathrm{d}z = L, \tag{6}$$

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

3.2 Equilibrium Path

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

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

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

$$\frac{\partial F(z,t)}{\partial t} = -\int_0^z \eta(x,t)(1-F(z|z\geq x,t))\mathrm{d}F(x,t). \tag{7}$$

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

The initial productivity distributions plays a significant role in the transition dynamics. Standard notion of tail index applies only to thick-tailed distributions.[20] Then, the initial productivity distribution is restricted to the class of

[20]See Section 6.2 for a definition of the tail index for general probability distributions.

thick-tailed distributions in order to obtain meaningful changes in the value of tail index.[21] Furthermore, I impose Assumption 1 that the initial productivity distribution is exactly Pareto. With this assumption, I am able to obtain a closed-form characterization of the complete transition dynamics.

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

$$F(z,0) = 1 - z^{-k_0} \tag{8}$$

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

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

Definition (Equilibrium)**.** A competitive equilibrium for this economy consists of wages $w(t)$, interest rates $r(t)$, firms' value functions $v(z,t)$, firms' search intensity $\eta(z,t)$, households' income $y(t)$ and 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), and $\eta(z,t)$ is the associated policy function;

(ii) Given $\{y(t), r(t)\}$, $c(t)$ solves the households' problem (4);

(iii) Both goods and labor markets clear, i.e., (5) and (6) are satisfied;

(iv) $F(z,t)$ solves the KFE (7) given $\eta(z,t)$ and satisfies the initial condition (8).

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

[21] Alternatively, in the presence of external learning, the initial productivity distribution could have a thin tail or even bounded support. In Supplemental Appendix K, I provide an example where firms can learn both internally from other firms in the economy and externally from an independent source of ideas. With a continuum of firms, the resulting productivity distribution immediately inherits the thicker tail of the two distributions—the initial firm productivity distribution and the external idea source distribution. Consequently, the initial productivity distribution could be bounded if ideas from the external source follow a thick-tailed distribution. Moreover, the right tail of the productivity distribution thickens over time, and changes in its tail index are solely driven by the rate of internal learning. In other words, the tail dynamics in this setting with both internal and external learning are identical to those in a model with only internal learning and an initially thick-tailed productivity distribution. This example suggests that focusing exclusively on internal learning with an initial thick-tailed productivity distribution may entail little loss of generality.

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

$$\lim_{t\to\infty} \frac{\dot{y}(t)}{y(t)} = g > 0.$$

3.3 Discussion on the search assumption

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

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

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

A more subtle and technical point about Gibrat's law is that it disciplines a wide class of idea search models. I elaborate on this point in Section 6 of Chen (2023) by showing that among those models, Gibrat's law holds if and only if the

right tail of the firm size distribution becomes thicker. With additional functional form assumptions, my model turns out to be the only one that is consistent with Gibrat's law and a thickening right tail.

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

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

Potential Mechanisms The reduced-form assumption that firms exclusively search among more productive firms can be intuitively understood through various mechanisms. Firstly, the availability of abundant public information allows firms to easily discern which firms are more productive than others. This information can include indicators such as market capitalization, employment size, number of patents, third-party rankings, and more. In a broader sense, it is common for individuals or organizations to focus their efforts on learning from peers who

are known to be superior in the profession. An analogous example in the realm of research is the organization of seminars. Researchers often invite more established speakers from higher-ranked institutions.

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

4 Transition Dynamics

In this section, I provide the main theoretical results of the paper. Specifically, I obtain a fully analytical and explicit characterization of the evolution of equilibrium firm size distributions. Based on these results, I explore the properties of firm growth and output growth along the transition path. All proofs are relegated to Appendix B.

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

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

This first result comes from the linearity of the HJB equation (3), which implies the gains from each search must be non-positive for all firms. Otherwise, firms will demand an infinite amount of labor. In addition, it is independent of the initial conditions on the productivity distribution. Nevertheless, assuming a Pareto initial distribution yields significant gains in tractability. As suggested by

the proposition below, the greatest convenience of having a Pareto initial distribution is that the productivity distribution remains Pareto along the equilibrium path. Consequently, the Pareto parameters of the distribution are sufficient to characterize the aggregate state.

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

(i) The search intensity is invariant with regard to productivity, i.e., $\eta(z,t) = \eta(t) > 0$*;*

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

$$F(z,t) = 1 - z^{-k(t)} \quad \text{for} \quad z \geq 1, \tag{9}$$

in which $k(t)$ *satisfies*

$$\frac{\dot{k}(t)}{k(t)} = -\eta(t) < 0, \tag{10}$$

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

This proposition states that despite heterogeneity in productivity, all firms search for ideas at the same intensity. The resulting productivity distribution is always a Pareto distribution, in which the scale remains constant, while the shape varies over time. Furthermore, the growth rate of the shape parameter is governed solely by the uniform search intensity. Proposition 1 delivers two important implications on the right tail thickness and firms' individual growth, respectively.[22]

A thickening right tail In line with previous empirical findings, the model predicts a positive relationship between the right tail thickness of the firm size distribution and output per capita. With Pareto distributions, the shape parameter k is also the tail index. Based on Equation (10), it follows that with positive $\eta(t)$, the tail index $k(t)$ exhibits a strict decrease at the given moment, resulting in a thicker right tail. That all firms search at a positive intensity ($\eta(t) > 0$) is also guaranteed in equilibrium as the labor market clears. Additionally, in the absence of population growth, the growth in output per capita is simply the growth in

[22]It should be noted that Proposition 1 describes a specific equilibrium of the model. The assumption of linear search cost allows for the existence of other equilibria. However, I discuss in Supplemental Appendix H that this particular equilibrium is preferred because it emerges as the limiting case of a broad class of equilibria with non-linear search costs. Each of these equilibria retains the properties described in Propositions 1 and 2.

aggregate output, which is the sum of all firms' output. Therefore, the aggregate output can be expressed as the mean productivity of firms, given a unit measure of firms and the production function. Due to the Pareto nature of the productivity distribution, the mean productivity is given by $\frac{k(t)}{k(t)-1}$, which decreases with respect to $k(t)$ and consequently grows over time. Collectively, the model suggests that the process of idea search can lead to a thickening right tail along the growth path since it generates growth by increasing the proportion of relatively more productive firms.

Gibrat's law The model implies that firms always have the same expected output growth, i.e., Gibrat's law holds. While Gibrat's law is a well-known empirical regularity of firm growth, existing idea search models, in general, are incompatible with it.[23] I demonstrate in below that Gibrat's law is consistent with this model.

In the current setting, the output growth of firms is equivalent to their productivity growth, which is jointly determined by the number and quality of firms' idea searches. On one hand, Lemma 1 shows that gains from each search are proportionate to firm productivity when the productivity distribution follows a Pareto distribution. This result leads to the first part of Proposition 1: all firms are indifferent to any level of search intensity in equilibrium and can agree upon the same search intensity.

On the other hand, firms engage in random draws during realized meetings, and the distribution of productivity growth from each draw remains the same for all firms. This property stems from the fact that Pareto distributions are preserved under left truncation. Let $\tilde{x}(z,t)$ represent the productivity growth of firm z in each meeting, i.e., $\tilde{x}(z,t) = \frac{\tilde{X}(z,t)}{z}$. It can be further demonstrated that $\text{Prob}(\tilde{x}(z,t) \leq y) = F(y,t)$. This idea search process has a similar implication to the traditional assumption in quality ladder models, where the size of quality improvement is independent of the current quality.

In summary, the expected productivity growth of firm z is the product of the search intensity and the expected productivity growth per search. Since both components are independent of firm productivity, the expected productivity growth is the same across firms and can be expressed as:

$$\lambda(z,t) \equiv \frac{\mathbb{E}[\mathrm{d}z]}{z} = \eta\left(\mathbb{E}[\tilde{x}(z,t)] - 1\right) = -\frac{\dot{k}(t)}{k(t)\left(k(t)-1\right)}, \tag{11}$$

[23] In Section 6 of the working paper Chen (2023), I provide a detailed discussion on why existing idea search models fail to obtain Gibrat's law.

where the last equality is obtained using equation (10).

With Proposition 1, it is sufficient to characterize the equilibrium path by the trajectory of the Pareto shape parameter. Equation (10) outlines a differential equation on $k(t)$ given the trajectory of $\eta(t)$. Besides, the labor market clearing condition (6) pins down the search intensity $\eta(t)$ as a function of the aggregate state. Integrating out the LHS of (6) using equation (9), I obtain that

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

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

$$-\frac{\dot{k}(t)}{k(t)-1} = L. \tag{12}$$

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

$$k(t) = 1 + (k_0 - 1)e^{-Lt}. \tag{13}$$

Based on the dynamics of the aggregate state given by equation (13), it is straightforward to derive the equilibrium trajectories of other variables. I then obtain the following two asymptotic results. They demonstrate that the model economy converges to a state in which firms' productivity follows Zipf's distribution, and the output per capita grows at a constant rate.

Zipf's law An immediate consequence of equation (13) is that $k(t)$ converges to 1. In other words, the productivity distribution converges to Zipf's distribution, which is a Pareto distribution with shape parameter 1. The so-called Zipf's law is a form taken by a remarkable number of regularities in economics and finance. If X denotes the variable of interest, Zipf's law states that $P(X > x) \approx 1/x$. In the context of firms, Zipf's law states that advanced economies such as the U.S. have firm size distributions very close to Zipf's distribution. This model then provides a growth-based theory of Zipf's law on firm size distributions. The process of idea search generates aggregate productivity growth by thickening the right tail of the productivity distribution. Rich countries are further displaced along the growth path, so their distributions are closer to the limit, which is Zipf's distribution. Additionally, it is worth noting that there are barely any restrictions on the model

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

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

$$g(t) \equiv \frac{\dot{y}(t)}{y(t)} = \underbrace{\frac{1}{k(t)}}_{\to 1} \underbrace{\left(-\frac{\dot{k(t)}}{k(t)-1}\right)}_{=L} = \frac{L}{k(t)} \implies g(t) \to g^* \equiv L, \tag{14}$$

i.e., $g(t)$ converges to the long-run growth rate g^* as $k(t) \to 1$. Therefore, the equilibrium path is an asymptotic balanced growth path, and the long-run growth rate is the average employment of researchers across firms.[24]

Two notable features of this asymptotic balanced growth path are worth emphasizing. First, the product market concentration rises steadily in an economy with nearly constant growth. Specifically, we may measure the product market concentration by a transformed right tail thickness measure, $k - 1$. The equilibrium trajectory of $k(t)$ implies that $k - 1$ decreases to zero at a constant rate. In addition, commonly-used concentration measures such as HHI decrease on k, given that the underlying distribution of firms' sales is Pareto. This aspect of the model is particularly noteworthy in that it reconciles the tension between rising market concentration and constant growth. In standard growth models, the firm size distribution remains stationary along an exactly balanced growth path. However, recent studies provide mounting evidence that market concentration has been increasing in the U.S.—an economy characterized by a stable long-term growth rate—even over extended time horizons.

Second, equation (14) reveals a growth mechanism that distinguishes this model from existing idea search growth models in the literature. In existing models, idea search processes enable the existence of balanced growth paths (BGPs)

[24]It may be tempting to conclude scale effects on the long-run growth rate, given that L presents the population size. However, the total employment of researchers coincides with the average employment per firm only because there is no firm entry, and the total measure of firms is normalized to one. In Section 7 and Appendix D, I present an extended model that incorporates population growth and firm entry to clarify this point explicitly. Since the measure of firms grows at the same rate as the population, scale effects remain absent in determining the long-run growth rate.

with stationary productivity distributions. Therefore, growth is sustained without affecting the productivity distribution. Namely, the relative weight between low-productivity and high-productivity firms remains constant. In contrast, the idea search process here increases aggregate output by relocating mass from relatively low-productivity firms to relatively high-productivity firms. This process, referred to as *tail growth*, sustains growth by thickening the right tail of the productivity distribution. In this context, Zipf's distribution emerges as a natural limit, as it represents the minimum level of thickness required for the mean productivity to approach infinity. Conversely, constant growth implies that the distribution must converge to Zipf's distribution; otherwise, growth will eventually cease.[25]

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

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

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

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

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

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

The parametric condition $\rho > L(1-\theta)$ ensures that the representative households' utility is finite and the relevant transversality condition is satisfied. While Proposition 2 does not pertain to prices, I characterize the equilibrium dynamics of $w(t)$ and $r(t)$ in Supplemental Appendix G. I show that wage $w(t)$ increases relative to household's income $y(t)$ over time. In other words, the share of R&D expenditure to GDP increases along the transition path.

5 Testing the model predictions

Despite its parsimony, the model in Section 3 makes theoretical predictions that are empirically testable. Specifically, the simple model makes two strong predic-

[25] It is worth pointing out that constant growth is only a sufficient condition for convergence to a Zipf's distribution. Convergence remains possible even with a declining growth rate. Section 6.4 of Chen (2023) provides a detailed characterization of the relationship between Zipf's law and growth rates.

tions: 1) output per capita y is proportional to $k/(k-1)$, and 2) $k-1$ decreases to 0 at a constant rate over time t.

In this section, I test both predictions using U.S. data, which offers the longest available time series for the tail index. It is important to emphasize that while tail estimates significantly influences the evaluation of model predictions, its estimation is independent of the model itself. Therefore, disciplining the choice of tail estimator is essential to ensure that appropriate values of model moments are being confronted with the data.

To this end, I use the Toda-Wang estimates k^{TW} for the tail index. Comparing to the simple estimator, the Toda-Wang estimator is more efficient since it utilizes all firm and employment shares in the right tail. In other words, k^{TW} represents a better "calibrated" model in terms of matching the right tail of the firm size distribution. In the following exercises, I present figures that compare model-based transformations of k^{TW} with either output per capita or time. Additionally, I report the respective correlation coefficients to quantify the alignment between the model predictions and the data.[26]

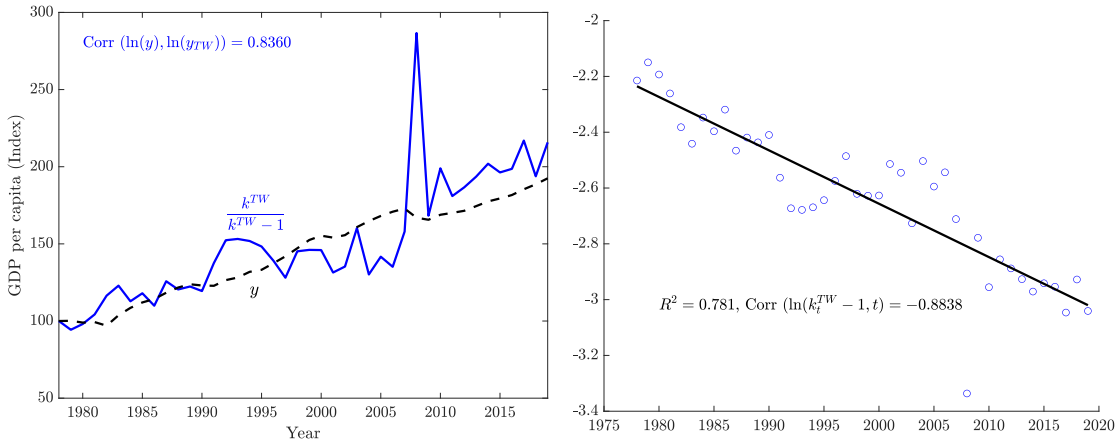
y vs. $k/(k-1)$ The model yields a simple relationship between output per capita and the tail index of the firm size distribution, expressed as:

$$\frac{y(t)}{y(t_0)} = \frac{k(t)}{k(t)-1} \Big/ \frac{k(t_0)}{k(t_0)-1}, \tag{15}$$

in which $k(t)$ is the Pareto shape parameter at time t, and t_0 is a reference date for normalization. Given estimates for $y(t)$ and $k(t)$, this equation allows for a direct comparison between the left-hand side and the right-hand side to assess whether they align.

In Figure 3(a), I present a comparison between y and $k/(k-1)$ in the U.S. over the period 1978 to 2019. y is measured by the U.S. real GDP per capita, taken from the Federal Reserve Economic Data (FRED). To facilitate comparison, both series are normalized to their respective 1978 levels, with an index value of 100 assigned to that year. The results reveal a remarkably close alignment between the two variables over the entire period. This strong relationship is further validated

[26]The correlation coefficient provides a normalized measure that controls for the variance of each variable, making it less sensitive to the choice of tail estimators compared to direct comparisons in figures. As a result, correlation coefficients present more robust results. This distinction is clearly demonstrated in Supplemental Appendix E.3, where I report the results of the same two exercises using the simple estimators k^f and k^{emp}. While direct comparisons may vary across tail estimators, the correlation coefficients remain highly similar and consistently high.



(a) y vs. $k/(k-1)$ (b) $\ln(k-1)$ vs. t

Figure 3: Testing the model predictions

Notes. This figure confronts two model predictions using tail estimates k^{TW} in the U.S. from 1978 to 2019. The left panel compares the relationship between y and $k/(k-1)$. Both data series are in levels and normalized with respect to their 1978 level. y is given by the real U.S. GDP per capita from FRED. The right panel compares the time series $\ln(k_t - 1)$ with its associated linear trend. Both the linear trend and R^2 are obtained from the regression $\ln(k_t - 1) = -2.24 - 0.019t$. Source: the U.S. BDS and FRED.

by the 0.8360 correlation coefficient between $\ln y$ and $\ln k/(k-1)$.

$\ln(k-1)$ vs. t Equation (12) in Section 4 makes it transparent that $k-1$ decreases to zero at a constant rate, indicating that $\ln(k_t - 1)$ decreases linearly with time (t). This implies that a transformation of the time series k_t in Figure 2 should closely follow a linear trend. In Figure 3(b), I present the transformed time series $\ln(k_t - 1)$ alongside its linear trend. The results demonstrate that $\ln(k_t - 1)$ aligns closely with the linear trend, with an R^2 value of 0.781 and a correlation coefficient of -0.8838.

6 Policy Implications

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

more searches by large firms.

Two policy exercises shed light on this policy implication. In the first exercise, I consider a scenario where a social planner selects a productivity threshold and implements a tax policy such that only firms above the threshold conduct search in the equilibrium. I prove that the equilibrium long-run growth rate increases with the level of productivity threshold. In the second exercise, I solve the social planner's problem and show that the optimal individual search intensity grows with individual productivity, following an approximate power rate function. Both exercises demonstrate that policies favoring large firms better capture the diffusion externality and improve welfare. Appendix C contains the omitted proofs and derivations.

6.1 Taxing small firms' idea search

In the first exercise, the policymaker chooses a threshold z^* and imposes a positive searching tax τ on firms below the threshold. Namely, if a firm has productivity less than z^*, its unit search cost becomes $z(1+\tau)w(t)$ rather than the original $zw(t)$. On the contrary, firms with productivity above that threshold retain the original unit search cost $zw(t)$. The tax revenue is further rebated to the households as a lump sum transfer. The following proposition characterizes the equilibrium with tax $\tau > 0$ and threshold z^*.

Proposition 3. *For any threshold productivity $z^* \geq 1$, there exists an equilibrium such that the equilibrium productivity distribution satisfies*

$$F(z,t) = \begin{cases} 1 - z^{-k_0} & \text{if } z \leq z^*, \\ 1 - (z^*)^{k(t)-k_0} z^{-k(t)} & \text{if } z > z^*, \end{cases} \tag{16}$$

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. Accordingly, the level of tax rate does not matter for the equilibrium. The other part of the distribution evolves just as before, as firms above the threshold still search at the same intensity. The only difference is that with fewer firms searching, each searching firm gets to hire more researchers and increases search intensity, accelerating output growth. Overall, taxing small firms raises the effective labor endowment per firm

by discouraging small firms' inefficient use of labor. Therefore, the policymaker can increase the long-run growth rate $L(z^*)^{k_0-1}$ by choosing higher thresholds z^*.[27]

6.2 The planner's problem

In the second policy exercise, I consider the following problem in which a utilitarian social planner maximizes the present value of the social welfare given an initial productivity density function f:

$$\begin{aligned} W(f) = \max_{\{c(t),\eta(z,t)\}} & \int_0^\infty e^{-\rho t} L u(c(t)) \mathrm{d}t \\ \text{s.t.} \quad & Lc(t) \leq \int_0^\infty z f(z,t) \mathrm{d}z, \\ & \int_0^\infty z\eta(z,t) f(z,t) \mathrm{d}z \leq L, \\ & \frac{\partial f(z,t)}{\partial t} = f(z,t) \left[\int_0^z \eta(x,t) \frac{f(x,t)}{1-F(x,t)} \mathrm{d}x - \eta(z,t) \right], \\ & f(z,0) = f(z). \end{aligned} \tag{17}$$

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

The optimal control problem (17) is challenging because the state variable is an infinite dimensional object–a distribution. Lucas and Moll (2014) and Nuño and Moll (2018) study similar planner's problem in heterogeneous agent models. 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. Following their strategy, I characterize the optimal search intensity $\eta(z,t)$ in the proposition below. I first present a definition of tail index for general distribution functions.

Definition (Regular Variation)**.** Let f be a positive measurable function, defined

[27] Proposition 3 also illustrates the multiplicity of equilibrium arising from the linear search cost. As shown in the proof, an equilibrium characterized by the distribution in (16) can be supported in the baseline economy ($\tau = 0$) for any threshold. More precisely, the model in Section 3 admits a continuum of equilibria indexed by the threshold $\underline{z} \in [1, +\infty)$. The equilibrium described in Proposition 1 corresponds to the case with the lowest threshold, $\underline{z} = 1$, which is favored for reasons discussed in Supplemental Appendix H. When a positive tax is imposed on firms below z^*, the model supports only a subset of threshold equilibria with $\underline{z} \in [z^*, +\infty)$, and Proposition 3 focuses on the equilibrium where $\underline{z} = z^*$.

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).

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

Proposition 4. *Given that $F(z,t)$ has tail index $k(t) > 1$, 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 4 indicates that under optimal search strategy, search intensity is an approximate power function of firm productivity. To illustrate this further, Figure 4 visualizes the comparison between search intensities in the competitive equilibrium and the planner's problem at time 0. With Assumption 1, Proposition 4 implies that the optimal search intensity $\eta^*(z,0) = Cz^{\frac{k_0-1}{2}}$ for some constant C. The labor market clearing condition pins down this constant to be $\frac{(k_0-1)L}{2k_0}$. Contrastingly, all firms search at the same intensity $\eta(z,0) = \frac{k_0-1}{k_0}L$ in the competitive equilibrium, as shown in Section 4. It is obvious that relative to the market outcome, a social planner would reallocate researchers from low-productivity firms to high-productivity firms.

A notable 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 $k(t) > 1$, its tail index drops instantly from $k(t)$ to $\frac{k(t)+1}{2}$, once the optimal search policy is in place. To see this, consider an infinitesimal time interval $[t, t+h]$. The law of motion of the productivity distribution implies that

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

in which $\tilde{L}(z,t) = L(z,t)zf(z,t)/\tilde{F}(z,t)$ and is slow varying. The integral is regularly varying with exponent $\frac{k(t)-1}{2}$.[28] $\tilde{F}(z,t+h)$ is then a regularly varying

[28]This result follows directly from Karamata's theorem in the theory of regular variation (see

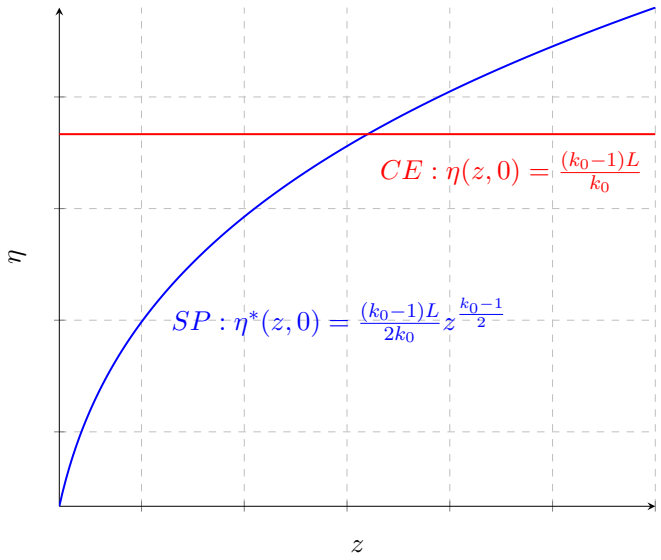


Figure 4: An illustration of search intensity comparison

Notes. This figure illustrates the difference between search intensity under the optimal policy and in the equilibrium. The blue line (SP) plots the optimal search intensity $\eta^*(z,0)$, and the red line (CE) the equilibrium search intensity $\eta(z,0)$.

function with exponent $-k(t) + \frac{k(t)-1}{2} = \frac{-k(t)-1}{2}$ for arbitrary h. Therefore, there is a jump in the tail index from $k(t)$ to $\frac{k(t)+1}{2}$. If $k(t) = 1$, there will be no changes in the tail index since the integral term is also slowly varying.

Given an initial tail index $k_0 > 1$, the sequence $\left\{1 + \frac{k_0-1}{2^n}\right\}$ characterizes the dynamics in the tail index under the optimal search policy. With countably many jumps in continuous time, the tail index immediately becomes one, and the output jumps into infinity.[29] Figure A.4 illustrates this process in the appendix. Albeit unusual, this result demonstrates the potential of the tail growth mechanism.

6.3 Discussion

Solving the social planner's problem provides a transparent illustration on the benefits of favoring large firms when searches are directed towards more productive firms. Nevertheless, it yields a growth explosion: the tail growth mechanism is so powerful such that it is possible for the social planner to achieve infinite output in the economy instantly. In the following discussion, I identify the assumptions underlying this unusual feature and propose a potential remedy.

Theorem 1.5.11 in Bingham et al. (1987)).

[29] If the output is finite for an interval of time, the social planner can always immediately reduce the tail index to below one (e.g. follow the equilibrium strategy) while respecting the labor market clearing condition. A tail index below one implies an infinite mean of the distribution and hence infinite output. It then constitutes a violation to the optimality of the policy.

One deviation from existing idea search models is the absence of an technological upper bound on how much information a firm can process in each instant.[30] Consequently, it is possible for infinitely productive firms to process infinitely many ideas immediately, resulting into jumps in the tail index.

However, imposing a boundedness condition alone is insufficient to prevent growth explosions. Suppose that $\bar{\eta}$ represents the maximum number of ideas a firm can process instantaneously. The arrival rate of ideas for a firm z is then given by $\eta(z,t) = \max\{l(z,t)/z, \bar{\eta}\}$, where $l(z,t)$ denotes the number of researchers hired by the firm. Evidently, the threshold equilibrium described in Section 6.1 remains feasible, as $\eta(t)$ declines to zero, eventually falling below $\bar{\eta}$ along the equilibrium path. To achieve even faster growth, the social planner can improve this strategy by selecting a time-dependent threshold $z^*(t)$ such that only firms with $z > z^*(t)$ search, and they always search at the maximum intensity $\bar{\eta}$. I leave a complete analysis in Appendix C.3 to show that given such strategy, the right tail of the productivity distribution is still Pareto, and the tail index $k(t)$ decreases at a constant rate $-\bar{\eta}$, i.e.,

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

Hence, $k(t)$ reaches unity at time $T = \ln(k_0)/\bar{\eta}$, at which point output also becomes infinite. In other words, the boundedness condition merely extends the time for the output to reach infinity from zero to a finite value.

This example illustrates well the nature of the growth explosion: whenever the search technology makes it feasible to vary the tail index, one could thicken the right tail as fast as possible to achieve higher growth. Conversely, this anomaly vanishes if the tail index remains constant for any feasible search strategy, as in the model of Lucas and Moll (2014).

I then consider the following generalization of the model in Section 3. Assume the same search process and cost as before except that adoption now is not guaranteed upon meeting. Specifically, I assume that conditional on meeting each other, the probability for a searching firm x to adopt the technology of the meeting firm y $(y > x)$ is given by

$$\left(\frac{1 - F(y,t)}{1 - F(x,t)}\right)^{\lambda},$$

[30] For example, models in Lucas and Moll (2014) and Perla and Tonetti (2014) assume endogenous technology adoption with bounded arrival rate of ideas.

where $\lambda \geq 0$. The baseline model is the case with $\lambda = 0$. In contrast, it becomes increasingly difficult for firms to adopt a higher-ranked technology if $\lambda > 0$.

The generalized model with $\lambda > 0$ effectively eliminates growth explosions in the planner's problem while closely approximating the equilibrium properties of the baseline model. I summarize two key results here, and a complete analysis can be found in Supplementary Appendix I. First, I prove that in any generalized model with $\lambda > 0$, given an upper bound $\bar{\eta}$ on the arrival rate of ideas, the tail index remains constant under any search strategy feasible for a social planner.[31] Consequently, the total output is always finite, and growth explosions never occur.

Second, the generalized model converges to the baseline model as λ approaches zero. For sufficiently small values of λ, the generalized model provides a close approximation of the baseline model. Numerically, I compare the equilibrium dynamics of the baseline model with those of the generalized model, setting $\lambda = 10^{-4}$. Under plausible parameter values, the evolution of the productivity distribution and the measured tail indices over a 400-year horizon are nearly indistinguishable between the two models.[32]

7 Extensions

The absence of four common elements in growth models makes the baseline model in Section 3 particularly tractable and its implications especially sharp. First, there are no input decisions for production. Second, there is no firm entry or exit, and the measure of firms is given exogenously. Third, there is no technological progress. Fourth, there is no population growth. Therefore, it is crucial to verify that the analytical results do not hinge on these simplifying assumptions.

In this section, I present an extended model that incorporates all these elements to demonstrate that the model's tractability is preserved. Moreover, the extended model not only retains all the key properties of the baseline model but also clarifies and generalizes its insights. In Section 7.1, I discuss the incorporation of additional elements into the extended model. In Section 7.2, I explain how these additional elements maintain and sharpen the insights derived from the baseline model. A

[31] I also demonstrate in Supplementary Appendix I that the boundedness condition is necessary. Without an upper bound on $\eta(z,t)$, the social planner can implement similar policies as in Section 6.2 in the generalized model with $\lambda > 0$ to thicken the right tail immediately.

[32] The measured tail index follows the construction of k^f, targeting the top 10% and 0.5% of firms. Thus, it differs from the tail index defined in terms of regular variation, as in Section 6.2. It should be noted that when $\lambda > 0$, idea search cannot sustain exponential growth indefinitely with the absence of tail growth. However, the numerical example shows that it can still approximate exponential growth over an extended period. See Supplemental Appendix I for further discussion.

detailed exposition of the full-blown model and equilibrium analysis is contained in Appendix D.

7.1 Additional elements

I extend the baseline model with input decisions, entrepreneurial choices, technological obsolescence, and population growth.

Input decisions The baseline model assumes homogeneous goods and perfect competition across firms. In there, firms' output y is given by their productivity without any flexible inputs. In contrast, I assume in the extend model that each firm produces a differentiated variety and engages in monopolistic competition, and households' intra-temporal preference over these varieties have a constant elasticity of substitution σ. Firms hire production workers l_p and have linear production functions $y = \tilde{z} l_p$. Their optimal pricing rule are then to charge a constant markup $\sigma/(\sigma - 1)$. Consequently, firms' profit π, revenue r, and employment of production worker l_p are all linear in $\tilde{z}^{\sigma-1}$. I then focus on the transformed productivity $z \equiv \tilde{z}^{\sigma-1}$ and its distribution $F(z,t)$, which also represents the employment distribution over firms.

Population growth The baseline model assumes a fixed measure L of households, whom supply labor for firms' idea search. In the extend model, the measure of households grows over time. I use $L(t)$ to denote its measure at time t, and $g_L(t)$ the associated growth rate. Then, $g_L(t) \geq 0$. Furthermore, each household consists of a unit continuum of members with s fraction of them being skilled labor. While unskilled labor can only engage in production, skilled labor can either work as researchers for incumbent firms or start up their own firms as entrepreneurs. In other words, households now supply unskilled labor for production and skilled labor for idea search or firm entry.

Technological obsolescence The baseline model assumes that firms can always upgrade their technology regardless of productivity. In the extended model, I introduce the obsolescence of technologies and assume that obsolete technologies cannot be upgraded or replicated. Consequently, incumbent firms using obsolete technologies cannot improve its productivity, and new firms cannot adopt obsolete technologies to enter. Specifically, there is an increasing productivity threshold $z^*(t)$ such that a technology becomes obsolete if its productivity z falls below $z^*(t)$.

A growing $z^*(t)$ also implies an exogenous technological progress since entrants only adopt technologies with productivity above $z^*(t)$. This technological progress makes the distribution over productivity z an asymptotic traveling wave. I define $g_z(t)$ as the growth rate of $z^*(t)$, $T^*(z)$ the inverse of $z^*(t)$, and $F^*(t) \equiv F(z^*(t), t)$ the share of incumbent firms using obsolete technologies.

Entrepreneurial choices The baseline model assumes a fixed unit measure of firms. In the extended model, I endogenize the measure of firms $M(t)$ through the entry of firms, which results from skilled labor's entrepreneurial choices. Skilled labor differ in their innate managerial ability e, which has a cumulative distribution function J over support $[0, +\infty)$. They become entrepreneurs if the expected value of startup activities $v_E(e, t)$ exceeds the after-tax(subsidy) wage of researchers $(1 - \tau_r)w_r(t)$, where τ_r is the tax or subsidy rate on researchers' wage

Since obsolete technologies cannot be replicated, entrepreneurs search for ideas over incumbent firms using non-obsolete technologies. For them, completing a unit search requires the efforts of $\psi z^*(t)$ researchers.[33] In other words, entrepreneurs use the same search technology as firms with productivity $z^*(t)$. However, they cannot hire researchers at the startup stage but instead rely on their managerial ability to search for ideas. The expected return of a startup v_E is given by:

$$v_E(e, t) = \frac{e}{\psi z^*(t)} \int_{z^*(t)}^{\infty} v(x, t) \mathrm{d}F(x | x \geq z^*(t), t),$$

where $e/\psi z^*(t)$ gives the arrival rate of ideas to an entrepreneur e. Then, there exists a threshold $e^*(t)$ that divides skilled labor into entrepreneurs and researchers. I use $E(t)$ to denote the measure of entrants in time t.

7.2 Implications

Proposition 1 establishes the tractability of the baseline model by summarizing the productivity distribution in a single variable $k(t)$. For the extended model, I obtain the following characterization of the equilibrium productivity distribution akin to Proposition 1.

Proposition 5. *With Assumption 1 and $z^*(0) = 1$, there exists an equilibrium with the following properties:*

[33] The parameter ψ measures the efficiency of the search technology and is common to all firms, both incumbents and entrants. See Appendix D.1 for further details.

i) The search intensity is invariant across productivity among search active firms, i.e., $\eta(z,t) = \eta(t) > 0$ for $z \geq z^*(t)$;

ii) The equilibrium productivity distribution at time t is

$$F(z,t) = \begin{cases} 1 - (1 - F^*(t)) \left(\frac{z}{z^*(t)}\right)^{-k(t)}, & \text{if } z \geq z^*(t), \\ F^*(T^*(z)) e^{-\int_{T^*(z)}^{t} g_E(s) \mathrm{d}s}, & \text{if } z < z^*(t), \end{cases}$$

in which

$$(F^*)'(t) = k(t) g_z(t) \left(1 - F^*(t)\right) - g_E(t) F^*(t),$$

$g_E(t) = E(t)/M(t)$, *and*

$$\frac{\dot{k}(t)}{k(t)} = -\eta(t) < 0.$$

Proposition 5 shows that the equilibrium productivity distribution comprises of two parts, divided by $z^*(t)$. The part above $z^*(t)$ evolves in the same way as in the baseline model, exhibiting Pareto distributions at all times. The part below $z^*(t)$ is shaped by the race between technological obsolescence and firms' entry. Proposition 5 makes it clear that along the transition path, the right tail thickens over time, and Gibrat's law always holds for firms in the right tail.

It is worth noting that the characterization of $F(z,t)$ holds along arbitrary paths of $g_z(t) \geq 0$ and $g_L(t) \geq 0$. A sharper characterization of the long-run equilibrium is possible by imposing additional structure on these exogenous processes. Specifically, I assume that $\lim_{t\to\infty} g_L(t) = g_L^*$, $\lim_{t\to\infty} g_z(t) = g_z^* > 0$, and that there is positive firm entry in the long run. These conditions ensure that the right tail converges to Zipf's distribution and that the equilibrium path represents an asymptotic balanced growth path. Notably, the long-run growth rate of household income is given by

$$g_y^* = \frac{g_k^* + g_L^*}{\sigma - 1}, \quad \text{where} \quad g_k^* \equiv \lim_{t\to\infty} -\frac{\dot{k}(t)}{k(t) - 1} = \frac{J(e^*)}{K(e^*)}(g_L^* + \delta), \tag{18}$$

$e^* \equiv \lim_{t\to\infty} e^*(t) = 1 - \tau_r$, and $K(e) = \int_e^\infty x \mathrm{d}J(x)$. That is, two sources sustain long-run growth: the expansion in varieties, $g_L^*/(\sigma-1)$, and tail growth, $g_k^*/(\sigma-1)$. I have then verified that all properties outlined in Proposition 2 for the baseline model are preserved in this extended model.

Moreover, the extended model offers two new perspectives to investigate the nature of tail growth. First, compared to the baseline model, the extended model explicitly confirms the absence of scale effects on the long-run growth rate. In (18), g_k^* is also independent from the share of skilled labor s or the efficiency of the search technology ψ. Consequently, any expansion in the effective supply of skilled labor—whether through population growth (higher $L(t)$), improved search technology (lower ψ), or increased education (higher s)—is exactly offset by an increase in the number of firms, leaving the growth rate unchanged.

In contrast, g_k^* increases with the ratio of researchers to total entrepreneurial efforts, $J(e^*(t))/K(e^*(t))$. Subsidizing researchers ($\tau_r < 0$) can increase the growth rate by raising the average number of researchers per firm through a higher e^*. Thus, the extended model falls within the class of endogenous growth models without scale effects such as Young (1998).

Second, the extended model provides new insights into the relationship between the stability of the firm size distribution and economic growth. Firm-based growth models typically emphasize stationary firm size distributions, often focusing on a traveling wave equilibrium where the productivity distribution $F(z,t)$ is stationary after scaling. Formally, $F(z,t)$ is a traveling wave if there exists a scaling factor $\gamma(t)$ and a distribution $\Gamma(z)$ such that $F(z\gamma(t),t) = \Gamma(z)$ for all z and t. Let $Z(t)$ represent the mean of $F(z,t)$. In this case, $Z(t)$ always grows at the same rate as $\gamma(t)$.[34] An exact traveling wave equilibrium, therefore, eliminates the possibility of gains from reallocation.

However, if one assumes that the firm size distribution is only asymptotically stationary, the extended model presents a possibility to obtain persistent gains from reallocation in the long run. In Appendix D.4, I show that $F(z,t)$ in Proposition 5 is an asymptotic traveling wave: it converges to a stationary distribution Γ after being rescaled by $z^*(t)$, i.e., $\lim_{t\to\infty} F(rz^*(t),t) = \Gamma(r)$. While $z^*(t)$ grows at a rate g_z^* in the long run, $Z(t)$ has a long-run growth rate of $g_z^* + g_k^*$, where the additional growth g_k^* is driven by the thickening right tail. This result reinforces the idea in the baseline model that tail growth is a form of persistent reallocation gain.

[34] This follows from $Z(t) \equiv \int_0^\infty z, \mathrm{d}F(z,t) = \gamma(t) \int_0^\infty z, \mathrm{d}\Gamma(z)$.

8 Conclusions

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

The model in this paper not only provides an explicit theory for the thickening of the right tail along the growth path but also highlights that the reallocation gain from thicker tails will persist if the firm size distribution converges to a limiting distribution with a tail index equal to one. Meanwhile, given appropriate initial conditions, standard firm dynamics models such as Luttmer (2007) can, in principle, generate a firm size distribution with a thickening right tail during the transition dynamics. In these models, the firm size distribution converges to a stationary distribution with a tail index strictly greater than one, implying that tail growth vanishes in the long run. Given the distinct implications for growth, it would be an interesting exercise for future research to first compare these theories based on their ability to account for the historical decline in the tail index and then quantify the contribution of tail growth.

Finally, the central analysis of this paper is carried out in a starkly simple context in order to elucidate the economic forces and the nature of their interactions. That said, additional firm characteristics, such as market power and innovation, are crucial for a comprehensive design of size-dependent industrial policies. Exploring the policy implications of the diffusion mechanism within a richer firm-based growth framework remains a promising avenue for future research.

References

Abreha, K. G., X. Cirera, R. N. Fattal Jaef, H. B. Maemir, E. Davies, and H. B. Maemir (2022): *Deconstructing the Missing Middle: Informality*

and Growth of Firms in Sub-Saharan Africa, Policy Research Working Papers, The World Bank.

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.

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

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.

Ayyagari, M., A. Demirguc-Kunt, and V. Maksimovic (2014): "Who creates jobs in developing countries?" *Small Business Economics*, 43, 75–99.

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

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

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

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

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

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

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

Chen, Y.-F., W.-T. Hsu, and S.-K. Peng (2023): "Innovation, firm size distribution, and gains from trade," *Theoretical Economics*, 18, 341–380.

Chen, Z. (2023): "Economic Growth and the Rise of Large Firms," *Working Paper*.

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

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

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

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

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

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

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

Gouin-Bonenfant, E. (2022): "Productivity Dispersion, Between-Firm Competition, and the Labor Share," *Econometrica*, 90, 2755–2793.

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

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

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

Jones, C. I. (2023): "Recipes and Economic Growth: A Combinatorial March Down an Exponential Tail," *Journal of Political Economy*, 131, 1994–2031.

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

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

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

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

Kwon, S. Y., Y. Ma, and K. Zimmermann (2024): "100 Years of Rising Corporate Concentration," *American Economic Review*, 114, 2111–2140.

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

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

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

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

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

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

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

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

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

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

Peters, M. (2020): "Heterogeneous Markups, Growth, and Endogenous Misallocation," *Econometrica*, 88, 2037–2073.

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

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

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

Toda, A. A. (2017): "Zipf's Law: A Microfoundation," *SSRN Electronic Journal*.

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

Tybout, J. (2014): "The Missing Middle: Correspondence," *Journal of Economic Perspectives*, 28, 235–236.

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

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

Online Appendix

Appendix A Full results in Section 2

A.1 Additional results in Section 2.2

A.1.1 Full regression results for the OECD countries

Table A.1: Tail index and log GDP per capita in OECD countries

	k^{TW}	k^{TW}	k^{f}	k^{f}	k^{emp}	k^{emp}
Log GDP p.c.	-0.02531^{a}	-0.06079^{a}	-0.05926^{a}	-0.07669^{a}	-0.05598^{a}	-0.02793^{a}
	(0.009)	(0.018)	(0.017)	(0.015)	(0.010)	(0.009)
Cons.	1.3974^{a}	1.7726^{a}	1.7629^{a}	1.9473^{a}	1.8573^{a}	1.5608^{a}
	(0.101)	(0.186)	(0.185)	(0.163)	(0.103)	(0.093)
Country FE	No	Yes	No	Yes	No	Yes
k mean	1.130	1.130	1.136	1.136	1.266	1.266
Obs.	247	247	299	299	297	297
R^2	0.024	0.880	0.056	0.969	0.124	0.978

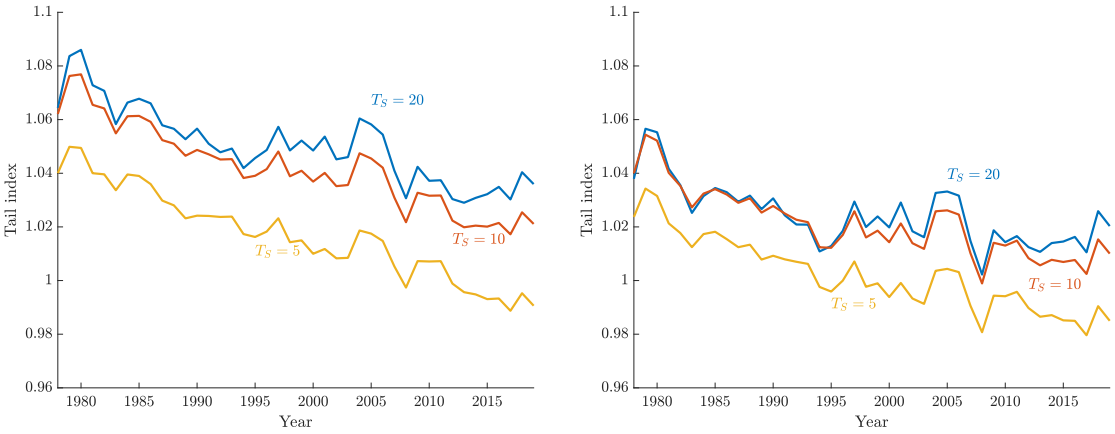
Notes This tables reports the results of regression (2) between the tail index k and log GDP per capita among OECD countries. Each observation is a country-year pair. The tail index are k^{TW} with $T = 10$, k^{f} and k^{emp} with $T_S = 10$ and $T_L = 250$. The log GDP per capita is in constant international dollars. Robust standard errors are in parentheses. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$. Sources: the OECD SBS and PWT 10.0

A.1.2 Alternative thresholds for the U.S. tail indices

A.2 Additional results in Section 2.3

A.2.1 Across sectors

In Table A.2, I present the results of regression (2) using sectoral k estimates as dependent variables. The findings confirm that both the manufacturing and service sectors exhibit thicker right tails as the economy grows. Notably, the sample size for the manufacturing sector is considerably smaller for k^{TW} (87) compared to k^{f} (294) and k^{emp} (295). This reduction arises because the tail indices for manufacturing firms are very close to 1, resulting in the constraint $k > 1.001$ frequently binding.



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

Figure A.1: Tail indices k^f in the U.S. (1978-2019) with alternative thresholds

Notes. This figure illustrates the tail index k of the size distribution for all U.S. business firms from 1978 to 2019. The tail indices are estimated using the simple estimator k^f, which is based on the top firm shares and the corresponding thresholds T_S and T_L as indicated in the labels. Data source: the U.S. BDS.

Table A.2: Tail index and log GDP per capita in OECD countries

	k^{TW}		k^f		k^{emp}	
	MFT	SEV	MFT	SEV	MFT	SEV
Log GDP p.c.	-0.06969^c	-0.1183^a	-0.1163^a	-0.1344^a	-0.05381^a	-0.03915^a
	(0.039)	(0.026)	(0.021)	(0.020)	(0.014)	(0.009)
Cons.	1.8096^a	2.4222^a	2.2067^a	2.6000^a	1.8150^a	1.6722^a
	(0.400)	(0.271)	(0.224)	(0.208)	(0.148)	(0.095)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
k mean	1.088	1.174	0.979	1.180	1.247	1.259
Obs.	87	275	294	297	295	295
R^2	0.850	0.700	0.971	0.952	0.976	0.975

Notes This tables reports the results of regression (2) between the tail index k and log GDP per capita among OECD countries by sectors. The tail index are k^{TW} with $T = 10$, k^f and k^{emp} with $T_S = 10$ and $T_L = 250$. They are estimated respectively using the size distributions of firms in the manufacturing (MFT) or service (SEV) sector. The log GDP per capita is in constant international dollars. Robust standard errors are in parentheses. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$. Sources: the OECD SBS and PWT 10.0

A.2.2 Across countries

In Table A.1, I show that the positive relationship between the right tail thickness and log GDP per capita still holds among OECD countries after removing country fixed effects. In the following, I examine the same cross-sectional relationship in developing countries using data from the World Bank Enterprise Survey (WBES).

The WBES is a collection of surveys conducted by the World Bank designed to provide a representative portray of a country's business economy. Specifically, I use the official aggregate statistics from the WBES—the Employment Indicators —to obtain employment shares across firms of different sizes. My sample covers over 130 countries between 2006 and 2019, of which 113 are classified as low-income, lower-middle-income, or upper-middle-income countries. Firms are grouped into three size categories based on the number of employees: 5–19 (small), 20–99 (medium), and 100+ (large). With $T_S = 5$ and $T_L = 100$, the employment-based simple estimator k^{emp} becomes a monotonic transformation of the employment share of large firms, consistent with the methodologies in Ayyagari et al. (2014) and García-Santana and Ramos (2015).[35]

In Figure A.2, I present the estimated tail index k^{emp} plotted against log GDP per capita for 113 low, lower-middle, and upper-middle-income countries included in the survey. The observed negative relationship corroborates the findings in Ayyagari et al. (2014) and García-Santana and Ramos (2015) that richer countries tend to exhibit thicker right tails. In Table A.3, I further show that this relationship remains robust to the inclusion of high-income countries or year fixed effects.

Notably, in Figure A.2, there remains substantial variation among countries with similar GDP per capita. In addition, many low-income countries exhibit very thick right tails. The R^2 for the cross-sectional regression of developing countries (country-year pairs) is only 0.036, considerably lower than that of the OECD countries, which stands at 0.124 (see Table A.1). This heterogeneity may be attributable to issues discussed in Section 2.3, such as measurement.

[35] Formally, $k^{emp} = 1 - \log(\text{employment share of large firms})/\log(100/5)$. The choice of $T_S = 5$ not only ensures the validity of the monotonic transformation, but also adequately captures the right tail in most developing countries. For example, Bento and Restuccia (2021) estimate the average establishment size for 91 countries using micro-level data on establishments and document a positive relationship between average establishment size and GDP per capita. According to their data, the average size of non-agricultural establishments is below five employees in 90% of countries with a log GDP per capita below 10.23, which is the highest value in my baseline WBES sample.

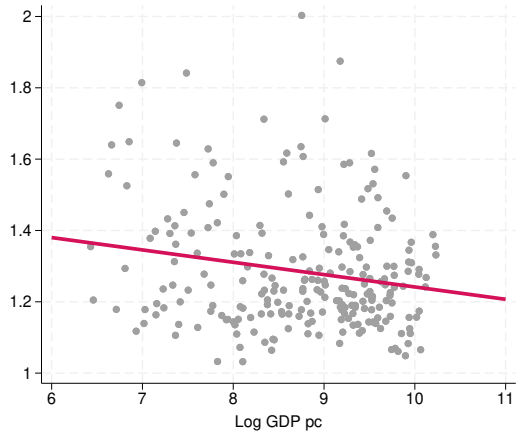


Figure A.2: Tail index and log GDP per capita in developing countries

Notes. This figure plots the tail index k^{emp} against the log GDP per capita for each developing country (country-year) in the WBES sample. The red line is the linear fit. The tail index k^{emp} is estimated using the simple estimator k^{emp} with $T_S = 5$ and $T_L = 100$. The log GDP per capita is in constant international dollars. Data sources: the WBES and PWT 10.0.

A.2.3 Across time

The primary data source used by Kwon et al. (2024) is the Statistics of Income (SOI) and the associated Corporation Source Book, both published annually by the IRS. The SOI was established under the Revenue Act of 1916, which mandates the IRS to report statistics derived from the tax returns filed each year. Kwon et al. (2024) digitized historical SOI publications and made the cleaned data series available online at http://businessconcentration.com. In particular, they provide data on the business share—by assets, receipts, or net income—held by the top 50%, 10%, 1%, and 0.1% of firms. Among these, the assets data series offers the most extensive time coverage (1931-2018), followed by receipts (1959-2018) and net income (1918-1974).

Consistent with the main analysis, I focus on the top 10% of firms to represent the right tail. Although the data series lack information on firm sizes, the formulas for the simple estimators k^f and k^{emp} suggest the following estimator based on the top firm and employment shares. By eliminating the term $\log T_L/T_S$, the formulas imply that

$$1 - \frac{1}{k} = \frac{\log \frac{E(T_L)}{E(T_S)}}{\log \frac{N(T_L)}{N(T_S)}} \equiv \frac{\log\left(\frac{q_L}{q_S}\right)}{\log\left(\frac{x_L}{x_S}\right)}, \tag{A.1}$$

where $x_i \equiv N(T_i)$ and $q_i \equiv E(T_i)$ for $i \in L, S$. In other words, I estimate k using

Table A.3: Tail index and log GDP per capita in developing countries

	k^{emp}	k^{emp}	k^{emp}	k^{emp}
Log GDP p.c.	-0.03461^a	-0.02567^b	-0.03794^a	-0.03164^a
	(0.013)	(0.010)	(0.013)	(0.011)
Cons.	1.5877^a	1.5140^a	1.6169^a	1.5674^a
	(0.113)	(0.096)	(0.119)	(0.105)
k mean	1.285	1.284	1.285	1.284
HI countries	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Obs.	232	270	232	270
R-sq	0.036	0.026	0.092	0.092

Notes This tables reports the results of regressing the tail index on log GDP per capita in developing countries. The tail index is k^{emp} with $T_S = 5$ and $T_L = 100$. The log GDP per capita is in constant international dollars. Robust standard errors are in parentheses. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$. Sources: the WBES and PWT 10.0.

equation (A.1) with the top firm shares x_L and x_S, and their corresponding top employment shares q_L and q_S.[36]

In Figure A.3, I plot the time series of the estimated tail index based on equation (A.1), using $x_S = 10\%$ and $x_L = 1\%$ or 0.1%. These results confirm that the right tail in the United States has been thickening over time. Furthermore, since k is estimated as a monotonic transformation of the ratio q_L/q_S, Figure A.3 serves as a tail index representation of Figure 2 in Kwon et al. (2024).

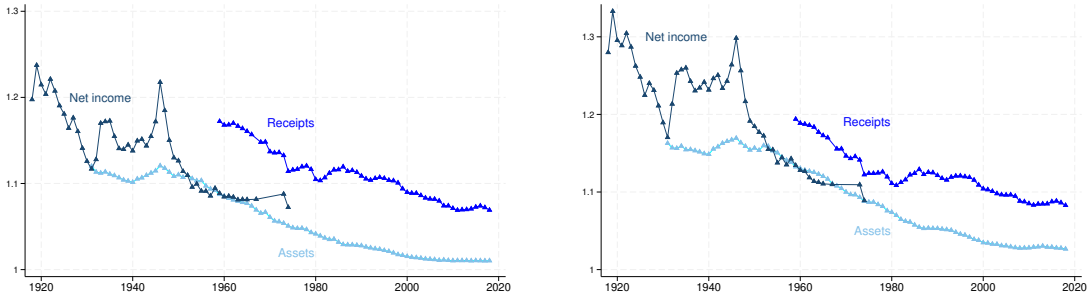
Appendix B Omitted Proofs in Section 4

B.1 Proof of Lemma 1

Proof. From the HJB equation (3), the gains from learning is linear in η. Hence, the gains from learning per unit arrival rate must be non-positive, otherwise firms will choose infinitely large η, i.e.,

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

[36] This simple estimator is widely used in the literature on income and wealth distributions where data report only top income percentiles and shares. In principle, the Toda-Wang estimator is also applicable with three thresholds. However, I find that the Toda-Wang estimator produces too many binding estimates ($k \approx 1.001$ for many observations).



(a) Top 1%

(b) Top 0.1%

Figure A.3: Tail index in the US (1918-2018)

Notes. This figure plots the 1918-2018 time series of the estimated tail index based on equation (A.1). I set $x_L = 1\%, x_S = 10\%$ for the left figure and $x_L = 0.1\%, x_S = 10\%$ for the right figure. Data source: Kwon et al. (2024).

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

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

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

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

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

B.2 Proof of Proposition 1

Proof. As in the proof of lemma 1, let $v(z,t) = v(t)z$, in which $v(t)$ is given by (A.2), and $w(t) = \frac{v(t)}{k(t)-1}$. I verify that $v(z,t)$ and $F(z,t)$ solve respectively firm's problem (3) and the law of motion on the productivity distribution (7). First, with $F(z,t)$ given by (9), 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) following the proof of lemma 1.

Second, with $\eta(z,t) = \eta(t)$, $F(z,t)$ solves the KFE (7). To see this, rewrite (7) 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 (10).

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

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

With $k(0) = k_0$, (10) 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. ■

Appendix C Omitted Proofs in Section 6

C.1 Proof of proposition 3

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

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

I show in the following that this search strategy is consistent with $\tau = 0$. Therefore, it holds trivially with $\tau > 0$ in which firms below the threshold are more discouraged to search. Proposition 3 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 $z > z^*$,

$$\begin{aligned} \tilde{F}(z,t+h) &= \tilde{F}(z,t) + \int_{z^*}^{z} \eta(x,t) h \frac{\tilde{F}(z,t)}{\tilde{F}(x,t)} f(x,t) \mathrm{d}x \\ \Rightarrow \frac{\partial \tilde{F}(z,t)}{\partial t} &= \tilde{F}(z,t) \int_{z^*}^{z} \eta(x,t) \frac{f(x,t)}{\tilde{F}(x,t)} \mathrm{d}x. \end{aligned} \tag{A.4}$$

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

To see that (16) satisfies all three conditions, it suffices to verify that it satisfies the law of motion (A.4). Note that for $z > z^*$,

$$\begin{aligned} \frac{\partial \tilde{F}(z,t)}{\partial t} \frac{1}{\tilde{F}(z,t)} &= \frac{\partial \ln \tilde{F}(z,t)}{\partial t} = -\dot{k}(t) \left(\ln z - \ln z^* \right), \\ \int_{z^*}^{z} \eta(x,t) \frac{f(x,t)}{\tilde{F}(x,t)} \mathrm{d}x &= \eta(t) \int_{z^*}^{z} \frac{k(t)}{x} \mathrm{d}x = \eta(t) k(t) \left(\ln z - \ln z^* \right). \end{aligned}$$

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

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

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

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

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 3. 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,

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

In equilibrium, we must have

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

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

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

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

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

it then suffices to compute the compare the value of the integral with $k(t)/(k(t)-$

1). For $z \leq z^*$,

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

The last equality uses that $(z^*/z)^{1-k_0} < 1$ and $k(t) < k_0$. For $z > z^*$, the truncated distribution $F(x|x \geq z, t)$ is exactly Pareto with shape parameter $k(t)$ and then has mean $zk(t)/(k(t)-1)$. Then, that all firms above the threshold search at the same intensity satisfies the optimality of firm's problem.

At this point, it is straightforward to see what a positive tax does. Let $\hat{z}$ be the threshold of a threshold equilibrium (16). With $\tau > 0$, the policy maker can choose the threshold z^* to eliminate threshold equilibria with $\hat{z} < z^*$.

■

C.2 Solving the planner's problem

To solve the planner's problem (17), I work with $w(f, z)$, which is the Gateaux derivative of $W(f)$ at point z with a Dirac delta function as increment.[37] $w(f, z)$ is then the marginal social value of a firm z with aggregate state f. Moreover, let $w(z, t) \equiv w(f(\cdot, t), z)$. $w(z, t)$ is then the marginal social value along the trajectory of the productivity distribution $f(z, t)$, which results from the optimal policies. I show that $w(z, t)$ satisfies the following partial differential equation and leave the

[37] 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.

derivations into Appendix C.2.2:

$$\rho w(z,t) = \hat{\lambda} z + \frac{\partial w(z,t)}{\partial t} + \max_{\eta} \left\{ \eta \int_z^{\infty} [w(y,t) - w(z,t)] \frac{f(y,t)}{1-F(z,t)} \mathrm{d}y - \hat{\mu} z \eta \right\}$$
$$+ \int_0^{\infty} \left\{ w(y,t) \int_0^{\min\{y,z\}} \eta^*(x,t) \frac{-\varphi(x,t)}{1-F(x,t)} \mathrm{d}x + w(z,t) \int_0^z \eta^*(x,t)\varphi(x,t)\mathrm{d}x \right\} f(y,t)\mathrm{d}y, \tag{A.5}$$

in which $\varphi(x,t) = f(x,t)/(1-F(x,t))$, $\hat{\lambda}$ and $\hat{\mu}$ are the respective Lagrangian multipliers of the goods and labor market clearing conditions.

Equation (A.5) is the counterpart of HJB equation (3) for the social planner. The LHS is the flow social value of firm z based on the discounting factor ρ rather than the interest rate r_t. Three items on the RHS compose this flow value. The first item is the value of static output with $\hat{\lambda}$ the shadow price of the output in utility. The second item is the option value which is the sum of an incremental change, $\partial w(z,t)/\partial t$, and the net return of idea searches. The third and additional item that only shows up in the social planner's HJB equation is the last term in (A.5), capturing the externality of other firms' learning. The conditional source distribution makes the externality term very complicated since any firm can affect others in various ways. To see this, consider an increase in the density $f(z)$ and a firm y with $y < z$. The density of the source distribution facing by firm y is $\frac{f(x)}{1-F(y)}$ for $x \geq y$. An increase in the portion of firm z increases the likelihood for firm y to meet a firm z. Yet it also decreases the likelihood of meeting other firms x with $x \neq z$, which could also benefit firm y.

In the following, I solve the policy function associated with HJB equation (A.5) to obtain Proposition 4.

C.2.1 Proof of Proposition 4

Proof. To solve for the optimal policy, I first differentiate equation (A.5) with respect to z.

$$\begin{aligned}
\rho w_z(z,t) =& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} + \eta^*(z,t) \frac{\partial}{\partial z} \left\{ \int_z^{\infty} [w(y,t) - w(z,t)] \frac{f(y,t)}{1-F(z,t)} \mathrm{d}y - \hat{\mu} z \right\} \\
&+ w(z,t) f(z,t) \int_0^z \eta^*(x,t) \frac{-\varphi(x,t)}{1-F(x,t)} \mathrm{d}x - w(z,t) f(z,t) \int_0^z \eta^*(x,t) \frac{-\varphi(x,t)}{1-F(x,t)} \mathrm{d}x \\
&+ \int_z^{\infty} w(y,t) f(y,t) \mathrm{d}y \eta^*(z,t) \frac{-\varphi(z,t)}{1-F(z,t)} + w_z(z,t) \int_0^z \eta^*(x,t)\varphi(x,t)\mathrm{d}x + w(z,t)\eta^*(z,t)\varphi(z,t) \\
=& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} - \eta^*(z,t)\varphi(z,t) \int_z^{\infty} (w(y,t) - w(z,t)) \frac{f(y,t)}{1-F(z,t)} \mathrm{d}y + w_z(z,t) \int_0^z \eta^*(x,t)\varphi(x,t)\mathrm{d}x \\
=& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} - \hat{\mu}\eta^*(z,t)\varphi(z,t)z + w_z(z,t) \int_0^z \eta^*(x,t)\varphi(x,t)\mathrm{d}x
\end{aligned} \tag{A.6}$$

The first equality is the result of an envelope theorem, and the second and the third use the first order condition (A.9), which holds for all z at any time t. Let $k(z,t) = zf(z,t)/(1-F(z,t))$, I obtain $w_z(z,t)$ and $\partial w_z(z,t)/\partial t$ by differentiating (A.9):

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

in which the last equality comes from the law of motion on $\varphi(z,t)$. Formally, the original law of motion (7) 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 (A.6), the following differential equation characterizes the optimal policy.

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

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

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

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

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

If $F(z,t)$ has tail index $k(t) > 1$, $\lim_{z\to\infty} k(z,t) = k(t)$. The Karamata's Representation Theorem implies that $e^{\int_0^z \frac{k(x,t)-1}{2x}dx}$ is a regularly varying function with

exponent $(k(t)-1)/2$. Furthermore, $\int_0^z \frac{1}{2y} e^{\int_0^y \frac{1-k(x,t)}{2x} dx} \mathrm{d}y$ converges as $z \to \infty$. $\eta^*(z,t)$ is then 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. With a Pareto initial distribution, $k(z,0) = k_0$ for $z \geq 1$, and

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

a power function. The proof is then complete.

■

C.2.2 Derivation of Equation (A.5) on $w(z,t)$

Following Lucas and Moll (2014), the corresponding HJB equation of problem (17) is

$$\begin{aligned} \rho W(f) = & \max_{\{c(\omega), \eta(y)\}} Lu(c) + \int_0^\infty \frac{\delta W(f)}{\delta f(y)} f(y) \left[\int_0^y \eta(x)\varphi(x)\mathrm{d}x - \eta(y) \right] \mathrm{d}y \\ \text{s.t.} \quad & Lc \leq \int_0^\infty y f(y) \mathrm{d}y, \quad \int_0^\infty y\eta(y) f(y) \mathrm{d}y \leq L, \end{aligned} \tag{A.7}$$

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

$$u'(c) = \hat{\lambda}. \tag{A.8}$$

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

$$\begin{aligned} & \int_z^\infty w(f,y) f(y) \varphi(z) \mathrm{d}y - w(f,z) f(z) - \hat{\mu} z f(z) = 0 \\ \Longrightarrow & \int_z^\infty [w(f,y) - w(f,z)] \frac{f(y)}{1-F(z)} \mathrm{d}y = \hat{\mu} z \end{aligned} \tag{A.9}$$

Differentiating both sides of the HJB equation (A.7) with respect to $f(z)$,

$$\rho w(f,z) = \int_0^\infty \frac{\delta w(f,z)}{\delta f(y)} f(y) \left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right] \mathrm{d}y + \int_0^\infty w(f,y) \frac{\delta}{\delta f(z)} \left\{ f(y) \left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right]\right\} \mathrm{d}y + \hat{\lambda} z - \hat{\mu} z \eta^*(z). \tag{A.10}$$

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 (A.10) is simply $\partial w(z,t)/\partial t$.
To calculate the second term, note that

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

Therefore,

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

Then,

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

Finally, rewriting (A.10) gives equation (A.5):

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

in which the max operator comes from first order condition (A.9).

C.2.3 The tail dynamics with optimal search policy

Figure A.4 illustrates the jumps in tail index with optimal search policy.

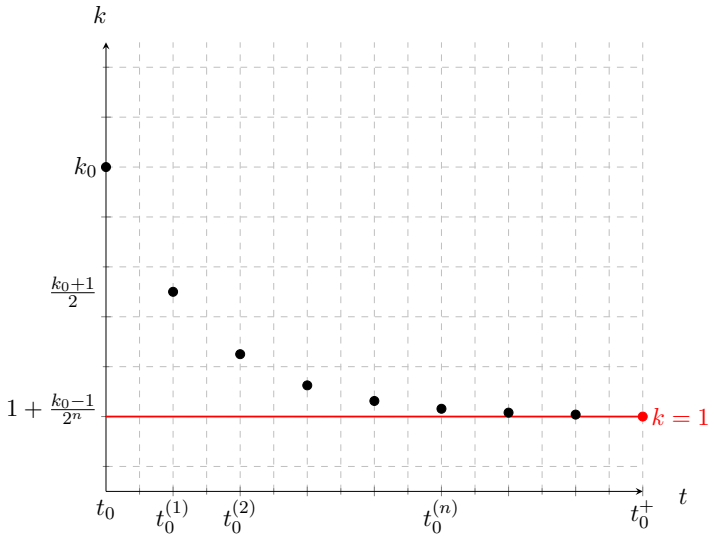


Figure A.4: Illustration of jumps in tail indices

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.

C.3 Bounded arrival rate

Given an upper bound on the arrival rate $\bar{\eta}$, the social planner chooses $z^*(t)$ such that only firms above $z^*(t)$ are active in search, and they search at the maximum intensity $\bar{\eta}$. Meanwhile, $z^*(t)$ is chosen such that researchers are fully employed by the search-active firms, i.e.,

$$\bar{\eta} \int_{z^*(t)}^{\infty} z f(z,t) \mathrm{d}z = L.$$

It worths noting that Proposition A.1 for the extended model characterizes

similar search activities and holds for arbitrary paths of $z^*(t)$ and $\eta(t)$. Hence, the resulting productivity distribution follows directly from Proposition A.1 with $g_E(t) = 0$ and $\eta(t) = \bar{\eta}$. Namely,

$$F(z,t) = \begin{cases} 1 - (1 - F^*(t)) \left(\frac{z}{z^*(t)}\right)^{-k_t}, & \text{if } z \geq z^*(t), \\ F^*(T^*(z)), & \text{if } z < z^*(t). \end{cases} \tag{A.11}$$

in which $T^*(z)$ is the inverse of $z^*(t)$, $g_z(t) \equiv \dot{z}^*(t)/z^*(t)$,

$$(F^*)'(t) = k(t) g_z(t) \left(1 - F^*(t)\right), \quad \text{and} \quad \frac{\dot{k}(t)}{k(t)} = -\bar{\eta}. \tag{A.12}$$

Next, I pin down the path of $z^*(t)$. Substituting (A.11) into the labor market clearing condition,

$$\left[1 - F^*(t)\right] z^*(t) \frac{k(t)}{k(t) - 1} \bar{\eta} = L. \tag{A.13}$$

Log differentiating the above equation, I obtain

$$g_z(t) = \frac{(F^*)'(t)}{1 - F^*(t)} + \frac{\dot{k}(t)}{k(t)(k(t) - 1)} = g_z(t) k(t) - \frac{\bar{\eta}}{k(t) - 1} \Rightarrow g_z(t) = \frac{\bar{\eta}}{(k(t) - 1)^2}, \tag{A.14}$$

where the second equality uses Equation (A.12). Hence, at time $T = \ln(k_0)/\bar{\eta}$, $k(T) = 1$, $g_z(T) = z^*(T) = \infty$, and $F^*(T) = 1$.

Finally, I show that the total output reaches infinity at time T. Using integration by parts,

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

When $t \to T$, $k(t) \to 1$, and (A.13) implies that $[1 - F^*(t)] z^*(t) \to 0$. It then suffices to show that the first integral diverges as $t \to T$. Notice that

$$\int_1^{z^*(t)} \left[1 - F(z,t)\right] \mathrm{d}z = \int_1^{z^*(t)} \left[1 - F^*(T^*(z))\right] \mathrm{d}z = \int_0^t \left[1 - F^*(s)\right] z^*(s) g_z(s) \mathrm{d}s,$$

where the first equality follows (A.11) and the second the substitution $z = z^*(s)$.

Using (A.13) and (A.14),

$$\int_0^t \left[1 - F^*(s)\right] z^*(s) g_z(s) \mathrm{d}s = \int_0^t \frac{L}{\bar{\eta}} \frac{k(s)-1}{k(s)} \frac{\bar{\eta}}{(k(s)-1)^2} \mathrm{d}s = L \int_0^t \frac{1}{k(s)\,(k(s)-1)} \mathrm{d}s.$$

When $t \to T$, $k(t) \to 1$, and the integral diverges.

Appendix D An extended model

In this section, I provide a full analysis on the extended model sketched in Section 7. In Section D.1 and D.2, I offer a detailed description of the extended model. In Section D.3 and D.4, I show that the extended model preserves all equilibrium properties of the baseline model. Omitted proofs can be found in Supplemental Appendix J.

D.1 Model setup

Demography The measure of representative households $L(t)$ now grows at an exogenous rate $g_L(t)$ over time, i.e., $g_L(t) = \dot{L}(t)/L(t)$. Moreover, each household consists of a unit continuum of people with s fraction of them being skilled labor.[38] Each person is endowed with one unit of indivisible time. While unskilled labor can only engage in production, skilled labor choose to either work as researchers for incumbent firms or start up their own firms as entrepreneurs. Moreover, skilled labor differ in their innate managerial ability e, which has a cumulative distribution function J over support $[0, +\infty)$. I use e to index entrepreneurs.

At time t, household members share the following utility function over intertemporal consumption:

$$\int_t^\infty e^{-\rho(\tau - t)} \frac{c(\tau)^{1-\theta} - 1}{1-\theta} \mathrm{d}\tau, \tag{A.15}$$

in which $c(\tau)$ is a Dixit-Stiglitz aggregator over variety consumption $q(\omega, \tau)$, i.e.,

$$c(\tau) = \left(\int_{\Omega(\tau)} q(\omega, \tau)^{\frac{\sigma-1}{\sigma}} \mathrm{d}\omega \right)^{\frac{\sigma}{\sigma-1}}.$$

$\Omega(\tau)$ denotes the set of all available varieties at time τ, and σ is the elasticity of substitution.

[38] Although s is exogenous, the long run growth rate in this model is also independent of s.

Market structure Firms are characterized by their technologies and produce distinct varieties. Each firm has a linear production function $y = \tilde{z}l$, in which l is the input of production workers and $\tilde{z}$ its productivity. Firms engage in monopolistic competition and maximize profits given households' demand and wage of production workers $w_p(t)$. I use $z \equiv \tilde{z}^{\sigma-1}$ to index firms and their technologies (productivity). Standard results give that firms' demand for production workers l_p, revenue r, and profit π have the following expressions and are linear in z:[39]

$$l_p(z,t) = \frac{\sigma-1}{w_p(t)}\pi(z,t), \quad r(z,t) = \sigma\pi(z,t), \quad \text{and} \quad \pi(z,t) = \pi(t)z, \tag{A.16}$$

in which $\pi(t) = \frac{1}{\sigma}\frac{R(t)}{P(t)^{1-\sigma}}\left(\frac{\sigma}{\sigma-1}w_p(t)\right)^{1-\sigma}$, $R(t)$ the total expenditure of households, and $P(t)$ the ideal price index. Furthermore, I use $M(t)$ and $F(z,t)$ to denote the respective measure and productivity distribution of firms at time t.

Technological obsolescence An exogenously increasing productivity threshold $z^*(t)$ captures the obsolescence of technology. Technologies are graded by their productivity z and become obsolete at time t if $z \leq z^*(t)$. Obsolete technologies cannot be upgraded or replicated.[40] That is, incumbents using obsolete technologies cannot improve its productivity, and new firms cannot adopt obsolete technologies to enter. I use $g_z(t)$ to denote the growth rate of $z^*(t)$, i.e., $g_z(t) = \dot{z}^*(t)/z^*(t) \geq 0$.

Incumbents A key departure from the baseline model is that firms with productivity $z < z^*(t)$ (obsolete technologies) can no longer increase their productivity through idea search and thus become "search-inactive". Let $w_r(t)$ be the wage of researchers and $r(t)$ the interest rate. All incumbents face an exogenous death shock at a rate δ. The HJB equation for search-inactive firms is given by:

$$(r(t)+\delta)\, v(z,t) = \pi(z,t) + \partial_t v(z,t), \quad \text{for } z < z^*(t). \tag{A.17}$$

The remaining "search-active" incumbent firms continue to engage in the same search process as in the baseline model, upgrading their technologies to enhance

[39] These results imply that the distributions of firms' revenue, profit, production employment are identical after scaling. The distribution of productivity z is constructed to be identical to the size distribution of firms after scaling. Hence, I focus on the productivity (z) distribution and use it interchangeably with firm size distribution. The distribution over productivity $\tilde{z}$, however, will generally be different unless $\sigma = 2$.

[40] Technologies often become obsolete when they are no longer supported. For example, suppliers may stop producing compatible parts. Therefore, these technologies will not be replicated.

productivity. For these firms, the HJB equation is given by:

$$(r(t)+\delta)\,v(z,t) = \pi(z,t) + \partial_t v(z,t) + \max_{\eta\geq 0} \eta \left\{ \int_z^\infty [v(x,t)-v(z,t)]\mathrm{d}F(x|x\geq z,t) - \psi z w_r(t) \right\}, \tag{A.18}$$

for $z \geq z^*(t)$. The only modification to the baseline model is the introduction of the parameter ψ, which measures the efficiency of the search technology. The unit search cost for a firm z in terms of researchers is now ψz.

Entrants The entry of new firms originates from skilled labor's entrepreneurial choice. Skilled labor become entrepreneurs if the expected value of startup activities $v_E(e,t)$ exceeds the after-tax(subsidy) wage of researchers $(1-\tau_r)w_r(t)$, where τ_r is the tax or subsidy rate on researchers' wage.

Since entrepreneurs cannot replicate obsolete technologies, they make random draws among search-active incumbents to adopt their technologies. For entrepreneurs, completing a unit search requires the efforts of $\psi z^*(t)$ researchers. In other words, entrepreneurs "stand on the shoulders of giants" by avoiding obsolete technologies and using the same search technology as firm $z^*(t)$. At the startup stage, entrepreneurs cannot hire researchers and instead rely on their managerial ability to search for ideas. The expected return of a startup v_E is given by:

$$v_E(e,t) = \frac{e}{\psi z^*(t)} \int_{z^*(t)}^\infty v(x,t)\mathrm{d}F(x|x\geq z^*(t),t), \tag{A.19}$$

where $e/\psi z^*(t)$ gives the arrival rate of ideas to an entrepreneur e.

Since $v_E(e,t)$ increases in e, there exists a threshold $e^*(t)$ such that skilled labor choose to become entrepreneurs if $e \geq e^*(t)$ and researchers if $e < e^*(t)$. With individual startup rates $e/\psi z^*(t)$, the measure of entrants $E(t)$ follows that

$$E(t) = \frac{sK(e^*(t))L(t)}{\psi z^*(t)}, \tag{A.20}$$

where $K(e) = \int_e^\infty x\mathrm{d}J(x)$.

Consumption Households choose consumption for all members by maximizing household utility subject to their household income. The household utility is the sum of individual members' utilities as defined in (A.15). Given that all members have identical CRRA utility functions, they always consume the same amount $c(t)$. Consequently, the household utility equals the utility of an individual member.

Household income $y(t)$ is the sum of labor income received by all members and the dividends from owning firms. Assuming the law of large numbers holds, each household receives the same dividend $\bar{\Pi}(t)$, which equals the total firm profits divided by the measure of households. Then, $y(t) = (1-s)w_p(t) + sJ(e^*(t))(1-\tau_r)w_r(t) + \bar{\Pi}(t) + T(t)$, in which $T(t)$ is a transfer that balances the government budget, defined as $T(t) = sJ(e^*(t))\tau_r w_r(t)$. Households then solve the following utility maximization problem given interest rate $r(t)$ and income $y(t)$:

$$\begin{aligned} &\max_{c(\tau)\geq 0} \int_t^\infty e^{-\rho(\tau-t)} \frac{c(\tau)^{1-\theta}-1}{1-\theta} \mathrm{d}\tau. \\ &\text{s.t.} \int_t^\infty e^{-\int_t^s r(\tau)\mathrm{d}\tau} c(s)\mathrm{d}s \leq \int_t^\infty e^{-\int_t^s r(\tau)\mathrm{d}\tau} y(s)\mathrm{d}s. \end{aligned} \tag{A.21}$$

D.2 Equilibrium path

Aggregate state The state of the economy is now summarized by the measure of firms $M(t)$ and productivity distribution $F(z,t)$. As described earlier, $M(t)$ evolves as follows:

$$\dot{M}(t) = E(t) - \delta M(t). \tag{A.22}$$

Let $g_M(t) \equiv \dot{M}(t)/M(t)$, and $g_E(t) \equiv E(t)/M(t) \geq 0$. The above equation could be rewritten as $g_M(t) = g_E(t) - \delta$.

With technological obsolescence, the productivity distribution evolves differently among search-active and search-inactive firms. For $z < z^*(t)$,

$$\frac{\partial F(z,t)}{\partial t} = -g_E(t)F(z,t). \tag{A.23}$$

This is intuitive as firms using obsolete technologies do not search and gradually die out. For $z \geq z^*(t)$,

$$\frac{\partial F(z,t)}{\partial t} = -\int_{z^*(t)}^z \eta(x,t)(1-F(z|z\geq x,t))\mathrm{d}F(x,t) - g_E(t)\frac{F(z^*(t),t)\,(1-F(z,t))}{1-F(z^*(t),t)}. \tag{A.24}$$

The first term is the same as the baseline case capturing incumbents' search, and the second term captures the entrants' productivity distribution. It is evident that the law of motion on $F(z,t)$ reduces to the baseline case (7) if $g_E(t) = 0$ and $z^*(t) = 1$. In addition, I use $F^*(t)$ to denote the fraction of search-inactive firms at time t, i.e., $F^*(t) = F(z^*(t),t)$. I further denote the time when $z^*(t)$ reaches z by $T^*(z)$, which is the inverse of z^*. Namely, $z^*(T^*(z)) = z$, and $F^*(T^*(z)) = F(z,t)$.

Initial conditions I assume that the initial state of the economy mirrors the baseline model. Specifically, Assumption 1 still holds, meaning that the initial distribution of productivity is given by $F(z,0) = 1 - z^{-k_0}$ for $z \geq 1$ with $k_0 > 1$. In addition, the initial productivity threshold $z^*(0) = 1$, and the initial measure of firms $M(0) = 1$.

Market clearing Labor markets for production workers and researchers clear in the equilibrium. Namely,

$$M(t) \int_1^\infty l_p(z,t) \mathrm{d}F(z,t) = (1-s)L(t), \tag{A.25}$$

$$M(t) \int_{z^*(t)}^\infty l_r(z,t) \mathrm{d}F(z,t) = sJ(e^*(t))L(t), \tag{A.26}$$

where firm z's demand for researchers $l_r(z,t)$ is given by $\psi z \eta(z,t)$. While the share of production workers in the population is always constant, the share of researchers is endogenous due to entrepreneurial choices. Additionally, goods market always clears since firms take households demand as given. I choose the ideal price index $P(t)$ as the numeraire, so $P(t) \equiv 1$.

Equilibrium concept With richer model elements, I modify the definition of an equilibrium for the extended model as follows. The definition of an asymptotic balanced growth path still follows that in the baseline model.

Definition (Equilibrium)**.** A competitive equilibrium for this economy consists of wages $w_r(t)$ and $w_p(t)$, interest rates $r(t)$, firms' value functions $v(z,t)$, firms' search intensity $\eta(z,t)$ and demand for production labor $l_p(z,t)$, households' income $y(t)$ and consumption $c(t)$, startup decisions $e^*(t)$, the measure of firms $M(t)$, and the productivity distribution $F(z,t)$ that satisfy the following:

(i) Given $\{w_p(t), w_r(t), r(t), F(z,t)\}$, $v(z,t)$ solves the HJB equation (A.17) and (A.18) with search intensity policy $\eta(z,t)$, and $l_p(z,t)$ is given by (A.16);

(ii) Given $F(z,t)$ and $w_r(t)$, $e^*(t)$ is the entrepreneurial choice threshold;

(iii) Given $\{y(t), r(t)\}$, $c(t)$ solves the households' problem (A.21);

(iv) Goods and labor markets clear, i.e., (A.25) and (A.26) are satisfied;

(v) $M(t)$ solves the ODE (A.22) given entrants $E(t)$ in (A.20) and $M(0)$. $F(z,t)$ solves the PDE (A.23) and (A.24) given $\eta(z,t)$ and $F(z,0)$.

D.3 Transition dynamics

The analysis of the transition dynamics closely follows that of the baseline model. First, I show that the equilibrium productivity distribution can be summarized by two variables: $F^*(t)$ and $k(t)$. I then derive a system of differential equations for $(M(t), F^*(t), k(t))$ that fully characterizes the economy's transition dynamics.

Productivity distribution To begin with, it is useful to note that Lemma 1 for the baseline model still holds. Since $\pi(z,t) = \pi(t)z$, the linearity of HJB equations (A.17) and (A.18) implies that $v(z,t)$ is also linear in z. I then write $v(z,t) = v(t)z$. The following proposition generalizes Proposition 1 from Section 4 to characterize the evolution of productivity distribution in the extended model.[41]

Proposition A.1. *There exists an equilibrium with the following properties:*

i) The search intensity is invariant across productivity among search active firms, i.e., $\eta(z,t) = \eta(t) > 0$ for $z \geq z^*(t)$;

ii) The equilibrium productivity distribution at time t is

$$F(z,t) = \begin{cases} 1 - (1 - F^*(t)) \left(\frac{z}{z^*(t)}\right)^{-k(t)}, & \text{if } z \geq z^*(t), \\ F^*(T^*(z)) e^{-\int_{T^*(z)}^{t} g_E(s)\mathrm{d}s}, & \text{if } z < z^*(t), \end{cases} \tag{A.27}$$

in which

$$(F^*)'(t) = k(t) g_z(t) \left(1 - F^*(t)\right) - g_E(t) F^*(t), \tag{A.28}$$

and

$$\frac{\dot{k}(t)}{k(t)} = -\eta(t) < 0.$$

Similar to the baseline model, search-active firms search for ideas at a common intensity $\eta(t)$. As a result, the portion of the distribution corresponding to search-active firms—the right tail—remains a Pareto distribution with a varying shape parameter $k(t)$, which continues to decline at the rate of $\eta(t)$. The equilibrium share of inactive firms $F^*(t)$ is shaped by the race between the speed of technological obsolescence $g_z(t)$ and the entry rate $g_E(t)$.

[41] Proposition 5 copies Proposition A.1 into the main text.

Proposition A.1 demonstrates that the productivity distribution in the extended model maintains a thickening right tail along the transition. Moreover, Gibrat's law also holds for firms in the right tail. With a Pareto right tail and a common search intensity, it follows from the arguments in Section 4 that search-active firms have the same expected growth rate.

Aggregate state Given Proposition A.1, it is sufficient to characterize the state of the economy by three variables: the measure of firms $M(t)$, the share of search-inactive incumbents $F^*(t)$, and tail index $k(t)$. I first derive the dynamics of $k(t)$. Given $F(z,t)$ in (A.27), the market clearing condition (A.26) for researchers implies that

$$\psi M(t)\eta(t)\left[1-F^*(t)\right]\frac{k(t)}{k(t)-1}z^*(t)=sJ(e^*(t))L(t). \tag{A.29}$$

Combining it with $\dot{k}(t)/k(t)=-\eta(t)$, I obtain that

$$-\frac{\dot{k}(t)}{k(t)-1}=\frac{sJ(e^*(t))}{1-F^*(t)}\frac{L(t)}{\psi M(t)z^*(t)}. \tag{A.30}$$

Equation (A.30) generalizes (12) in the baseline model, but it now relies on endogenous variables $F^*(t)$, $e^*(t)$, and $M(t)$. I define $g_k(t)\equiv-\dot{k}(t)/(k(t)-1)$.

I then derive the dynamics of $M(t)$ and $F^*(t)$. Equation (A.28) in Proposition A.1 characterizes the evolution of $F^*(t)$. Meanwhile, the dynamics of $M(t)$ follows directly from (A.22) and (A.20):

$$\frac{\dot{M}(t)}{M(t)}+\delta=\frac{sK(e^*(t))L(t)}{\psi M(t)z^*(t)}. \tag{A.31}$$

Next, I express $e^*(t)$ as a function of $M(t)$, $F^*(t)$, and $k(t)$. In equilibrium, the payoff from a unit search equals its cost for all search-active firms. Using the expression for $F(z,t)$ in (A.27), I derive the following condition:

$$\frac{v(t)}{k(t)-1}=\psi w_r(t), \tag{A.32}$$

where the left-hand side represents the payoff from a unit search, the right-hand side corresponds to the cost of a unit search, and both are adjusted by productivity.

Since J has support $[0,+\infty)$, (A.19) implies that $e^*(t)$ is always interior so that $v_E(e^*(t),t)=(1-\tau_r)w_r(t)$. Substituting (A.19), (A.27), and (A.32) into this

equation, I obtain the following expression for $e^*(t)$:

$$e^*(t) = \frac{1-\tau_r}{k(t)}. \tag{A.33}$$

Finally, Equation (A.28), (A.30), (A.31), and (A.33) constitute a system of ODE on $M(t)$, $F^*(t)$, and $k(t)$. With initial conditions $F^*(0) = 0$, $M(0) = 1$, $k(0) = k_0$, this ODE system completely characterizes the transition path of the economy.

D.4 Long-run equilibrium

So far, the results on transition dynamics do not depend on specific restrictions on the exogenous processes of population growth and technological obsolescence. To discipline the long-run behavior of the economy, I impose Assumption 2, which assumes that both processes exhibit asymptotically exponential growth. Additionally, the condition $g_L^* + \delta > g_z^*$ ensures that firm entry remains positive in the long run. I then characterize the long-run equilibrium in Proposition A.2.

Assumption 2. $\lim_{t\to\infty} g_L(t) = g_L^*$, $\lim_{t\to\infty} g_z(t) = g_z^* > 0$, *and* $g_L^* + \delta > g_z^*$.

Proposition A.2. *With Assumption 2 and* $\rho > (1-\theta)g_y^*$, *there exists an equilibrium path features the following limiting properties as* $t \to \infty$*:*

i) The tail index $k(t)$ *decreases to* 1, *and the startup threshold* $e^*(t)$ *increases to* $e^* = 1 - \tau_r$. *Moreover,*

$$g_k^* \equiv \lim_{t\to\infty} g_k(t) = \frac{J(e^*)}{K(e^*)}(g_L^* + \delta); \tag{A.34}$$

ii) The measure of firms growth $g_M(t)$ *converges to* $g_M^* = g_L^* - g_z^*$;

iii) The share of search-inactive firms $F^*(t)$ *converges to* $f = \frac{g_z^*}{g_L^*+\delta}$;

iv) $F(z,t)$ *is an asymptotic traveling wave with velocity* $z^*(t)$, *i.e.,*

$$\lim_{t\to\infty} F(rz^*(t),t) = \Gamma(r) \equiv \begin{cases} fr^{\varsigma}, & \text{for } r \in (0,1), \\ 1-(1-f)r^{-1}, & \text{for } r \geq 1, \end{cases} \tag{A.35}$$

where $\varsigma = g_E^*/g_z^*$, *and* $g_E^* = g_L^* - g_z^* + \delta > 0$;

v) The growth rate of mean productivity, $Z(t) \equiv \int_1^\infty z\mathrm{d}F(z,t)$, *converges to* $g_z^* + g_k^*$;

vi) The household income growth $g_y(t) \equiv \dot{y}(t)/y(t)$ *converges to* $g_y^* = \frac{g_k^* + g_L^*}{\sigma - 1}$.

The parametric condition $\rho > (1-\theta)g_y^*$ ensures that the household utility is finite and the relevant transversality condition is satisfied. Proposition A.2 states that the right tail of the productivity distribution still converges to a Zipf's distribution. Moreover, the equilibrium path features asymptotic balanced growth in household income. Based on the previous discussions on thickening right tails and Gibrat's law, I verify that the extended model preserves all the properties outlined in Proposition 2 for the baseline model.

References

Ayyagari, M., A. Demirguc-Kunt, and V. Maksimovic (2014): "Who creates jobs in developing countries?" *Small Business Economics*, 43, 75–99.

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

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

Kwon, S. Y., Y. Ma, and K. Zimmermann (2024): "100 Years of Rising Corporate Concentration," *American Economic Review*, 114, 2111–2140.

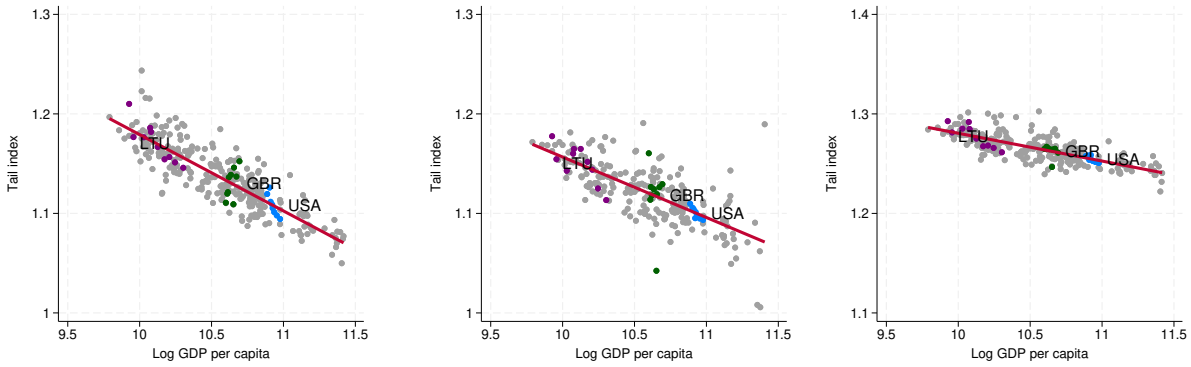
Lucas, R. E. and B. Moll (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122, 1–51.

Supplemental Appendix (Not For Publication)

Appendix E Employment-based simple estimator

E.1 Results with employment-based simple estimator

The OECD countries Together with k^f and k^{TW}, I present the regression results with k^{emp} in Table A.1 and visualize them in Figure A.5.



(a) k^f (b) k^{TW} (c) k^{emp}

Figure A.5: Tail index and log GDP per capita in OECD countries

Notes. This figure plots the tail index k against log GDP per capita among a panel of OECD countries. The scatter dots are readjusted by country fixed effects, and the red lines are the linear fits. The tail index is k^f with $T_S = 10$ and $T_L = 250$ in the left panel, k^{TW} with $T = 10$ in the middle, and k^{emp} with $T_S = 10$ and $T_L = 250$ in the right. The log GDP per capita is in constant international dollars. Three annotated countries are Lithuania (LTU, purple), the UK (GBR, green) and the USA (blue). Data sources: OECD SBS and PWT 10.0.

The United States In Figure A.6, I present the time series of k^{emp} for the U.S. together with k^f and k^{TW}. Furthermore, I verify in Figure A.7 that the negative trend in k^{emp} is robust to alternative choices of thresholds.

E.2 Comparing tail estimators

In this section, I compare the tail estimates obtained using k^f, k^{emp}, and k^{TW}. In the U.S. sample, Figure A.6 provides clear evidence that while tail indices based on k^f and k^{TW} are well aligned, those derived from k^{emp} are systematically higher and exhibit lower volatility over time.

In the OECD sample, Figure A.5 shows that the semi-elasticity of k^{emp} with respect to GDP per capita is significantly lower than that of k^f or k^{TW}. This finding similarly suggests that the within-country over-time volatility of tail indices based on k^{emp} is lower than those based on k^f or k^{TW}. Although the tail indices in this figure also appear higher when estimated using

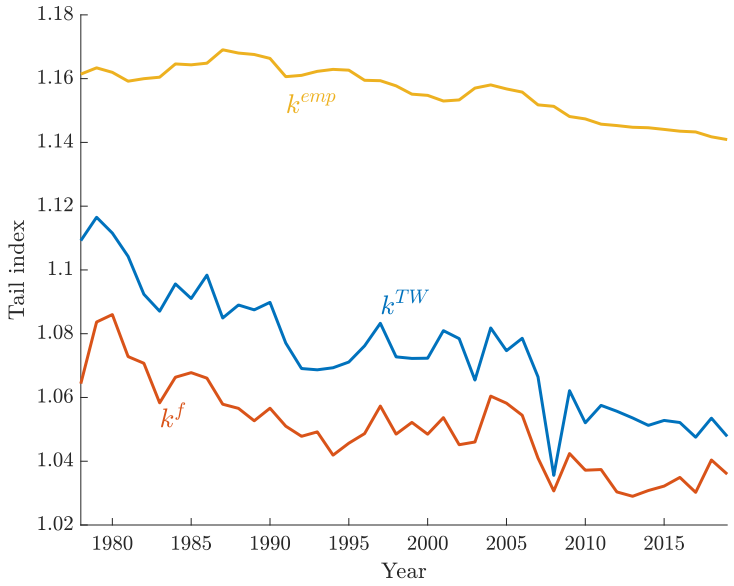
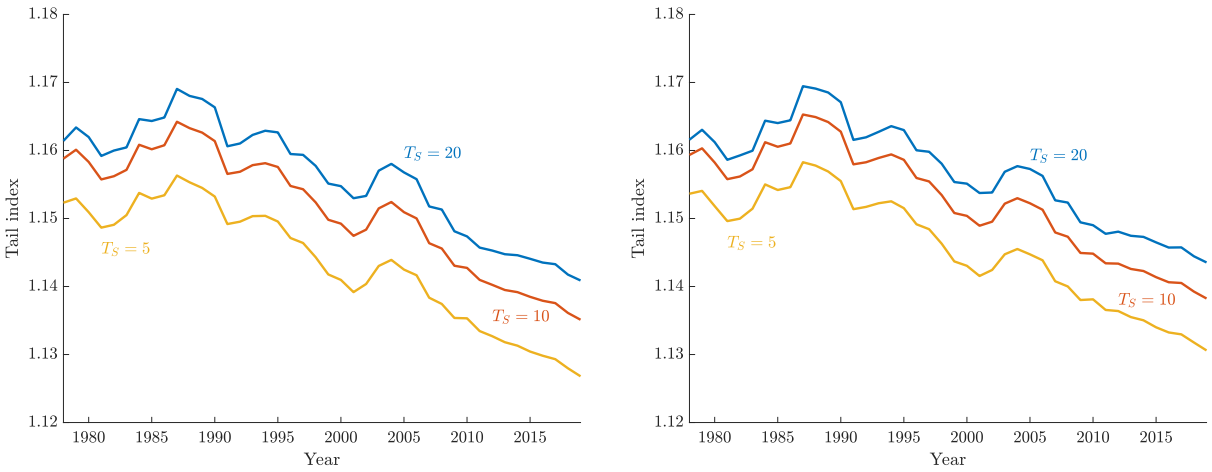


Figure A.6: Tail index in the US (1978-2019)

Notes. This figure plots the tail index k of the size distribution of all US business firms from 1978 to 2019. The tail indices are obtained by either the Toda-Wang estimator k^{TW} with $T = 20$ (in blue) or the simple estimator k^{f} (in red) or k^{emp} (in yellow) with $T_S = 20$ and $T_L = 500$. Data source: the U.S. Census BDS.



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

Figure A.7: Tail indices k^{emp} in the U.S. (1978-2019) with alternative thresholds

Notes. This figure illustrates the tail index k of the size distribution for all U.S. business firms from 1978 to 2019. The tail indices are estimated using the simple estimator k^{f}, which is based on the top employment shares and the corresponding thresholds T_S and T_L as indicated in the labels. Data source: the U.S. BDS.

k^{emp}, they are purged of country fixed effects. In the following analysis, I directly compare tail indices based on k^{emp} with those based on k^f or k^{TW} for OECD countries.

Since k^{TW} is an efficient estimator, I use it as a benchmark and compare estimates based k^f and k^{emp} with k^{TW}. For sharper comparison, I focus on the OECD sample in which country-year pairs with binding k^{TW} estimates are excluded. Figure A.8 presents the results. It is obvious that k^{emp} are systematically higher than k^{TW}, while k^f are mostly consistent with k^{TW}. The only exception is Turkey, which generates the cloud of outliers on the upper left corner in Figure A.8(a). Regression results also confirm this point. Regressing k^f and k^{emp} on k^{TW}, the coefficient of k^{TW} are 1.007 and 0.466 respectively.

The low R^2 in the regression between k^f and k^{TW} is due to the outlier country Turkey. Figure A.8(c) and A.8(d) present the same results excluding Turkey. The R^2 increases from 0.456 to 0.811, and there are no obvious outliers. In contrast, Figure A.8(d) barely changes from Figure A.8(b). The R^2 remains low in both samples: 0.221 vs. 0.239.

Distinctions in the properties of the tail estimators anticipate these comparison results. First, it is unsurprising that the tail indices based on k^{emp} systematically diverge from those based on k^f in the observed direction. Unlike k^f, the construction of k^{emp} ensures that the resulting tail estimates are strictly greater than one, leading to systematically higher values. Moreover, the time series of top employment shares is inherently less volatile than that of top firm shares, as firms near the threshold have a greater impact on the latter. This property makes the time series of k^{emp} mechanically less volatile.

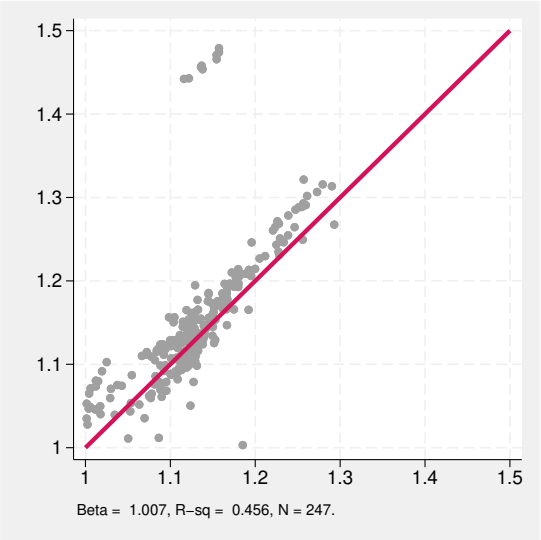
Second, the differences between k^{emp} and k^{TW} highlight the relative inefficiency of k^{emp}. As noted in Footnote 8, k^{emp} is as restrictive as k^{TW} but less efficient. While both require $k > 1$, k^{emp} relies on only two size thresholds, making it less informative about the tail behavior.

On another note, it is particularly reassuring that k^f produces highly consistent estimates with k^{TW}, despite the substantial ex ante differences between them, as discussed in Section 2.1. Although k^f uses only two size thresholds, it demonstrates reasonable efficiency, further supporting its use for the main results. At least for firm size distributions, the comparison between k^f and k^{emp} suggests that firm shares are more informative about tail behavior than employment shares.

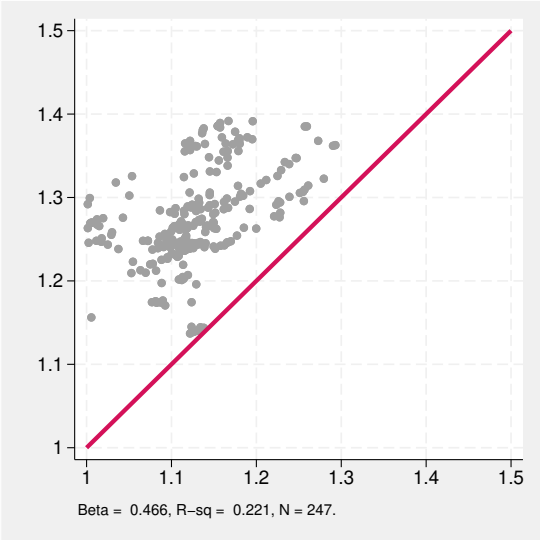
E.3 Testing model predictions with simple estimators

y vs. $\frac{k}{k-1}$ In Figure A.9, I compare real GDP per capita, y, with $k/(k-1)$ calculated using the simple estimators k^f and k^{emp}. On the one hand, $k/(k-1)$ based on k^{emp} shows significantly less growth compared to its counterparts based on k^f or k^{TW}, both of which align well with the growth in y. These discrepancies are expected, as k^{emp} is systematically higher and less volatile than the other estimators.

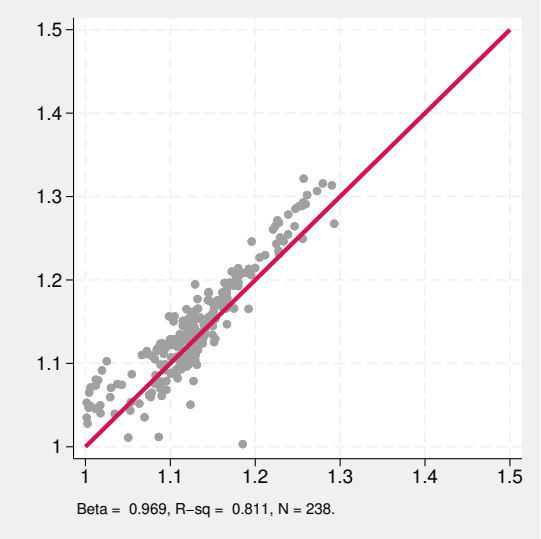
On the other hand, the correlation coefficient between $\ln y$ and $\ln k/(k-1)$ remains consistently high and nearly identical across all tail estimators, suggesting a robust linear relationship between growth in y and growth in $k/(k-1)$. Moreover, the differences observed in direct comparisons between y and $k/(k-1)$ are primarily driven by variations in the variance of k estimates. When regressing $\ln y$ on $\ln k/(k-1)$, the elasticity of growth in y with respect to



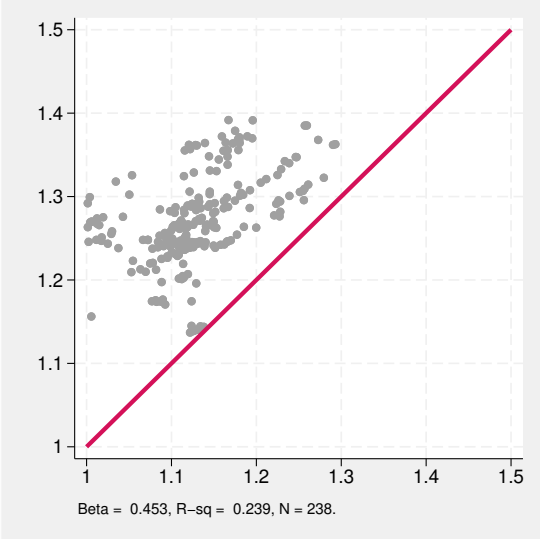
(a) k^f vs. k^{TW}, with Turkey



(b) k^{emp} vs. k^{TW}, with Turkey



(c) k^f vs. k^{TW}, without Turkey



(d) k^{emp} vs. k^{TW}, without Turkey

Figure A.8: Comparison between k^f and k^{emp}: the OECD countries

Notes. These two figures plot the estimates based on k^f and k^{emp} against those based on k^{TW} using OECD samples with and without Turkey. In each figure, the gray dots represent (k^{TW}, k^f) (or (k^{TW}, k^{emp})) with k^{TW} on the x-axis, and k^f (or k^{emp}) are on the y-axis. Red lines are 45° lines. Source: the OECD SBS.

growth in $k/(k-1)$ is measured by the regression coefficient. Given the almost identical correlation coefficients, the variations in these growth elasticities are explained by differences in the variance of $\ln k/(k-1)$.[42]

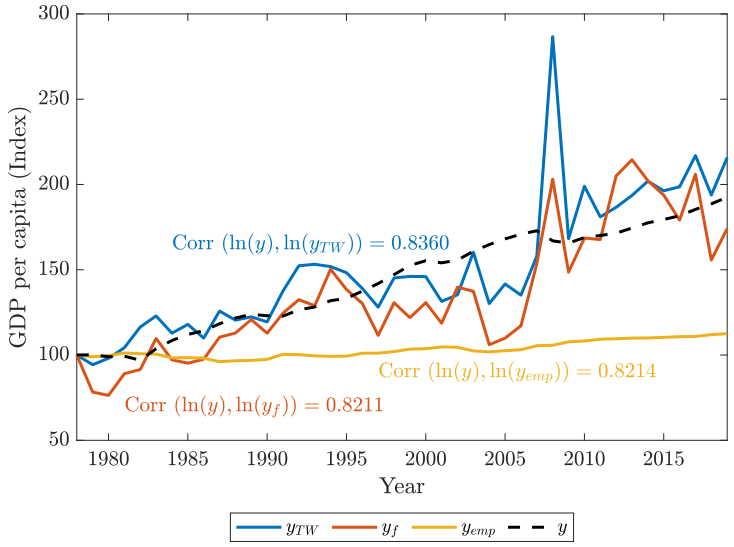


Figure A.9: y vs. $k/(k-1)$

Notes. This figure compares the relationship between y and $k/(k-1)$ for the US between 1978 and 2019. The tail index is estimated respectively by k^{TW}, k^f, and k^{emp}. The output per capita is given by the GDP per capita from FRED. Both data series are in level and normalized with respect to their 1978 level. $y_x = k^x/(k^x-1)$, for $x \in \{TW, f, emp\}$. Source: the U.S. Census BDS and FRED.

$\ln(k-1)$ vs. t In Figure A.10, I present the time series $\ln(k_t - 1)$ estimated using k^f and k^{emp}, along with their respective linear trends. Similar to Figure 3(b), the linear trends provide a good fit to the time series. The alignment between the trend and data is further supported by the high R^2 values and correlation coefficients.

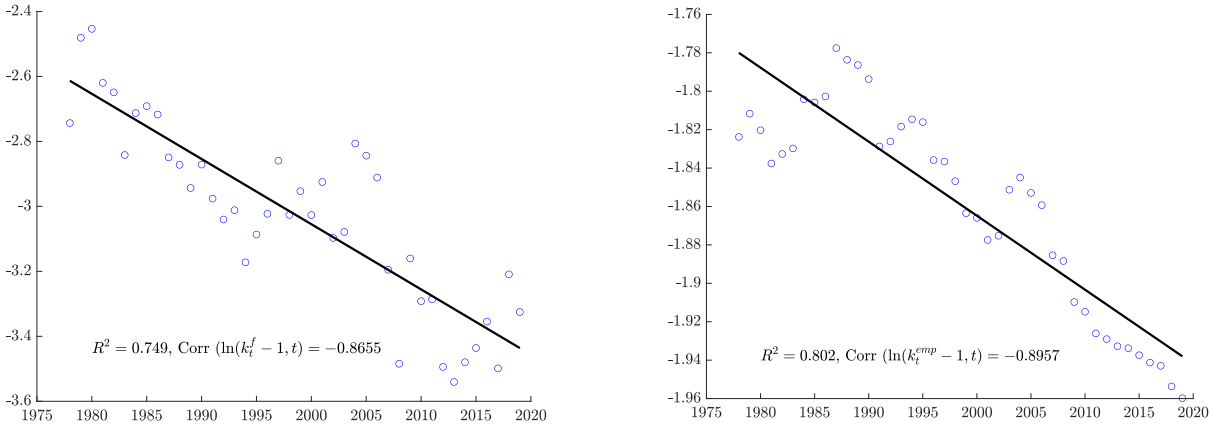
Appendix F Idea search with technological choices

In the baseline model of Section 3, the unit search cost is assumed to be linear in firms' productivity. This assumption could be motivated as the outcome of firms' choices on idea search technologies. In the following, I detail firms' technological choices and show that the resulting search behaviors are observationally equivalent to those in the baseline model. It then provides a micro-foundation for the linear search cost.

Firms face the same set of search technology but may make potential adjustment according to their own technological level. They choose a search technology before deciding on the search

[42] To see this, notice that the regression coefficient α between $\ln y$ and $\ln k/(k-1)$ can be expressed as:

$$\alpha = \text{Corr}(\ln y, \ln k/(k-1)) \sqrt{\frac{\text{Var}(\ln y)}{\text{Var}(\ln k/(k-1))}}.$$



(a) k^f (b) k^{emp}

Figure A.10: $\ln(k-1)$ vs. t

Notes. Both figures compare the time series $\ln(k_t - 1)$ with a linear trend in the U.S. between 1978 and 2019. The tail index is estimated respectively using k^f or k^{emp}. The linear trend and R^2 for k^f or k^{emp} are obtained from the respective regressions: $\ln\left(k_t^f - 1\right) = -2.61 - 0.020t$, and $\ln(k_t^{emp} - 1) = -1.78 - 0.004t$. Source: the U.S. Census BDS.

intensity. A firm z faces a continuum of search technology $\chi_z(m)$ ordered by efficiency m, where $m \in [1, +\infty)$. The trade off in deciding the search technology is that while a more efficient search technology increases the probability of meeting high productivity firms in each search, it also requires more researchers to complete one search. Specifically, a search technology is a pair of source distribution and unit search cost:

$$\chi_z(m) = \left\{F(x|x \geq m), \max\left\{m^{\chi} z^{1-\chi}, m\right\}\right\},$$

where $\chi > 1$. The first component describes the targeting threshold: search technology $\chi_z(m)$ targets on firms with productivity at least m and makes random draws among them. The second component is the unit search cost: it takes $\max\left\{m^{\chi} z^{1-\chi}, m\right\}$ researchers to complete one such search. Similar to standard capital adjustment costs, there is a convex adjustment cost ($\chi > 1$) to use search technology that is more efficient than firms' own productivity z. This specification aims to capture the following two features. First, it shares the same spirit as the truncated sampling assumption: more productive firms have some advantages in meeting top firms. Second, targeting more productive firms becomes increasingly more difficult.

With technology choices, the HJB equation of firms' value now becomes

$$\begin{aligned} r(t)v(z,t) &= z + \partial_t v(z,t) \\ &+ \max_{\eta \geq 0} \eta \left[\max_{m \in [1,+\infty)} \left\{ \int_{\max\{m,z\}}^{\infty} [v(x,t) - v(z,t)] \mathrm{d}F(x|x > m, t) - \max\left\{m^{\chi} z^{1-\chi}, m\right\} w(t) \right\} \right]. \end{aligned} \tag{A.36}$$

Comparing to the baseline case, firms first choose a targeting threshold m to maximize the return of each search and then choose search intensity η to maximize the total return. Let $m(z,t)$ be the policy function of the technological choice. The following proposition states that

given appropriate parametric restrictions, firms' optimal targeting thresholds are always their own productivity, and this policy function supports the equilibrium in the baseline model.

Proposition A.3. *When $k_0 < \chi$, $m(z,t) = z$ for all t and z.*

Proof. We show that $m(z,t) = z$ is consistent with the transition path in the baseline equilibrium. In that equilibrium, productivity distribution $F(z,t)$ is a Pareto distribution with varying shape parameter $k(t)$ and value function $v(z,t) = v(t)z$. We verify that those properties imply a policy function $m(z,t) = z$ and are preserved under such policy. Note first that

$$\begin{aligned}
&\int_{\max\{m,z\}}^{\infty} (v(x,t) - v(z,t))\mathrm{d}F(x|x > m, t) \\
=&\frac{v(t)}{1-F(m,t)} \int_{\max\{m,z\}}^{\infty} (x-z)\mathrm{d}F(x,t) \\
=&\frac{v(t)}{1-F(m,t)} \left[(\max\{m,z\} - z)\left[1 - F(\max\{m,z\},t)\right] + \int_{\max\{m,z\}}^{\infty} \left[1-F(x,t)\right]\mathrm{d}x \right]
\end{aligned}$$

Using that $F(z,t) = 1 - z^{-k(t)}$ for $z \geq 1$, I obtain that for $m < z$, the above equation reduces into

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

As in the baseline equilibrium, let the equilibrium wage be $w(t) = \frac{v(t)}{k(t)-1}$. The net return of one search with search technology $m < z$ is then

$$\frac{v(t)}{k(t)-1}\frac{z^{1-k(t)}}{m^{-k(t)}} - mw(t) = \frac{v(t)}{k(t)-1}\left(\frac{z^{1-k(t)}}{m^{-k(t)}} - m\right) < 0,$$

where the negative sign comes from $m < z$ and $k(t) > 1$. Therefore, I verify that search technology $m < z$ is dominated by search technology z, under which the return is zero.

Now I consider the case $m > z$. The payoff of each search is given by

$$\frac{v(t)}{1-F(m,t)}\left[(m-z)\left[1-F(m,t)\right] + \int_m^{\infty}\left[1-F(x,t)\right]\mathrm{d}x\right] = v(t)\left[(m-z) + \frac{m}{k(t)-1}\right].$$

The net return is then

$$v(t)\left[(m-z) + \frac{m}{k(t)-1}\right] - m^{\chi}z^{1-\chi}w(t) = v(t)\left[m - z + \frac{m - m^{\chi}z^{1-\chi}}{k(t)-1}\right].$$

Consider the function $h(m,z,t) = m - z + \frac{m-m^{\chi}z^{1-\chi}}{k-1}$. Note that

$$h_m(m,z,t) = 1 + \frac{1}{k-1}\left(1 - \chi\left(\frac{m}{z}\right)^{\chi-1}\right) < 1 + \frac{1-\chi}{k-1} = \frac{k-\chi}{k-1}$$

for all $m > z$. With $k_0 < \chi$ and $k(t) \leq k_0$, we know that $h_m(m,z,t) < 0$ for all $m > z$ and t. Hence, $h(m,z,t) < h(z,z,t) = 0$. That is, firms also do not choose search technology $m > z$.

In sum, we have verified that firms $m(z,t) = z$ will be the dominant technological choice, which also implies that the expected return per search is zero. Hence, firms will still be indifferent across the search intensity. The baseline equilibrium is then supported. ■

Appendix G The dynamics of equilibrium prices

In this section, I discuss the transition dynamics of prices in the baseline model of Section 3.

G.1 Price dynamics

Two prices, the wage of researchers and the interest rate, clear the goods and labor markets. I first characterize the dynamics of the interest rate. The Euler equation implies that $r(t) = \theta\frac{\dot{c}(t)}{c(t)} + \rho$. Since the goods market clears and output per capita growth satisfies equation (14), the equilibrium interest rate can be written as follows as a function of the aggregate state $k(t)$, i.e.,

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

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

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

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

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

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

This equation is intuitive since each unit of productivity yields one unit of output at each instant. The value of an additional unit of productivity is then the discounted sum of its output flow, or equivalently, the inverse of average interest rate $\tilde{r}(t)$.[43] As $r(t)$ converges to r^*, the value of unit productivity $v(t)$ converges to $1/r^*$. The growth rate of the researchers' wage then converges to a constant, as the transformed right tail thickness $k(t) - 1$ decreases at a constant rate to zero. Therefore, I consider instead a normalized wage $\tilde{w}(t)$, which is the ratio between

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

the wage of researchers and households' income. Formally,

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

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

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

It is evident to see that the long-run normalized wage is the ratio of the long-run growth rate (L) and long-run interest rate (r^*). The monotonicity, however, is less evident since both the growth rate and average interest rate are increasing. Note that the instantaneous interest rate $r(t)$ is the sum of the discount rate (ρ) and a linear term of the instantaneous growth rate ($\theta g(t)$). Then, $r(t)$ grows less than $g(t)$, i.e., $g(t)/r(t)$ increases over time. When t is sufficiently large, the average interest rate $\tilde{r}(t)$ and the instantaneous interest rate $r(t)$ are very close to each other. The above intuition carries over to see that the normalized wage increases over time. In Lemma A.1, I prove that the same intuition holds with average interest rate $\tilde{r}(t)$. As the economy grows over time, the value of unit productivity is more discounted with higher interest rates, but the size of productivity improvement from each meeting increases with a thicker right tail. Together, the return of each idea search increases over time at a growth rate faster than the output per capita. In words, idea search becomes more valuable with thicker right tails of the firm size distribution.

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

G.2 Proof of Lemma A.1

Proof. From Section G, 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. [44]

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

in which F is a hypergeometric function such that

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

Then,

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

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

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,[45]

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

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 $\tilde{w}(k;\alpha)$. ■

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

Appendix H A discussion on equilibrium selection

In Proposition 1, I construct an equilibrium of a particular form and prove its existence. However, it does not rule out other types of equilibrium. With a linear cost function of idea search, firms are indifferent between any level of search intensity in the baseline equilibrium. That is the reason why Gibrat's law holds, and the productivity distribution remains Pareto. Conversely, an asymmetric equilibrium arises if not all firms agree on the same search intensity. Some firms could search strictly more intensively than other firms, despite that all of them face zero net gain of search in that time, and the labor market still clears. The productivity distribution will lose its Pareto shape afterwards but still evolve according to the KFE (7).

H.1 Equilibria with an adjustment cost in idea search

I argue that the baseline symmetric equilibrium should be preferred to those equilibria. While this set of alternative equilibria is essentially the product of linear search cost, the following shows that the baseline equilibrium is the limiting case of a class of equilibria with non-linear search cost. To see this, I focus on the case that an additional adjustment cost is required for firms to complete the same amount of searches. Particularly, a firm z now has to hire $z(\eta+g(\eta))$ researchers to achieve a search intensity of η. g is assumed to be a strictly increasing and strictly convex function and satisfies $g(0) = g'(0) = 0$. Immediately, Lemma 1 does not always hold with adjustment cost: a linear value function is not a feature of all equilibria. Nevertheless, Proposition A.4 presents that there still exists an equilibrium satisfying all properties stated in Lemma 1 and Proposition 1.

Proposition A.4. *With Assumption 1 and adjustment cost $g(\eta)$, there exists an equilibrium with the following properties:*

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

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

(iii) The equilibrium productivity distribution at time t is

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

in which $k(t)$ satisfies

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

and $k(0) = k_0$.

The proof of Proposition A.4 in Section H.2 follows very closely the arguments for proving Proposition 1. In below, I verify that this equilibrium with adjustment cost inherits all properties stated in Proposition 2 and discuss its connections with the baseline equilibrium. First, Gibrat's law holds immediately in this equilibrium. With a linear value function, the first order condition

on η implies a unique optimal search intensity, i.e.,

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

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

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

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

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

$$k(t) = 1+(k_0-1)\exp\left(-\int_0^t \tilde{L}(s)\mathrm{d}s\right), \tag{A.37}$$

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

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

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

where the second equality is obtained using equation (A.37). With $k(t)$ decreasing to one, $\eta(t)$ is decreasing to zero to equalize the supply and demand of labor. $g(\eta(t))/\eta(t)$ also decreases to zero since $g'(0) = 0$. As a result, $\tilde{L}(t)$ converges to L. It is also worth noting that the wedge between the gains from each search and the researchers' wage vanishes as η goes to zero. Following the arguments in Supplemental Appendix G, both wages and interest rates converge to those in the baseline equilibrium.

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

limiting economy, where the adjustment cost approaches zero.[46]

H.2 Proof of Proposition A.4

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 (7) implies that it is consistent to have Pareto $F(z,t)$ at all times. To see this, rewrite (7) as follows for any $z \geq 1$,

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

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

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

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

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

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

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

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

[46] Readers should be alerted that even with adjustment costs, the equilibrium in Proposition A.4 is not guaranteed to be unique. This is because my proof only shows the uniqueness of equilibrium with value functions linear in productivity. There might be equilibria with non-linear value functions and non-Pareto productivity distributions. To prove whether this equilibrium is unique *or not* is beyond the scope of this paper. It is identified as an open mathematical challenge in Achdou et al. (2014) to show the uniqueness of a solution of coupled PDE systems (the HJB and the KFE) in idea search models.

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 I Generalized learning function

In this section, I extend the learning process in the baseline model with imperfect adoption. Similarly, firms still target other more productive firms. Whereas upon meeting other firms, they are not guaranteed to replicate the other's technology. Specifically, the probability of adoption depends on the relative ranking between the searching and meeting firms. If the searching firm has productivity x and the meeting firm has productivity y ($y > x$), then the probability of adoption for the searching firm is

$$\left(\frac{1 - F(y,t)}{1 - F(x,t)}\right)^{\lambda},$$

where $\lambda \geq 0$. In this formulation, the prevalence of technology (ranking $1 - F(z,t)$) measures the level of technological sophistication and determines the difficulty of technology absorption.

I use $m(z,x,t)$ to denote the probability that a firm with productivity x increases its productivity to at least $z > x$ upon a meeting. Then,

$$m(z,x,t) = \int_z^{\infty} \left(\frac{1 - F(y,t)}{1 - F(x,t)}\right)^{\lambda} \frac{f(y,t)}{1 - F(x,t)} \mathrm{d}y = \frac{(1 - F(z,t))^{\lambda+1}}{\lambda + 1} \frac{1}{(1 - F(x,t))^{1+\lambda}}. \quad \text{(A.38)}$$

Given search intensity $\eta(x,t)$, the law of motion on productivity distribution $F(z,t)$ is as follows:

$$\frac{\partial F(z,t)}{\partial t} = -(1 - F(z,t))^{\lambda+1} \int_0^z \frac{\eta(x,t)}{\lambda + 1} \frac{f(x,t)}{(1 - F(x,t))^{1+\lambda}} \mathrm{d}x$$

It is straightforward to verify that Equation (7) of the baseline model is a special case with $\lambda = 0$. I use tilde to denote the complement CDF, e.g., $\tilde{F}(z,t) = 1 - F(z,t)$. Therefore, the above equation can be rewritten as:

$$\frac{\partial \ln \tilde{F}(z,t)}{\partial t} = \tilde{F}(z,t)^{\lambda} \int_0^z \eta(x,t) \frac{\tilde{F}(x,t)^{-(1+\lambda)}}{\lambda + 1} f(x,t) \mathrm{d}x. \quad \text{(A.39)}$$

I.1 Bounded search intensity

Suppose that the initial distribution $F(z,0)$ has tail index k_0, then $\lim_{z\to\infty}\frac{\ln\tilde{F}(z,0)}{\ln z}=-k_0$.[47] If the search intensity is bounded by a $\bar{\eta}<\infty$, i.e., $\eta(x,t)\leq\bar{\eta}$, productivity distribution $F(z,t)$ will always have tail index k_0 at all times for all $\lambda>0$. To see this, let $T(z,t;\eta)$ denote the right-hand side of (A.39) given an arbitrary path $\eta(x,t)$. Then,

$$T(z,t;\eta)=\tilde{F}(z,t)^{\lambda}\int_0^z\eta(x,t)\frac{\tilde{F}(x,t)^{-(1+\lambda)}}{\lambda+1}f(x,t)\mathrm{d}x$$

Integrating (A.39) over time gives

$$\ln\tilde{F}(z,t)=\ln\tilde{F}(z,0)+\int_0^t T(z,s;\eta)\mathrm{d}s.$$

Note that for $\lambda>0$,

$$T(z,t;\eta)\leq\frac{\bar{\eta}}{\lambda+1}\tilde{F}(z,t)^{\lambda}\int_0^z\tilde{F}(x,t)^{-(1+\lambda)}f(x,t)\mathrm{d}x=\frac{\bar{\eta}}{\lambda+1}\frac{1-\tilde{F}(z,t)^{\lambda}}{\lambda}\leq\frac{\bar{\eta}}{\lambda(\lambda+1)}.$$

Therefore, $\int_0^t T(z,s;\eta)\mathrm{d}s\leq\frac{\bar{\eta}t}{\lambda(\lambda+1)}$ for all z. Hence, for any t,

$$\lim_{z\to\infty}\frac{\ln\tilde{F}(z,t)}{\ln z}=\lim_{z\to\infty}\frac{\ln\tilde{F}(z,0)}{\ln z}+\lim_{z\to\infty}\frac{\int_0^t T(z,s;\eta)\mathrm{d}s}{\ln z}=-k_0.$$

In other words, the tail index remains the same under bounded search intensity.[48]

If $k_0>1$, it is straightforward to show that the total output, $Y(t)=\int_0^\infty zf(z)\mathrm{d}z$, is always finite at all times. Using integration by parts, $Y(t)=\int_0^\infty[1-F(z,t)]\,\mathrm{d}z$. This integral always converges with a tail index $k_0>1$ as a result of the comparison test.

I.2 Common search intensity

I intend to compare the evolution of equilibrium productivity distribution between the baseline and the generalized model. To sharpen the insight, I focus on a similar search strategy as in

[47] The Karamata's theorem (e.g., Theorem 1.5.11 and 1.6.1 of Bingham et al. (1987)) states that $f(x)$ varies regularly with index $-k-1$ if and only if $\lim_{x\to\infty}\frac{xf(x)}{\tilde{F}(x)}=k$. Using L'Hospital rule, a tail index k then implies $\lim_{x\to\infty}\frac{\ln\tilde{F}(x)}{\ln x}=-k$. While $\lim_{x\to\infty}\frac{\ln\tilde{F}(x)}{\ln x}=-k$ does not necessarily imply $\lim_{x\to\infty}\frac{xf(x)}{\tilde{F}(x)}=k$ without further smoothness conditions, one can define a weaker version of tail index k using $\lim_{x\to\infty}\frac{\ln\tilde{F}(x)}{\ln x}=k$. I slightly abuse the definition of tail index here and adopt both definitions.

[48] It is also useful to see that there will be no jumps in the tail index in a baseline model with bounded search intensity. When $\lambda=0$,

$$\frac{\partial}{\partial t}\frac{\ln\tilde{F}(z,t)}{\ln z}=\frac{T(z,t;\eta)}{\ln z}\leq-\frac{\bar{\eta}}{\lambda+1}\frac{\ln\tilde{F}(z,t)}{\ln z}.$$

With a finite tail index, the change in the tail index will be bounded, eliminating jumps. Meanwhile, it differs from the case with $\lambda>0$ where the change in tail index is always zero.

the equilibrium of the baseline model.[49] In particular, all firms adopt the same search intensity $\eta(t)$, which clears the labor market. Substituting $\eta(x,t)=\eta(t)$ into (A.39), I obtain that

$$\frac{\partial \tilde{F}(z,t)}{\partial t}=\frac{\eta(t)}{\lambda+1}\tilde{F}(z,t)\left[\frac{1-\tilde{F}(z,t)^{\lambda}}{\lambda}\right]. \tag{A.40}$$

Fixing z, the solution to this differential equation is a generalized logistic function.[50] For $\lambda>0$, the above differential equation admits the following solution given initial distribution $F(z,0)$ and a path of $\eta(t)$:

$$1-F(z,t)=\left\{1+\left[(1-F(z,0))^{-\lambda}-1\right]e^{-\frac{1}{\lambda+1}\int_0^t\eta(s)\mathrm{d}s}\right\}^{-\frac{1}{\lambda}}. \tag{A.41}$$

Next, I derive the full transition dynamics. Let $Z(t)$ be the aggregate productivity, i.e., $Z(t)=\int_0^{\infty}zf(z,t)\mathrm{d}z$. Differentiating it with respect to t and using (A.40),

$$\dot{Z}(t)=\frac{\eta(t)}{(\lambda+1)\lambda}\left\{Z(t)-\int_0^{\infty}(1-F(z,t))^{\lambda+1}\,\mathrm{d}z\right\}.$$

Meanwhile, differentiating the labor market clearing condition $\eta(t)Z(t)=L$ and combining with the above equation, I obtain that

$$\frac{\dot{\eta}(t)}{\eta(t)}=-\frac{\eta(t)}{(\lambda+1)\lambda}\left[1-\frac{\int_0^{\infty}(1-F(z,t))^{\lambda+1}\,\mathrm{d}z}{\int_0^{\infty}(1-F(z,t))\,\mathrm{d}z}\right]. \tag{A.42}$$

Given the expression of $1-F(z,t)$ in (A.41) and an initial distribution $F(z,0)$, (A.42) is a second-order ODE on $X(t)$, where $X(t)=\int_0^t\eta(s)\mathrm{d}s$. The initial conditions are $X(0)=0$, and $X'(0)=\eta(0)=L/Z(0)$. It completely characterizes the evolution of the search intensity and then the productivity distribution.

I.3 Numerical comparison

In the section, I show that the generalized model provides good approximation for the evolution of productivity distribution in the baseline equilibrium when λ is small. Despite that the mathematical tail index never changes, commonly-used tail index measures will usually vary along the transition in the generalized model. Thus, I also compare the trajectories of measured tail indices between the two models.

In practice, I use tail estimator k^f as a measured tail index, which uses two size thresholds (z_i) and their respective rankings $(1-F(z_i))$. I use z_x to denote the size thresholds corre-

[49] For a fully developed equilibrium, one needs to solve firms' HJB equation given the generalized learning function and to include non-linear search cost, e.g., the adjustment cost in Section H. However, with a Pareto tail, one can show that the expected productivity growth per unit search are asymptotically constant as $z\to\infty$. Hence, firms have similar incentives to search and choose similar search intensities. I then consider the path with common search intensity as a proxy for the full equilibrium path.

[50] It becomes a standard logistic differential equation when $\lambda=1$. Moreover, this PDE coincides with the evolution of productivity distribution in a standard random search idea flow model, such as that in Buera and Lucas (2018).

sponding to the top $x\%$ firms. Given a known CDF, I can now fix the rankings and choose the corresponding thresholds. To align with the tail estimates k^f for the U.S., I specifically target the top 0.5% and 10% firms. The measured tail index then has the following expression:

$$k^m = -\frac{\ln 0.5/10}{\ln z_{0.5}/z_{10}} \tag{A.43}$$

For the numerical exercise, I adopt the initial Pareto assumption so that the equilibrium analysis in Section 4 applies directly. Namely, I set $F(z,0) = 1 - z^{-k_0}$ for $z \geq 1$. Given this setup, it remains to determine appropriate values for the parameters k_0 and L. I calibrate the model to match the U.S. economy from 1978 to 2019. To ensure consistency with the measured tail index, I use the 1978 estimate of k^f for k_0, setting $k_0 = 1.07$. Following the approach in Section 5, I estimate $L = 0.02$ by regressing $\ln\left(k^f - 1\right)$ on t (see Figure A.10).[51] The labor market clearing condition implies that $\eta(0) = L\frac{k_0-1}{k_0} = 0.0013$.

I then consider a small deviation from the baseline model by setting $\lambda = 10^{-4}$. At this value, the probability for a top 10% firm to adopt the technology of a top 1% will be 99.98%, a slight deviation from complete certainty. Given the initial conditions, I simulate the respective economies in both the baseline and generalized models for 400 years. The evolution of productivity distribution in the baseline model ($\lambda = 0$) follows Equation (9) and (13), while it follows Equation (A.41) and (A.42) in the generalized model ($\lambda = 10^{-4}$).

I present the comparison in Figure A.11. The left panel depicts the evolution of the measured tail index k^m over 400 years in both models. The trajectories align closely, with k^m differing by less than 0.005 after 400 years. The right panel displays the productivity distributions at year 400. It plots the complementary CDF $1 - F(z, 400)$ for $z \in [0.1^{-\frac{1}{k_0}}, 0.001^{-\frac{1}{k_0}}]$. These endpoints correspond to the respective productivity thresholds of the top 0.1% and top 10% firms in the initial distribution. Notably, even after 400 years, the productivity distributions in the two models remain virtually indistinguishable.

Lastly, I compare the dynamics in log output between the two models in Figure A.12. Unsurprisingly, the generalized model eventually lags behind the baseline model due to the absence of tail growth. However, the growth trajectories remain closely aligned for the first 150 years.[52] This suggests that significant differences between the two models only emerge after 150 years. Therefore, the generalized model provides a reasonable approximation of the baseline model in terms of output growth over a substantial period.

[51] The current value of L is intended to be relatively demanding, as it implies a long-run growth rate of 2%. The numerical results would be even more compelling if one considers a lower long-run growth rate, such as the 0.3% suggested by Jones (2022). If $L = 0.003$, the growth trajectories in Figure A.12 would be nearly identical between the two models. These additional results are available upon request.

[52] The baseline model features asymptotically exponential growth. One can further prove that the generalized model features asymptotically additive growth.

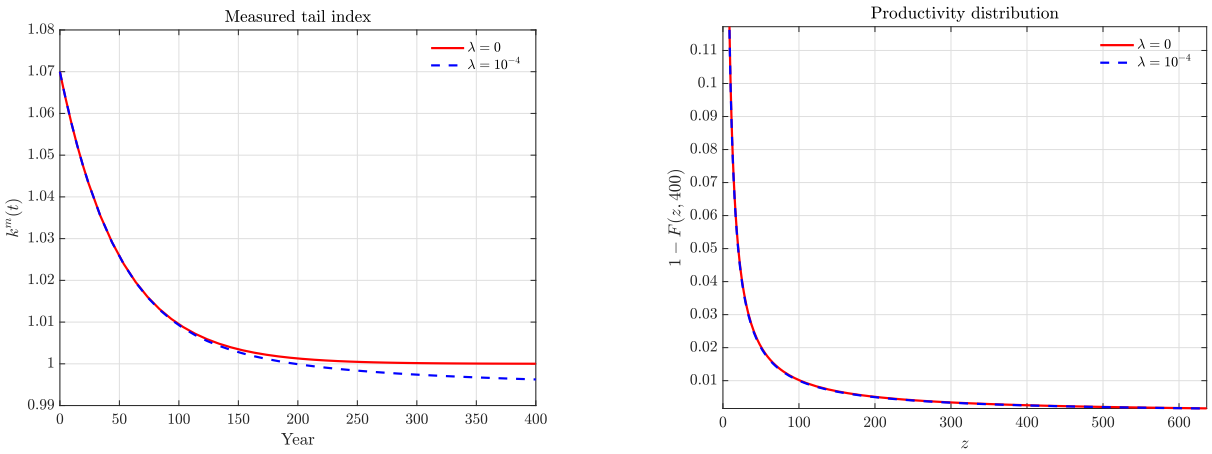


Figure A.11: Comparing productivity distributions

Notes. The left panel shows the respective time series of measured tail index k^m, as in (A.43), for the two models, whereas the right panel displays the complement CDF $1-F(z)$ of the productivity distributions in year 400. The red solid line represents the baseline model ($\lambda = 0$), and the blue dashed line represents the generalized model ($\lambda = 10^{-4}$). The initial distribution $F(z,0) = 1 - z^{-k_0}$ for $z \geq 1$. $k_0 = 1.07$, and $L = 0.02$.

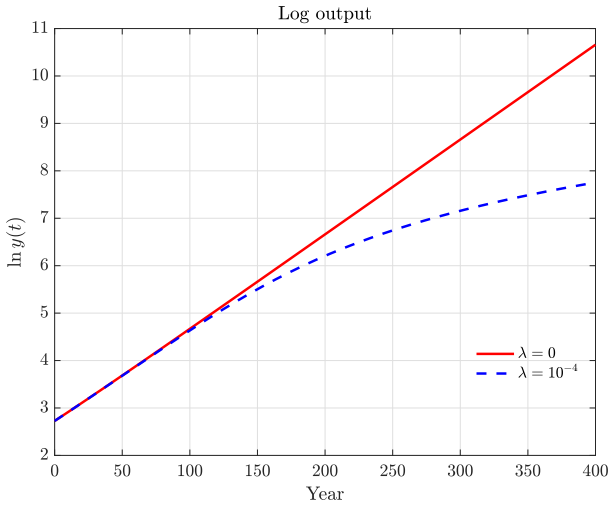


Figure A.12: Comparing growth

Notes. This figure compares the time series of log output in the two models. The red solid line represents the baseline model ($\lambda = 0$), and the blue dashed line represents the generalized model ($\lambda = 10^{-4}$). $\ln y(t) = \ln \int_1^\infty z f(z,t) \mathrm{d}z$. The initial distribution $F(z,0) = 1 - z^{-k_0}$ for $z \geq 1$. $k_0 = 1.07$, and $L = 0.02$.

I.4 Jumps in the tail index

Without an upper bound on $\eta(z,t)$, a social planner can still implement search policies that induce discrete jumps in the tail index, as observed in the baseline model. Following the approach in Section 6.2, I illustrate changes in the right tail by considering the following approximation of $\tilde{F}(z, t+h)$, where h represents an infinitesimal time lapse after implementing the policy $\eta(z,t)$

at time t. Using (A.39),

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

Meanwhile, the labor market clearing condition implies that

$$\int_0^{\infty}x\eta(x,t)f(x,t)\mathrm{d}x\leq L.$$

To obtain a sharper characterization, I restrict the distribution at the time t to be a standard Pareto distribution, i.e., $F(z,t)=1-z^{-k}$ with $z\geq 1$ and $k>1$. I also consider a specific form of search intensity: $\eta(x,t)=\eta_t x^p$, where η_t is a constant that clear the labor market at time t. $p<k-1$ so that the labor market clearing condition is respected. With these functional forms,

$$\tilde{F}(z,t+h)=\begin{cases}z^{-k}\left(1+\frac{h\eta_t k}{\lambda+1}\left[\frac{z^p-z^{-k\lambda}}{p+k\lambda}\right]\right) & \text{if } p+k\lambda\neq 0,\\ z^{-k}\left(1+\frac{h\eta_t k}{\lambda+1}\ln z\right) & \text{if } p+k\lambda=0.\end{cases}$$

It is straightforward that the tail index jumps from k to $k-p$ whenever $p>0$. It is then feasible to thicken the right tail by choosing a $p>0$ whenever $k>1$, e.g., $p=\frac{k-1}{2}$. Hence, the policy implications in Section 6.2 hold more generally with $\lambda>0$.

Appendix J Omitted proofs in Appendix D

J.1 Proof of Proposition A.1

Proof. I guess and verify that $\eta(z,t)=\eta(t)$ for $z\geq z^*(t)$ is an equilibrium search strategy. Let $Q(r,t)$ be as follows:

$$1-Q(r,t)=\frac{1-F(rz^*(t),t)}{1-F(z^*(t),t)},\quad r\geq 1.$$

Namely, $1-Q(r,t)$ is the fraction of firms with productivity at least r times the current vintage productivity among all firms whose technology are not obsolete. Alternatively, for $z\geq z^*(t)$

$$1-F(z,t)=\left(1-Q\left(\frac{z}{z^*(t)},t\right)\right)(1-F^*(t)) \tag{A.44}$$

Then,

$$\frac{\partial\ln(1-F(z,t))}{\partial t}=-\frac{\partial}{\partial r}\ln\left(1-Q\left(\frac{z}{z^*(t)},t\right)\right)\frac{z}{z^*(t)}g_z(t)+\frac{\partial}{\partial t}\ln\left(1-Q\left(\frac{z}{z^*(t)},t\right)\right)+\frac{\mathrm{d}\ln(1-F^*(t))}{\mathrm{d}t}.$$

Noticing that

$$\frac{\mathrm{d}\ln(1-F^*(t))}{\mathrm{d}t}=\frac{\partial\ln(1-F(z^*(t),t))}{\partial z}z^*(t)g_z(t)+\frac{\partial\ln(1-F(z^*(t),t))}{\partial t}$$

and

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

then

$$\frac{\mathrm{d}\ln\left(1-F^*(t)\right)}{\mathrm{d}t}=\frac{\partial \ln\left(1-Q\left(1,t\right)\right)}{\partial r}g_z(t)+\frac{\partial \ln\left(1-F(z^*(t),t)\right)}{\partial t} \tag{A.45}$$

Rearranging terms and using $r=z/z^*(t)$

$$\begin{aligned}\frac{\partial \ln\left(1-Q(r,t)\right)}{\partial t}=&\frac{\partial \ln\left(1-Q(r,t)\right)}{\partial r}rg_z(t)+\frac{\partial \ln\left(1-F(z,t)\right)}{\partial t}-\frac{\partial \ln\left(1-F(z^*(t),t)\right)}{\partial t}-\frac{\partial \ln\left(1-Q\left(1,t\right)\right)}{\partial r}g_z(t),\\ =&\left[\frac{\partial \ln\left(1-Q(r,t)\right)}{\partial r}r-\frac{\partial \ln\left(1-Q\left(1,t\right)\right)}{\partial r}\right]g_z(t)-\eta(t)\ln\frac{1-F(z,t)}{1-F(z^*(t),t)},\\ =&\left[\frac{\partial \ln\left(1-Q(r,t)\right)}{\partial r}r-\frac{\partial \ln\left(1-Q\left(1,t\right)\right)}{\partial r}\right]g_z(t)-\eta(t)\ln\left(1-Q(r,t)\right).\end{aligned}$$

The second equality comes from (A.24). It is straightforward to verify that $Q(r,t)=1-r^{-k(t)}$ and $\dot{k}(t)=-\eta(t)k(t)$ satisfy the above the partial differential equation and the initial condition $F(z,0)=Q(z,0)$. In other words, the productivity distributions in the baseline economy is the relative productivity distribution among search active firms in this economy. This is obvious given that no technology becomes obsolete in the baseline. Consequently, I have shown that $F(z,t)$ is given by (A.44) for $z\geq z^*(t)$.

With HJB equation (A.18) and (A.17), same arguments lead to that $v(z,t)=v(t)z$. As $Q(r,t)$ is always Pareto, it is straightforward to verify that the expect payoff of each search is zero with $w_r(t)=\frac{v(t)}{\psi(k(t)-1)}$. Namely,

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

I then verify that $\eta(z,t)=\eta(t)$ is an equilibrium strategy. The market clearing condition for researchers implies that $\eta(t)>0$.

Using $\frac{\partial \ln(1-Q(r,t))}{\partial r}=-\frac{k(t)}{r}$, I obtain (A.28) from (A.45) and (A.24):

$$\frac{\mathrm{d}\ln\left(1-F^*(t)\right)}{\mathrm{d}t}=-k(t)g_z(t)+g_E(t)\frac{F^*(t)}{1-F^*(t)}.$$

The proof is then complete. ■

J.2 Proof of Proposition A.2

Proof. This proof consists of three steps. First, I derive the limits k^*, e^*, f, g_M^*, and g_k^*. Then, I show that $F(z,t)$ is an asymptotic traveling wave. Finally, I obtain the long-run growth rates of $Z(t)$ and $y(t)$.

Step 1: Limits k^*, e^*, f, g_M^*, and g_k^* I start with g_M^*. Proposition A.1 indicates that $k(t)$ is decreasing and bounded from zero, it then converges to a $k^* \geq 0$. Furthermore, ODE (A.30) ensures that $k(t) \geq 1$ since $k(0) > 1$, and the RHS of (A.30) is positive. Then, $k^* \geq 1$. Solving ODE (A.31) with $e^*(t) \to e^* = (1-\tau_r)/k^* \in ((1-\tau_r)/k_0, 1-\tau_r)$ and Assumption 2, I obtain that

$$g_M^* \equiv \lim_{t\to\infty} g_M(t) = g_L^* - g_z^*, \quad \text{and} \quad \lim_{t\to\infty} \frac{L(t)}{\psi M(t) z^*(t)} = \frac{g_L^* - g_z^* + \delta}{sK(e^*)}.$$

Next, solving (A.28) given $k(t)$, $g_z(t)$, and $g_E(t)$ gives that $F^*(t)$ converges to the following stationary point f:

$$f \equiv \lim_{t\to\infty} F^*(t) = \frac{k^* g_z^*}{k^* g_z^* + g_E^*}, \tag{A.46}$$

where $g_E^* = g_M^* + \delta = g_L^* - g_z^* + \delta$.

To obtain k^*, notice that the ODE on $g_k(t)$ (A.30) implies that

$$\begin{aligned}
\lim_{t\to\infty} -\frac{\dot{k}(t)}{k(t)-1} &= \lim_{t\to\infty} \frac{sJ(e^*(t))}{1-F^*(t)} \lim_{t\to\infty} \frac{L(t)}{\psi M(t) z^*(t)} \\
&= \frac{sJ(e^*))}{1-f} \frac{g_L^* - g_z^* + \delta}{sK(e^*)} \\
&= \frac{J(e^*)}{K(e^*)} (k^* g_z^* + g_E^*),
\end{aligned}$$

where I use $1 - f = \frac{g_E^*}{k^* g_z^* + g_E^*}$. Since $k^* \geq 1$ and $g_z^* > 0$, $k^* g_z^* + g_E^* > 0$. Then, $k(t) - 1$ decreases to 0 at an asymptotically constant and positive rate. Consequently,

$$k^* = 1, \quad e^* = 1 - \tau_r, \quad f = \frac{g_z^*}{g_L^* + \delta}, \quad \text{and} \quad g_k^* = \frac{J(e^*)}{K(e^*)} (g_L^* + \delta).$$

Step 2: Asymptotic traveling wave I define $Q^l(r,t)$ and $Q^r(r,t)$ as follows:

$$Q^l(r,t) = \frac{F(rz^*(t),t)}{F(z^*(t),t)}, \quad \text{and } 1 - Q^r(r,t) = \frac{1 - F(rz^*(t),t)}{1 - F(z^*(t),t)}.$$

Namely, Q^l for $r \in (0,1)$ and Q^r for $r \geq 1$ are the respective distributions for search-active and search-inactive firms. The following lemma characterizes their limiting distributions.

Lemma A.2. *With Assumption 2,*

$$\lim_{t\to\infty} Q^l(r,t) = r^\varsigma \quad \textit{for } r \in (0,1), \tag{A.47}$$

$$\lim_{t\to\infty} Q^r(r,t) = 1 - r^{-1} \quad \textit{for } r \geq 1, \tag{A.48}$$

where $\varsigma = \frac{g_E^*}{g_z^*}$.

Proof. It is evident that (A.48) follows directly from Proposition A.1 and $k^* = 1$. Meanwhile, Proposition A.1 also implies that[53]

$$Q^l(r,t) = \frac{F^*(T^*(rz^*(t)))}{F^*(T^*(z))} e^{-\int_{T^*(rz*(t))}^{t} g_E(s)\mathrm{d}s} = \frac{F^*(T^*(rz^*(t)))}{F^*(T^*(z))} e^{-\int_{T^*(rz^*(t))}^{T^*(z*(t))} g_E(s)\mathrm{d}s},$$

where the last equality uses that $t = T^*(z^*(t))$. Since $g_z^* > 0$, $z^*(t) \to \infty$ as $t \to \infty$. $T^*(rz^*(t))$ also goes to infinity since $rz^*(t) \to \infty$. Consequently,

$$\lim_{t\to\infty} F^*(T^*(rz^*(t))) = \lim_{t\to\infty} F^*(T^*(z^*(t))) = \lim_{t\to\infty} F^*(t) = f.$$

On the other hand, the mean value theorem implies that

$$\int_{T^*(rz^*(t))}^{T^*(z*(t))} g_E(s)\mathrm{d}s = \int_{T^*(rz^*(t))}^{T^*(z*(t))} \frac{g_E(s)}{g_z(s)} g_z(s)\mathrm{d}s = \frac{g_E(\tilde{t})}{g_z(\tilde{t})} \int_{T^*(rz^*(t))}^{T^*(z*(t))} g_z(s)\mathrm{d}s$$

for some $\tilde{t} \in (T^*(rz^*(t)), T^*(z^*(t)))$. MVT applies since g_E and g_z are both positive and continuous and T^* is monotonic. Hence,

$$e^{-\int_{T^*(rz^*(t))}^{T^*(z*(t))} g_E(s)\mathrm{d}s} = \left(e^{\int_{T^*(rz^*(t))}^{T^*(z*(t))} g_z(s)\mathrm{d}s}\right)^{-\frac{g_E(\tilde{t})}{g_z(\tilde{t})}} = \left(\frac{e^{\int_0^{T^*(z*(t))} g_z(s)\mathrm{d}s}}{e^{\int_0^{T^*(rz*(t))} g_z(s)\mathrm{d}s}}\right)^{-\frac{g_E(\tilde{t})}{g_z(\tilde{t})}} = r^{\frac{g_E(\tilde{t})}{g_z(\tilde{t})}},$$

where the last equality follows that $z = e^{\int_0^{T^*(z)} g_z(s)\mathrm{d}s}$. Moreover, $\tilde{t} \to \infty$ as both $T^*(rz^*(t))$ and $T^*(z^*(t))$ go to infinity. Then, $\frac{g_E(\tilde{t})}{g_z(\tilde{t})} \to \varsigma \equiv \frac{g_E^*}{g_z^*}$. In sum,

$$\lim_{t\to\infty} Q^l(r,t) = \frac{\lim_{t\to\infty} F^*(T^*(rz^*(t)))}{\lim_{t\to\infty} F^*(T^*(z))} \lim_{t\to\infty} e^{-\int_{T^*(rz^*(t))}^{T^*(z*(t))} g_E(s)\mathrm{d}s} = r^{\varsigma}.$$

■

Given the limiting distributions of Q^l and Q^r, it is straightforward that $F(z,t)$ is a traveling wave with velocity $z^*(t)$ and takes a limiting distribution Γ in (A.35).

Step 3: Long-run growth rates of $Z(t)$ and $y(t)$ I first characterize the growth in $Z(t)$. Note that

$$\begin{aligned}
Z(t) =& \int_1^{z^*(t)} z\mathrm{d}F(z,t) + \int_{z^*(t)}^{\infty} z\mathrm{d}F(z,t) \\
=& z^*(t)F(z^*(t),t) - \int_1^{z^*(t)} F(z,t)\mathrm{d}z + z^*(t)\left[1 - F(z^*(t),t)\right] + \int_{z^*(t)}^{\infty} \left[1 - F(z,t)\right]\mathrm{d}z \\
=& z^*(t) - \int_1^{z^*(t)} F(z,t)\mathrm{d}z + \int_{z^*(t)}^{\infty} \left[1 - F(z,t)\right]\mathrm{d}z
\end{aligned}$$

[53] Technically, the lower support of $Q^l(r,t)$ is $\frac{1}{z^*(t)}$. However, for any $r \in (0,1)$, there exists T such that r is in the support of $Q^l(r,t)$ if $t > T$.

Using $z = rz^*(t)$ for substitution,

$$\int_1^{z^*(t)} F(z,t)\mathrm{d}z = z^*(t)\int_{\frac{1}{z^*(t)}}^1 F(rz^*(t),t)\mathrm{d}r$$

Since $\lim_{t\to\infty} F(rz^*(t),t) = fr^\varsigma$, dominated convergence theorem implies that

$$\lim_{t\to\infty}\int_{\frac{1}{z^*(t)}}^1 F(rz^*(t),t)\mathrm{d}r = \int_0^1 fr^\varsigma \mathrm{d}r = \frac{f}{\varsigma+1}$$

On the other hand, Proposition A.1 implies that

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

Combining all these terms,

$$\frac{Z(t)}{z^*(t)} = 1 - \int_{\frac{1}{z^*(t)}}^1 F(rz^*(t),t)\mathrm{d}r + \frac{1-F^*(t)}{k(t)-1} \sim 1 - \frac{f}{\varsigma+1} + \frac{1-f}{k(t)-1} \sim \frac{1-f}{k(t)-1}$$

as $t=\infty$.[54] So $Z(t)/z^*(t)$ behaves like $1/(k(t)-1)$ as $t\to\infty$. To formally establish the growth rate $g_Z \equiv \dot{Z}(t)/Z(t)$, I note that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{Z(t)}{z^*(t)} = -\int_{\frac{1}{z^*(t)}}^1 \frac{\mathrm{d}F(rz^*(t),t)}{\mathrm{d}t}\mathrm{d}r - \frac{\mathrm{d}F^*(t)}{k(t)-1} - \frac{[1-F^*(t)]\,\dot{k}(t)}{(k(t)-1)^2}$$

I show that the first term vanishes since $\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}F(rz^*(t),t)}{\mathrm{d}t} \to 0$ as $t\to\infty$. To see this, totally differentiating $F(rz^*(t),t)$ in (A.27) with respect to t gives us

$$\begin{aligned}\frac{\mathrm{d}F(rz^*(t),t)}{\mathrm{d}t} &= (F^*)'\left(T^*(rz^*(t))\right)(T^*)'(rz^*(t))r(z^*)'(t)e^{-\int_{T^*(rz^*(t))}^t g_E(s)\mathrm{d}s} \\ &\quad + F\left(T^*\left(rz^*(t)\right)\right)e^{-\int_{T^*(rz^*(t))}^t g_E(s)\mathrm{d}s}\left[-g_E(t) + g_E\left(T^*(rz^*(t))\right)(T^*)'(rz^*(t))r(z^*)'(t)\right]\end{aligned}$$

To simplify the above derivative, I first obtain the following by totally differentaiting $z = e^{\int_0^{T^*(z)} g_z(s)\mathrm{d}s}$ w.r.t. z:

$$1 = zg_z(T^*(z))(T^*)'(z). \tag{A.49}$$

Then,

$$\begin{aligned}\lim_{t\to\infty}(T^*)'(rz^*(t))r(z^*)'(t) &= \lim_{t\to\infty}(T^*)'(rz^*(t))rz^*(t)g_z(t) \\ &= \lim_{t\to\infty}(T^*)'(rz^*(t))rz^*(t)g_z(T^*(rz^*(t))) = 1.\end{aligned}$$

The second equality uses that $g_z(t) = g_z(T^*(z^*(t))) \sim g_z(T^*(rz^*(t)))$ at $t=\infty$. The last

[54] It is tempting to conclude from here that $Z(t)/z^*(t)$ has asympotitic growth rate $g_k(t)$. I need to verify first that the derivative of the constant part (the left tail) is well behaved so that $\lim_{t\to\infty}\dot{Z}(t)$ is well defined.

equality uses the above identity (A.49) at $z = rz^*(t)$. Next, I obtain from ODE (A.28) that $\lim_{t\to\infty}(F^*)'(t) = 0$.

Therefore, the derivative has a limit of zero:

$$\begin{aligned}\lim_{t\to\infty}\frac{\mathrm{d}F(rz^*(t),t)}{\mathrm{d}t} =& 0\times 1\times\lim_{t\to\infty}e^{-\int_{T^*(rz^*(t))}^{t}g_E(s)\mathrm{d}s}+f\times\lim_{t\to\infty}e^{-\int_{T^*(rz^*(t))}^{t}g_E(s)\mathrm{d}s}\left[-g_E^*+g_E^*\times 1\right]\\ =&0,\end{aligned}$$

since I show in the last step that $\lim_{t\to\infty}e^{-\int_{T^*(rz^*(t))}^{t}g_E(s)\mathrm{d}s}=r^{\varsigma}$. From here,

$$\frac{\frac{\mathrm{d}}{\mathrm{d}t}Z(t)/z^*(t)}{Z(t)/z^*(t)}\sim\frac{-\frac{(F^*)'(t)}{k(t)-1}-\frac{[1-F^*(t)]\dot{k}(t)}{(k(t)-1)^2}}{1-\frac{f}{\varsigma+1}+\frac{1-f}{k(t)-1}}\sim-\frac{(F^*)'(t)}{1-f}-\frac{\dot{k}(t)}{k(t)-1}\sim g_k(t)$$

at $t=\infty$. Then,

$$\lim_{z\to\infty}\frac{\dot{Z}(t)}{Z(t)}-g_z(t)=g_k^*.$$

Lastly, I derive the growth rate of household income $g_y(t)$. Notice that

$$\begin{aligned}&R(t)=M(t)\int_1^{\infty}r(z,t)\mathrm{d}F(z,t)=M(t)R(t)\left(\frac{\sigma}{\sigma-1}w_p(t)\right)^{1-\sigma}Z(t)\\ \Rightarrow\quad&M(t)Z(t)=\left(\frac{\sigma}{\sigma-1}w_p(t)\right)^{\sigma-1}\\ \Rightarrow\quad&w_p(t)=\frac{\sigma-1}{\sigma}\left(M(t)Z(t)\right)^{\frac{1}{\sigma-1}}.\end{aligned}$$

The first line follows from firms' static choices in (A.16). Firms' pricing rule also implies that $\frac{\sigma-1}{\sigma}R(t)=w_p(t)(1-s)L(t)$. Since households' total income is equal to firms' total revenue $R(t)$,

$$y(t)=\frac{R(t)}{L(t)}=\frac{\sigma}{\sigma-1}\frac{w_p(t)(1-s)L(t)}{L(t)}=\frac{\sigma(1-s)}{\sigma-1}w_p(t)=(1-s)\left(M(t)Z(t)\right)^{\frac{1}{\sigma-1}}.$$

Given that $\lim_{t\to\infty}\frac{\dot{Z}(t)}{Z(t)}=g_z^*+g_k^*$, and $g_M^*=g_L^*-g_z^*$,

$$g_y^*=g_{w_p}^*\equiv\lim_{t\to\infty}\frac{\dot{w}_p(t)}{w_p(t)}=\frac{g_k^*+g_L^*}{\sigma-1}.$$

In addition, given that $M(t)\sim\frac{L(t)}{\psi z^*(t)}\frac{sK(e^*)}{g_L^*-g_z^*+\delta}$ and $\frac{Z(t)}{z^*(t)}\sim 1-\frac{f}{\varsigma+1}+\frac{1-f}{k(t)-1}$,

$$y(t)\sim(1-s)\left(\frac{sK(e^*)L(t)}{\psi\left(g_L^*-g_z^*+\delta\right)}\left((1-f)\frac{k(t)}{k(t)-1}+\frac{\varsigma}{\varsigma+1}f\right)\right)^{\frac{1}{\sigma-1}}.$$

■

Appendix K Learning from an external source

I now introduce an external source on which firms can make productivity draws. To illustrate the mechanics of the learning process, I follow Lucas and Moll (2014) to consider a special case of passive learning. Namely, ideas come to all firms at the same rate α_t if they are from an internal source, and at rate β_t if an external source. The productivity distribution $F(z,t)$ then has the following law of motion:

$$\begin{aligned}\frac{\partial F(z,t)}{\partial t} =& -\alpha_t\left[1-F(z,t)\right]\int_0^z \frac{f(x,t)}{1-F(x,t)}\mathrm{d}x - \beta_t\left[1-G(z)\right]F(z,t)\\ =&\alpha_t\left[1-F(z,t)\right]\ln\left(1-F(z,t)\right) - \beta_t\left[1-G(z)\right]F(z,t) \end{aligned} \tag{A.50}$$

The first part is the same as before: firms only draw among more productive firms when opportunities realize. The second part indicates that firms innovate by randomly sampling from a pool of ideas with productivity distribution G ("the universe of ideas") as in Kortum (1997).

We define that a distribution F has tail index k if

$$\lim_{x\to\infty}\frac{\ln\left(1-F(x)\right)}{\ln x} = -k, \quad k\in\mathbb{R}^+\cup\{+\infty\}.$$

It is equivalent to the exponent of a regular varying function if $\lim_{x\to\infty}\frac{xf(x)}{1-F(x)}$ exists. We assume that the initial distribution $F(z,0)$ and external source $G(z)$ have respective tails k_0 and k_e. The following proposition characterizes the tail index of the resulting productivity distributions.

Proposition A.5. *$F(z,t)$ has tail index $\min\{k_0,k_e\}e^{-\int_0^t\alpha_s\mathrm{d}s}$.*

Proof. I establish the tail index of $F(\cdot,t)$ in several steps. It is difficult to compute the tail index of $F(z,t)$ directly, so the idea is to find upper and lower bounds (functions) of $F(z,t)$ and show that these bounds have the same tail index.

Step 0: In this step, I contruct the lower bound $F^l(z,t;\tau)$ and $F^u(z,t)$ as follows. $F^l(z,t;\tau)$ is the solution to the following differential equation:

$$\frac{\partial F(z,t)}{\partial t} = \begin{cases} -\beta_t\left[1-G(z)\right]F(z,t), & \text{if } t\in[0,\tau]\\ \alpha_t\left[1-F(z,t)\right]\ln\left(1-F(z,t)\right) - \beta_t\left[1-G(z)\right]F(z,t), & \text{if } t>\tau, \end{cases}$$

with initial distribution $F(z,0)$, and $\tau>0$. In words, F^l corresponds to the process in which there is only external learning in the first τ years. It is straightforward that $F^l(z,t;\tau)\geq F(z,t)$ for any z,t and τ. Furthermore, $F^l(z,t;\tau)$ decreases stochastically in τ.

The upper bound $F^u(z,t)$ is the solution to the same differential equation in (A.50) but with a different initial distribution

$$F^u(z,0) = F(z,0)G(z).$$

Intuitively, $F^u(z,0)$ implies that the initial distribution is the maximum statistics of the internal and external sources. It follows that it stochastically dominates $F(z,0)$, a distribution from only internal source. Since (A.50) implies that $\partial F(z,t)/\partial t < 0$ for any z and t, then $F^u(z,t) \leq F(z,t)$ for any z and t.

Step 1: In this step, I establish the tail index of $F^l(z,t;\tau)$ through Lemma A.3 and A.4.

Lemma A.3. *$F^l(z,t;\tau)$ has tail index $\min\{k_0,k_e\}$ on $t \in [0,\tau]$.*

Proof. For $t \in [0,\tau]$,

$$\frac{\partial F^l(z,t;\tau)}{\partial t} = -\beta_t \left[1-G(z)\right] F^l(z,t;\tau) \Rightarrow F^l(z,t;\tau) = F(z,0)\exp\left(-\left[1-G(z)\right]\int_0^t \beta_s \mathrm{d}s\right).$$

On the one hand,

$$1-F^l(z,t;\tau) = \left[1-G(z)\right]\left[\frac{1-F(z,0)}{1-G(z)} + F(z,0)\frac{1-\exp\left(-\left[1-G(z)\right]\int_0^t \beta_s \mathrm{d}s\right)}{1-G(z)}\right] \tag{A.51}$$

$$= \left[1-F(z,0)\right]\left[1 + F(z,0)\frac{1-\exp\left(-\left[1-G(z)\right]\int_0^t \beta_s \mathrm{d}s\right)}{1-G(z)}\frac{1-G(z)}{1-F(z,0)}\right] \tag{A.52}$$

On the other hand,

$$\ln\frac{1-F(z,0)}{1-G(z)} \sim -(k_0-k_e)\ln z$$

when $z \to \infty$. Hence, I discuss the cases whenever $\frac{1-F(z,0)}{1-G(z)}$ is bounded or not.

If it is bounded, it must be $k_0 \geq k_e$. Note that when $z \to \infty$,

$$F(z,0)\frac{1-\exp\left(-\left[1-G(z)\right]\int_0^t \beta_s \mathrm{d}s\right)}{1-G(z)} \to \int_0^t \beta_s \mathrm{d}s.$$

(A.51) implies that

$$\lim_{z\to\infty}\frac{\ln\left(1-F^l(z,t;\tau)\right)}{\ln z} = \lim_{z\to\infty}\frac{\ln\left(1-G(z)\right)}{\ln z} = -k_e = -\min\{k_0,k_e\}.$$

Conversely, an unbounded $\frac{1-F(z,0)}{1-G(z)}$ implies that $k_0 \leq k_e$. Notice that

$$\lim_{z\to\infty}\frac{\ln\left\{1+F(z,0)\frac{1-\exp\left(-\left[1-G(z)\right]\int_0^t \beta_s \mathrm{d}s\right)}{1-G(z)}\frac{1-G(z)}{1-F(z,0)}\right\}}{\ln z} = 0. \tag{A.53}$$

This equation is trivial if $\frac{1-G(z)}{1-F(z,0)}$ is also bounded. Otherwise, I must have $k_0 = k_e$, and $\lim_{z\to\infty}\frac{\ln\left\{\frac{1-G(z)}{1-F(z,0)}\right\}}{\ln z} = 0$. It follows that (A.53) holds from its asymptotic equivalence with

$\frac{\ln\left\{\frac{1-G(z)}{1-F(z,0)}\right\}}{\ln z}$. In all cases, (A.52) gives

$$\lim_{z\to\infty}\frac{\ln\left(1-F^l(z,t;\tau)\right)}{\ln z}=\lim_{z\to\infty}\frac{\ln\left(1-F(z,0)\right)}{\ln z}=-k_0=-\min\{k_0,k_e\}.$$

The proof is then complete. ■

Lemma A.4. *For $t\geq\tau$, $\liminf_{z\to\infty}\frac{\ln\left(1-F^l(z,t;\tau)\right)}{\ln z}\geq-\min\{k_0,k_e\}e^{-\int_\tau^t\alpha_s\mathrm{d}s}$.*

Proof. For $t\geq\tau$, the dynamics follows the differential equation (A.50). Rewriting it gives

$$w'(t)=-\alpha_t w(t)+\beta_t u\left(e^{-w(t)}-1\right) \tag{A.54}$$

in which z is fixed and suppressed, $w(t)=\ln[1-F(z,t)]$, and $u=1-G(z)$. Then, for any $t>\tau$,

$$w(t)=e^{-\int_\tau^t\alpha_s\mathrm{d}s}w(\tau)+u\int_\tau^t e^{-\int_s^t\alpha_r\mathrm{d}r}\beta_s\left(e^{-w(s)}-1\right)\mathrm{d}s.$$

Substituting $G(z)$ and $F^l(z,t;\tau)$ into it,

$$\frac{\ln\left(1-F^l(z,t;\tau)\right)}{\ln z}=e^{-\int_\tau^t\alpha_s\mathrm{d}s}\frac{\ln\left(1-F^l(z,\tau;\tau)\right)}{\ln z}+\frac{1-G(z)}{\ln z}\int_\tau^t e^{-\int_s^t\alpha_r\mathrm{d}r}\beta_s\frac{F^l(z,s;\tau)}{1-F^l(z,s;\tau)}\mathrm{d}s.$$

Then,

$$\liminf_{z\to\infty}\frac{\ln\left(1-F^l(z,t;\tau)\right)}{\ln z}\geq\lim_{z\to\infty}e^{-\int_\tau^t\alpha_s\mathrm{d}s}\frac{\ln\left(1-F^l(z,\tau;\tau)\right)}{\ln z}=-\min\{k_e,k_0\}e^{-\int_\tau^t\alpha_s\mathrm{d}s}$$

■

Step 2: In this step, I establish the tail index of the upper bound $F^u(z,t)$ in Lemma A.5.

Lemma A.5. *$F^u(z,t)$ has tail index $\min\{k_0,k_e\}e^{-\int_0^t\alpha_s\mathrm{d}s}$.*

Proof. It follows from the previous proof that

$$\frac{\ln\left(1-F^u(z,t)\right)}{\ln z}=e^{-\int_0^t\alpha_s\mathrm{d}s}\frac{\ln\left(1-F^u(z,0)\right)}{\ln z}+\frac{1-G(z)}{\ln z}\int_0^t e^{-\int_s^t\alpha_r\mathrm{d}r}\beta_s\frac{F^u(z,s)}{1-F^u(z,s)}\mathrm{d}s. \tag{A.55}$$

First, notice that

$$\frac{\ln\left(1-F^u(z,0)\right)}{\ln z}=\frac{\ln\left[1-F(z,0)+F(z,0)\left(1-G(z)\right)\right]}{\ln z}.$$

The same arguments that prove Lemma A.3 apply to show that $F^u(z,0)$ has tail index $\min\{k_e,k_0\}$.

Next, (A.55) implies that

$$\liminf_{z\to\infty}\frac{\ln\left(1-F^u(z,t)\right)}{\ln z}\geq\lim_{z\to\infty}e^{-\int_0^t\alpha_s\mathrm{d}s}\frac{\ln\left(1-F^u(z,0)\right)}{\ln z}=-\min\{k_e,k_0\}e^{-\int_0^t\alpha_s\mathrm{d}s}$$

Alternatively, $\frac{F^u(z,s)}{1-F^u(z,s)}$ decreases on s since $F^u(z,s)$ decreases on s. Then,

$$\frac{\ln\left(1-F^u(z,t)\right)}{\ln z} \leq e^{-\int_0^t \alpha_s \mathrm{d}s} \frac{\ln\left(1-F^u(z,0)\right)}{\ln z} + \frac{F^u(z,0)}{\ln z} \frac{1-G(z)}{1-F^u(F,0)} \int_0^t e^{-\int_s^t \alpha_r \mathrm{d}r} \beta_s \mathrm{d}s.$$

By construction, $1-F^u(z,0) > 1-G(z)$. Then, for any t,

$$\frac{F^u(z,0)}{\ln z} \frac{1-G(z)}{1-F^u(F,0)} \int_0^t e^{-\int_s^t \alpha_r \mathrm{d}r} \beta_s \mathrm{d}s = O\left(\frac{1}{\ln z}\right) = o(1)$$

at $z=\infty$. Consequently,

$$\limsup_{z\to\infty} \frac{\ln\left(1-F^u(z,t)\right)}{\ln z} \leq e^{-\int_0^t \alpha_s \mathrm{d}s} \lim_{z\to\infty} \frac{\ln\left(1-F^u(z,0)\right)}{\ln z} = -\min\{k_e, k_0\} e^{-\int_0^t \alpha_s \mathrm{d}s}.$$

Combining both bounds, I conclude that $\lim_{z\to\infty} \frac{\ln(1-F^u(z,t))}{\ln z} = -\min\{k_e, k_0\} e^{-\int_0^t \alpha_s \mathrm{d}s}$. ■

Step 3: I return to the tail index of $F(z,t)$ in this last step. From Lemma A.3 and A.4, I obtain that

$$\liminf_{z\to\infty} \frac{\ln\left(1-F(z,t)\right)}{\ln z} \geq \liminf_{z\to\infty} \frac{\ln\left(1-F^l(z,t;\tau)\right)}{\ln z} \geq -\min\{k_0, k_e\} e^{-\int_\tau^t \alpha_s \mathrm{d}s}$$

Since τ is arbitrary,

$$\liminf_{z\to\infty} \frac{\ln\left(1-F(z,t)\right)}{\ln z} \geq -\min\{k_0, k_e\} e^{-\int_0^t \alpha_s \mathrm{d}s}.$$

Conversely, Lemma A.5 implies that

$$\limsup_{z\to\infty} \frac{\ln\left(1-F(z,t)\right)}{\ln z} \leq \lim_{z\to\infty} \frac{\ln\left(1-F^u(z,t)\right)}{\ln z} = -\min\{k_e, k_0\} e^{-\int_0^t \alpha_s \mathrm{d}s}.$$

The proof is complete by combining these two bounds. ■

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.

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

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

Jones, C. I. (2022): "The Past and Future of Economic Growth: A Semi-Endogenous Perspective," *Annual Review of Economics*, 14, 125–152.

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

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

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