Economic Growth and the Rise of Large Firms

Zhang Chen*

January 2026

Abstract

I document that the right tail of the firm size distribution systematically thickens with economic development. To rationalize this fact, I develop a parsimonious idea search model in which both aggregate growth and the firm size distribution are endogenously determined. The model features an asymptotic balanced growth path along which Gibrat's law holds at each date, and the right tail of the firm size distribution thickens monotonically toward Zipf's law. The model also implies that policies favoring large firms can improve welfare by better utilizing the diffusion externalities arising from idea search.

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

*Department of Economics, HKUST, Hong Kong. I am greatly indebted to 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, various seminar participants, and particularly Manuel García-Santana. This research received generous financial support from the International Economics Section (IES) at Princeton University and the STEG Ph.D. Grant. I thank the Einaudi Institute for Economics and Finance (EIEF) for its hospitality, where significant revisions of the paper were made during my time as a postdoc. 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 often viewed as symbols of economic success, a defining feature of developing economies is the prevalence of small firms. Understanding how firm size distributions evolve with development is therefore a central question of economic growth. This paper focuses on one salient but understudied aspect of this evolution: the right tail of the firm size distribution.

I contribute to this issue with new evidence, theory, and policy implications. Empirically, I document that the right tail of the firm size distribution systematically thickens with economic development. Theoretically, I develop an endogenous growth model based on idea search that rationalizes this relationship as a feature of the transition dynamics. The model features an asymptotic balanced growth path in which Gibrat's law holds at all times, and the firm size distribution converges to Zipf's distribution with a progressively thickening right tail. On the policy side, the model sheds light on how policies favoring large firms can improve social welfare in the presence of diffusion externalities associated with idea search.

I begin by establishing the empirical relationship between economic development and the right tail thickness. My primary datasets are firm size tabulations from the OECD Structural Business Statistics and the U.S. Census Business Dynamics Statistics. I estimate the (Pareto) tail index of the firm size distribution at the country-year level using methods designed for tabulated data. For OECD countries, the tail index declines with log GDP per capita within countries over time. A similar pattern holds for the United States, where the tail index has fallen steadily over the past four decades. These results indicate that the right tail of the firm size distribution systematically thickens as an economy grows. Additional robustness checks confirm that this relationship holds across major sectors, across countries, over extended time horizons, and within the subset of large firms.

The persistence of this relationship suggests that thickening right tails might be an intrinsic feature of economic growth. To pursue this idea, I first propose a novel learning mechanism that explains how firms' social learning can thicken the right tail of the productivity distribution. I then embed this mechanism into a parsimonious growth model and show that the right tail of the firm size distribution thickens along an asymptotic balanced growth path.

The model economy consists of a unit continuum of firms with heterogeneous productivity. Firms increase productivity by learning from more productive firms via random meetings. Meetings between firms occur as Poisson events: firms decide on how much to invest in idea search, which determines the arrival rate of meeting opportunities. When a meeting occurs, 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 drives aggregate growth, and the collective search activities of firms continuously reshape the productivity distribution.

The model possesses a tractable property: if the distribution of firm productivity is Pareto at time t, it remains Pareto at $t+h$ for $h>0$. Assuming a Pareto initial distribution, I obtain a closed-form characterization for the entire 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 independent of firm size. Third, the productivity distribution remains Pareto with a time-varying tail index $k(t)$, i.e., $F(z,t) = 1 - z^{-k(t)}$ for productivity $z \geq 1$. Fourth, the tail index $k(t)$ decreases toward one over time. In other words, the productivity distribution develops progressively thicker right tails and converges to Zipf's distribution, a Pareto distribution with tail index equal to one.

Gibrat's law highlights the key distinction between the learning assumption in my model and those in existing idea flow models. Earlier models such as Lucas and Moll (2014) and Perla and Tonetti (2014) assume *random sampling*, where firms draw ideas from all firms in the economy and adopt them only if the sampled firm is more productive. Consequently, more productive (larger) firms are less likely to encounter superior ideas and exhibit lower expected growth. In contrast, Gibrat's law is restored in my model through *truncated sampling*, whereby firms sample ideas exclusively from more productive firms. Learning thus remains efficient even for large firms, as they draw from better sources.

This departure of my model yields three new insights into the relationship between firm size distributions and economic growth. First, the model uncovers a novel source of endogenous growth, which I term *tail growth*: growth sustained by the thickening of the right tail. In standard firm-based growth models, productivity distributions are stationary along the balanced growth path. Therefore, productivity gains from reallocation dynamics that reshapes the distribution are purely transitory and do not contribute to the long-run growth. In contrast, learning through truncated sampling provides a concrete mechanism through which reallocation gains persist and contribute directly to long-run economic growth.

Second, the model rationalizes rising market concentration as a secular trend.

Recent evidence by Kwon et al. (2024) suggests that market concentration in the United States has increased persistently over the past century. This pattern poses a theoretical challenge, as existing growth models typically feature balanced growth paths with stationary firm size distributions, making it difficult to reconcile rising concentration with stable aggregate growth. The present model captures this pattern through the thickening of the right tail, demonstrating that rising market concentration can coexist with nearly constant aggregate growth.

Third, the model provides a growth-based theory of Zipf's law. In the context of firms, Zipf's law states that advanced economies exhibit firm size distributions close to Zipf's distribution. With tail growth, economic development is driven by the thickening of the right tail, leading the firm size distribution to converge toward Zipf's law. Richer economies are further displaced along the growth path, so their distributions lie closely to the limit. Moreover, convergence to Zipf's law requires only minimal restrictions on model parameters.

Despite its parsimony, the model makes theoretical predictions consistent with the data. Two predictions about the tail index $k(t)$ are notable: (1) output per capita $y(t)$ is proportional to $k(t)/(k(t)-1)$, and (2) $k(t)-1$ decreases to 0 at a constant rate. Both predictions can be readily tested using the estimated tail indices. For the U.S. from 1978 to 2019, the observed growth in output per capita closely tracks changes in $k(t)/(k(t)-1)$. Plotting $\ln(k(t)-1)$ against time t also reveals a clear linear trend, further confirmed by a high correlation coefficient.

The model also delivers new implications for policies. Individual idea search generates a diffusion externality since it shapes the future productivity distribution and thus the efficiency of subsequent learning. Search by large firms disproportionately thickens the right tail and benefits the entire economy, whereas search by small firms has limited spillovers. As a result, large firms under-invest in idea search relative to the social optimum.

Policies that promote idea search among large firms can thus improve welfare. In particular, restricting search by small firms reallocates resources toward high-productivity firms and raises the long-run growth rate. Solving the planner's problem further reveals that the socially optimal search intensity increases with firm productivity, in contrast to the uniform search intensity in equilibrium.

Finally, I explore several extensions that relax the simplifying 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 the major equilibrium properties. Moreover, the extended model confirms that scale effects can be absent with tail growth and further generalizes the key insight from the baseline model that tail growth represents a persistent source of reallocation gains that contribute directly to long-run growth.

The remainder of the paper is organized as follows. Section 2 presents the empirical evidence. Section 3 develops the simple learning model that illustrates the mechanism of tail thickening. Section 4 introduces the growth model and characterizes its equilibrium. Section 5 confronts the growth model's predictions with data. Section 6 studies the policy implications. Section 7 discusses extensions, and Section 8 concludes. The supplement, Chen (2025), contains all the appendices.

Related Literature This paper contributes to four strands of literature.

First, it relates to studies on the relationship between economic development and firm size distributions.[1] Most existing work focuses on cross-country comparisons of average firm size and attributes the prevalence of small firms in poorer economies to distortions. Closely related, Poschke (2018) documents that firm size dispersion increases with economic development and offers an explanation based on technological change. This paper extends that literature by examining the right tail of the firm size distribution and developing a microfounded model of technological progress that drives growth through the thickening of right tail.

Second, it connects to recent research documenting the rise in market concentration in the United States over the past several decades (Gutiérrez and Philippon, 2017; De Loecker et al., 2020; Autor et al., 2020). In particular, Cao et al. (2022) and Chen et al. (2023) provide both theory and evidence of thickening right tails in firm size distributions in the U.S. and North America. In their models, the tail index of the stationary distribution is endogenized through firms' innovation decisions and shifts in response to exogenous technological shocks.[2] By contrast, this paper focuses on the transition dynamics of the firm size distribution rather than steady state comparisons. The model proposed here rationalizes thickening right tails as an inherent feature of the growth process. Kwon et al. (2024) provide perhaps the most directly related empirical evidence, showing that the rise in market concentration in the United States has persisted for roughly a century.

Third, this paper builds upon a burgeoning literature of growth models based on idea flows. Buera and Lucas (2018) provide a comprehensive overview of this

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

[2] Relatedly, Oberfield (2018) and Gouin-Bonenfant (2022) develop tractable models of firm dynamics that explicitly link lower labor shares to thicker right tails.

literature.[3] Following the seminal contribution of Kortum (1997), balanced growth paths in most existing models arise from the random sampling of ideas from Pareto distributions and feature stationary productivity distributions. The model developed here offers an alternative perspective: an asymptotic balanced growth path can emerge from the truncated sampling of ideas over Pareto distributions, yielding a nonstationary productivity distribution with progressively thickening right tails. In this setting, Gibrat's law is also restored. A related exception is Jones (2023), who shows that the combinatorics of ideas drawn randomly from thin-tailed distributions can likewise sustain balanced growth.

Fourth, this paper contributes to the theory of Zipf's law. Gabaix (1999) and Luttmer (2007, 2011, 2012) establish canonical theories of Zipf's law based on individual random growth processes. They derive explicit expressions for the tail index of the stationary distribution and show that it is close to one under reasonable calibration. Toda (2017) further explores modeling assumptions that preserve tail index's proximity to one for a broad range of parameters. In contrast, this paper interprets Zipf's law as the limiting case of a nonstationary distribution with thickening right tails. This perspective links Zipf's law to the observed thickening of the right tail over the course of economic development and interprets it as an outcome of the aggregate growth itself.

2 Stylized Facts

In this section, I present empirical evidence supporting a positive relationship between the right tail thickness of the firm size distribution and the level of economic development. I begin by describing the measure of tail thickness—the tail index—and the estimation methods. I then document the main relationship using separate datasets for OECD countries and the United States, and assess the robustness of the findings across multiple empirical settings. Taken together, the evidence suggests that the thickening of the right tail of the firm size distribution is an inherent feature of the process of economic growth.

[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).

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.[4] Let $F(x)$ denote the cumulative distribution function (CDF) of firm size in the economy. Throughout the paper, I use a tilde to denote the complementary CDF, defined as $\tilde{F}(x) \equiv 1 - F(x)$. 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 positive constants. The parameter k, known as the (Pareto) tail index, summarizes the degree of concentration in the upper tail: a smaller k implies a thicker right tail.

Standard methods for estimating the tail index k rely on micro-level data on individual firms.[5] In contrast, national statistical agencies typically publish only tabulated data on the firm size distribution, reporting the number of firms and total employment within predefined size bins, where firm size is measured by employment. I therefore propose two complementary estimators of the tail index that are suitable for use with such tabulated data.

The following notations are useful for defining the tail estimators. Let $\{T_i\}_{i=0}^{n}$ denote a sequence of increasing size thresholds that define the size bins in the tabulated data.[6] Based on these thresholds, I construct two measures of top shares. The top firm share at threshold T_i, denoted N_i, is the proportion of firms with size at least T_i. Similarly, the top employment share at threshold T_i, denoted E_i, is the proportion of total employment in firms with size at least T_i. Since T_0 corresponds to the minimum size in the data, $N_0 = E_0 = 1$. Finally, let $\mathbf{N}_s \equiv (N_n, N_{n-1}, \cdots, N_s)$ denote the vector of top firm shares for thresholds above T_s, and define $\mathbf{E}_s$ analogously for top employment shares.

The first estimator obtains the tail index k by directly inverting the top firm shares. Consider two thresholds $T_L > T_S$ from $\{T_i\}_{i=0}^{n}$, with corresponding top firm shares $N_L < N_S$. The smaller threshold T_S defines the right tail. Under the

[4] Axtell (2001) famously documents that the size distribution of U.S. firms follows Zipf's law. Power-law behavior in the right tail of firm size distribution has also been reported for the U.K., France, and Italy (Fujiwara et al., 2004; di Giovanni et al., 2011; Cirillo and Hüsler, 2009).

[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.

[6] For example, in the OECD data, firm size is categorized into four classes based on employment: 1-9, 10-49, 50-249, and 250 or more employees. Then, $\{T_0, T_1, T_2, T_3\} = \{1, 10, 50, 250\}$.

Pareto assumption in equation (1), $N_x = cT_x^{-k}$, the tail index can be recovered as

$$k^f = -\ln \frac{N_L}{N_S} \Big/ \ln \frac{T_L}{T_S}.$$

I refer to k^f as the firm-based simple estimator, or simply the *simple estimator*, as it requires only two top firm shares.[7]

The second estimator adopts the efficient minimum distance estimator developed by Toda and Wang (2021) for estimating the tail index from top shares. Appendix A.1 provides full details. Let T_R denote the threshold defining the right tail. I first construct the data moments from all top employment shares in the right tail, $\mathbf{E}_R$, forming a vector of normalized top employment shares $\bar{s}(\mathbf{E}_R)$. I then construct the corresponding model moments $r(k; \mathbf{N}_R)$, implied by a Pareto distribution with tail index k and top firm shares $\mathbf{N}_R$. The *Toda-Wang estimator* k^{TW} minimizes the distance between the two:

$$k^{TW} = \arg\min_{k>1} \quad (r(k; \mathbf{N}_R) - \bar{s}(\mathbf{E}_R))^T \Omega(k; \mathbf{N}_R)^{-1} (r(k; \mathbf{N}_R) - \bar{s}(\mathbf{E}_R)),$$

where Ω is the optimal weighting matrix that achieves efficiency. The restriction $k > 1$ is required for the model moments $r(k; \mathbf{N}_R)$ to be well defined.

The two estimators entail a trade-off between robustness and efficiency. The Toda-Wang estimator k^{TW} is more efficient, as it exploits all top firm and employment shares in the right tail. However, because Zipf's law implies that k is close to one, the constraint $k > 1$ introduces numerical corner cases when applying the estimator to firm size distributions.[8] By contrast, the simple estimator k^f imposes no such restriction, though it is less efficient since it relies on only two top firm shares. I therefore report all empirical results based on both estimators.

2.2 Main results

I examine the relationship between the tail index of firm size distributions and the level of economic development across OECD countries and the United States. Details on the data and estimation results are provided in Appendix A.2.

[7] In Appendix A.1, I show how to invert the top employment shares to obtain an employment-based simple estimator, k^{emp}. Appendix A.4 verifies that all results extend to k^{emp} and provides evidence that k^f is the preferred simple estimator.

[8] Toda and Wang (2021) impose a lower bound of $k = 1.001$ in estimation. In practice, I exclude k^{TW} estimates between 1.001 and 1.0012 to mitigate attention bias in the regression results, which substantially reduces the sample size.

2.2.1 The OECD countries

The OECD Structural Business Statistics (SBS) database provides harmonized information on the industrial composition of OECD economies. The sample covers a panel of 33 countries observed from 2008 to 2017. For each economy, the SBS reports the number of firms and total employment by sector and firm size class. Firm size is categorized into four classes based on employment: 1-9, 10-49, 50-249, and 250 or more employees.

I set the threshold of 10 employees to define the right tail, capturing approximately the top 10% of firms in a typical OECD economy.[9] Consequently, I set $T_R = 10$ for the Toda-Wang estimator and $T_S = 10$ for the simple estimator. For the latter, I choose $T_L = 250$ to align with the OECD's definition of large firms. Using these thresholds, I obtain tail indices $k^f_{c,t}$ and $k^{TW}_{c,t}$ for each country-year pair (c, t), estimated using the simple and Toda-Wang estimators, respectively.

I complement these with data on log real GDP per capita, $\text{logGDPpc}_{c,t}$, from the Penn World Table (PWT) version 10.0, and investigate the relationship between the tail index and log GDP per capita through the following regression:

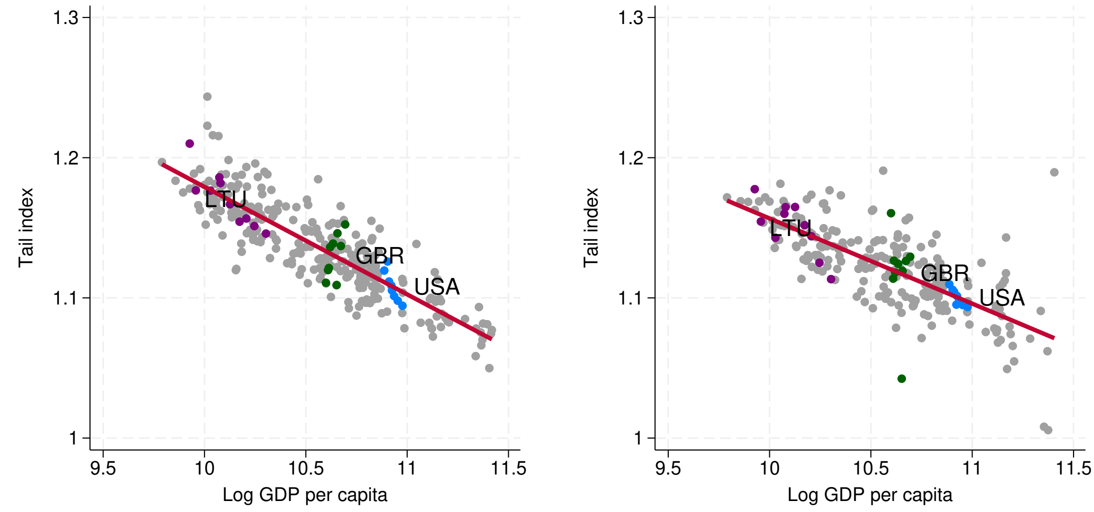
$$k^i_{c,t} = \alpha^i + \beta^i \text{logGDPpc}_{c,t} + \gamma^i_c + \varepsilon^i_{c,t}, \qquad \text{for } i \in \{f, TW\}, \tag{2}$$

where γ^i_c denotes country fixed effects that control for cross-country differences. Thus, the identification of β^i relies solely on within-country over-time variations.

Figure 1 summarizes the main regression results in three layers. First, each gray dot represents a $(\text{logGDPpc}_{c,t}, \hat{k}^i_{c,t})$ pair, where $\hat{k}^i_{c,t} = k^i_{c,t} - \gamma^i_c + \bar{\gamma}^i$ and $\bar{\gamma}^i \equiv \frac{\sum_{c=1}^{N_c} \gamma^i_c}{N_c}$ denotes the mean of the country fixed effects. This scatter plot illustrates that, net of cross-country differences, there is a negative over-time relationship between the tail index and log GDP per capita. Second, the slopes of the fitted lines correspond to the estimated coefficients $\hat{\beta}^f$ and $\hat{\beta}^{TW}$, both of which are 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. Table A.1 reports the full regression results.

In summary, among OECD countries, the right tail of the firm size distribution systematically thickens as economies grow. This finding suggests that explanations

[9] Across all country-year pairs in the sample, the median share of firms with fewer than 10 employees is 91.6%. Only three countries have fewer than 80% of firms below this threshold: the U.S. (79.2%), Switzerland (68.1%), and New Zealand (79.5%). Moreover, 10 is the only threshold that meets the minimum requirement of the Toda-Wang estimator, which requires at least three size thresholds (excluding the minimum size) for estimation (10, 50, and 250).



(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 for a panel of OECD countries. The scatter points are adjusted for country fixed effects, and the red lines represent fitted linear trends. The estimated slopes are $\hat{\beta}^f = -0.077(0.015)$ and $\hat{\beta}^{TW} = -0.061(0.018)$. The left panel reports results using the simple estimator k^f with $T_S = 10$ and $T_L = 250$, while the right panel reports results using the Toda-Wang estimator k^{TW} with $T_R = 10$. Log GDP per capita is measured in constant international dollars. The three annotated countries are Lithuania (LTU, purple), the United Kingdom (GBR, green), and the United States (USA, blue). See Table A.1 for the complete regression results. Data sources: OECD SBS and PWT 10.0.

based solely on cross-country differences in exogenous factors are insufficient to account for the observed thickening of the right tail with economic development.

2.2.2 The United States

The Business Dynamics Statistics (BDS) from the U.S. Census Bureau provide detailed data on the distribution of firm sizes in the United States over the past four decades (1978-2019). In the BDS, all U.S. business firms are grouped into ten size classes based on employment: 1-4, 5-9, 10-19, 20-99, 100-499, 500-999, 1,000-2,499, 2,500-4,999, 5,000-9,999, and 10,000 or more employees. The dataset reports, for each size class, both the number of firms and total employment.

To ensure comparability with the OECD analysis, I set the right tail threshold at 20 employees to capture roughly the top 10% of firms in the United States. Accordingly, $T_R = 20$ for the Toda-Wang estimator and $T_S = 20$ for the simple estimator. For the latter, I set $T_L = 500$ to maintain the same ratio T_L/T_S as in the previous analysis.[10]

[10]The thresholds 20, 500, and 1,000 correspond to approximately the top 11.4%, 0.4%, and

Unlike the OECD analysis, I plot the estimated tail indices k^f and k^{TW} against calendar years rather than log GDP per capita. Since the United States represents an archetypal growing economy, the time-series perspective provides a more direct depiction of how the tail index evolves with economic development.

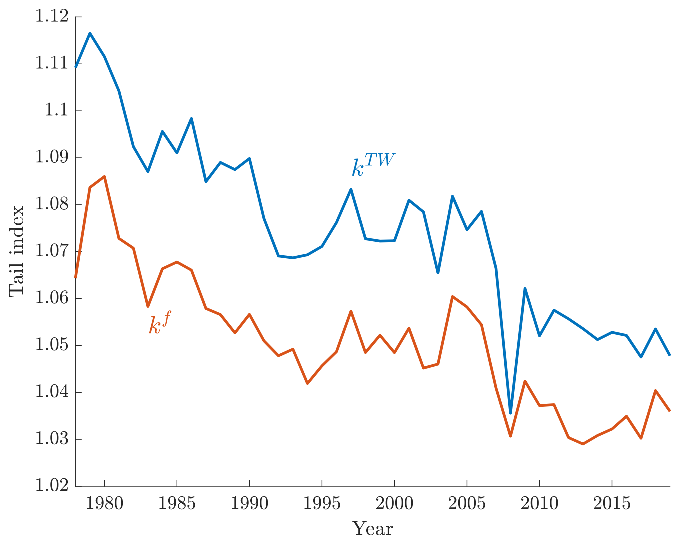


Figure 2: Tail index in the US (1978-2019)

Notes. This figure plots the tail index k of the U.S. firm size distribution from 1978 to 2019. The tail indices are estimated using either the Toda-Wang estimator, k^{TW}, with $T_R = 20$ (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.

Figure 2 shows a clear declining trend in the tail index of the U.S. firm size distribution between 1978 and 2019. This thickening of the right tail is consistent with recent evidence from Autor et al. (2020) and Kwon et al. (2024), who document rising product market concentration in the U.S. over the past four decades.

2.3 Additional evidence

Evidence from OECD countries and the United States indicates that the right tail of the firm size distribution grows thicker with economic development. In this section, I provide additional evidence showing that this relationship between tail thickness and development remains robust across sectors, countries, time pe-

0.2% of U.S. firms, respectively. In the OECD sample, $T_L = 250$ captures the top 0.2% of firms in a typical economy. Thus, $T_L = 1,000$ in the BDS case provides a percentile-aligned alternative to $T_L = 250$ in the OECD analysis. The BDS's granularity also allows for a sensitivity analysis of the simple estimator to threshold choices. Figure A.1 plots the time series of k^f estimates based on $T_S = 5, 10, 20$ and $T_L = 500, 1,000$, showing qualitatively similar trends across all specifications.

riods, and among large firms. Details of these robustness analyses are provided in Appendix A.3.

First, Table A.2 shows that the negative relationship between the tail index and log GDP per capita holds within both manufacturing and service sectors in the OECD data. These results help address the concern that the observed thickening of the aggregate right tail might be a compositional effect driven by structural transformation across sectors. Moreover, the thickening of the right tail within the service sector suggests that the mechanism is not solely tied to international trade, as it extends to less-tradable sectors as well.

Second, I show that this negative relationship also holds in the cross-section: richer countries tend to exhibit thicker right tails in their firm size distributions. For OECD economies, Table A.1 shows that both $\hat{\beta}^f$ and $\hat{\beta}^{TW}$ in regression (2) remain negative even when country fixed effects are excluded. Beyond the OECD, the World Bank Enterprise Survey (WBES) provides harmonized information on business environments across a wide range of developing economies. In particular, the Employment Indicators of the WBES report the World Bank's estimates of employment shares by firm size (number of employees) for each country. Using this dataset, I estimate the tail index for more than 130 countries surveyed between 2006 and 2019. Figure A.2 and Table A.3 confirm that the tail index remains negatively correlated with log GDP per capita in this broader global sample.

Third, I present historical evidence suggesting that the thickening of the right tail in the U.S. firm size distribution may date back a century. Kwon et al. (2024) study corporate concentration in the U.S. using historical IRS corporate income data and compile annual firm size distributions from 1918 to 2018, measuring firm size by assets, receipts, and net income. Using their dataset, I estimate the historical tail index of U.S. firms in Figure A.3 and confirm that the thickening of the right tail is a persistent, long-term feature of U.S. economic growth.

Fourth, I draw on firm-level evidence from the literature showing that the thickening right tail also holds among the largest firms. Cao et al. (2022) estimate the tail index for the top 1% of U.S. firms (by employment) using confidential Census data and find that the index declined from 1.17 in 1995 to 0.99 in 2014. Similarly, using Compustat data, Chen et al. (2023) document a decreasing trend in the tail index of the largest publicly listed firms (by revenue) in North America over 1970-2019. These findings alleviate the concern that aggregate statistics might fail to capture changes among the largest firms.

In conclusion, the evidence presented in this section shows that the thickening of the right tail of the firm size distribution is a pervasive feature of economic development. This suggests that the underlying mechanism driving this phenomenon is generic to the growth process itself. The next section develops such a mechanism.

3 A simple model of thickening right tails

In this section, I develop a productivity process, driven by firms' social learning, that generates a progressively thicker right tail of the productivity distribution. This process serves as a building block for the growth model in the next section.

Consider an economy with a unit continuum of firms indexed by productivity $z \in \mathbb{R}_+$. Let $F(z,t)$ and $f(z,t)$ denote the CDF and PDF of productivity at time t. Productivity evolves through meetings: firms encounter more productive firms and adopt their technology. Meetings occur as Poisson events with an exogenous rate $\eta(t) > 0$, which is now common to all firms. In the growth model, I will allow firms to choose their own meeting rates. Conditional on a meeting, firm x samples only from firms with productivity $z' > x$ and adopts z'. I refer to this assumption as *truncated sampling*, since the sampling distribution for firm x is the truncated distribution $F(z|z \geq x, t)$.[11] I provide a detailed discussion of this assumption at the end of this section.

With truncated sampling, the probability that firm x meets a firm with productivity at least $z > x$ between t and $t + \mathrm{d}t$ is

$$\frac{1 - F(z,t)}{1 - F(x,t)} \eta(t) \mathrm{d}t.$$

The fraction of firms above z then evolves as

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

That is, firms above z at $t + \mathrm{d}t$ are either those already above z at t or those that were below z but met firms above z during $[t, t + \mathrm{d}t]$. Letting $\mathrm{d}t \to 0$, the law of motion for $F(z,t)$ is

$$\frac{\partial F(z,t)}{\partial t} = -\eta(t) \left[1 - F(z,t)\right] \int_0^z \frac{f(x,t)}{1 - F(x,t)} \mathrm{d}x. \tag{3}$$

[11] I thank an anonymous referee for suggesting this terminology.

Since $F(0,t)=0$, this simplifies to

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

I then obtain a linear ODE in t for variable $\ln(1-F(z,t))$, which admits the following solution given initial condition $F(z,0)$:

$$\ln(1-F(z,t)) = \ln(1-F(z,0))\exp\left(-\int_0^t \eta(s)\mathrm{d}s\right). \tag{4}$$

Equation (4) fully characterizes the productivity distribution under truncated sampling. To illustrate the evolution of the right tail, I impose the following assumption that the initial productivity distribution is Pareto with a finite mean.[12]

Assumption 1. *The initial productivity distribution satisfies that*

$$F(z,0) = 1 - z^{-k_0} \quad \textit{for} \quad z \geq 1, \quad \textit{where} \quad k_0 > 1. \tag{5}$$

With Assumption 1, Equation (4) implies that for $z \geq 1$,

$$F(z,t) = 1 - z^{-k(t)}, \qquad \text{where} \quad k(t) = k_0 \exp\left(-\int_0^t \eta(s)\mathrm{d}s\right). \tag{6}$$

Thus the distribution remains Pareto, with a declining tail index $k(t)$. In words, the right tail of the productivity distribution is getting thicker over time. Equation (6) further indicates that the tail index $k(t)$ decreases at the following rate:

$$-\frac{\dot{k}(t)}{k(t)} = \eta(t) > 0. \tag{7}$$

The meeting rate $\eta(t)$ thus governs the thickening speed of the right tail: higher meeting rates imply faster reallocation into the right tail and greater concentration at the top.

What are the economic determinants of $\eta(t)$, and how might it change over development stages? Answering these questions is one primary goal of the following growth model.

[12]It is worth noting that the characterization of the tail index $k(t)$ in (6) does not depend on the Pareto assumption. In Appendix B.1, I show that Equation (6) extends to settings with general initial distributions and external learning. In those more general environment, the right tail of the productivity distribution may still thicken over time when the initial distribution is thin-tailed or has bounded support.

3.1 Discussions on the learning assumption

The property of thickening right tails in this model hinges critically on the truncated sampling assumption. In contrast, most existing idea flow models assume *random sampling*, where a firm with productivity x samples from all firms z' in the economy and adopts the idea if $z' > x$. I conclude this section by discussing the plausibility of truncated sampling in relation to Gibrat's law, empirical evidence, and possible microfoundations.

First, truncated sampling helps restore Gibrat's law in idea flow models where firm growth is driven by social learning. As one of the most well-known empirical regularities on firm dynamics, Gibrat's law states that firm growth rates are independent of firm size. Under random sampling, the efficiency of social learning declines with firm productivity: since all firms search within the same pool of ideas, more productive (larger) firms are less likely to encounter superior ones and therefore exhibit lower expected growth. By contrast, truncated sampling allows more productive firms to continue learning efficiently by drawing from a higher-quality subset of ideas. In the growth model developed below, I show that firms share the same equilibrium expected growth rate regardless of productivity, namely, Gibrat's law holds under truncated sampling.[13]

Next, recent studies on firms' technology adoption present evidence in favor of the truncated sampling assumption. A key implication of random sampling is frequent "leapfrogging", whereby low-productivity firms adopt frontier technologies as often as others. Truncated sampling instead predicts gradual upgrading, with high-productivity firms more likely to adopt frontier technologies. Cirera et al. (2022) carefully examine firm-level technology adoption decisions using comprehensive data from the Firm-level Adoption of Technology (FAT) survey. Consistent with the implications of truncated sampling, they find that leapfrogging is rare, and the probability of adopting frontier technologies increases with firm size.

Finally, several mechanisms help explain why firms' learning is directed upward rather than entirely random. First, the spatial distribution of firms implies that knowledge diffusion is inherently directed. A large literature documents that knowledge flows are highly localized (Baum-Snow et al., 2024) and that firms are spatially sorted by productivity across cities (Combes et al., 2012). Consequently,

[13] Truncated sampling is not merely a sufficient condition for Gibrat's law or for thickening right tails. In the working paper Chen (2023), I show that within a broad class of idea flow models, Gibrat's law holds if and only if the right tail of the firm size distribution thickens. With additional structural assumptions, truncated sampling emerges as the only functional form consistent with both Gibrat's law and thickening right tails.

the set of potential learning partners depends on firms' productivity levels, consistent with truncated sampling. Second, firms may identify more productive peers through observable signals—such as market shares or patent records—and target them strategically. An analogous example in academia reflects similar directions of knowledge exchange: seminar organizers typically invite speakers from higher-ranked institutions.

4 A growth model of thickening right tails

This section develops a parsimonious growth model that embeds the firm productivity process described above. In the model, firms endogenously choose their meeting rates, which in turn thicken the right tail of the productivity distribution while sustaining asymptotically balanced growth. The analysis proceeds in three steps. I first describe the economic environment, then characterize the equilibrium, and finally discuss the model's key properties.

4.1 The environment

As before, I consider a continuous-time economy with a unit continuum of firms. The notation for firm productivity follows that introduced in Section 3.

Firms I make several simplifying assumptions to isolate the growth mechanism driven by firms' active social learning. First, there is no firm entry or exit, so the total measure of firms is constant and normalized to one. Second, firms produce a homogeneous final good and compete under perfect competition. Third, firms make no input decisions. Each firm is a fixed unit of capital that produces final goods in proportion to its productivity: firm z produces z final goods per unit of time. The final good serves as the numeraire.[14]

The firm's only decision concerns improving productivity through social learning. Learning follows the truncated sampling process described earlier, except that the arrival rate of meetings is now determined endogenously by the firm's research effort. Firms choose their own meeting rates by hiring researchers. To distinguish from the passive learning case in the previous section, I refer to this mechanism

[14]In the absence of production labor, firm size is measured by firm revenue. In this case, the productivity distribution is isomorphic to the firm size distribution, since productivity equals firm revenue (output). In Section 7, I introduce production labor and show that employment is proportional to firm revenue. Consequently, the firm size distribution—whether measured by employment or revenue—remains isomorphic to a transformed productivity distribution.

as *idea search*, and the firm-specific meeting rate as search intensity. Following Atkeson and Burstein (2010), I assume that the search cost scales with firm sales: it takes $z\eta$ researchers for a firm with productivity z to achieve a meeting rate η. This scaling neutralizes the effect of firm size on idea search decisions, thereby preserving Gibrat's law in equilibrium.[15]

Let $w(t)$ denote the wage of researchers and $r(t)$ the interest rate. Taking these as given, firms optimally choose search intensities to maximize their present value. Let $v(z,t)$ denote the value of a firm with productivity z at time t. It satisfies the following Hamilton-Jacobi-Bellman (HJB) equation:

$$r(t)v(z,t) = z + \max_{\eta \geq 0} \eta \left\{ \int_z^\infty [v(x,t) - v(z,t)] \mathrm{d}F(x|x \geq z, t) - zw(t) \right\} + \partial_t v(z,t). \quad (8)$$

Let $\eta(z,t)$ denote the optimal search policy. Equation (8) states that the flow value of the firm (LHS) equals the sum of three components (RHS): the flow profit, the expected gains from learning net of the total search cost, and the option value associated with changes in the aggregate state.

Households There is a fixed continuum of infinitely lived representative households with measure L. Each household inelastically supplies one unit of labor, which is employed as a researcher. In the absence of population growth, aggregate labor supply remains constant at L. Households own the firms, so the flow income of each household is given by the output per capita $y(t)$. Taking the interest rate $r(t)$ and income $y(t)$ as given, households choose consumption $c(t)$ to maximize the present value of their CRRA utility subject to a lifetime budget constraint:

$$\begin{aligned} &\max_{c(\tau)\geq 0} \quad \int_t^\infty e^{-\rho(\tau - t)} \frac{c(\tau)^{1-\theta} - 1}{1-\theta} \mathrm{d}\tau. \\ &\text{s.t.} \quad \int_t^\infty e^{-\int_t^s r(\tau)\mathrm{d}\tau} c(s)\mathrm{d}s \leq \int_t^\infty e^{-\int_t^s r(\tau)\mathrm{d}\tau} y(s)\mathrm{d}s. \end{aligned} \quad (9)$$

Market clearing The market for the final good clears when total household consumption equals total output across firms $Y(t)$:

$$Lc(t) = Y(t) = \int_0^\infty z f(z,t) \mathrm{d}z. \quad (10)$$

[15]The assumption that innovation cost scales with firm size is standard in firm-based innovation models such as Klette and Kortum (2004), Atkeson and Burstein (2010), and Peters (2020). In these frameworks, the scaling factor ensures that endogenous R&D intensity is independent of firm size, thereby preserving Gibrat's law. In the context of idea search, I provide a micro-foundation for this assumption based on firms' choice of search technologies in Appendix B.2.

Similarly, the labor market for researchers clears when the total labor demand of firms equals the fixed aggregate labor supply:

$$\int_0^{\infty} z\eta(z,t)f(z,t)\mathrm{d}z = L. \tag{11}$$

Law of motion of the aggregate state The distribution of firm productivity summarizes the aggregate state of the economy. Given the search policy $\eta(z,t)$, the same derivation as in Section 3 yields the following Kolmogorov Forward Equation (KFE) for the productivity distribution $F(z,t)$:

$$\frac{\partial F(z,t)}{\partial t} = -\left[1 - F(z,t)\right] \int_0^{z} \eta(x,t) \frac{f(x,t)}{1 - F(x,t)} \mathrm{d}x. \tag{12}$$

The initial distribution $F(z,0)$ follows Assumption 1, with $F(z,0) = 1 - z^{-k_0}$ for $z \geq 1$ and $k_0 > 1$. This assumption allows for a closed-form characterization of the full transition dynamics and ensures finite aggregate output at all times.

Equilibrium I conclude the description of the growth model by defining an equilibrium and an asymptotic balanced growth path (BGP). While an exact BGP requires that output per capita grows at a constant rate at all times, an asymptotic BGP only requires that this growth rate converges to a constant in the limit.

Definition (Equilibrium)**.** A competitive equilibrium for this economy consists of researchers' 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 (8), and $\eta(z,t)$ is the associated policy function;

(ii) Given $\{y(t), r(t)\}$, $c(t)$ solves the households' problem (9);

(iii) Both goods and labor markets clear, i.e., (10) and (11) are satisfied;

(iv) $F(z,t)$ solves the KFE (12) given $\eta(z,t)$ and the initial condition (5).

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.,

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

4.2 Solving the equilibrium

In this section, I guess and verify a model equilibrium in which all firms choose the same search intensity, and the productivity distribution evolves identically to that in the simple model in Section 3. In particular, there exist an interest rate $r(t)$ and a researcher wage $w(t)$ that support such an equilibrium.

Suppose, as in the simple model, that the meeting rate is common across firms at all times, i.e., $\eta(z,t) = \eta(t) > 0$ for all z and t. With a Pareto initial productivity distribution as in (5), the resulting productivity distribution $F(z,t)$ follows directly from Equations (6) and (7) in Section 3:

$$F(z,t) = 1 - z^{-k(t)} \quad \text{for} \quad z \geq 1, \tag{13}$$

in which $k(t)$ satisfies

$$\frac{\dot{k}(t)}{k(t)} = -\eta(t) < 0, \tag{14}$$

and $k(0) = k_0$.

Since the productivity distribution remains Pareto at all times, the tail index $k(t)$ serves as a sufficient statistic for the aggregate state, and its evolution fully characterizes the transition dynamics of the economy. Using the expression for $F(z,t)$ in (13), the labor market clearing condition (11) can be rewritten as

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

Substituting $\eta(t)$ from (14) into the above equation yields a first-order linear ordinary differential equation (ODE) for $k(t)$:

$$-\frac{\dot{k}(t)}{k(t)-1} = L, \tag{15}$$

which admits the solution, given $k(0) = k_0 > 1$,

$$k(t) = 1 + (k_0 - 1)e^{-Lt}. \tag{16}$$

The associated equilibrium search intensity is therefore

$$\eta(z,t) = \eta(t) = L\frac{k(t)-1}{k(t)} = \frac{L}{e^{Lt}/(k_0-1)+1}. \tag{17}$$

It follows that $k(t)$ declines monotonically toward one, while $\eta(t)$ converges to zero over time. The convergence of $k(t)$ to unity is intuitive from the labor market clearing condition: the thickening of the tail ceases once $k(t)/(k(t)-1)$ diverges to infinity and the search intensity $\eta(t)$ falls to zero.

Given the equilibrium path of $k(t)$, it is straightforward to characterize the dynamics of output per capita and the corresponding interest rate. With the productivity distribution $F(z,t)$ given in (13), aggregate output satisfies $Y(t) = k(t)/(k(t)-1)$. Hence, output per capita $y(t)$ is given by

$$y(t) = \frac{1}{L}\frac{k(t)}{k(t)-1} = \frac{1}{L}\left[\frac{e^{Lt}}{k_0 - 1} + 1\right] \tag{18}$$

The corresponding output per capita growth rate, $g(t) \equiv \dot{y}(t)/y(t)$, is

$$g(t) = \frac{1}{k(t)}\left(-\frac{\dot{k}(t)}{k(t)-1}\right) = \frac{L}{1 + (k_0 - 1)e^{-Lt}}. \tag{19}$$

Thus, the equilibrium features an asymptotic balanced growth path in which $g(t)$ converges to L. From the goods market clearing condition (10), consumption growth equals output per capita growth. The household Euler equation then implies the equilibrium interest rate:

$$r(t) = \theta\frac{\dot{c}(t)}{c(t)} + \rho = \theta g(t) + \rho, \tag{20}$$

which converges to $\theta L + \rho$ as $t \to \infty$. I further impose the parametric condition $\rho > (1-\theta)L$ to ensure that the representative household's utility is finite and the transversality condition is satisfied.

I then show that there exists a researcher wage $w(t)$ under which it is optimal for all firms to choose the same search intensity. Let $S(z,t)$ denote the net payoff per idea search for firm z at time t. From the HJB equation (8),

$$S(z,t) = \int_z^{\infty} [v(x,t) - v(z,t)]\mathrm{d}F(x|x \geq z, t) - zw(t). \tag{21}$$

In what follows, I first characterize the firm value function $v(z,t)$ and then identify $w(t)$ such that $S(z,t) = 0$ for all z and t, ensuring the optimality of search policy.

To solve for $v(z,t)$, note that the total search payoff $\eta S(z,t)$ is linear in η, and since $\eta \geq 0$, the equilibrium condition $\eta(z,t)S(z,t) = 0$ must hold. Substituting this condition into the HJB equation (8) yields

$$r(t)\tilde{v}(z,t) = 1 + \partial_t \tilde{v}(z,t),$$

where $\tilde{v}(z,t) \equiv v(z,t)/z$. Integrating forward over time, $\tilde{v}(z,t)$ is independent of z and equals the present value of a flow of unit output:

$$v(t) \equiv \tilde{v}(z,t) = \int_t^{\infty} e^{-\int_t^x r(s)\mathrm{d}s} \mathrm{d}x. \tag{22}$$

Given convergence of $r(t)$, $v(t)$ is finite, and $v(z,t) = v(t)z$ is then well-defined.

Finally, let the researcher wage be $w(t) = v(t)/(k(t)-1)$, which is well-defined since $k(t) > 1$ by (16). Substituting $v(z,t)$, $F(z,t)$, and $w(t)$ into (21) gives

$$S(z,t) = z\left[\frac{v(t)}{k(t)-1} - w(t)\right] = 0.$$

Hence, the uniform search policy $\eta(z,t) = \eta(t) > 0$ is optimal, and all equilibrium variables are solved with explicit solutions. The following proposition summarizes the key characterization of this equilibrium.[16]

Proposition 1. *With Assumption 1 and $\rho > (1-\theta)L$, there exists an equilibrium in which (1) the search intensity is uniform across firms, $\eta(z,t) = \eta(t) > 0$, and (2) the productivity distribution $F(z,t)$ remains Pareto, with only its tail index $k(t)$ declining over time at rate $-\dot{k}(t)/k(t) = \eta(t) > 0$.*

Remarks on equilibrium selection The assumption that idea search costs are linear in η is a convenient simplification that allows for an explicit characterization of the equilibrium. However, it also implies that $\eta(z,t) = \eta(t)$ is not the firm's unique optimal choice, and alternative equilibria may therefore exist (Section 6.1). In Appendix B.4, I relax this linearity assumption by introducing a non-linear adjustment cost in η. I show that the resulting equilibrium not only yields unique firm-level optimal search intensities but also preserves the characterization in Proposition 1. Moreover, I demonstrate that the equilibrium in this section is the limiting case of those with non-linear search costs, which justifies its selection as the preferred benchmark.

[16]In Appendix B.3, I further characterize the equilibrium dynamics of prices, $w(t)$ and $r(t)$, and show that the researcher wage $w(t)$ rises relative to household income $y(t)$ over time. In other words, the share of R&D expenditure in GDP increases along the transition path.

4.3 Discussions

With the equilibrium fully characterized, I now discuss the model's main properties and their connection to empirical regularities. Specifically, the model accounts for four salient facts about firm dynamics and economic growth: the thickening of the right tail of the firm size distribution, Gibrat's law, Zipf's law, and the existence of a balanced growth path. The first feature represents the key empirical finding of this paper, while the latter three are well-established in the literature.

Thickening right tails The expressions for $k(t)$ and $y(t)$ in (16) and (18) show that the right tail of the productivity distribution thickens as the economy grows. Consistent with the empirical evidence presented in Section 2, the model generates the negative relationship between the tail index and output per capita as an equilibrium feature of the growth process. This relationship arises naturally from firms' learning through truncated sampling, which drives growth by increasing the share of high-productivity firms.

Gibrat's law The model implies that firms have the same expected productivity growth rates, i.e., Gibrat's law holds at all times. To see this, note that a firm's expected productivity growth equals the product of its search intensity and the expected productivity gain per search. Since all firms choose the same search intensity in equilibrium, it suffices to show that the expected productivity gain in each meeting is identical across firms.

Let z' denote the productivity of the other firm in a meeting, and let $\hat{F}(z'; z, t)$ be the CDF of z' in meetings facing firm z at t. Given the Pareto productivity distribution $F(z, t)$ in (13), truncated sampling implies that $\hat{F}(\cdot; z, t)$ is also Pareto, with $\hat{F}(z'; z, t) = 1 - (z'/z)^{-k(t)}$ for $z' \geq z$. Let $\hat{z}$ denote the productivity growth ratio z'/z. Its distribution in meetings facing firm z at t is then

$$\text{Prob}(\hat{z} \leq x; z, t) = \hat{F}(zx; z, t) = 1 - x^{-k(t)} = F(x, t),$$

for $x \geq 1$. Hence, all firms face the same distribution of productivity growth ratio $\hat{z}$ in each meeting. The expected productivity growth per search is therefore identical across firms and equals $k(t)/(k(t) - 1)$. This result indicates that truncated sampling over a Pareto distribution may be viewed as a microfoundation to the standard assumption in quality-ladder models that the step size of quality improvement is independent of current quality.

Finally, let $\lambda(z, t)$ denote the expected productivity growth rate of firm z at time t, i.e., $\lambda(z,t) = \mathbb{E}[\mathrm{d}z|z,t]/z$. Given $\eta(z,t) = \eta(t)$ in equilibrium, $\lambda(z,t)$ equals the aggregate output per capita growth rate $g(t)$:

$$\lambda(z,t) = \eta(z,t)\left(\mathbb{E}[\hat{z}|z,t] - 1\right) = \frac{\eta(t)}{k(t)-1} = g(t),$$

where the last equality follows from (16), (17), and (19).

Interestingly, while the existing literature often emphasizes that Gibrat's law implies a power law (Gabaix, 2009), this model suggests that the reverse may also hold. Under truncated sampling, the Pareto distribution ensures that idea search efficiency is identical across firms, allowing them to choose the same search intensity and thus satisfy Gibrat's law in equilibrium.

Zipf's law As $k(t)$ converges to 1 by (16), the equilibrium productivity distribution converges to Zipf's distribution, a Pareto distribution with tail index 1. In the context of firms, Zipf's law states that advanced economies such as the U.S. exhibit firm size distributions very close to Zipf's distribution.[17] This model thus provides a growth-based theory for Zipf's law in firm size distributions. Firms' learning through truncated sampling generates aggregate productivity growth by progressively thickening the right tail of the productivity distribution. Richer economies are further along this growth path, and their distributions lie closer to the limit, which is Zipf's distribution. The convergence to Zipf's law requires minimal restrictions on model parameters, helping explain its empirical prevalence.

Asymptotic balanced growth path The expression for $g(t)$ in (19) indicates that the equilibrium constitutes an asymptotic balanced growth path, with the long-run growth rate equal to the population size L.[18] Two features distinguish this model from standard firm-based growth frameworks, where balanced growth paths correspond to traveling wave equilibria with stationary (scaled) productivity distributions.

First, the model provides an explicit example in which product market concentration rises along an (asymptotic) balanced growth path. Existing firm-based growth models typically feature balanced growth paths where market concentration remains constant with stationary productivity distributions. However, recent

[17] More generally, Zipf's law is a pervasive empirical regularity observed across the natural and social sciences. If X denotes the variable of interest, Zipf's law states that $P(X > x) \approx 1/x$.

[18] With a fixed population and measure of firms, the model features scale effects in the growth rate. In Section 7, I extend the model to incorporate both population growth and firm entry. Scale effects vanish in the extended model, since the measure of firms grows at the same rate as the population in the long-run equilibrium.

evidence by Kwon et al. (2024) shows that market concentration has been increasing in the U.S.—an economy characterized by a stable long-term growth rate—over the past century. The present model captures this pattern through the thickening of the right tail of the productivity distribution, demonstrating that rising market concentration can coexist with nearly constant aggregate growth.

Second, the model uncovers a novel source of endogenous growth. Standard frameworks regard balanced growth paths with stationary productivity distributions as the long-run equilibrium. In other words, reallocation dynamics that alter the shape of the productivity distribution do not contribute to long-run growth. In contrast, growth in this model is driven exclusively by firms' learning, which continuously thickens the right tail and generates persistent reallocation gains in aggregate productivity. Constant output growth implies that Zipf's distribution is the natural limiting case of the productivity distribution, as it has the thinnest possible tail consistent with an infinite mean. Otherwise, growth would cease as output converges to the finite mean of the limiting distribution.[19]

To conclude, firms' learning through truncated sampling not only generates thickening right tails of the productivity distribution along an asymptotic balanced growth path but also serves as an important driver of long-run economic growth. I refer to this mechanism as *tail growth*.

5 Testing the model predictions

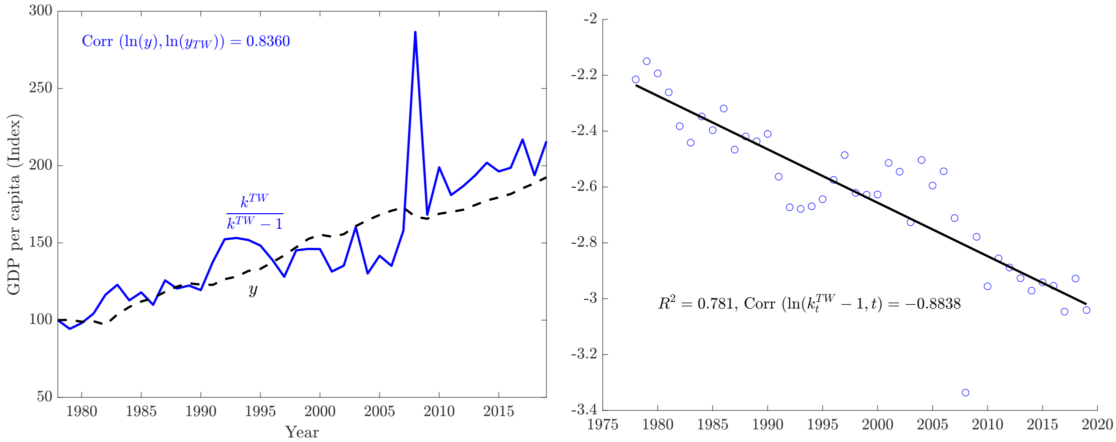
Despite its parsimony, the model in Section 4 yields empirically testable predictions. In particular, it implies two sharp theoretical relationships: 1) output per capita y is proportional to $k/(k-1)$, and 2) $k-1$ decreases to 0 at a constant rate over time t.

In this section, I test both predictions using U.S. data, which provide the longest available time series for the tail index. It is important to emphasize that, while the tail estimates play a crucial role in assessing the model's quantitative implications, their estimation is entirely independent of the model itself. Disciplining the choice of tail estimator is therefore essential to ensure a meaningful data-model comparison.

To this end, I use the Toda-Wang estimator k^{TW} for the tail index. Comparing to the simple estimator, the Toda-Wang estimator is much more efficient because it

[19]Constant growth is only a sufficient condition for convergence to Zipf's distribution; convergence may still occur under declining growth. In the working paper verision, Chen (2023) provides a detailed characterization of the relationship between Zipf's law and growth rates.

exploits all available information on firm and employment shares in the right tail. In this sense, k^{TW} represents a better "calibrated" model in terms of matching the right tail of the firm size distribution. In the following exercises, I compare model-based transformations of k^{TW} with either output per capita or time, and report the corresponding correlation coefficients to quantify the alignment between the model predictions and the data.[20]



(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 $y_{TW} \equiv k^{TW}/(k^{TW}-1)$. 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 $\ln\left(k_t^{TW}-1\right)$ with its associated linear trend. Both the linear trend and R^2 are obtained from the regression $\ln\left(k_t^{TW}-1\right) = -2.24 - 0.019t$. Source: the U.S. Census BDS and FRED.

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:

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

in which $k(t)$ is the tail index 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.

[20] The correlation coefficient provides a normalized measure that controls for the variance of each variable, making it less sensitive to the choice of tail estimator than direct visual comparisons. Consequently, correlation coefficients yield a more robust assessment of model-data fit. Appendix A.4.3 reports analogous results using the simple estimators k^f and k^{emp}. While direct comparisons may vary across tail estimators, the correlation coefficients remain similarly high and consistent.

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

$\ln(k-1)$ vs. t Equation (15) implies that $k(t) - 1$ declines toward zero at a constant rate, so that $\ln(k(t) - 1)$ decreases linearly over time. Figure 3(b) plots the time series of $\ln(k-1)$ for the U.S. between 1978 and 2019, together with a fitted linear trend. The results show that $\ln(k-1)$ closely follows the fitted line, with an R^2 of 0.781 and a correlation coefficient of -0.8838.

6 Policy Implications

In this section, I show that the growth model in Section 4 delivers a novel perspective for evaluating size-dependent industrial policies. Idea search by individual firms generates a knowledge diffusion externality: each firm's search effort shapes the productivity distribution, which in turn governs future search efficiency. Searches by large firms contribute disproportionately to thickening the right tail and generate positive spillovers for the entire economy, whereas searches by small firms produce limited externalities for larger firms. Consequently, relative to the first-best allocation, large firms underinvest in idea search, implying that policy should encourage greater search activity among them.

I illustrate this implication through two policy experiments. In the first, I consider a tax policy under which only firms above a targeted productivity threshold engage in idea search in equilibrium and show that the resulting long-run growth rate increases with the threshold level. In the second, I solve the social planner's problem and show that the optimal search intensity approximately follows a power function that rises with productivity. Together, these exercises demonstrate that policies favoring large firms enhance welfare by better exploiting the diffusion externality. Appendix C provides the omitted proofs and derivations.

6.1 Taxing small firms' idea search

In the first exercise, the policymaker sets a productivity threshold z^* and imposes a positive search tax τ on firms below it. Specifically, firms with productivity $z < z^*$ face a unit search cost of $z(1+\tau)w(t)$, whereas those with productivity $z \geq z^*$ retain the original unit cost $zw(t)$. The tax revenue is rebated to households as a lump sum transfer. The following proposition characterizes the equilibrium for a positive tax $\tau > 0$ and threshold z^*.

Proposition 2. *For any threshold productivity $z^* \geq 1$, there exists an equilibrium such that the equilibrium productivity distribution satisfies*

$$F(z,t) = \begin{cases} 1 - z^{-k_0} & \text{if } z < z^*, \\ 1 - (z^*)^{k(t)-k_0} z^{-k(t)} & \text{if } z \geq z^*, \end{cases} \tag{23}$$

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 engage in search, so that part of the productivity distribution remains constant. Consequently, the tax rate itself is irrelevant for the equilibrium outcome. The upper part of the distribution evolves as before, since firms above the threshold continue to search at a uniform intensity. The key difference is that, with fewer firms searching, each active firm hires more researchers and increases its search intensity, thereby accelerating output growth. Overall, taxing small firms discourages their inefficient use of labor and expands the effective labor supply available to large firms. The policymaker can then increase the long-run growth rate $L(z^*)^{k_0-1}$ by choosing higher thresholds z^*.[21]

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

[21] Appendix C.1 shows that Proposition 2 also holds for the case $\tau = 0$, which corresponds to the tax-free economy in Section 4. That is, for any $z^* \in [1, +\infty)$, (23) describes an equilibrium productivity distribution of the baseline model. Proposition 2 then illustrates the multiplicity of equilibria that arises from the linear search cost.

productivity density function f:

$$\begin{aligned} W(f) = \max_{\{c(t),\eta(z,t)\}} & \int_0^\infty e^{-\rho t} L u(c(t)) \mathrm{d}t \\ \text{s.t.} \quad & Lc(t) \leq \int_0^\infty z f(z,t) \mathrm{d}z, \\ & \int_0^\infty z\eta(z,t) f(z,t) \mathrm{d}z \leq L, \\ & \frac{\partial f(z,t)}{\partial t} = f(z,t) \left[\int_0^z \eta(x,t) \frac{f(x,t)}{1-F(x,t)} \mathrm{d}x - \eta(z,t) \right], \\ & f(z,0) = f(z). \end{aligned} \tag{24}$$

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 (24) 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 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 3. *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.,*

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

in which $\mathcal{L}(z,t)$ represents a slowly varying function. Moreover, $\mathcal{L}(z,t)$ is constant over z if $f(z,t)$ is exactly Pareto.

Proposition 3 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 3 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 4.2. It is obvious that relative to the market outcome, a social planner would reallocate researchers from low-productivity firms to high-productivity firms.

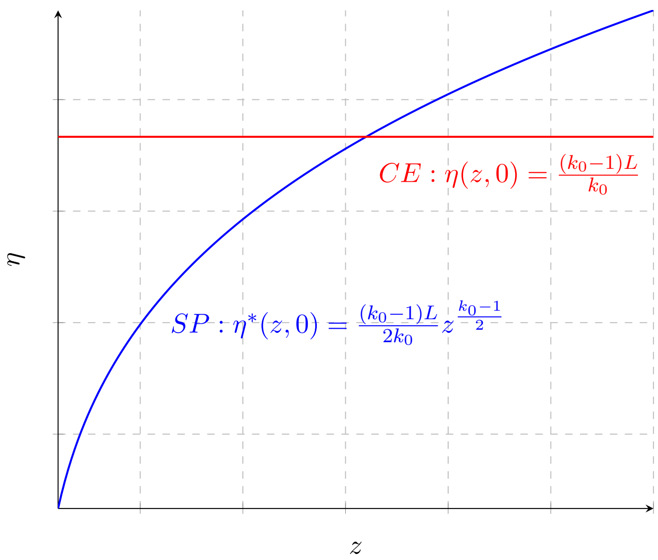


Figure 4: An illustration of search intensity comparison

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 $\eta(z,0)$.

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

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

in which $\tilde{\mathcal{L}}(z,t) = \mathcal{L}(z,t)zf(z,t)/\tilde{F}(z,t)$ and is slowly varying. The integral is regularly varying with exponent $\frac{k(t)-1}{2}$.[22] $\tilde{F}(z,t+h)$ is then a regularly varying function with exponent $-k(t)+\frac{k(t)-1}{2} = \frac{-k(t)-1}{2}$ for arbitrary h. Therefore, there is a jump in the tail index from $k(t)$ to $\frac{k(t)+1}{2}$. If $k(t)=1$, there will be no changes in the tail index since the integral term is also slowly varying.

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.[23] Figure A.9 illustrates this process in the appendix. Albeit unusual, this result demonstrates the potential of tail growth.

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.

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.[24] 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 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 $z > z^*(t)$

[22] This result follows directly from Karamata's theorem in the theory of regular variation (see Theorem 1.5.11 in Bingham et al. (1987)).

[23] If output remains finite for some interval $t \in [0, t_0]$ with $k(t) = 1$, the social planner can always lower the tail index to below one before t_0—for example, by following the equilibrium strategy—while still satisfying the labor market clearing condition. Since a tail index below one implies an infinite mean of the productivity distribution, and thus infinite output, the existence of any finite-output interval would contradict the optimality of the policy.

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

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 4. 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

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

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 Appendix C.4. 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.[25] 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

[25] I also demonstrate in Appendix C.4 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.

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.[26]

7 Extensions

The model in Section 4 is intentionally simplified to ensure tractability and yield sharp analytical insights. It omits four common elements of growth models. First, firms make no input decisions in production. Second, there is no entry or exit, and the measure of firms is given exogenously. Third, technological progress is absent. Fourth, the population remains constant. It is therefore important to verify that the results in Section 4 do not rely critically on these simplifying assumptions.

In this section, I extend the model to incorporate all four elements. I show that the extended framework not only preserves the key properties of the baseline model but also generalizes its main insights. Proofs and derivations are omitted for brevity but are provided in Appendix D.

7.1 The environment

Table 1 summarizes the economic environment of the extended model.

Production Each firm now produces a differentiated intermediate good ω under monopolistic competition. The intermediate good output q is a linear function of production labor l_p: $q = \tilde{z} l_p$, where $\tilde{z}$ denotes firm specific efficiency. The unique final good $Y(t)$ is produced competitively by aggregating over intermediate goods $q(\omega, t)$:

$$Y(t) = \left(\int_{\Omega(t)} q(\omega, t)^{\frac{\sigma-1}{\sigma}} \mathrm{d}\omega \right)^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1, \tag{25}$$

where $\Omega(t)$ denotes the set of intermediate goods at time t, and σ is the elasticity of substitution across varieties. The final good serves as the numeraire, implying

[26] 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 Appendix C.4 for further discussions.

Table 1: The economic environment of the extended model

Utility	$\int_t^\infty e^{-\rho(\tau-t)} \frac{c(\tau)^{1-\theta}-1}{1-\theta} \mathrm{d}\tau$
Population growth	$\dot{L}(t)/L(t) = g_L(t) \geq 0$
Managerial ability e	$e \sim J$ for $e \geq 0$, $\quad K(e) = \int_e^\infty x \mathrm{d}J$
Final good	$Y(t) = \left(\int_{\Omega(t)} q(\omega,t)^{\frac{\sigma-1}{\sigma}} \mathrm{d}\omega \right)^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1$
Intermediate good	$q = \tilde{z} l_p, \quad z \equiv \tilde{z}^{\sigma-1} \sim F(z,t)$
Technological obsolescence	$\dot{z}^*(t)/z^*(t) = g_z(t) \geq 0$
Idea search:	
Firm $z \geq z^*(t)$	random draws from $F(x \mid x \geq z, t)$ using l_r
Firm $z < z^*(t)$	no draws
Entrants	random draws from $F(x \mid x \geq z^*(t), t)$ using e
Resource constraints:	
Final good	$Lc(t) = Y(t)$
Production workers l_p	$M(t) \int_1^\infty l_p(z,t) \mathrm{d}F(z,t) = (1-s)L(t)$
Researchers l_r	$M(t) \int_{z^*(t)}^\infty l_r(z,t) \mathrm{d}F(z,t) = sJ(e^*(t))L(t)$
Entrepreneurs	$E(t)\psi z^*(t) = sK(e^*(t))L(t)$

that the ideal price index $P(t) = 1$, so that $Y(t)$ also equals total expenditure.

Firms choose prices to maximize static profits given residual demand from (25) and the wage of production workers $w_p(t)$. Defining $z \equiv \tilde{z}^{\sigma-1}$, standard results give that firms' production employment l_p, revenue r, and profit π have the following expressions and are linear in z:

$$l_p(z,t) = \frac{\sigma-1}{w_p(t)}\pi(z,t), \quad r(z,t) = \sigma\pi(z,t), \quad \text{and} \quad \pi(z,t) = \pi(t)z, \tag{26}$$

where $\pi(t) = \frac{1}{\sigma}Y(t)\left(\frac{\sigma}{\sigma-1}w_p(t)\right)^{1-\sigma}$. I therefore use z to denote firm productivity and focus on the distribution $F(z,t)$.[27]

Demography Households are representative and have measure $L(t)$, which now grows at an exogenous rate $g_L(t) \equiv \dot{L}(t)/L(t) \geq 0$. Each household contains a unit continuum of members, of whom $1-s$ are unskilled and s are skilled. Unskilled workers are homogeneous and only engage in production. Whereas, skilled labor choose between employment as researchers in incumbent firms or entrepreneurship by founding new firms. They are homogeneous as researchers but heterogeneous in

[27] Because of the linear relationships in (26), the distributions of firms' revenues, profits, production employment, and productivity z are identical up to scale. Hence, $F(z,t)$ can represent the firm size distribution as in the baseline model.

the managerial ability e as entrepreneurs, which follows a continuous distribution J with support $[0, +\infty)$. This occupation choice gives rise to firms' entry.

Household members consume final good $c(t)$ and have the same intertemporal utility as in (9). With perfect insurance within the household, the individual utility maximization problem is the same as in the baseline model, given the income flow $y(t)$. Assume that households own the firms and the government runs balanced budget, the household income equals to output per capita, i.e., $y(t) = Y(t)/L(t)$.

Technological progress In the baseline economy, the median firm productivity remains constant in the long run, and the least productive technology $z = 1$ is always the most prevalent among firms. This is because with only tail growth, there lacks forces that shift the overall productivity distribution and make it a traveling wave as in many firm-based growth models.

I now introduce a technological progress that directly shifts the productivity distribution. Specifically, I model it through a productivity threshold $z^*(t)$ that grows exogenously at a continuous rate $g_z(t) \equiv \dot{z}^*(t)/z^*(t) \geq 0$. Entrant firms draw their productivity randomly from incumbents with productivity above $z^*(t)$. This captures the notion that each new technology vintage, embodied in the entry cohort of firms, builds upon and enhances the previous ones. Meanwhile, technologies with productivity below $z^*(t)$ are deemed obsolete, and firms using them can no longer upgrade through idea search. This yields technological obsolescence that low-productivity technologies gradually vanish over time. I will later show that the resulting productivity distribution evolves as an asymptotic traveling wave.

Incumbent firms Let $w_r(t)$ denote the wage of researchers and $r(t)$ the interest rate. All firms face an exogenous exit shock with Poisson arrival rate δ. The HJB equation for firms with productivity $z < z^*(t)$ is given by:

$$(r(t) + \delta)\, v(z, t) = \pi(z, t) + \partial_t v(z, t). \tag{27}$$

In words, firms with obsolete technologies are *search-inactive* and only earn static profits until they exit the market. Contrastingly, firms with productivity $z \geq z^*(t)$ remain *search-active* and choose search intensities optimally as in the baseline case. Their HJB equation is given by:

$$\begin{aligned}(r(t) + \delta)\, v(z, t) &= \pi(z, t) + \partial_t v(z, t) \\ &+ \max_{\eta \geq 0} \eta \left\{ \int_z^\infty [v(x, t) - v(z, t)] \mathrm{d}F(x | x \geq z, t) - \psi z w_r(t) \right\},\end{aligned} \tag{28}$$

where the new parameter $\psi > 0$ measures the efficiency of the search technology, and the unit search cost (in terms of researchers) is now ψz.

Firm entry As described earlier, entrepreneurs found new firms by sampling ideas from incumbent firms with non-obsolete technologies. At the startup stage, they cannot hire researchers but have to rely on their managerial ability for idea search. The unit search cost in terms of managerial ability is $\psi z^*(t)$, so entrepreneurs with ability e have an arrival rate of ideas equal to $e/(\psi z^*(t))$. The expected firm value v_E is then

$$v_E(e,t) = \frac{e}{\psi z^*(t)} \int_{z^*(t)}^{\infty} v(x,t) \mathrm{d}F(x|x \geq z^*(t), t). \tag{29}$$

Skilled labor chooses to be entrepreneurs if and only if the the expected firm value $v_E(e,t)$ exceeds the after-tax(subsidy) wage of researchers $(1 - \tau_r)w_r(t)$, where τ_r is the tax (subsidy) rate on researchers' wage. The government also makes lump sum transfer to households to balance the budget. Since J is continuous and has support $[0, +\infty)$, there exists a threshold $e^*(t)$ such that only skilled labor with $e \geq e^*(t)$ becomes entrepreneurs. The measure of total entrants $E(t)$ is then

$$E(t) = \frac{sK(e^*(t))L(t)}{\psi z^*(t)}, \tag{30}$$

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

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

$$M(t) \int_1^{\infty} l_p(z,t) \mathrm{d}F(z,t) = (1-s)L(t), \tag{31}$$

$$M(t) \int_{z^*(t)}^{\infty} l_r(z,t) \mathrm{d}F(z,t) = sJ(e^*(t))L(t), \tag{32}$$

where firms' demand for researchers $l_r(z,t) = \psi z \eta(z,t)$. While the share of production workers in the population is always constant, the share of researchers is endogenous and governed by the startup threshold $e^*(t)$. With representative households, goods market clearing requires that $L(t)c(t) = Y(t)$.

Law of motion of the aggregate state The measure of incumbent firms $M(t)$ and productivity distribution $F(z,t)$ summarize the endogenous aggregate state of the extended model. Following the previous description on firm entry and exit,

$$\dot{M}(t) = E(t) - \delta M(t). \tag{33}$$

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

$$\frac{\partial F(z,t)}{\partial t} = -g_E(t)F(z,t). \tag{34}$$

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

$$\frac{\partial F(z,t)}{\partial t} = -\int_{z^*(t)}^{z} \eta(x,t)(1 - F(z|z \geq x,t))\mathrm{d}F(x,t) - g_E(t)\frac{F(z^*(t),t)\,(1-F(z,t))}{1-F(z^*(t),t)}. \tag{35}$$

The first term is the same as in the baseline case reflecting incumbents' idea search, while the second term captures the entry of new firms. Let $F^*(t) \equiv F(z^*(t),t)$ denote the share of search-inactive firms at time t. Moreover, define $T^*(z)$ as the stopping time when the obsolescence threshold $z^*(t)$ reaches productivity z, i.e., $z^*(T^*(z)) = z$ and $F^*(T^*(z)) = F(z, T^*(z))$.

The initial state of the economy mirrors the baseline model. Specifically, the productivity distribution $F(z,0) = 1 - z^{-k_0}$ for $z \geq 1$ with $k_0 > 1$, the obsolescence threshold $z^*(0) = 1$, and the measure of incumbent firms $M(0) = 1$.

Equilibrium The notions of competitive equilibrium and asymptotic balanced growth path are defined analogously to those in the baseline model. A full characterization is provided in Appendix D.1. In addition, I introduce the concept of a traveling wave equilibrium to capture balanced growth in the evolution of the productivity distribution.[28]

Definition (Traveling wave equilibrium)**.** An asymptotic traveling wave equilibrium is an equilibrium in which the productivity distribution $F(z,t)$ admits a velocity function $\gamma(t) > 0$ and a stationary distribution $\Gamma(z)$ such that for all z,

$$\lim_{t\to\infty} F(z\gamma(t),t) = \Gamma(z).$$

[28]This definition generalizes the BGP definition in Buera and Lucas (2018) that $F(ze^{vt},t) = \Gamma(z)$ for all $t \geq 0$ and some constant $v > 0$.

The equilibrium becomes an (exact) traveling wave equilibrium if the above equality also holds for all $t \geq 0$.

7.2 Solving the equilibrium

I follow the strategy in Section 4.2 to solve the equilibrium of the extended model. I first present firms' idea search policy and the resulting productivity distribution. I then derive the associated transition dynamics of the aggregate state and, finally, analyze the long-run behavior of the equilibrium.

To begin with, the following proposition extends the logic of the baseline analysis by showing that there exists an equilibrium in which all firms above the obsolescence threshold $z^*(t)$ search at the same intensity. The productivity distribution of search-active firms (the right tail) evolves analogously to the baseline model, maintaining a Pareto form with a declining tail index. Whereas, the distribution of search-inactive firms (the left tail) is shaped by technological obsolescence and firm entry, captured respectively by $g_z(t)$ and $g_E(t)$.

Proposition 4. *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

$$F(z,t) = \begin{cases} 1 - (1 - F^*(t)) \left(\frac{z}{z^*(t)}\right)^{-k(t)}, & \textit{if } z \geq z^*(t), \\ F^*(T^*(z))e^{-\int_{T^*(z)}^{t} g_E(s)\mathrm{d}s}, & \textit{if } z < z^*(t), \end{cases} \tag{36}$$

in which $\dot{k}(t)/k(t) = -\eta(t) < 0$, and $\dot{F}^*(t) = k(t)g_z(t)(1 - F^*(t)) - g_E(t)F^*(t)$.

With Proposition 4, the productivity distribution $F(z,t)$ can be summarized by two variables: the share of search-inactive firms $F^*(t)$ and the tail index $k(t)$. Together with the measure of incumbents $M(t)$, these variables constitute sufficient statistics for the endogenous aggregate state, whose evolution can be expressed as a system of first-order ordinary differential equations (ODEs):

Proposition 5. *In the above equilibrium, the dynamics of $(M(t), F^*(t), k(t))$ is* *governed by the following system of first-order ODEs:*

$$\frac{\dot{M}(t)}{M(t)} + \delta = g_E(t), \tag{37}$$

$$\dot{F}^*(t) = k(t) g_z(t) \left(1 - F^*(t)\right) - g_E(t) F^*(t), \tag{38}$$

$$g_{TG}(t) \equiv -\frac{\dot{k}(t)}{k(t)-1} = \frac{J(e^*(t))}{K(e^*(t))} \frac{g_E(t)}{1-F^*(t)}, \tag{39}$$

where $e^*(t) = (1-\tau_r)/k(t)$*, and* $g_E(t) = \frac{sK(e^*(t))L(t)}{\psi M(t) z^*(t)}$.

The first two equations follow directly from Equation (33) and Proposition 4. The startup threshold $e^*(t)$ and corresponding entry rate $g_E(t)$ are determined by the individual entrepreneurial choice. Equation (39) generalizes the tail dynamics in (15) from the baseline model and follows from the market clearing condition for researchers. $g_{TG}(t)$ then represents the rate of tail growth. Given exogenous paths of $L(t)$ and $z^*(t)$, the equilibrium trajectories of $(M(t), F^*(t), k(t))$ solve the ODE system above with initial conditions $M(0) = 1$, $F^*(0) = 0$, and $k(0) = k_0 > 1$.

Finally, I analyze the behavior of the equilibrium when time approaches infinity. To obtain an asymptotic balanced growth path, I focus on cases where population and the obsolescence threshold grow asymptotically at constant rates g_L^* and g_z^*, respectively, with $g_L^* + \delta > g_z^*$ to ensure positive firm entry in the limit. The following proposition summarizes the asymptotic features of the equilibrium.

Proposition 6. *Assume* $\lim_{t\to\infty} g_L(t) = g_L^*$*,* $\lim_{t\to\infty} g_z(t) = g_z^* > 0$*, and* $g_L^* + \delta > g_z^*$*. Suppose further that* $\rho > (1-\theta)g^*$*, the above equilibrium has the following features as* $t \to \infty$*:*

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

$$g_{TG}^* \equiv \lim_{t\to\infty} g_{TG}(t) = \frac{J(e^*)}{K(e^*)}(g_L^* + \delta); \tag{40}$$

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.,*

$$\lim_{t\to\infty} F(rz^*(t), t) = \Gamma(r) \equiv \begin{cases} fr^{\varsigma}, & \textit{for } r \in (0,1), \\ 1-(1-f)r^{-1}, & \textit{for } r \geq 1, \end{cases} \tag{41}$$

where $\varsigma = g_E^*/g_z^*$, *and* $g_E^* \equiv \lim_{t\to\infty} g_E(t) = g_L^* - g_z^* + \delta > 0$;

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

vi) The output per capita growth $g(t) \equiv \dot{y}(t)/y(t)$ *converges to* $g^* = \frac{g_{TG}^* + g_L^*}{\sigma - 1}$.

The condition $\rho > (1-\theta)g^*$ ensures finite household utility and that the transversality condition is satisfied. Proposition 6 establishes that the extended model features an asymptotic traveling wave equilibrium, in which the limiting productivity distribution exhibits a Zipfian right tail. The economy converges to a balanced growth path where the measure of incumbents, mean productivity, and output per capita all grow at asymptotically constant rates.

7.3 Discussions

Building on the equilibrium analysis above, I now discuss how the extended model generalizes and deepens the insights from the baseline framework.

First, the extended model preserves the four key properties established in Section 4.3: thickening right tails, Gibrat's law, Zipf's law, and an asymptotic balanced growth path. The first two follow directly from Proposition 4, although Gibrat's law now holds only within the right tail among search-active firms. The limiting results in Proposition 6 confirm the latter two: the right tail converges to Zipf's law, and output per capita exhibits asymptotic balanced growth. Moreover, the productivity distribution evolves as an asymptotic traveling wave, representing a form of "balanced growth" in the distribution itself.

Second, the extended model clarifies the nature of tail growth along two dimensions. The first concerns scale effects. Equation (40) shows that any expansion in the effective supply of skilled labor—whether through population growth (higher $L(t)$), improved search efficiency (lower ψ), or greater educational attainment (higher s)—is fully offset by increased firm entry. Consequently, the long-run growth rate remains unchanged. The absence of scale effects aligns with the second-generation endogenous growth models exemplified by Peretto (1998) and Young (1998).

The second dimension concerns the comparison between tail growth and variety expansion. With a love-of-variety preference, the startup threshold $e^*(t)$ embodies a trade-off between the variety gains from more entrepreneurship and tail growth generated by more intensive idea search among incumbents. The former yields only a level effect: the long-run variety growth rate g_M^* is determined by exogenous population growth and technological obsolescence. In contrast, tail growth remains a key driver of long-run growth. Consequently, policies that promote idea search by incumbent firms—for instance, subsidizing researchers ($\tau_r < 0$)—can still enhance the long-run growth rate.[29]

Finally, the extended model offers new perspectives into the link between firm size distribution and economic growth. Let $Z(t)$ denote the mean of the productivity distribution $F(z,t)$. Conventional firm-based growth models typically focus on equilibria where $F(z,t)$ forms an exact traveling wave, implying that $Z(t)$ grows at the same rate as the wave's velocity.[30] This restriction, however, rules out growth driven by reallocation dynamics. In contrast, the extended model features an *asymptotic* traveling wave equilibrium in which the growth rate of $Z(t)$ converges to $g_z^* + g_{TG}^*$, exceeding the velocity growth rate g_z^*. The additional component g_{TG}^* originates from the progressive thickening of the right tail toward Zipf's law. In this sense, tail growth provides a concrete mechanism through which reallocation gains can persist and contribute directly to long-run economic growth.

8 Conclusions

This paper develops new theory and evidence on the dynamics of firm size distributions along the process of economic growth. Empirically, I document that the right tail of the firm size distribution systematically thickens as economies grow. Theoretically, I construct a growth model based on idea search that rationalizes this pattern as an equilibrium feature of the growth path. The model further generates a secular rise in market concentration and Zipf's law in the firm size distribution. On the policy side, it suggests that policies favoring large firms can better utilize the diffusion externality resulting from idea search, leading to improved welfare.

Beyond providing an explicit theory for the thickening of the right tail along the growth path, the model highlights that the reallocation gains from tail growth can persist if the firm size distribution converges to a limiting distribution with tail index equal to one. By contrast, given appropriate initial conditions, standard

[29] Notably, the long-run output per capita growth rate g^* is independent of the exogenous technological progress $z^*(t)$. Since growth is jointly driven by expanding varieties and productivity improvements, $g^* = \frac{g_M^* + g_Z^*}{\sigma - 1} = \frac{g_L^* - g_z^*}{\sigma - 1} + \frac{g_z^* + g_{TG}^*}{\sigma - 1}$, where the effects of g_z^* cancel out.

[30] As an exact traveling wave, the productivity distribution is stationary (after scaling), i.e., $F(z\gamma(t), t) = \Gamma(z)$ with $\gamma(t)$ its velocity. Therefore, its mean $Z(t) \equiv \int_0^\infty z \, \mathrm{d}F(z,t) = \gamma(t) \int_0^\infty z \, \mathrm{d}\Gamma(z)$, growing at the same rate as $\gamma(t)$. While the productivity distribution is almost stationary in an asymptotic traveling wave equilibrium, $Z(t)$ needs not to grow at the same rate as the wave's velocity, as the extended model shows.

firm dynamics models such as Luttmer (2007) can, in principle, generate a firm size distribution with thickening right tails during the transition dynamics. In those 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 long-run implications, an interesting direction for future work is to compare these theories based on their ability to account for the historical decline in the tail index and to 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

Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva, and V. Sterzi (2018): "Dancing with the Stars: Innovation Through Interactions," Tech. Rep. w24466, National Bureau of Economic Research, Cambridge, MA.

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.

Baum-Snow, N., N. Gendron-Carrier, and R. Pavan (2024): "Local Productivity Spillovers," *American Economic Review*, 114, 1030–1069.

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*.

——— (2025): "Supplement to 'Economic Growth and the Rise of Large Firms'," *Supplemental Materials*.

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.

Cirillo, P. and J. Hüsler (2009): "On the upper tail of Italian firms' size distribution," *Physica A: Statistical Mechanics and its Applications*, 388, 1546–1554.

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.

di Giovanni, J., A. A. Levchenko, and R. Rancière (2011): "Power laws in firm size and openness to trade: Measurement and implications," *Journal of International Economics*, 85, 42–52.

Fujiwara, Y., C. Di Guilmi, H. Aoyama, M. Gallegati, and W. Souma (2004): "Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms," *Physica A: Statistical Mechanics and its Applications*, 335, 197–216.

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.

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

Gutiérrez, G. and T. Philippon (2017): "Declining Competition and Investment in the U.S." .

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.

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.

——— (2011): "On the Mechanics of Firm Growth," *The Review of Economic Studies*, 78, 1042–1068.

——— (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.

Peretto, P. F. (1998): "Technological Change and Population Growth," *Journal of Economic Growth*, 3, 283–311.

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. 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.

Supplement to
"Economic Growth and the Rise of Large Firms"

Zhang Chen
January 2026

Appendix A Additional empirical results

A.1 Details of tail estimation in Section 2.1

A.1.1 The Employment-based simple estimator

One can construct a similar simple estimator by directly inverting the top employment shares. Consider two thresholds $T_L > T_S$ from $\{T_i\}_{i=0}^n$, with corresponding top employment shares $E_L < E_S$. The smaller threshold T_S defines the right tail. Under the Pareto assumption in equation (1), $E_x = (T_x/T_0)^{1-k}$, the tail index can be recovered as

$$k^{emp} = 1 - \ln\frac{E_L}{E_S} / \ln\frac{T_L}{T_S}.$$

I verify that all empirical findings are robust to the employment-based simple estimator k^{emp}. For the OECD countries, Table A.1 reports the regression results from equation (2) using k^{emp}, and Figure A.4 illustrates the negative relationship between the tail index (based on k^{emp}) and log GDP per capita. For the United States, Figure A.5 shows that the tail index constructed from k^{emp} exhibits a similar downward trend over time. A detailed comparison of the three tail estimators is deferred to Section A.4.

A.1.2 The Toda-Wang estimator

I first introduce the original tail estimator in Toda and Wang (2021) for estimating the tail index of income distributions and then discuss its translation to the context of firm size distributions.

Toda and Wang (2021) consider a tabulated income distribution defined by a sequence of top income percentiles $0 < p_1 < \cdots < p_K < p_{K+1}$ and the corresponding top income shares $S_1 < \cdots < S_K < S_{K+1}$. They construct the following vector of self-normalized, non-overlapping top income shares $\bar{s}$:

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

A well-defined $\bar{s}$ requires $K \geq 2$, which implies the availability of at least three size thresholds.

If the underlying distribution follows a Pareto distribution with tail index α, then $S_k = p_k^{1-\frac{1}{\alpha}}$, and the normalized top income shares $\bar{s}$ are given by

$$\tilde{r}(\xi) \equiv (\tilde{r}_1(\xi), \ldots, \tilde{r}_{K-1}(\xi))^T \equiv \left(\frac{p_2^{1-\xi} - p_1^{1-\xi}}{p_{K+1}^{1-\xi} - p_K^{1-\xi}}, \ldots, \frac{p_K^{1-\xi} - p_{K-1}^{1-\xi}}{p_{K+1}^{1-\xi} - p_K^{1-\xi}} \right)^T,$$

where $\xi \equiv 1/\alpha$. Since S_k must be finite, $\alpha > 1$ is required. The estimator $\xi^{TW} \in (0, 1)$ minimizes the weighted distance between the model-implied moments $\tilde{r}(\xi)$ and the data moments $\bar{s}$:

$$\min_{\xi \in (0,1)} (\tilde{r}(\xi) - \bar{s})^T \tilde{\Omega}(\xi)^{-1} (\tilde{r}(\xi) - \bar{s}),$$

where $\tilde{\Omega}(\xi)$ is the optimal weighting matrix that achieves efficiency.[1] The estimated tail index is then $\alpha^{TW} = 1/\xi^{TW}$.

In the context of firm size distributions defined in Section 2.1, the top firm shares N_i correspond to the top income percentiles p_k, and the top employment shares E_i correspond to the top income shares S_k. These two sets of top shares can be translated as follows.

Let $\mathbf{p}_K = (p_k)_{k=1}^{K+1}$ and $\mathbf{S}_K = (S_k)_{k=1}^{K+1}$ denote the vectors of top income percentiles and shares, respectively. The moments defined above can be expressed as functions of these data vectors: $\bar{s} = \bar{s}(\mathbf{S}_K)$, $\tilde{r}(\xi) = \tilde{r}(\xi; \mathbf{p}_K)$, and $\tilde{\Omega}(\xi) = \tilde{\Omega}(\xi; \mathbf{p}_K)$. Following Toda and Wang (2021), I define $r(\alpha) \equiv \tilde{r}(1/\alpha)$ and $\Omega(\alpha) \equiv \tilde{\Omega}(1/\alpha)$. Now consider a right tail threshold T_R for the firm size distribution. I define the following mapping from $\mathbf{N}_R$ to $\mathbf{p}_K$ (simultaneously, $\mathbf{E}_R$ to $\mathbf{S}_K$):

$$K = n - R, \quad p_k = N_{-k+n+1}, \text{ and } S_k = E_{-k+n+1}.$$

Then, $(p_1, \cdots, p_K, p_{K+1}) = (N_n, \cdots, N_{R+1}, N_R)$, and $(S_1, \cdots, S_K, S_{K+1}) = (E_n, \cdots, E_{R+1}, E_R)$. Therefore, I define $\bar{s}(\mathbf{E}_R) = \bar{s}(\mathbf{S}_K)$, $r(k; \mathbf{N}_R) = \tilde{r}(1/k; \mathbf{p}_K)$, and $\Omega(k; \mathbf{N}_R) = \tilde{\Omega}(1/k; \mathbf{p}_K)$. The Toda-Wang estimator in Section 2.1 is then well-defined.

A.2 Details of the results in Section 2.2

A.2.1 Data sources

The OECD SBS The sample covers all OECD member countries as of 2018, excluding Chile, Mexico, and Korea due to data inconsistencies or incomplete series. Although the original database spans 2005-2018, I focus on the period 2008-2017. Few countries report data before 2008, and both 2007 and 2018 display systematic discontinuities, likely reflecting changes in measurement practices.

Firm size is measured by the number of employees, defined as the total number of persons working in or for the enterprise during the reference year, including both full-time and part-time workers. The raw dataset contains heterogeneous size classifications across countries and years—such as 1-249, 1-9, 10-19, 10-49, 20-49, 50-249, 250+, and Total. To enhance cross-country comparability, I standardize these categories into four uniform size classes: 1-9, 10-49, 50-249, and 250+.

The U.S. BDS The sample covers the universe of U.S. business firms. Firm size is measured by total employment, defined as the number of full- and part-time workers on the payroll during the reference year.

A.2.2 Full regression results for the OECD countries

Table A.1 reports the full regression results from equation (2) for all three estimators—k^f, k^{emp}, and k^{TW}. Notably, the sample size based on the Toda-Wang estimator (247) is substantially

[1] $\tilde{\Omega}$ depends only on ξ and $\{p_k\}_{k=1}^{K+1}$. See Proposition 3.1 of Toda and Wang (2021) for its exact functional form.

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 table 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 $T_R = 10$, 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

smaller than that based on the simple estimators (299 and 297), reflecting the frequency with which the constraint $k > 1.001$ binds during estimation.

A.2.3 Alternative thresholds for the U.S. tail indices

As shown in Figure A.1, the downward trend in the tail index remains consistent across alternative choices of thresholds.

A.3 Robustness analyses in Section 2.3

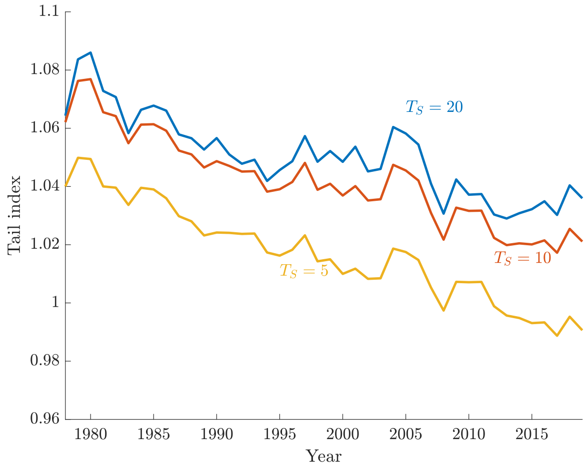
A.3.1 Across sectors

Table A.2 reports the results from regression (2) using the sectoral tail index as the dependent variable. The estimates show that both manufacturing and services exhibit thicker right tails as economies grow. If anything, the right tail tends to be thicker in manufacturing than in services. This implies that structural transformation—toward a greater share of services—would raise rather than lower the aggregate tail index.

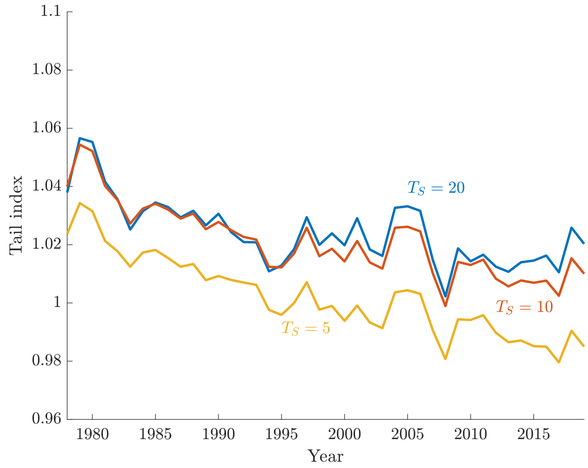
Notably, the sample size for the manufacturing sector is considerably smaller for the Toda-Wang estimator (87 observations) than for the simple estimators k^f (294) and k^{emp} (295). A plausible explanation is that manufacturing exhibits a thicker right tail, leading to more frequent binding of the constraint $k > 1.001$ in estimation.

A.3.2 Across countries

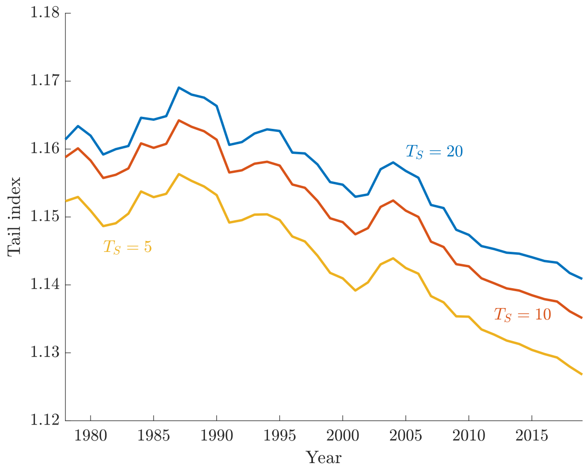
Table A.1 confirms that the negative relationship between the tail index and log GDP per capita holds among OECD countries even without controlling for country fixed effects. I next examine whether a similar cross-sectional relationship holds in developing economies using data from the World Bank Enterprise Survey (WBES).



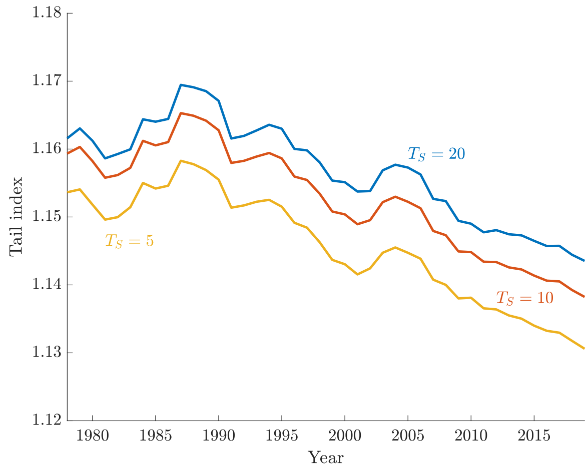
(a) k^f, $T_L = 500$



(b) k^f, $T_L = 1000$



(c) k^{emp}, $T_L = 500$



(d) k^{emp}, $T_L = 1000$

Figure A.1: Tail indices k^f and 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 either the firm-based or employment-based simple estimator (k^f or k^{emp}), with corresponding thresholds T_S and T_L as labeled in the labels. Data source: the U.S. Census 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 regression results of equation (2) between the tail index k and log GDP per capita among OECD countries by sectors. The tail index are k^{TW} with $T_R = 10$, 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

The WBES is a collection of surveys conducted by the World Bank designed to provide a representative portray of a country's business economy. Using an early WBES sample (2006-2010), both Ayyagari et al. (2014) and García-Santana and Ramos (2015) document that the employment share of large firms (100+ employees) is higher in high-income countries relative to low-income countries.

I extend this analysis using a more recent WBES sample covering 2006–2019, which yields broader country and time coverage. Specifically, I use the Employment Indicators—the official harmonized aggregate employment statistics—to obtain employment shares by firm size for each country-year pair. The sample covers more than 130 countries, of which 113 are classified as low-, lower-middle-, 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). Setting $T_S = 5$ and $T_L = 100$, the employment-based simple estimator k^{emp} becomes a monotonic transformation of the employment share of large firms, aligning with the methodologies in Ayyagari et al. (2014) and García-Santana and Ramos (2015).[2]

Figure A.2 plots the estimated tail index k^{emp} against log GDP per capita for the 113 developing countries in the sample.[3] The observed negative relationship shows that richer economies tend to display thicker right tails in their firm size distributions. Table A.3 reports the underlying regression results and further shows that this relationship remains robust when

[2]Formally, $k^{emp} = 1 - \log(\text{employment share of large firms})/\log(100/5)$. The choice of $T_S = 5$ ensures both the validity of the monotonic transformation and adequate coverage of the right tail in most developing economies. Bento and Restuccia (2021) show that the average non-agricultural establishment size is below five employees in 90% of countries with a log GDP per capita below 10.23—the highest value observed in my baseline WBES sample.

[3]The estimators k^f and k^{TW} cannot be computed here because the WBES Employment Indicators do not report firm counts by size.

extending the sample to include high-income countries or when adding year fixed effects.

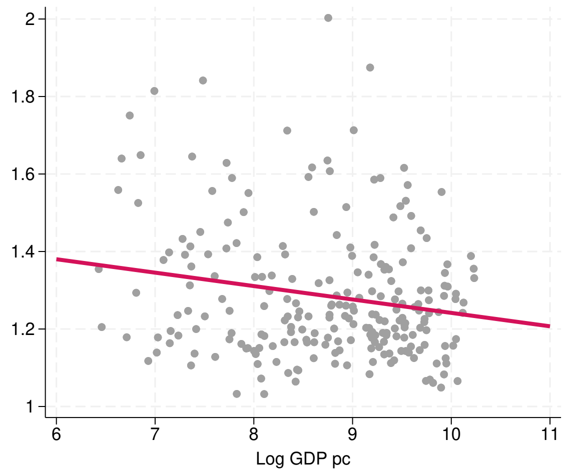


Figure A.2: Tail index and log GDP per capita in developing countries

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 Employment Indicators and PWT 10.0.

It is important to highlight several data challenges in extending the relationship between the tail index and economic development to the case of developing countries. First, commonly used datasets such as the WBES capture only formal firms, while informal firms—prevalent and typically smaller in developing economies—are excluded. This omission likely induces a downward bias in estimated tail indices. Second, a longstanding but debated hypothesis suggests that firm size distributions in developing countries may be bimodal, with relatively few mid-sized firms compared with advanced economies. This potential "missing middle" casts doubt on whether the Pareto distribution is still a reasonable approximation for the firm size distribution.[4] Third, cross-sectional evidence for developing countries does not directly speak to tail dynamics. The development literature has emphasized greatly the role of institutional distortions in shaping firm size distribution in developing countries, which are contrastingly low in developed countries. As economies grow and distortions evolve, the time path of the tail index may therefore differ from that observed in advanced economies. These challenges underscore the need for comprehensive panel data based on business censuses that cover both formal and informal sectors.[5]

Consistent with these data challenges, Figure A.2 shows substantial variation among countries with similar GDP per capita. Notably, many low-income countries exhibit very thick right tails. The R^2 for the cross-sectional regression of developing countries (country-year pairs) is

[4] Tybout (2000, 2014) and Hsieh and Olken (2014) provide contrasting views on the existence of a "missing middle" among mid-sized firms in developing countries. Notably, Tybout (2014) show that mid-sized firms account for a much smaller share of employment than would be implied by a Pareto distribution.

[5] A notable attempt is Abreha et al. (2022), who study the "missing middle" hypothesis using establishment-level censuses from Sub-Saharan African countries.

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 Employment Indicators and PWT 10.0.

only 0.036, considerably lower than that of the OECD countries, which stands at 0.124 (see Table A.1). In conclusion, future research are encouraged to utilize comprehensive firm-level panels to further investigate the relationship between tail index and economic development in developing countries.

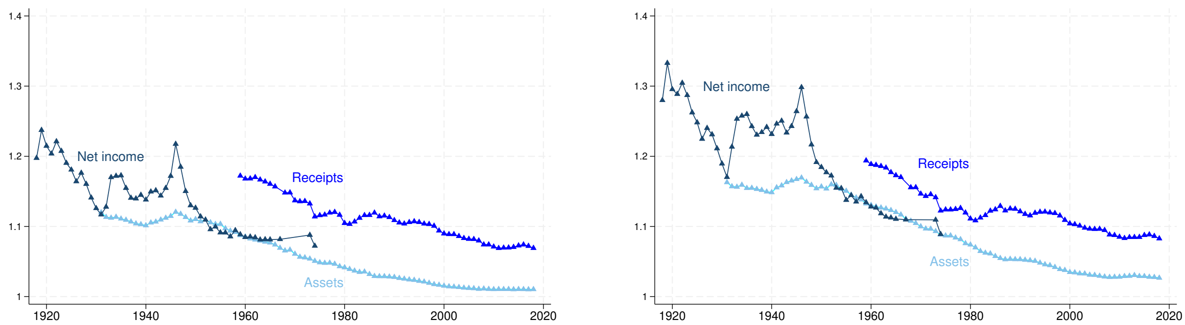
A.3.3 Across time

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 k^f and k^{emp} are not directly applicable due to missing information on size thresholds, their formulas imply an alternative simple estimator based solely on top firm and employment shares. Eliminating the term $\log(T_L/T_S)$ yields the following estimator for the tail index:[6]

$$1 - \frac{1}{k} = \frac{\log \frac{E_L}{E_S}}{\log \frac{N_L}{N_S}}. \tag{A.1}$$

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



(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 $N_L = 1\%, N_S = 10\%$ for the left figure and $N_L = 0.1\%, N_S = 10\%$ for the right figure. Data source: Kwon et al. (2024).

A.4 Comparing the estimators

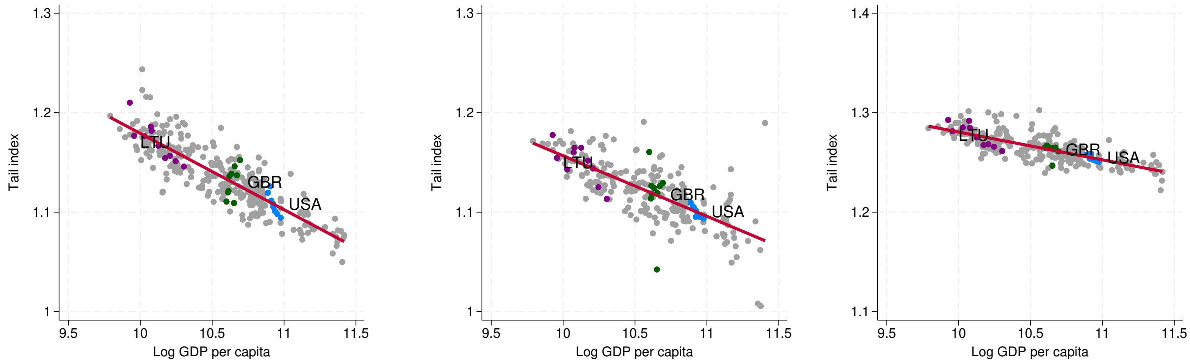
In this section, I compare the tail estimates obtained from the three estimators—k^f, k^{emp}, and k^{TW}. I first show that the main empirical findings in Section 2.2 are robust across all the three estimators. I then assess the relative performance of the two simple estimators, k^f and k^{emp}, using the Toda–Wang estimator k^{TW} as a benchmark. Finally, I examine how the choice of tail estimator affects the quantitative exercises in Section 5.

A.4.1 Comparing the results

Figures A.4 and A.5 compare the main results in Section 2.2 for the OECD and U.S. samples across tail indices estimated using k^f, k^{emp}, and k^{TW}. Two observations are immediate. First, the qualitative patterns are consistent across all estimators: the tail index is negatively associated with economic development in the OECD sample and declines over calendar time in the

[6] This estimator is widely used in studies of income and wealth inequality, where only top percentiles and shares are typically available. In principle, the Toda–Wang estimator is also applicable with three top thresholds. In practice, however, it produces too many binding estimates ($k \approx 1.001$ for many observations), rendering it unsuitable in this setting.

U.S. sample. Second, although the tail indices based on k^f and k^{TW} are also quantitatively well aligned, those based on k^{emp} exhibit noticeable departures. In particular, the k^{emp} series is systematically higher and displays substantially less variation over time.



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

Figure A.4: Tail index and log GDP per capita in OECD countries

Notes. This figure plots the tail index k against log GDP per capita for a panel of OECD countries. The scatter points are adjusted for country fixed effects, and the red lines represent fitted linear trends. The estimated slopes are $\hat{\beta}^f = -0.077(0.015)$, $\hat{\beta}^{TW} = -0.061(0.018)$, and $\hat{\beta}^{emp} = -0.028(0.009)$. The left and right panels report results using the respective simple estimator k^f and k^{TW} with $T_S = 10$ and $T_L = 250$, while the middle panel reports results using the Toda-Wang estimator k^{TW} with $T_R = 10$. Log GDP per capita is measured in constant international dollars. The three annotated countries are Lithuania (LTU, purple), the United Kingdom (GBR, green), and the United States (USA, blue). See Table A.1 for the complete regression results. Data sources: OECD SBS and PWT 10.0.

The properties of the simple estimator k^{emp} anticipates such departures. 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. Second, the time series of top employment shares is mechanically less volatile than that of top firm shares, as firms near the threshold have a greater impact on the latter.[7] As k^{emp} relies solely on top employment shares, this property makes its time series inherently less volatile.

More importantly, the construction of k^{emp} implies its inefficiency. Like k^f, it relies on only two size thresholds, rendering it less efficient than the Toda–Wang estimator k^{TW}, which exploits all top shares in the right tail. Moreover, k^{emp} not only shares the same lower-bound restriction $k > 1$ as k^{TW} but also guarantees that estimates remain strictly above one. This feature makes it at least as restrictive as k^{TW}, if not more so. Consequently, k^{emp} appears strictly dominated, as it inherits the inefficiency of k^f and the restrictiveness of k^{TW}. In what follows, I present direct evidence on the relative inefficiency of k^{emp}.

[7] Let $N_i(t)$ and $E_i(t)$ denote the top firm and employment shares at threshold T_i and year t, respectively. Using the U.S. BDS, I find that the variance of $\log N_i(t)$ over time, $\text{Var}_t(\log N_i(t))$, is roughly two to four times larger than that of $\log E_i(t)$, $\text{Var}_t(\log E_i(t))$, across all thresholds during 1978–2019.

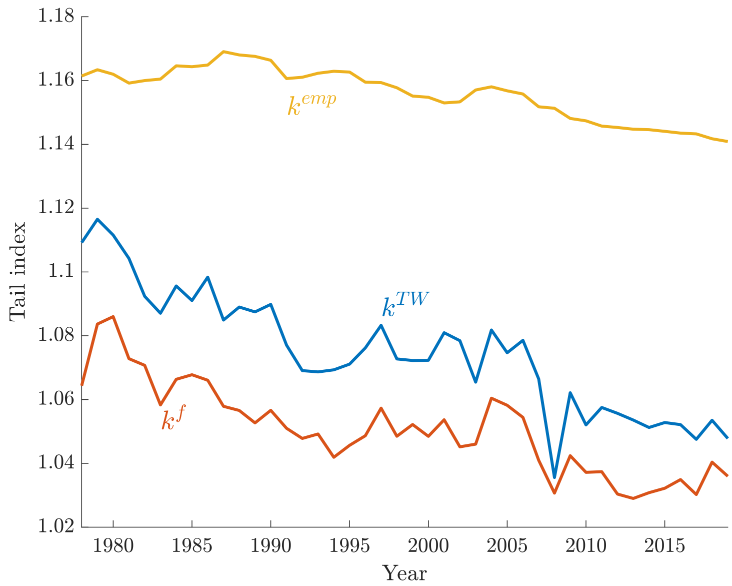


Figure A.5: Tail index in the US (1978-2019)

Notes. This figure plots the tail index k of the U.S. firm size distribution from 1978 to 2019. The tail indices are estimated using the Toda-Wang estimator, k^{TW}, with $T_R = 20$ (blue), and the simple estimators k^f and k^{emp}, with $T_S = 20$ and $T_L = 500$ (red and yellow). Data source: the U.S. Census BDS.

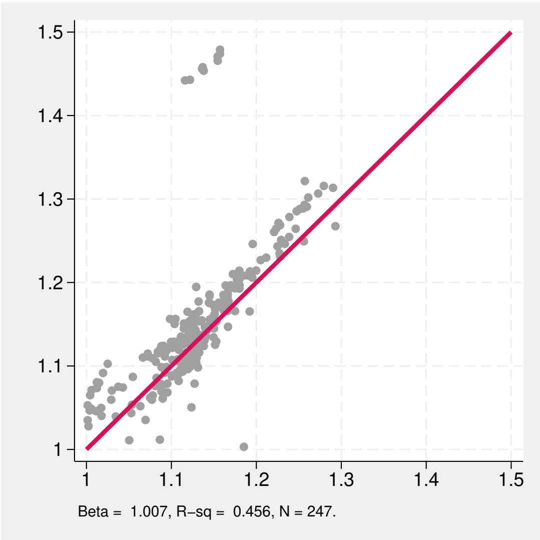
A.4.2 Comparing the tail indices

The Toda–Wang estimator is more efficient than the simple estimators whenever it is not binding. I therefore use it as an efficiency benchmark and compare the tail indices obtained from k^f and k^{emp} with those from k^{TW}, focusing only on observations where k^{TW} does not bind. For the United States, Figure A.5 shows that no binding cases occur over the sample period. The tail indices based on k^f closely track k^{TW}, while those based on k^{emp} are systematically larger and notably less volatile, exhibiting significant departures.

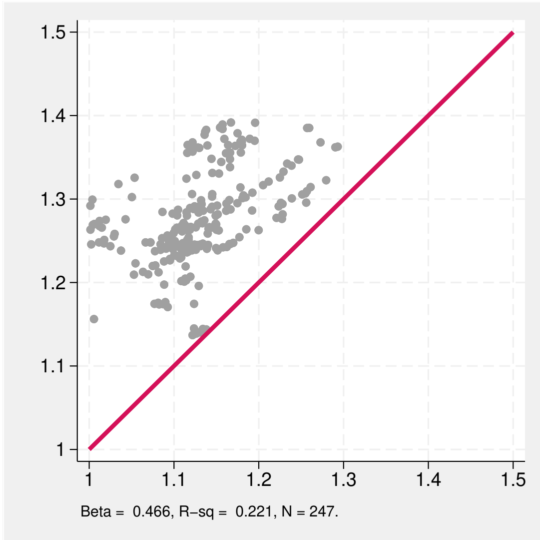
For OECD countries, I restrict the sample to country–year pairs where the k^{TW} constraint does not bind. Figure A.6 plots the tail indices based on k^f and k^{emp} against k^{TW} in this restricted sample. Proximity to the 45-degree line (in red) indicates the degree of alignment. The figure shows that k^{emp} systematically exceeds k^{TW}, while k^f aligns closely with it. The main exception is Turkey, which produces the cluster of outliers in the upper-left corner of Figure A.6(a). Regression results confirm this pattern: regressing k^f and k^{emp} on k^{TW} yields coefficients of 1.007 and 0.466, respectively.

The relatively low R^2 in the regression of k^f on k^{TW} is driven by the Turkish outliers. Excluding Turkey raises the R^2 from 0.456 to 0.811 (Figure A.6(c)), with no remaining outliers. By contrast, excluding Turkey barely changes the relationship between k^{emp} and k^{TW} (Figure A.6(d)), as the R^2 remains low—0.221 versus 0.239.

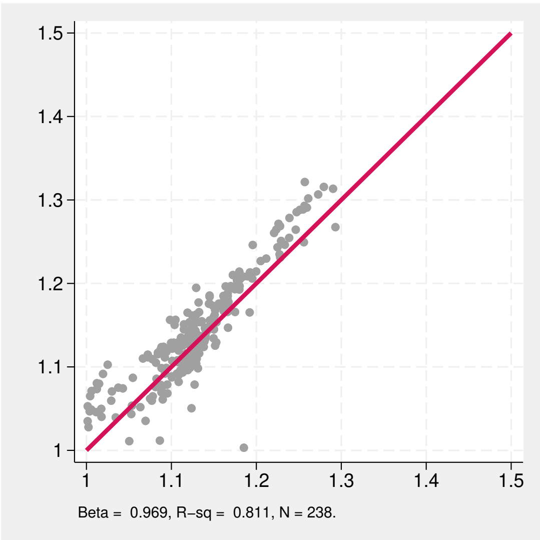
Finally, it is noteworthy that k^f produces estimates highly consistent with k^{TW} in both samples, despite its simplicity. Although k^f relies only on two top firm shares, it performs remarkably well, demonstrating reasonable efficiency and justifying its use as the preferred estimator for the main analysis. At least for firm size distributions, this comparison between



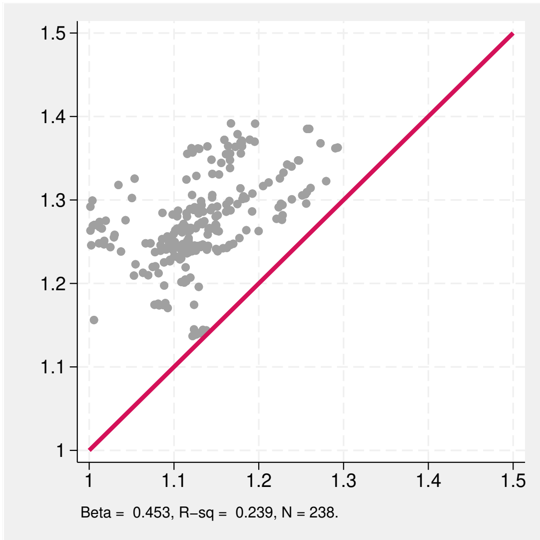
(a) k^f vs. k^{TW}, with Turkey



(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.6: 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-degree lines. Source: the OECD SBS.

k^f and k^{emp} seems to suggest that firm shares are more informative about tail behavior than employment shares.

A.4.3 Testing model predictions with simple estimators

y vs. $\frac{k}{k-1}$ In Figure A.7, 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 since the previous analysis has suggested that changes in employment shares do not accurately capture changes in overall tail behavior: k^{emp} is systematically higher and less volatile than k^{TW}. The distinction between estimates targeting firm shares (k^f) and employment shares (k^{emp}) also highlights the importance of using the Toda–Wang estimator k^{TW} for model–data comparison, as it more accurately captures the overall tail behavior and better disciplines moment selection.

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 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 solely by differences in the variance of $\ln k/(k-1)$.[8]

$\ln(k-1)$ vs. t In Figure A.8, I present the time series $\ln(k-1)$ 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.

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

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

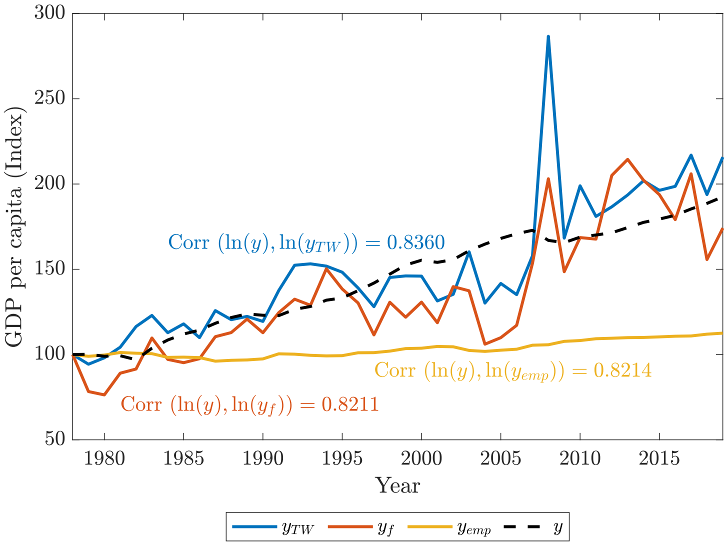
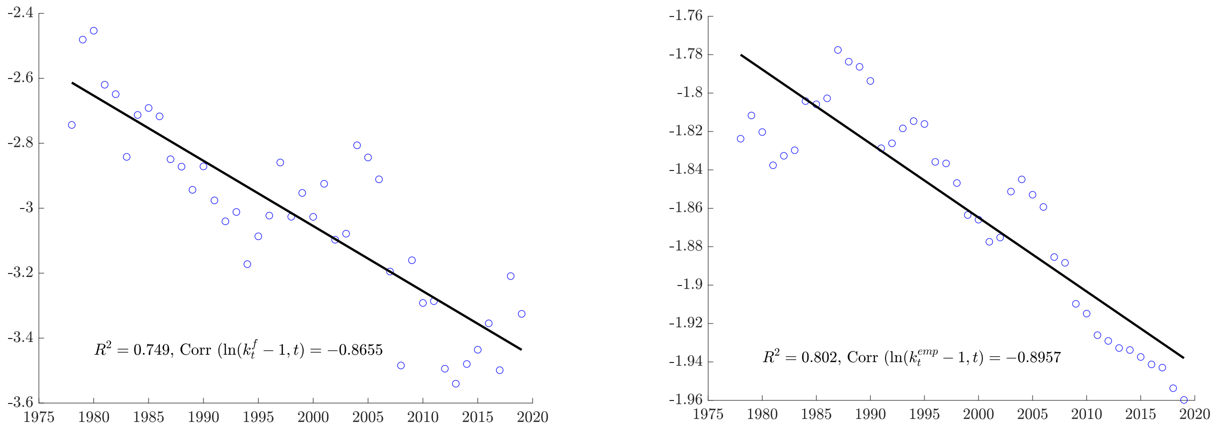


Figure A.7: y vs. $k/(k-1)$

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.



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

Figure A.8: $\ln(k-1)$ vs. t

Notes. Both figures compare the time series $\ln(k-1)$ 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.

Appendix B Additional analytical results

This section presents additional analytical results for the simple model in Section 3 and the growth model in Section 4. Proofs omitted from the main text are provided in Section B.5.

B.1 Initial conditions and external learning

In this section, I revisit the tail thickening mechanism in Section 3 within a more general environment to address two potential concerns. First, while the initial Pareto distribution in Assumption 1 renders important tractability, it is unclear how general initial conditions affect the thickening of the right tail. Second, in addition to "internal learning"—that is, learning from other firms—firms may also innovate by exploring the unknown. Following the seminal contribution of Kortum (1997), such innovation is typically modeled as the sampling of ideas from an external source, referred to as "external learning." It remains an open question how external learning interacts with internal learning in shaping the evolution of the productivity distribution's tail.

To address these concerns, I extend the model in Section 3 to incorporate both Kortum (1997)'s external learning and general initial distributions. In addition to sampling ideas from other firms, firms can now randomly sample ideas from a common external source and adopt those with productivity higher than firms' current level. One may interpret this external source as "the universe of ideas" as in Kortum (1997). I denote the productivity distribution of external ideas by G.

As in Section 3, the arrival rates of ideas—whether from internal or external sources—are common across firms and exogenously given. Let η_t denote the arrival rate of meetings between firms and β_t the arrival rate of external ideas. The evolution of the productivity distribution is then given by

$$\begin{aligned}\frac{\partial F(z,t)}{\partial t} &= -\eta_t\left[1-F(z,t)\right]\int_0^z \frac{f(x,t)}{1-F(x,t)}\mathrm{d}x - \beta_t\left[1-G(z)\right]F(z,t)\\ &= \eta_t\left[1-F(z,t)\right]\ln\left(1-F(z,t)\right) - \beta_t\left[1-G(z)\right]F(z,t), \end{aligned} \tag{A.2}$$

where the first item on the right hand side captures firms' internal learning through truncated sampling, and the second item captures firms' external learning through random sampling.

Meanwhile, I introduce the following definition of the tail index for general distributions. A distribution F has tail index k if

$$\lim_{x\to\infty}\frac{\ln\left(1-F(x)\right)}{\ln x} = -k, \qquad \text{where} \qquad k\in\mathbb{R}^+\cup\{+\infty\}. \tag{A.3}$$

This definition is slightly weaker than the regular variation definition used in Section 6.2, and the two are equivalent whenever $\lim_{x\to\infty}\frac{xf(x)}{1-F(x)}$ exists. I refer to a distribution as thin-tailed if $k=+\infty$ and thick-tailed if $k<+\infty$. Distributions with bounded support or exponential decay are typical examples of thin-tailed distributions, whereas Pareto distributions are representative thick-tailed cases. I then impose the following assumption on the productivity distributions of initial firms and external ideas.

Assumption 2. *The initial productivity distribution $F(z,0)$ and the external source $G(z)$ have respective tail indices k_0 and k_e.*

Assumption 2 is mild, requiring only that these distributions admit well-defined limits as productivity approaches infinity. It imposes no restriction on whether the tails are thick or thin, nor on whether the supports are bounded. Given these generalizations, the following proposition characterizes the tail index of the productivity distribution at all times.

Proposition A.1. *Suppose Assumption 2 holds. If firms' learning is purely internal, i.e., $\beta_s = 0$ for $s \in [0,t]$, then $F(z,t)$ has tail index*

$$k(t) = k_0 e^{-\int_0^t \eta_s \mathrm{d}s}.$$

If external learning is also present, i.e., $\beta_s > 0$ for $s \in [0,t]$, then

$$k(t) = \min\{k_0, k_e\} e^{-\int_0^t \eta_s \mathrm{d}s}.$$

Proposition A.1 shows that the tail dynamics derived in the simple model remain robust along three dimensions. First, the result does not rely on the Pareto assumption for the initial distribution. With only internal learning, the productivity process in (A.2) reduces to that in Section 3, and the same expression for $k(t)$ holds for general initial distributions. In particular, a productivity distribution that begins thin-tailed remains thin-tailed over time.

Second, when external learning is present, a thin-tailed initial distribution can develop a thickening right tail. Whenever external source has a thick-tailed distribution, the resulting productivity distribution behaves as if it originated from a thick-tailed initial condition with tail index $\min\{k_e, k_0\}$. Internal learning through truncated sampling then continues to thicken the tail in the same manner as in the simple model.

Third, the evolution of the tail index is governed entirely by internal learning, regardless of whether external learning is present. When the productivity distribution is thick-tailed, the rate of tail thickening, $-\dot{k}(t)/k(t)$, is determined solely by the internal idea arrival rate η_t.

In sum, if the productivity of external ideas follows a thick-tailed distribution, as in Kortum (1997), the tail dynamics in a model with both internal and external learning and a general initial productivity distribution are equivalent to those in a model with only internal learning and a thick-tailed initial distribution. Since the analysis in this paper focuses on tail behavior, it entails little loss of generality to restrict attention to models with internal learning only and a thick-tailed initial productivity distribution.

B.2 Idea search with technological choices

In the baseline model of Section 4, the unit search cost is assumed to scale linearly with firm productivity. This assumption can be microfounded as the equilibrium outcome of firms' endogenous choices over idea search technologies. In what follows, I formalize firms' technology choices and show that the induced search behavior is observationally equivalent to that assumed in the baseline model.

Firms face a common set of search technologies but may adjust their choice according to their own productivity level. Each firm selects a search technology before deciding on its search intensity. A firm with productivity z faces a continuum of search technologies $\chi_z(m)$ indexed by efficiency $m \in [1, +\infty)$. The trade-off in this choice is that a more efficient search technology increases the probability of meeting high-productivity firms in each search but also requires more researchers per search. Formally, a search technology is represented by a pair consisting of a source distribution and a 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 specifies the targeting threshold: search technology $\chi_z(m)$ targets firms with productivity at least m and samples randomly from this set. The second component defines the unit search cost, i.e., it requires $\max\{m^{\chi} z^{1-\chi}, m\}$ researchers to complete one such search. Analogous to standard capital adjustment costs, the convexity parameter $\chi > 1$ captures the increasing cost of adopting a search technology more efficient than the firm's own productivity level z. Finally, firms never adopt technologies from less productive firms in their meetings.

This specification is designed to capture two key features. First, it mirrors the spirit of the truncated sampling assumption: more productive firms have an advantage in meeting top firms. Second, targeting higher-productivity firms becomes increasingly difficult due to the convexity of the unit search cost.

With technology choice, the HJB equation for the firm's value becomes

$$\begin{aligned} r(t)v(z,t) =& z + \partial_t v(z,t) \\ &+ \max_{\eta \geq 0} \eta \Big[\max_{m \in [1,+\infty)} \Big\{ \int_{\max\{m,z\}}^{\infty} [v(x,t) - v(z,t)] \mathrm{d}F(x|x \geq m, t) \\ &\qquad - \max\left\{ m^{\chi} z^{1-\chi}, m \right\} w(t) \Big\} \Big], \end{aligned} \tag{A.4}$$

with $m(z,t)$ denotes the optimal technological choice. Relative to the baseline case, firms first select a targeting threshold m to maximize the expected return from each search and then choose the search intensity η to maximize the total return. The following proposition characterizes the optimal targeting threshold:

Proposition A.2. *When $k_0 < \chi$, $m(z,t) = z$ for all t and z.*

Under appropriate parameter restrictions, firms optimally target peers with similar productivity, i.e., $m(z,t) = z$, and this policy supports the equilibrium characterized in the baseline model.

B.3 The dynamics of equilibrium prices

I discuss here the transition dynamics of two prices, $w(t)$ and $r(t)$, in the equilibrium developed in Section 4.2. First, the dynamics of interest rate $r(t)$ follows directly from (20):

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

where $k(t) = 1 + (k_0 - 1)e^{-Lt}$. 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. Recall that

$$w(t) = \frac{v(t)}{k(t) - 1}, \qquad \text{where} \qquad v(t) = \int_t^\infty e^{-\int_t^x r(s)\mathrm{d}s} \mathrm{d}x.$$

As $r(t) \to r^*$, $v(t) \to 1/r^*$, and $w(t) \sim \frac{e^{Lt}}{r(k_0-1)}$. In other words, wage growth follows the output per capita growth and also converges to L in the limit.

To further characterize wage dynamics, I normalize researchers' wages by household income. Formally, the normalized wage is defined as

$$\tilde{w}(t) \equiv \frac{w(t)}{y(t)} = \frac{w(t)}{\frac{k(t)}{k(t)-1}\frac{1}{L}} = \frac{v(t)L}{k(t)} = \frac{g(t)}{\tilde{r}(t)}, \tag{A.5}$$

where $\tilde{r}(t) = 1/v(t)$. Since $1/\tilde{r}(t) = \int_t^\infty e^{-\int_t^x \tilde{r}(t)\mathrm{d}s}\mathrm{d}x$, $\tilde{r}(t)$ represents the effective average interest rate from time t onward. Hence, the normalized wage is governed by the relative dynamics between output per capita growth and the average interest rate. The following proposition characterizes the evolution of the normalized wage.

Proposition A.3. *The normalized wage $\tilde{w}(t)$ increases to $\tilde{w}^* = 1/(\rho/L + \theta) < 1$.*

Since $v(t) \to 1/r^*$, the long-run normalized wage equals the ratio of the long-run growth rate (L) to the long-run interest rate (r^*). The monotonicity of $\tilde{w}(t)$, however, is less straightforward, as both the growth rate and the average interest rate increase over time. Recall that the instantaneous interest rate satisfies $r(t) = \rho + \theta g(t)$, the sum of the discount rate and a linear function of the instantaneous growth rate. It follows that $r(t)$ grows more slowly than $g(t)$, so that $g(t)/r(t)$ increases over time. When t is sufficiently large, the average interest rate $\tilde{r}(t)$ converges to the instantaneous rate $r(t)$, suggesting that the normalized wage $g(t)/\tilde{r}(t)$ should also rise over time. Proposition A.3 confirms that this intuition holds exactly.

Proposition A.3 also 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.

B.4 Non-linear idea search costs

I introduce nonlinearity into the cost function of firms' idea search in the growth model of Section 4. Specifically, a firm with productivity z must now hire $z(\eta + g(\eta))$ researchers to achieve a search intensity of η. The adjustment cost function g is twice differentiable and satisfies $\lim_{\eta \to 0} \eta g''(\eta) = 0$. Moreover, g is strictly increasing and convex, with $g(0) = g'(0) = 0$. I refer to the equilibrium in Section 4.2 as the baseline equilibrium. Below, I show that the characterization of the baseline equilibrium extends to this case with the adjustment cost.

Solving the equilibrium As in Section 4.2, I first assume that all firms search at the same intensity and then verify the optimality of this strategy. Suppose $\eta(z,t) = \eta(t) > 0$ for all firms z and t. It follows from Section 3 that the resulting productivity distribution is $F(z,t) = 1 - z^{-k(t)}$ for $z \geq 1$, where $\dot{k}(t)/k(t) = -\eta(t) < 0$.

Next, I solve for the equilibrium paths of $k(t)$ and $\eta(t)$. The labor market clearing condition implies

$$(\eta(t) + g(\eta(t))) \frac{k(t)}{k(t) - 1} = L. \tag{A.6}$$

Since g is strictly increasing, Equation (A.6) yields a unique $\eta(t)$ for each $k(t)$, defining an implicit function $\eta(k)$ rather than the explicit one obtained in the baseline equilibrium. Substituting $\eta(k)$ into the law of motion for $k(t)$, $\dot{k}(t)/k(t) = -\eta(t)$, again gives a first-order ordinary differential equation in $k(t)$ with initial condition $k(0) = k_0$. Its solution determines the equilibrium path of $k(t)$ and the associated $\eta(t)$.

To characterize the equilibrium properties of $k(t)$, note that

$$k(t) = 1 + (k_0 - 1) \exp\left(- \int_0^t \tilde{L}(s) \mathrm{d}s \right), \tag{A.7}$$

in which $\tilde{L}(t) = \frac{L}{1 + g(\eta(t))/\eta(t)}$. This expression mirrors Equation (16) in the baseline equilibrium, except that $\tilde{L}(t) = L$ in that case. Since $k(t)$ is decreasing and $g'(\eta) > 0$, Equation (A.6) implies that $\eta(t)$ declines over time. Moreover, because $g(0) = 0$ and $g''(\eta) > 0$, the ratio $g(\eta)/\eta$ increases with η. Consequently, $\tilde{L}(t)$ rises over time, satisfying $\tilde{L}(t) \geq \tilde{L}(0) > 0$. As the integral in (A.7) diverges, $k(t)$ converges to 1. Meanwhile, $\eta(t)$ converges to 0, and $\tilde{L}(t)$ converges to L because $\lim_{\eta \to 0} g(\eta)/\eta = g'(0) = 0$. Convergence to Zipf's law is then preserved.

Given the equilibrium path of $k(t)$, output per capita is $y(t) = \frac{k(t)}{k(t)-1} \frac{1}{L}$. The corresponding growth rate is

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

Since $\tilde{L}(t)$ converges to L, $g(t)$ also converges to its baseline counterpart with limit L. The equilibrium thus features asymptotic balanced growth. Similarly, the interest rate $r(t) = \theta g(t) + \rho$ converges to $\theta L + \rho$. The same parametric condition, $\rho > (1 - \theta)L$, guarantees that the transversality condition holds.

Lastly, I verify the optimality of the search strategy $\eta(z,t) = \eta(t)$. I guess and verify that a value function of the form $v(z,t) = v(t)z$ solves the firm's HJB equation and identify the corresponding $w(t)$ that supports this equilibrium strategy. Suppose $v(z,t) = v(t)z$. The expected return from one idea search is 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 each firm satisfies

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

Hence, $w(t)$ is independent of z and can be expressed as a function of $v(t)$, $k(t)$, and $\eta(t)$. Conversely, the convexity of g ensures that $\eta(t)$ is the unique optimal search intensity given $v(t)$, $w(t)$, and $k(t)$.

I now solve for the per-productivity value $v(t)$. The net payoff from idea search under the optimal choice $\eta(t)$ is

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

which is positive (since g is strictly convex) and linear in z. Then $v(z,t) = v(t)z$ satisfies the firm's HJB equation if

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

Given known paths for $k(t)$, $\eta(t)$, and $r(t)$, this is a first-order ordinary differential equation in $v(t)$, which admits the solution

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

where

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

Moreover, $\bar{r}(t)$ converges to $r(t)$. Using η as shorthand for $\eta(t)$ and noting that $\eta(t) \to 0$,

$$\begin{aligned}\lim_{t\to 0} \frac{\eta g'(\eta) - g(\eta)}{1+g'(\eta)} \frac{1}{k(t)-1} =& \lim_{t\to 0} \frac{\eta g'(\eta) - g(\eta)}{1+g'(\eta)} \frac{1}{(\eta+g(\eta))k(t)} \frac{(\eta+g(\eta))k(t)}{k(t)-1} \\ =& L \lim_{t\to 0} \frac{\eta g'(\eta) - g(\eta)}{1+g'(\eta)} \frac{1}{(\eta+g(\eta))k(t)} \\ =& L \lim_{t\to\infty} \frac{1}{k(t)(1+g'(\eta))} \lim_{t\to\infty} \frac{\eta g'(\eta) - g(\eta)}{\eta+g(\eta)} \\ =& L \lim_{\eta\to 0} \frac{\eta g''(\eta)}{1+g'(\eta)} = 0,\end{aligned}$$

where the second equality follows from the labor market clearing condition, the fourth from L'Hôpital's rule, and the last from $\lim_{\eta\to 0} \eta g''(\eta) = 0$. Therefore, both $v(t)$ and $w(t)$ are finite and converge to the same limits as in the baseline equilibrium. All equilibrium conditions are thus satisfied.

Discussions The above equilibrium analysis clarifies two main points. First, the linear search cost assumption is imposed primarily for analytical tractability. As the construction above shows, the equilibrium with adjustment costs retains all four key properties of the baseline equilibrium in Section 4.3. Relative to the baseline case, this equilibrium features a unique optimal search intensity for each firm but no longer admits closed-form solutions.

Second, while multiple equilibria may exist under linear search costs, the baseline equilibrium is preferred because it emerges as the limiting case of the equilibrium with adjustment costs. On one hand, for any equilibrium with the specified adjustment cost, the equilibrium path converges to that of the baseline model in the long run, since the adjustment cost becomes negligible as search intensity approaches zero. On the other hand, the baseline equilibrium can be interpreted as that of a limiting economy in which the adjustment cost approaches zero.[9]

B.5 Omitted proofs

B.5.1 Proof of Proposition A.1

Proof. With only internal learning, I obtain the same expression for $F(z,t)$ as in Section 3 following the same derivation. It is straightforward to verify that the expression for $k(t)$ remains valid under the general definition of the tail index in (A.3). I then turn to the case with external learning. Because it is difficult to compute the tail index of $F(z,t)$ directly, the approach is to construct upper and lower bounds for $F(z,t)$ and show that these bounding distributions share the same tail index. The proof proceeds in four steps.

[9]Although the adjustment cost eliminates equilibrium multiplicity arising from linear search costs, the resulting equilibrium is not necessarily unique. This construction demonstrates uniqueness only when the firm's value function is linear in productivity. Other equilibria may exist in which the HJB equation admits nonlinear value functions as solutions. Determining whether the equilibrium is unique 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 flow models.

Step 0: In this step, I construct 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:

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

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.2) 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.2) 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.1 and A.2.

Lemma A.1. *$F^l(z,t;\tau)$ 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,

$$1 - F^l(z,t;\tau) = \left[1 - G(z)\right]\left[\frac{1 - F(z,0)}{1 - G(z)} + F(z,0)\frac{1 - \exp\left(-\left[1 - G(z)\right]\int_0^t \beta_s \mathrm{d}s\right)}{1 - G(z)}\right] \tag{A.8}$$

$$= \left[1 - F(z,0)\right]\left[1 + F(z,0)\frac{1 - \exp\left(-\left[1 - G(z)\right]\int_0^t \beta_s \mathrm{d}s\right)}{1 - G(z)}\frac{1 - G(z)}{1 - F(z,0)}\right] \tag{A.9}$$

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.8) 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 $k_0\le k_e$. Notice that

$$\lim_{z\to\infty}\frac{\ln\left\{1+F(z,0)\frac{1-\exp\left(-[1-G(z)]\int_0^t\beta_s\mathrm{d}s\right)}{1-G(z)}\frac{1-G(z)}{1-F(z,0)}\right\}}{\ln z}=0. \tag{A.10}$$

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.10) holds from its asymptotic equivalence with $\frac{\ln\left\{\frac{1-G(z)}{1-F(z,0)}\right\}}{\ln z}$. In all cases, (A.9) 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.2. *For $t\ge\tau$,* $\liminf_{z\to\infty}\frac{\ln\left(1-F^l(z,t;\tau)\right)}{\ln z}\ge-\min\{k_0,k_e\}e^{-\int_\tau^t\eta_s\mathrm{d}s}$.

Proof. For $t\ge\tau$, the dynamics follows the differential equation (A.2). Rewriting it gives

$$w'(t)=-\eta_t w(t)+\beta_t u\left(e^{-w(t)}-1\right) \tag{A.11}$$

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

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

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

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

Then,

$$\liminf_{z\to\infty}\frac{\ln\left(1-F^l(z,t;\tau)\right)}{\ln z}\ge\lim_{z\to\infty}e^{-\int_\tau^t\eta_s\mathrm{d}s}\frac{\ln\left(1-F^l(z,\tau;\tau)\right)}{\ln z}=-\min\{k_e,k_0\}e^{-\int_\tau^t\eta_s\mathrm{d}s}$$

■

Step 2: In this step, I establish the tail index of the upper bound $F^u(z,t)$ in Lemma A.3.

Lemma A.3. *$F^u(z,t)$ has tail index* $\min\{k_0,k_e\}e^{-\int_0^t\eta_s\mathrm{d}s}$.

Proof. It follows from the previous proof that

$$\frac{\ln\left(1-F^u(z,t)\right)}{\ln z}=e^{-\int_0^t\eta_s\mathrm{d}s}\frac{\ln\left(1-F^u(z,0)\right)}{\ln z}+\frac{1-G(z)}{\ln z}\int_0^t e^{-\int_s^t\eta_r\mathrm{d}r}\beta_s\frac{F^u(z,s)}{1-F^u(z,s)}\mathrm{d}s. \tag{A.12}$$

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.1 apply to show that $F^u(z,0)$ has tail index $\min\{k_e,k_0\}$.

Next, (A.12) implies that

$$\liminf_{z\to\infty}\frac{\ln\left(1-F^u(z,t)\right)}{\ln z}\geq\lim_{z\to\infty}e^{-\int_0^t\eta_s\mathrm{d}s}\frac{\ln\left(1-F^u(z,0)\right)}{\ln z}=-\min\{k_e,k_0\}e^{-\int_0^t\eta_s\mathrm{d}s}$$

Alternatively, $\frac{F^u(z,s)}{1-F^u(z,s)}$ decreases on s since $F^u(z,s)$ decreases on s. Then,

$$\frac{\ln\left(1-F^u(z,t)\right)}{\ln z}\leq e^{-\int_0^t\eta_s\mathrm{d}s}\frac{\ln\left(1-F^u(z,0)\right)}{\ln z}+\frac{F^u(z,0)}{\ln z}\frac{1-G(z)}{1-F^u(F,0)}\int_0^t e^{-\int_s^t\eta_r\mathrm{d}r}\beta_s\mathrm{d}s.$$

By construction, $1-F^u(z,0)>1-G(z)$. Then, for any t,

$$\frac{F^u(z,0)}{\ln z}\frac{1-G(z)}{1-F^u(F,0)}\int_0^t e^{-\int_s^t\eta_r\mathrm{d}r}\beta_s\mathrm{d}s=O\left(\frac{1}{\ln z}\right)=o(1)$$

at $z=\infty$. Consequently,

$$\limsup_{z\to\infty}\frac{\ln\left(1-F^u(z,t)\right)}{\ln z}\leq e^{-\int_0^t\eta_s\mathrm{d}s}\lim_{z\to\infty}\frac{\ln\left(1-F^u(z,0)\right)}{\ln z}=-\min\{k_e,k_0\}e^{-\int_0^t\eta_s\mathrm{d}s}.$$

Combining both bounds, I conclude that

$$\lim_{z\to\infty}\frac{\ln\left(1-F^u(z,t)\right)}{\ln z}=-\min\{k_e,k_0\}e^{-\int_0^t\eta_s\mathrm{d}s}.$$

■

Step 3: I return to the tail index of $F(z,t)$ in this last step. From Lemma A.1 and A.2, I obtain that

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

Since τ is arbitrary,

$$\liminf_{z\to\infty}\frac{\ln\left(1-F(z,t)\right)}{\ln z}\geq-\min\{k_0,k_e\}e^{-\int_0^t\eta_s\mathrm{d}s}.$$

Conversely, Lemma A.3 implies that

$$\limsup_{z\to\infty}\frac{\ln\left(1-F(z,t)\right)}{\ln z}\leq\lim_{z\to\infty}\frac{\ln\left(1-F^u(z,t)\right)}{\ln z}=-\min\{k_e,k_0\}e^{-\int_0^t\eta_s\mathrm{d}s}.$$

The proof is complete by combining these two bounds. ■

B.5.2 Proof of Proposition A.2

Proof. I show that $m(z,t)=z$ is consistent with the baseline equilibrium, where productivity distribution $F(z,t)$ is a Pareto distribution with varying tail index $k(t)$, and value function $v(z,t)=v(t)z$. I 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 $k(t)\leq k_0$, I 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, I 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. ■

B.5.3 Proof of Proposition A.3

Proof. Recall from (A.5) that 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.[10]

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

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,

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

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 $\alpha > 0$,

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

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

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,[11]

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

As $k \to 1$, $F(\nu+1,1;\nu+\alpha+1;1-\frac{1}{k}) \to 1$. Then, with again $\rho > L(1-\theta)$,

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

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 $\tilde{w}(k;\alpha)$. ■

[11] One can also obtain the same $\tilde{w}'(k)$ from here, but the computation will be much harder.

Appendix C Additioanl results for policies

C.1 Proof of Proposition 2

Proof. Consider the policy in Section 6.1 that firms with productivity below z^* face a search tax of $\tau \geq 0$. Obviously, $\tau = 0$ corresponds to the tax-free economy in Section 4. I show that the following search strategy $\eta(z,t)$ constitutes an equilibrium under this policy. Namely,

$$\eta(z,t) = \begin{cases} 0, & \text{if } z < z^*, \\ \eta(t), & \text{if } z \geq z^*. \end{cases} \tag{A.13}$$

In words, firms with productivity below z^* do not search, and those with productivity above z^* search at the same intensity $\eta(t)$.

To begin with, it is straightforward that $F(z,t) = F(z,0)$ for all $z < z^*$ and t since there are no actions for firms in this region, and firms above z^* will not drop into this region. For the part that is above the threshold, I substitute the search strategy in (A.13) into the KFE (12) of the productivity distribution in Section 4 and obtain that

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

Similar to Section 3, I derive a first-order linear ODE in variable $\ln \tilde{F}(z,t) - \ln \tilde{F}(z^*,0)$ using $\tilde{F}(z^*,t) = \tilde{F}(z^*,0)$. Analogously, it has the solution

$$\ln\left(\frac{\tilde{F}(z,t)}{\tilde{F}(z^*,0)}\right) = \ln\left(\frac{\tilde{F}(z,0)}{\tilde{F}(z^*,0)}\right)\exp\left(-\int_0^t \eta(s)\mathrm{d}s\right).$$

Using the initial condition $F(z,0) = 1 - z^{-k_0}$,

$$\tilde{F}(z,t) = \tilde{F}(z^*,0)\left(\frac{z}{z^*}\right)^{-k(t)} = (z^*)^{k(t)-k_0} z^{-k(t)}, \tag{A.14}$$

where $k(t) = k_0 \exp\left(-\int_0^t \eta(s)\mathrm{d}s\right)$, or equivalently, $\dot{k}(t)/k(t) = -\eta(t)$.

Next, I solve for $\eta(t)$ from labor market clearing condition. With both (A.14) and (A.13),

$$\eta(t)\int_{z^*}^{\infty} y f(y,t)\mathrm{d}y = L \Rightarrow \eta(t)\frac{k(t)}{k(t)-1} = L(z^*)^{k_0-1}.$$

Combining it with $\dot{k}(t)/k(t) = -\eta(t)$, I obtain the equilibrium $k(t)$ under threshold z^*:

$$-\frac{\dot{k}(t)}{k(t)-1} = L(z^*)^{k_0-1} \Rightarrow k(t) = 1 + (k_0 - 1)\exp\Big(-L(z^*)^{k_0-1}t\Big).$$

(23) in the proposition then follows from (A.14). Moreover,

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

Therefore, $\dot{y}(t)/y(t)\to L(z^*)^{k_0-1}$. While small firms never grow, the aggregate economy benefits from the faster growth among large firms.

Finally, I verify the optimality of the strategy in (A.13). With tax, the HJB equation (8) becomes

$$r(t)v(z,t)=z+\max_\eta \eta\left\{\int_z^\infty [v(x,t)-v(z,t)]\,\mathrm{d}F(x|x\ge z,t)-z(1+\tau(z))w(t)\right\}+\partial_t v(z,t),$$

in which $\tau(z)=\tau\mathbf{1}\{z<z^*\}$. As argued in Section 4.2, the equilibrium net payoff per idea search $S(z,t)$ must be non-positive:

$$S(z,t)=\int_z^\infty [v(x,t)-v(z,t)]\,\mathrm{d}F(x|x\ge z,t)-z(1+\tau(z))w(t)\le 0.$$

The same strategy shows that in equilibrium, value function $v(z,t)=v(t)z$, where

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

with $r(t)$ the equilibrium interest rate determined by output growth. Given $w(t)=v(t)/(k(t)-1)$ and $F(z,t)$ in (23), I show that $S(z,t)\le 0$ for $z<z^*$ and $S(z,t)=0$ for $z\ge z^*$ to justify the search intensity $\eta(z,t)$. Note that

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

it then suffices to compare the value of the integral with $\frac{k(t)+\tau(z)}{k(t)-1}$. For $z<z^*$,

$$\begin{aligned}
\int_z^\infty \frac{x}{z}\mathrm{d}F(x|x\ge z,t) &= \frac{1}{z[1-F(z,t)]}\int_z^\infty x\mathrm{d}F(x,t),\\
&=\frac{1}{z^{1-k_0}}\left\{\int_z^{z^*} k_0x^{-k_0}\mathrm{d}x+(z^*)^{k(t)-k_0}\int_{z^*}^\infty k(t)x^{-k(t)}\mathrm{d}x\right\},\\
&=\frac{1}{z^{1-k_0}}\left\{\frac{k_0}{k_0-1}\left[z^{1-k_0}-(z^*)^{1-k_0}\right]+(z^*)^{k(t)-k_0}\frac{k(t)}{k(t)-1}(z^*)^{1-k(t)}\right\},\\
&=\frac{k_0}{k_0-1}\left[1-\left(\frac{z^*}{z}\right)^{1-k_0}\right]+\frac{k(t)}{k(t)-1}\left(\frac{z^*}{z}\right)^{1-k_0},\\
&\le \frac{k(t)}{k(t)-1}\le\frac{k(t)+\tau}{k(t)-1}.
\end{aligned}$$

where the inequality follows that $(z^*/z)^{1-k_0}<1$, $k(t)\le k_0$, and $\tau\ge 0$. The equality holds only if $t=0$ and $\tau=0$, i.e., at the initial date in a tax-free economy. When $\tau>0$, firms below z^* never find it optimal to search at or after $t=0$, whereas without tax, they can search at $t=0$, leading to a path in which they can always search as in the baseline equilibrium. Hence, taxation operates off the equilibrium to discourage low-productivity firms to search.

For $z \geq z^*$, the truncated distribution $F(x|x \geq z, t)$ is exactly Pareto with tail index $k(t)$ and has mean $zk(t)/(k(t)-1)$. Then, $S(z,t) = 0$ with $\tau(z) = 0$. The optimality of this search strategy is thus confirmed.

I conclude the proof by discussing the special case $\tau = 0$. Since the above proof holds for $\tau = 0$, the tax-free economy in Section 4 admits a set of equilibria supported by the above search strategy for all thresholds $z^* \in [1, +\infty)$. This construction further illustrates the equilibrium multiplicity arising from the linear search cost. ■

C.2 Solving the planner's problem

To solve the planner's problem (24), I work with $w(f,z)$, which is the Gateaux derivative of $W(f)$ at point z with a Dirac delta function as increment.[12] $w(f,z)$ 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 derivations into Section C.2.2:

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

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.15) is the counterpart of HJB equation (8) 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.15), capturing the externality of other firms' learning. The conditional source distribution makes the externality term very complicated since any firm can affect others in various ways. To see this, consider an increase in the density $f(z)$ and a firm y with $y < z$. The density of the source distribution facing by firm y is $\frac{f(x)}{1-F(y)}$ for $x \geq y$. An increase in the portion of firm z increases the likelihood for firm y to meet a firm z. Yet it also decreases the likelihood of meeting other firms x with $x \neq z$, which could also benefit firm y.

[12] 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.

In the following, I solve the policy function associated with HJB equation (A.15) to obtain Proposition 3.

C.2.1 Proof of Proposition 3

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

$$\begin{aligned}
\rho w_z(z,t) =& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} + \eta^*(z,t)\frac{\partial}{\partial z}\left\{\int_z^\infty [w(y,t)-w(z,t)]\frac{f(y,t)}{1-F(z,t)}\mathrm{d}y - \hat{\mu}z\right\} \\
&+ w(z,t)f(z,t)\int_0^z \eta^*(x,t)\frac{-\varphi(x,t)}{1-F(x,t)}\mathrm{d}x - w(z,t)f(z,t)\int_0^z \eta^*(x,t)\frac{-\varphi(x,t)}{1-F(x,t)}\mathrm{d}x \\
&+ \int_z^\infty w(y,t)f(y,t)\mathrm{d}y\eta^*(z,t)\frac{-\varphi(z,t)}{1-F(z,t)} + w_z(z,t)\int_0^z \eta^*(x,t)\varphi(x,t)\mathrm{d}x + w(z,t)\eta^*(z,t)\varphi(z,t) \\
=& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} - \eta^*(z,t)\varphi(z,t)\int_z^\infty (w(y,t)-w(z,t))\frac{f(y,t)}{1-F(z,t)}\mathrm{d}y + w_z(z,t)\int_0^z \eta^*(x,t)\varphi(x,t)\mathrm{d}x \\
=& \frac{\partial w_z(z,t)}{\partial t} + \hat{\lambda} - \hat{\mu}\eta^*(z,t)\varphi(z,t)z + w_z(z,t)\int_0^z \eta^*(x,t)\varphi(x,t)\mathrm{d}x
\end{aligned} \tag{A.16}$$

The first equality is the result of an envelope theorem, and the second and the third use the first order condition (A.19), 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.19):

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

in which the last equality comes from the law of motion on $\varphi(z,t)$. Formally, the original law of motion (12) 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.16), 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.,

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

in which $\mathcal{L}(z,t)$ is a slowly 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.15) on $w(z,t)$

Following Lucas and Moll (2014), the corresponding HJB equation of problem (24) is

$$\begin{aligned}\rho W(f) = \max_{\{c(\omega),\eta(y)\}} Lu(c) + \int_0^\infty \frac{\delta W(f)}{\delta f(y)} f(y)\left[\int_0^y \eta(x)\varphi(x)\mathrm{d}x - \eta(y)\right]\mathrm{d}y \\ \text{s.t.} \quad Lc \leq \int_0^\infty y f(y)\mathrm{d}y, \quad \int_0^\infty y\eta(y)f(y)\mathrm{d}y \leq L,\end{aligned} \tag{A.17}$$

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

$$u'(c) = \hat{\lambda}. \tag{A.18}$$

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

$$\begin{aligned}&\int_z^\infty w(f,y)f(y)\varphi(z)\mathrm{d}y - w(f,z)f(z) - \hat{\mu}zf(z) = 0 \\ \Longrightarrow &\int_z^\infty [w(f,y) - w(f,z)]\frac{f(y)}{1-F(z)}\mathrm{d}y = \hat{\mu}z\end{aligned} \tag{A.19}$$

Differentiating both sides of the HJB equation (A.17) with respect to $f(z)$,

$$\begin{aligned}\rho w(f,z) = &\int_0^\infty \frac{\delta w(f,z)}{\delta f(y)} f(y) \left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right] \mathrm{d}y \\ &+ \int_0^\infty w(f,y) \frac{\delta}{\delta f(z)} \left\{ f(y) \left[\int_0^y \eta^*(x)\varphi(x)\mathrm{d}x - \eta^*(y)\right]\right\} \mathrm{d}y + \hat{\lambda} z - \hat{\mu} z \eta^*(z).\end{aligned} \tag{A.20}$$

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

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

Therefore,

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

Then,

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

Finally, rewriting (A.20) gives equation (A.15):

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

in which the max operator comes from first order condition (A.19).

C.2.3 The tail dynamics with optimal search policy

Figure A.9 illustrates the jumps in tail index with optimal search policy.

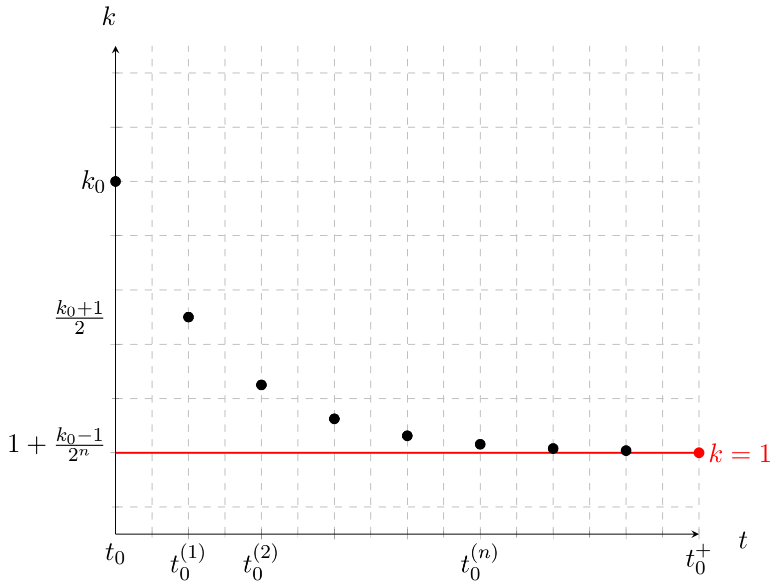


Figure A.9: Illustration of jumps in tail indices

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.,

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

It worths noting that Proposition 4 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 4 with $g_E(t)=0$ and $\eta(t)=\bar{\eta}$. Namely,

$$F(z,t) = \begin{cases} 1-(1-F^*(t))\left(\frac{z}{z^*(t)}\right)^{-k_t}, & \text{if } z \geq z^*(t), \\ F^*(T^*(z)), & \text{if } z < z^*(t). \end{cases} \tag{A.21}$$

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

$$(F^*)'(t) = k(t)g_z(t)\left(1-F^*(t)\right), \quad \text{and} \quad \frac{\dot{k}(t)}{k(t)} = -\bar{\eta}. \tag{A.22}$$

Next, I pin down the path of $z^*(t)$. Substituting (A.21) into the labor market clearing condition,

$$[1 - F^*(t)]\, z^*(t) \frac{k(t)}{k(t)-1} \bar{\eta} = L. \tag{A.23}$$

Log differentiating the above equation, I obtain

$$g_z(t) = \frac{(F^*)'(t)}{1 - F^*(t)} + \frac{\dot{k}(t)}{k(t)(k(t)-1)} = g_z(t)k(t) - \frac{\bar{\eta}}{k(t)-1} \Rightarrow g_z(t) = \frac{\bar{\eta}}{(k(t)-1)^2}, \tag{A.24}$$

where the second equality uses Equation (A.22). 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,

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

When $t \to T$, $k(t) \to 1$, and (A.23) 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

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

where the first equality follows (A.21) and the second the substitution $z = z^*(s)$. Using (A.23) and (A.24),

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

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

C.4 Generalized learning function

In this section, I extend the learning process in the baseline model in Section 4 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,

$$m(z,x,t) = \int_z^\infty \left(\frac{1-F(y,t)}{1-F(x,t)}\right)^\lambda \frac{f(y,t)}{1-F(x,t)}\mathrm{d}y = \frac{(1-F(z,t))^{\lambda+1}}{\lambda+1}\frac{1}{(1-F(x,t))^{1+\lambda}}. \tag{A.25}$$

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 (12) 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:

$$\frac{\partial \ln \tilde{F}(z,t)}{\partial t} = \tilde{F}(z,t)^\lambda \int_0^z \eta(x,t)\frac{\tilde{F}(x,t)^{-(1+\lambda)}}{\lambda+1}f(x,t)\mathrm{d}x. \tag{A.26}$$

C.4.1 Bounded search intensity

Suppose that the initial distribution $F(z,0)$ has tail index k_0, then $\lim_{z\to\infty}\frac{\ln \tilde{F}(z,0)}{\ln z} = -k_0$.[13] 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.26) 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.26) 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,

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

[13] 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.

In other words, the tail index remains the same under bounded search intensity.[14]

If $k_0 > 1$, it is straightforward to show that the total output, $Y(t) = \int_0^\infty z f(z) \mathrm{d}z$, 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.

C.4.2 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 the equilibrium of the baseline model.[15] In particular, all firms adopt the same search intensity $\eta(t)$, which clears the labor market. Substituting $\eta(x,t) = \eta(t)$ into (A.26), I obtain that

$$\frac{\partial \tilde{F}(z,t)}{\partial t} = \frac{\eta(t)}{\lambda+1} \tilde{F}(z,t) \left[\frac{1 - \tilde{F}(z,t)^{\lambda}}{\lambda} \right]. \tag{A.27}$$

Fixing z, the solution to this differential equation is a generalized logistic function.[16] For $\lambda > 0$, the above differential equation admits the following solution given initial distribution $F(z,0)$ and a path of $\eta(t)$:

$$1 - F(z,t) = \left\{ 1 + \left[(1 - F(z,0))^{-\lambda} - 1 \right] e^{-\frac{1}{\lambda+1} \int_0^t \eta(s) \mathrm{d}s} \right\}^{-\frac{1}{\lambda}}. \tag{A.28}$$

Next, I derive the full transition dynamics. Let $Z(t)$ be the aggregate productivity, i.e., $Z(t) = \int_0^\infty z f(z,t) \mathrm{d}z$. Differentiating it with respect to t and using (A.27),

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

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

$$\frac{\dot{\eta}(t)}{\eta(t)} = -\frac{\eta(t)}{(\lambda+1)\lambda} \left[1 - \frac{\int_0^\infty (1 - F(z,t))^{\lambda+1} \, \mathrm{d}z}{\int_0^\infty (1 - F(z,t)) \, \mathrm{d}z} \right]. \tag{A.29}$$

[14] 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.

[15] 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 B.4. 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.

[16] 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).

Given the expression of $1 - F(z,t)$ in (A.28) and an initial distribution $F(z,0)$, (A.29) 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.

C.4.3 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 corresponding 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:

$$k^m = -\frac{\ln 0.5/10}{\ln z_{0.5}/z_{10}} \tag{A.30}$$

For the numerical exercise, I adopt the initial Pareto assumption so that the equilibrium analysis in Section 4.2 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.8).[17] 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 (13) and (16), while it follows Equation (A.28) and (A.29) in the generalized model ($\lambda = 10^{-4}$).

I present the comparison in Figure A.10. 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%

[17] 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.11 would be nearly identical between the two models. These additional results are available upon request.

firms in the initial distribution. Notably, even after 400 years, the productivity distributions in the two models remain virtually indistinguishable.

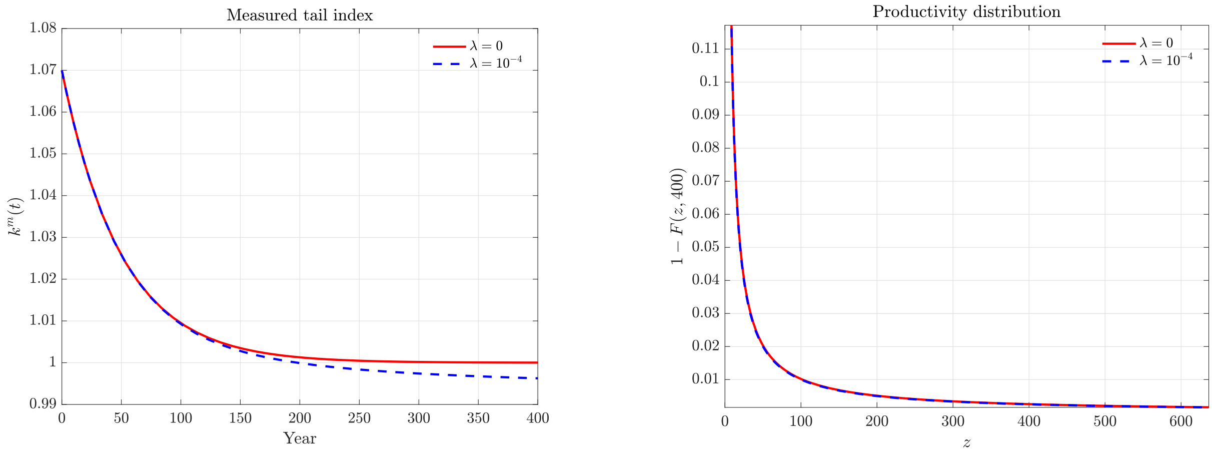


Figure A.10: Comparing productivity distributions

Notes. The left panel shows the respective time series of measured tail index k^m, as in (A.30), 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 $L = 0.02$.

Lastly, I compare the dynamics in log output between the two models in Figure A.11. 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.[18] 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.

C.4.4 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.26),

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

Meanwhile, the labor market clearing condition implies that

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

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

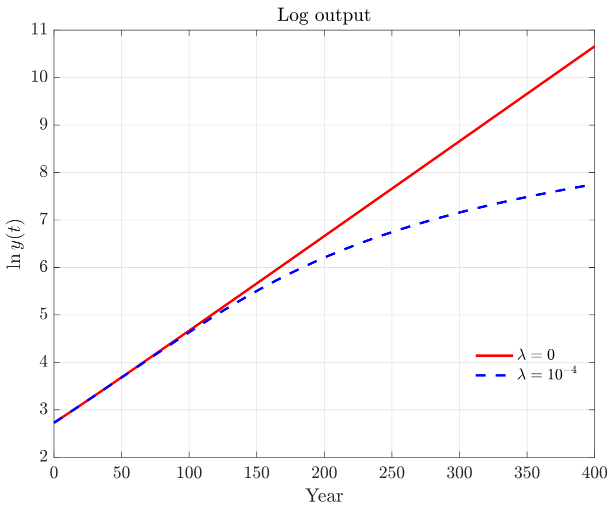


Figure A.11: Comparing growth

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 $L = 0.02$.

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 D Additional results for the extended model

D.1 Equilibrium definitions

Definition (Equilibrium)**.** A competitive equilibrium for the extended 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 thresholds $e^*(t)$, total output $Y(t)$, the measure of incumbent firms $M(t)$, and the productivity distribution $F(z,t)$ that satisfy the following:

(i) Given $\{w_p(t), w_r(t), r(t), Y(t), F(z,t)\}$, $v(z,t)$ solves the HJB equation (27) and (28) with search policy $\eta(z,t)$, and $l_p(z,t)$ is given by (26);

(ii) Given $\{F(z,t), w_r(t), v(z,t)\}$, $e^*(t)$ solves the entrepreneurial choice problem;

(iii) Given $\{y(t), r(t)\}$, $c(t)$ solves the households' utility maximization problem;

(iv) Goods and labor markets clear;

(v) $M(t)$ solves (33) given entrants $E(t)$ in (30) and $M(0)$. $F(z,t)$ solves the KFE (34) and (35) given $\eta(z,t)$ and $F(z,0)$.

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

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

D.2 Proof of Proposition 4

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 obsolescence threshold $z^*(t)$. For $z \geq z^*(t)$

$$1 - F(z,t) = \left(1 - Q\left(\frac{z}{z^*(t)}, t\right)\right)(1 - F^*(t)) \tag{A.31}$$

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

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

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

$$\frac{\mathrm{d}\ln\left(1-F^*(t)\right)}{\mathrm{d}t}=\frac{\partial\ln\left(1-Q\left(1,t\right)\right)}{\partial r}g_z(t)+\frac{\partial\ln\left(1-F(z^*(t),t)\right)}{\partial t} \tag{A.32}$$

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 (35). 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 distribution in the baseline economy is the relative productivity distribution among search active firms in the extended economy. The baseline model is a special case in which no technology becomes obsolete. Consequently, I have shown that $F(z,t)$ is given by (A.31) for $z\geq z^*(t)$.

With the linearity of $\pi(z,t)$ on z in HJB equations (27) and (28), same arguments in Section 4.2 lead to that $v(z,t)=v(t)z$. As $Q(r,t)$ is always Pareto, it is straightforward to verify that the net payoff per idea search $S(z,t)$ is zero if $w_r(t)=\frac{v(t)}{\psi(k(t)-1)}$. 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 the law of motion on $F^*(t)$ from (A.32) and (35):

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

Finally, the left tail is derived directly by integrating the KFE (34) for the search-inactive firms over time t from the stopping time $T^*(z)$ to the current time:

$$\ln F(z,t)-\ln F(z,T^*(z))=-\int_{T^*(z)}^t g_E(s)\mathrm{d}s.$$

I then obtain (36) using $F^*(T^*(z)) = F(z, T^*(z))$. The proof is then complete. ■

D.3 Proof of Proposition 5

Proof. As discussed in the main text, the law of motion for $M(t)$ follow directly from Equation (33) and the law of motion for $F^*(t)$ from Proposition 4. The expression for $g_E(t)$ follows the definition of entry rate and the measure of entrants in (30). It suffices to derive the law of motion for $k(t)$ and the equilibrium startup threshold $e^*(t)$.

First, I derive the ODE for $k(t)$. Given $F(z,t)$ in (36), the market clearing condition (32) for researchers implies that

$$\psi M(t)\eta(t)\left[1 - F^*(t)\right] \frac{k(t)}{k(t)-1} z^*(t) = sJ(e^*(t))L(t). \tag{A.33}$$

Combining it with $\dot{k}(t)/k(t) = -\eta(t)$, I obtain that the ODE (39) for $k(t)$:

$$-\frac{\dot{k}(t)}{k(t)-1} = \frac{sJ(e^*(t))}{1-F^*(t)} \frac{L(t)}{\psi M(t) z^*(t)} = \frac{J(e^*(t))}{K(e^*(t))} \frac{g_E(t)}{1-F^*(t)}.$$

I further define $g_{TG}(t) \equiv -\dot{k}(t)/(k(t)-1)$.

Next, I derive the equilibrium startup threshold $e^*(t)$. As shown in the previous proof of Proposition 4 in Section D.2, the equilibrium researchers' wage $w_r(t) = \frac{v(t)}{\psi(k(t)-1)}$. Since J has support $[0, +\infty)$, (29) implies that $e^*(t)$ is always interior so that $v_E(e^*(t), t) = (1-\tau_r)w_r(t)$. Substituting (29), (36), and the expression for $w_r(t)$ into this equation, I obtain the following expression for $e^*(t)$:

$$e^*(t) = \frac{1-\tau_r}{k(t)}. \tag{A.34}$$

The proof is then complete. ■

D.4 Proof of Proposition 6

Proof. This proof consists of three steps. First, I derive the limits k^*, e^*, f, g_M^*, and g_{TG}^*. 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_{TG}^* I start with g_M^*. Proposition 4 indicates that $k(t)$ is decreasing and bounded from zero, it then converges to a limit $k^* \geq 0$. Furthermore, ODE (39) ensures that $k(t) \geq 1$ since $k(0) > 1$, and the RHS of (39) is positive. Then, $k^* \geq 1$. Solving ODE (37) with $e^*(t) \to e^* = (1-\tau_r)/k^* \in ((1-\tau_r)/k_0, 1-\tau_r)$ and the assumptions on $g_L(t)$ and $g_z(t)$, 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 (38) given $k(t)$, $g_z(t)$, and $g_E(t)$ gives that $F^*(t)$ converges to the following stationary point f:

$$f \equiv \lim_{t\to\infty} F^*(t) = \frac{k^* g_z^*}{k^* g_z^* + g_E^*}, \tag{A.35}$$

where $g_E^* = g_M^* + \delta = g_L^* - g_z^* + \delta > 0$.

To obtain k^*, notice that the ODE (39) on $g_{TG}(t)$ 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,

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

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

$$\lim_{t\to\infty} Q^l(r,t) = r^\varsigma \quad \text{for } r \in (0,1), \tag{A.36}$$

$$\lim_{t\to\infty} Q^r(r,t) = 1 - r^{-1} \quad \text{for } r \geq 1, \tag{A.37}$$

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

Proof. It is evident that (A.37) follows directly from Proposition 4 and $k^* = 1$. Meanwhile, Proposition 4 also implies that[19]

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

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

[19]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 $t > T$.

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, $\frac{g_E(\tilde{t})}{g_z(\tilde{t})} \to \varsigma \equiv \frac{g_E^*}{g_z^*}$ with $g_E^* > 0$. 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 (41).

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

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

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 4 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 $t = \infty$.[20] 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 (36) with respect to t gives us

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

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:

$$1 = zg_z(T^*(z))(T^*)'(z). \tag{A.38}$$

Then,

$$(T^*)'(rz^*(t))r(z^*)'(t) = \frac{r(z^*)'(t)}{rz^*(t)g_z(T^*(rz^*(t)))} = \frac{g_z(t)}{g_z(T^*(rz^*(t)))} = \frac{g_z(T^*(z^*(t)))}{g_z(T^*(rz^*(t)))} \to 1,$$

where the second equality uses (A.38), the third the definition of $g_z(t)$, and the fourth $t = T^*(z^*(t))$. The ratio converges to 1 since $g_z(t) \to g_z^*$. The convergence of $g_z(t)$ also implies that $0 < \inf_t g_z(t) \leq \sup_t g_z(t) < \infty$.[21] Therefore, $\frac{g_z(T^*(z^*(t)))}{g_z(T^*(rz^*(t)))}$ is bounded for any r and t.

Let $A(t) = e^{-\int_{T^*(rz^*(t))}^{t} g_E(s)\mathrm{d}s} < 1$. I have shown in the previous step that $\lim_{t\to\infty} A(t) = r^{\varsigma}$. Also, let $B(t) = (T^*)'(rz^*(t))r(z^*)'(t)$, so $\lim_{t\to\infty} B(t) = 1$. Then,

$$\frac{\mathrm{d}F(rz^*(t), t)}{\mathrm{d}t} = (F^*)'\,(T^*(rz^*(t)))\,B(t)A(t) + F^*\,(T^*\,(rz^*(t)))\,A(t)\,[-g_E(t) + g_E\,(T^*(rz^*(t)))\,B(t)]\,.$$

Using that $\lim_{t\to\infty}(F^*)'(t) = 0$, $\lim_{t\to\infty}(F^*)(t) = f$, $\lim_{t\to\infty} g_E(t) = g_E^*$, $\frac{\mathrm{d}F(rz^*(t),t)}{\mathrm{d}t}$ is bounded

[20] To show that $Z(t)/z^*(t)$ has asymptotic growth rate $g_{TG}(t)$, 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.

[21] The lower bound holds since $g_z(t)$ is continous, and $g_z(t) > 0$, which follows from the assumed differentiability of $T^*(z)$.

and converges to

$$\lim_{t\to\infty} \frac{\mathrm{d}F(rz^*(t),t)}{\mathrm{d}t} = 0 \times 1 \times r^{\varsigma} + f \times r^{\varsigma} \times [-g_E^* + g_E^* \times 1] = 0.$$

The dominated convergence theorem then applies to show that $\int_{\frac{1}{z^*(t)}}^{1} \frac{\mathrm{d}F(rz^*(t),t)}{\mathrm{d}t}\mathrm{d}r \to 0$.

From here,

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

at $t = \infty$. Then,

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

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

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

The first line follows from firms' static choices in (26). Firms' pricing rule also implies that $\frac{\sigma-1}{\sigma}Y(t) = w_p(t)(1-s)L(t)$. Then, households' income or output per capita is

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

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

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

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

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

■

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.

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.

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.

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.

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

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

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.

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

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

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.