NTC
Joint Research

Discussion Paper Series

Rapid Growth and the Emergence of Pareto Tails

Visiting Professor, National Tax College, Japan
Fellow, Research institute of Economy, Trade and Industry
Yoshiyuki Arata

Emeritus Professor, University of Tokyo
Honorary President, Policy Research Institute, Ministry of Finance
Hiroshi Yoshikawa

Assistant Professor, Research Dept., National Tax College, Japan
Shingo Okamoto

250601–02ST

税務大学校
National Tax College
https://www.nta.go.jp

Rapid Growth and the Emergence of Pareto Tails

Yoshiyuki Arata*†
RIETI
National Tax College

Hiroshi Yoshikawa
University of Tokyo

Shingo Okamoto
National Tax College

Date: February 9, 2026

Abstract

Many models have been proposed to explain the Pareto-tailed behavior observed in the upper tails of firm-size and individual-income distributions. However, recent studies have pointed out that these models imply unrealistically long periods for firms or individuals to reach the upper tail, that is, to become very large firms or high-income earners. Moreover, while existing models typically predict that Pareto tails are primarily generated by older firms or individuals, empirical evidence shows that Pareto tails already emerge within the distributions of relatively young firms and individuals. This paper develops an alternative explanation for the emergence of Pareto tails that resolves these empirical inconsistencies. Focusing on the heavy-tailed nature of growth rate distributions, we show that firm growth and income growth are characterized by short episodes of exceptionally high growth. We demonstrate that such rapid-growth events give rise to a Pareto tail.

Keywords: Pareto tails; Weibull-tail distributions; Principle of a single big jump
JEL codes: D22; D31; L11

*Corresponding author: y.arata0325@gmail.com

†This study is part of the "Joint Statistical Research Program" conducted in collaboration with the National Tax College (NTC), under the approval of the National Tax Agency (NTA), in accordance with "Guideline on the Utilization of National Tax Data in the Joint Statistical Research Program." This study is also part of the project "Heterogeneity of Economic Agents and Challenges for the Japanese Economy" conducted at the Research Institute of Economy, Trade and Industry (RIETI). The views expressed herein are those of the authors and do not necessarily reflect the views of NTC, NTA, or RIETI. An earlier version of this paper circulated under the titles "Zipf's Law without the Stationarity Assumption" and "Explaining Zipf's Law by Rapid Growth." Arata gratefully acknowledges financial support from the Japan Society for the Promotion of Science (JSPS) (Grant Number: 21K01438). The authors would like to thank Alexis Akira Toda for his insightful and valuable comments. The authors are grateful to Kazuhiro Inaba and Takafumi Nakagawa for their excellent research assistance. The authors used OpenAI's ChatGPT to improve the clarity and fluency of the English in this manuscript. Conflict of interest: None.

1 Introduction

The Pareto-tailed nature of the size distribution of economic agents is one of the most important stylized facts in economics. Pareto tails have been documented across various fields, including firm size distributions and the distributions of income and wealth (see, e.g., Axtell (2001); Gabaix (2009); Luttmer (2010)). Many economists have studied the mechanisms that give rise to this statistical regularity, with early contributions by Champernowne (1953), Wold and Whittle (1957), Simon and Bonini (1958), and Ijiri and Simon (1977). More recently, renewed interest in this topic has led to the development of more refined theoretical models (e.g., Gabaix (1999); Reed (2001); Luttmer (2007); Luttmer (2011); Gabaix et al. (2016); Beare and Toda (2022)).

However, existing models exhibit inconsistencies with empirical data. A growing body of research has pointed out that these models imply excessively long periods for agents to reach the upper tail of the size distribution, where a Pareto tail is observed. For example, Luttmer (2011) shows that the time required for firms to grow into giant firms is far longer in the theoretical models than in the data. Similarly, Gabaix et al. (2016) demonstrate that the time for income distributions to converge to their stationary distributions is unrealistically long, preventing such models from accounting for observed fluctuations in income inequality. The root of this empirical inconsistency is that existing models implicitly assume that the upper tail is dominated by older agents. However, as shown in Section 2.2, our empirical evidence indicates that in both firm-size and individual-income distributions, Pareto tails are in fact generated by relatively young agents. These observations therefore call for a new theory capable of explaining why relatively young agents play the central role in generating Pareto tails.

This paper proposes a new explanation for the emergence of Pareto tails that resolves the inconsistencies with empirical data. The core of our analysis lies in examining the most likely patterns through which upper-tail agents, such as giant firms and high-income earners, emerge. Existing models assume that upper-tail agents arise as the cumulative outcome of gradual growth over long periods. In contrast, our theory shows that the emergence of upper-tail agents is driven by rapid, short-term growth—i.e., through jumps. Such jumps enable agents to reach the tail region of the distribution within a short time, thereby resolving the inconsistencies in existing models.

This paper builds on a series of recent empirical findings regarding the shape of growth rate distributions. For firm sales, it has been well documented since Stanley et al. (1996) that growth rate distributions deviate from a Gaussian distribution (for surveys, see Coad (2009); Dosi et al. (2017)). Specifically, these distributions exhibit high kurtosis and heavier tails, yielding distributions closer

to a Laplace distribution than to a Gaussian distribution (e.g., Bottazzi and Secchi (2006); Arata (2019)).[1] Moreover, subsequent studies show that growth rate distributions are even heavier-tailed than Laplace distributions, whose tails decay exponentially (Bottazzi et al. (2011); Dosi et al. (2020)). These features are not unique to firm sales growth rates but also characterize individual income growth rates. A pioneering study by Guvenen et al. (2021) shows that the distribution of income growth rates in the United State deviates from a Gaussian distribution, exhibiting high kurtosis and heavy tails. Similar distributional shapes have also been observed across many countries (see Guvenen et al. (2022)).[2]

In this paper, we show that the mechanism by which upper-tail agents emerge differs qualitatively depending on whether the growth rate distribution is heavy-tailed or light-tailed. We assume that (log) growth rates in each period are independent and identically distributed (i.i.d.), and that cumulative growth over n periods is given by the sum of n i.i.d. random variables. When the growth rate distribution is light-tailed, large deviations of the sum (i.e., high growth over n periods) arise from approximately equal contributions of all n variables. In this case, reaching the upper tail requires the accumulation of a large number of moderate growth rates and therefore takes a long time. This is precisely the growth pattern assumed in existing models. By contrast, when the growth rate distribution is heavy-tailed, a large deviation of the sum is typically generated by a single term that takes an extremely large value (i.e., a jump). That is, a single period of rapid growth can propel an agent into the upper tail immediately. A heavy-tailed growth rate distribution reflects the presence of such jumps and does not necessarily require a long time for an agent to reach the upper tail. This is the growth pattern underlying our theory.

Another feature of our theory is that it does not require the size distribution to be in a stationary state. Rather than analyzing the limit $n \to \infty$, we focus on how the distribution of the sum of n i.i.d. random variables evolves as n increases. In particular, when n is not sufficiently large, the Gaussian approximation based on the central limit theorem, which is often employed in existing models, does not provide an accurate approximation in the tail region. Instead, we show that the shape of the distribution in the non-Gaussian tail region is the key to explaining Pareto tails.

The study most closely related to our analysis is Beare and Toda (2022). Several previous

[1] In Arata (2019), the term *jump* was used to describe the discontinuous sample paths of continuous-time stochastic processes. In this paper, *jump* refers to situations in which cumulative growth over n periods is dominated by a single period of exceptionally large growth. Throughout the paper, the terms *rapid growth* and *jump* are used interchangeably.

[2] As another example of heavy-tailed behavior, Toda and Walsh (2015) show that the distribution of individual *consumption* growth rates is heavy-tailed to the extent that higher-order moments do not exist.

studies have sought to address the problem that the time required to become a giant firm or a super-rich individual is excessively long (e.g., Luttmer (2011); Gabaix et al. (2016)). To reduce the time needed to reach the upper tail of the size distribution, these studies introduce multiple types of agents and assume that some types have, on average, higher growth rates than others, so that agents of those types can reach the upper tail more quickly. Beare and Toda (2022) develop this approach in a more general framework and represent the state of the art in the literature.

Compared with such multi-type models, our theory has several advantages. First, multi-type models require the existence of agents whose average growth rates are persistently higher than those of others over long periods. However, empirical studies on high-growth firms (HGFs) show that high growth is not a persistent, firm-specific characteristic but is instead driven by brief episodes of exceptionally high growth; outside these episodes, the growth process of HGFs is indistinguishable from that of other firms, a growth pattern predicted by our theory (see Coad et al. (2014); Goswami et al. (2019)).[3] Second, multi-type models explain Pareto tails in the aggregate size distribution but do not account for Pareto tails within age-specific distributions. By contrast, our theory explains Pareto tails both in the aggregate and within each age group, including among young agents. Third, as in existing models, multi-type models assume light-tailed growth rate distributions for each type. In contrast, our empirical evidence strongly suggests that growth rate distributions are in fact heavy-tailed. Our theory shows that introducing multiple types is not necessary to overcome the unrealistically long time required for agents to reach the upper tail of the size distribution. With heavy-tailed growth rates, a single-type framework suffices to generate upper-tail agents within a realistically short time.

The main contribution of this paper is to demonstrate the empirical validity of our theory of Pareto tails. For firm sales, we use the Tokyo Shoko Research (TSR) dataset, which covers more than one million Japanese firms. For individual income, we use Japanese tax return data provided by the National Tax College, which covers tax records for more than 20 million individuals annually.[4] Using these two datasets, we test the two key assumptions underlying our theory. The first assumption is the random walk assumption, namely agents' (log) growth rates are i.i.d. Our results show that while growth rates in consecutive periods are not perfectly independent, dependence in

[3] In this regard, Commonwealth of Australia (2017) succinctly states that "[i]n general, HGFs do not appear to be a type of firm, but rather a phase that some firms go through during their life cycle" (vi).

[4] In our analysis, the term *gross income* refers to income derived from labor earnings and business activities before deducting expenses or tax deductions. Further details are provided in Section 4.1. Robustness checks using alternative definitions of income—namely, net income, defined as income after subtracting expenses and deductions, as well as earnings from wages and salaries only—are reported in the Supplemental Appendix.

the tail region is weak, making the independence assumption empirically reasonable. The second assumption is that growth rates have tails heavier than an exponential tail. Our results indicate that growth rate distributions indeed satisfy this property and, in particular, resemble a Weibull tail. Taken together, these empirical results provide support for the two assumptions underlying our theory.

Finally, we demonstrate that the predictions of our theory are indeed observed in the data. First, we show that the tail exponent of the age-specific size distribution is determined by the tail exponents of the growth rate distribution and the initial size distribution. In particular, the difference in the tail exponents for firm sales and individual incomes is explained, in line with our theory, by corresponding differences in the tail exponents of their growth rate distributions and initial size distributions. Second, we show that Pareto tails do not arise from aggregating agents of different ages, but instead are already present within each age group, especially among younger agents—a key prediction of our theory. Moreover, in the firm sales data, size distributions for older age groups become increasingly close to a Gaussian distribution, which is consistent with our theory. Third, we examine how agents achieve high growth over n periods. Our theory predicts that a large deviation in cumulative growth over n periods is generated by a single jump, whereas existing models attribute it to systematically higher growth rates across all n periods.[5] The data are consistent with the former pattern and thus support our theory.

The remainder of this paper is organized as follows. Section 2 describes the properties of existing models and discusses their inconsistencies with the data. Section 3 presents our alternative theory of Pareto tails. Section 4 empirically tests the two key assumptions underlying our theory. Section 5 examines whether the predictions of our theory are consistent with the data. Section 6 concludes the paper.

2 Existing Models and Their Empirical Inconsistencies

This section explains why existing models require an unrealistically long time for agents to reach the upper tail of the size distribution. Section 2.1 focuses on Reed (2001)—a model with a single agent type and a specified growth rate distribution, which can be viewed as a simplified version of Beare and Toda (2022). This model predicts that the right tail of the size distribution is predominantly populated by older agents. In Section 2.2, using data on firm sales and individual

[5]In a supplementary paper to this study, Arata et al. (2023), we use an alternative dataset with data on firm mergers and show that jumps play an important role in the growth process even after excluding firms involved in mergers.

incomes, we examine size distributions by age and show that this prediction is inconsistent with the data.

2.1 Theoretical Predictions of Existing Models

We begin by introducing the notation. Throughout the paper, we refer to economic entities such as firms and individuals as *agents*. The logarithm of an agent's scale—such as sales or income—is referred to simply as its *size* and is denoted by S.[6] A Pareto tail means that the right tail of the size distribution follows a power law. Specifically, when the y-axis is plotted on a logarithmic scale, the right tail of the distribution can be represented as a straight line with slope $-a$:

$$\log \mathbb{P}(S > x) = -ax + b.$$

We refer to a as the *tail exponent*.

The model proposed by Reed (2001) consists of two components: the size distribution conditional on age and the age distribution of agents in the population. Let S_n denote the size of an agent of age n. Reed (2001) assumes that the conditional distribution $\mathbb{P}(S_n > x \mid \text{age} = n)$ is Gaussian with variance $n\sigma^2$. For simplicity, we take the mean to be zero. When x is sufficiently large, the tail probability of the Gaussian distribution can be approximated using Mills' ratio:[7]

$$\mathbb{P}(S_n > x \mid \text{age} = n) \approx \frac{\sigma\sqrt{n}}{x\sqrt{2\pi}} e^{-\frac{x^2}{2n\sigma^2}}.$$

Note that the size distribution within each age group does not have a Pareto tail. Given these age-specific size distributions, the overall size distribution $\mathbb{P}(S > x)$ is obtained by summing over all ages. Reed (2001) further assumes that the age distribution follows a geometric distribution. This is justified by the idea that an agent exits with a constant probability p each year, so the survival probability at age n, denoted by p_n, is geometrically distributed as $p_n = p(1-p)^{n-1}$. Accordingly,

$$\mathbb{P}(S > x) = \sum_n \mathbb{P}(S_n > x, \text{age} = n)$$

$$\mathbb{P}(S_n > x, \text{age} = n) = p_n \mathbb{P}(S_n > x \mid \text{age} = n) \approx p(1-p)^{n-1} \cdot \frac{\sigma\sqrt{n}}{x\sqrt{2\pi}} e^{-\frac{x^2}{2n\sigma^2}}$$

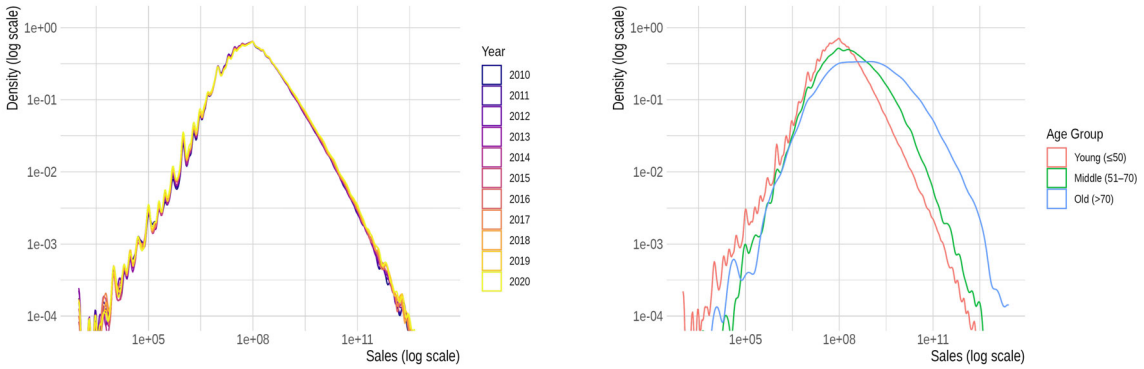
In this model, which age group of agents contributes most to the tail probability $\mathbb{P}(S > x)$? To examine this, consider the following ratio of tail probabilities for agents of two different ages:

$$\frac{\mathbb{P}(S_{n_2} > x, \text{age} = n_2)}{\mathbb{P}(S_{n_1} > x, \text{age} = n_1)} \approx (1-p)^{n_2 - n_1} \cdot \sqrt{\frac{n_2}{n_1}} \cdot e^{\frac{x^2(n_2 - n_1)}{2\sigma^2 n_1 n_2}}$$

where $n_2 > n_1$. The right-hand side is increasing in x (and, for fixed x and n_1, also increasing in

[6] Note that S refers to size without conditioning on the agent's age.

[7] For Mills' ratio, see Lemma 2 in Chapter 7 of Feller (1968).



(a) Aggregate distributions

(b) By age groups

Figure 1: Density estimates of firm size distributions. Panel (a) reports the density estimate of firm size S—defined as the logarithm of firm sales—using all firms from the 2010–2020 sample period. Panel (b) presents density estimates for three age groups (young, middle, and old) based on the 2020 cross-section. The number of observations in each group is 884,334 for young, 149,776 for middle, and 22,671 for old.

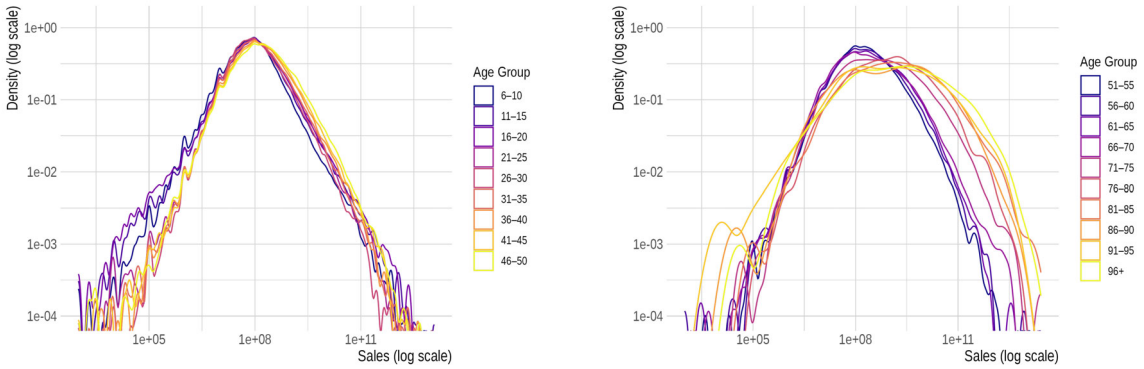
n_2). Thus, this model predicts that as x becomes large, the tail of the size distribution is increasingly dominated by older agents—that is, the Pareto tail is driven primarily by older agents.

This feature of existing models arises from two opposing effects as age n increases. On the one hand, if agents are born and exit at constant rates, the number of age-n agents—i.e., $\mathbb{P}(\text{age} = n)$—declines geometrically with n. On the other hand, as age increases, the variance of the size distribution within each age group becomes larger, leading to larger tail probabilities for S_n. In the tail region, this latter effect dominates. Consequently, the upper tail of the overall size distribution S is predicted to be populated mainly by older agents as x becomes large. This mechanism is common across many existing models: A Pareto tail emerges in the aggregate even though the size distributions for each age group do not themselves follow a Pareto tail. In the next section, we examine whether this theoretical prediction is supported by empirical data.

2.2 Empirical Evidence from Age-Specific Distributions

We begin by examining the distribution of firm size using Japanese firm-level data, where S is defined as the logarithm of firm sales.[8] To define firm age, we treat the date of incorporation as the birth year of each firm. **Figure 1**(a) presents a density estimate for firm size S on a logarithmic

[8]Further details on the data are provided in Section 4.1.



(a) Ages 6-50 **(b)** Ages above 50

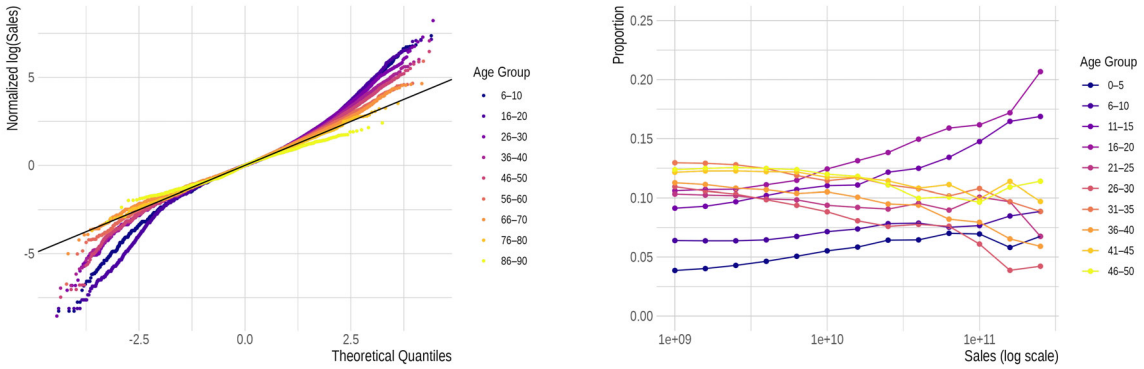
Figure 2: Density estimates of firm size distributions by age group. The sample is divided into five-year age groups, and the density of firm size is estimated for each group. Panel (a) presents the age groups from 6 to 50 years, while Panel (b) presents the age groups above 50 years.

scale, aggregated over all firms without conditioning on age.[9] As shown in the figure, the right tail of the distribution appears approximately linear. This indicates that the distribution of firm sales exhibits a Pareto tail, consistent with findings in the existing literature.

Which age group of firms contributes most to the formation of the Pareto tail observed in the aggregate size distribution? **Figure 1**(b) compares firms in three age groups: those younger than 50 years, those aged 50 to 70, and those older than 70. As shown in the figure, the size distribution of younger firms exhibits a Pareto tail, whereas that of older firms deviates from it. To examine the size distribution by age more closely, we divide the sample into five-year age bins and compare the distributions across groups. **Figure 2**(a) shows that younger firms exhibit a Pareto tail over a wide range, with each age group sharing a similar tail slope. In contrast, as **Figure 2**(b) illustrates, the size distributions of older firms deviate from the Pareto tail and instead resemble a bell-shaped curve, close to a Gaussian distribution.

To assess this pattern more rigorously, **Figure 3**(a) presents QQ-plots for each age group. If the size distribution were close to Gaussian, the points would lie along the reference line. As the figure shows, the tails of the distributions for young firms deviate from the line, whereas the distributions approach the line as firm age increases. These results indicate that, contrary to the predictions of existing models, the Pareto tail observed in the aggregate size distribution is in fact

[9] In our analysis, we use both the tail probability $\mathbb{P}(S > x)$ and the density $\mathbb{P}(S \in dx)$, depending on the context. Note that in the case of Pareto tails, both appear as straight lines on a log-scaled y-axis.



(a) QQ-plots of size distributions by age group

(b) Age-group proportions in the tail probability

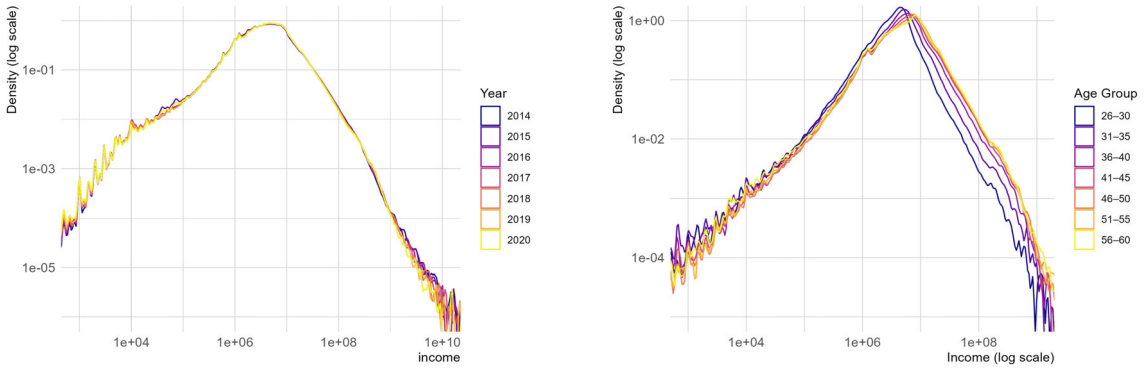
Figure 3: QQ-plots and age-group proportions in the tail of the firm sales distribution. Panel (a) presents QQ-plots of the size distributions for each age group. The straight line represents the reference line corresponding to a Gaussian distribution. For visual clarity, only a subset of the five-year age groups (e.g., 6-10, 16-20, and 26-30) is displayed. Panel (b) shows, as a function of x, the proportion of firms in each age group that contributes to the tail probability $\mathbb{P}(S > x)$. We restrict the sample to firms aged 50 or younger: within this subsample, we compute $\mathbb{P}(S > x)$ and report the share attributable to each age group.

shaped primarily by younger firms.

To clarify the inconsistency between the predictions of existing models and the data, we compute, for each age n, the proportion of the tail probability $\mathbb{P}(S > x)$ attributable to firms of that age, namely, $\mathbb{P}(S_n > x, \text{age} = n)/\mathbb{P}(S > x)$. **Figure 3**(b) reports these proportions across values of x. As the figure shows, the contribution of younger firms either remains stable or increases with x. This is in sharp contrast to the prediction of existing models, which imply that the tail should become increasingly dominated by older firms as x increases. These findings underscore the need for a new theory of Pareto tails—one that accounts for the fact that Pareto tails are generated primarily by younger firms.

A similar pattern is observed in the distribution of individual incomes. In the following analysis, we use individual-level income data and define the logarithm of income as size S. **Figure 4** presents both the aggregate distribution of size S and the distributions by five-year age groups. As in the case of firm sales, the aggregate size distribution exhibits a Pareto tail. Moreover, a Pareto tail is also observed within the age-specific distributions, each of which shares a similar slope in the tail region.

Figure 5(a) presents QQ-plots of the size distributions for each age group. While the tendency for the distribution to approach a Gaussian with age is less pronounced than in the firm-sales data,



(a) Aggregate distributions

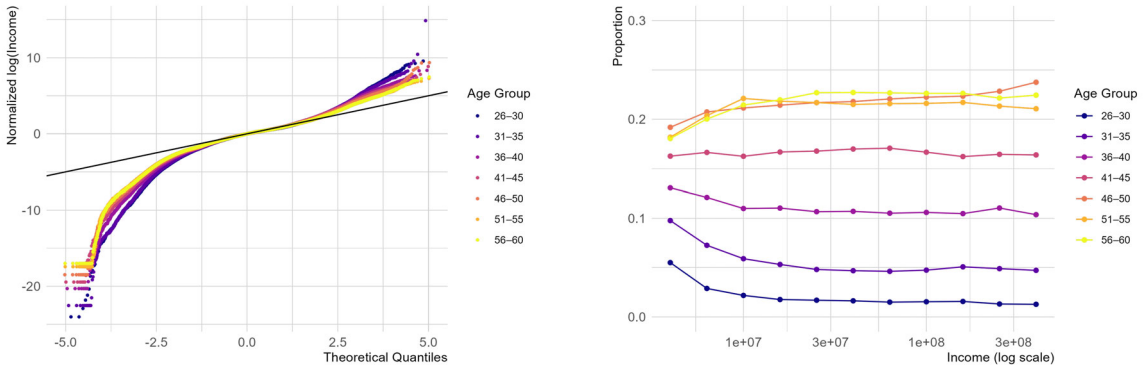
(b) By age groups

Figure 4: Density estimates of individual income distributions. Here, we define the logarithm of individual income as size S. Panel (a) presents the density estimate of size S for the years 2014-2020. Panel (b) presents density estimates of size S by five-year age groups using the 2020 cross-section. For clarity, Panel (b) displays only individuals aged 26 to 60.

the deviation from the Gaussian distribution is evident for the younger age groups. **Figure 5**(b) shows the proportion of each age group contributing to the tail probability $\mathbb{P}(S > x)$. These proportions remain stable as x increases, and there is no observable trend for older age groups to become more dominant in the tail. These findings indicate that the prediction of existing models—that the upper tail is dominated by older agents—does not hold. In particular, a Pareto tail does not arise as a consequence of aggregating across age groups; instead, it is already observed within each age group, especially among the younger ones. The next section develops an alternative theory of Pareto tails that is consistent with these observations.

3 An Alternative Theory of Pareto Tails

In this section, we develop an alternative theory of Pareto tails. Section 3.1 shows that the tail of the distribution of the sum of n i.i.d. random variables does not admit a Gaussian approximation. Section 3.2 examines how this distribution evolves as n increases. Section 3.3 investigates how the initial size distribution influences the subsequent size distribution. Section 3.4 analyzes how large deviations in the sum of n i.i.d. random variables arise.



(a) QQ-plots of size distributions by age group

(b) Age-group proportions in the tail probability

Figure 5: QQ-plots and age-group proportions in the tail of the individual income distribution. Panel (a) presents QQ-plots of the size distributions by age group. The straight line represents the reference line corresponding to a Gaussian distribution. Panel (b) shows, as a function of x, the proportion of individuals in each age group contributing to the tail probability $\mathbb{P}(S > x)$. For clarity, Panel (b) displays only individuals aged 26 to 60.

3.1 Setup

We begin by introducing the notation. Let S_0 denote the initial size (i.e., the logarithm of a firm's sales or an individual's income), and let X_k denote the growth rate in period k (i.e., the log-difference in size). Thus, cumulative growth over n periods is given by the sum of n growth rates,

$$\widetilde{S}_n := X_1 + \cdots + X_n,$$

and the size at time n is expressed as

$$S_n := S_0 + \widetilde{S}_n.$$

We make the following assumption regarding the process S_n:

Assumption 3.1. The size process S_n follows a random walk with an initial value S_0; that is, the growth rates $X_1, \ldots, X_n$ are independent and identically distributed with mean 0 and variance σ^2.

Under this assumption, $\widetilde{S}_n$ can be viewed as the sum of n i.i.d. random variables. A natural question is: what is the distribution of $\widetilde{S}_n$? According to the central limit theorem, the normalized sum of n i.i.d. random variables converges to the standard Gaussian distribution as $n \to \infty$ (see Chapter 5 of Petrov (1995)). More precisely, using the notation

$$Z_n = \sigma^{-1} n^{-1/2} \widetilde{S}_n, \quad F_n(x) = \mathbb{P}(Z_n < x)$$

the central limit theorem states that, for any fixed x,

$$\frac{1-F_n(x)}{1-\Phi(x)} \to 1, \ \frac{F_n(-x)}{\Phi(-x)} \to 1 \quad \text{as } n \to \infty \tag{1}$$

where Φ denotes the standard Gaussian distribution. Based on this result, one might conclude that, for sufficiently large n, the distribution of $\widetilde{S}_n$ can be well approximated by a Gaussian distribution over the entire range. Indeed, many existing models (e.g., Reed (2001)) implicitly assume that the size distribution of agents of the same age can be approximated by a Gaussian distribution. However, this reasoning is generally incorrect, because the convergence in the central limit theorem holds only for *fixed* x.

Now, what happens when x is allowed to depend on n? The most widely known result in probability theory concerning this question is due to Harald Cramér.

Theorem 3.1 (Theorem 5.23 in Petrov (1995))**.** *Suppose that Cramer's condition holds.*[10] *Then for* $x \geq 0, x = o(n^{1/2})$,

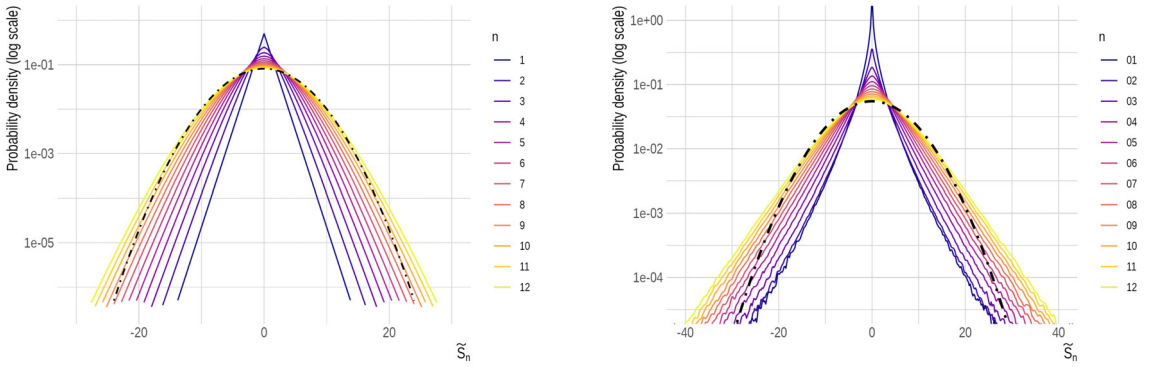
$$\begin{aligned} \frac{1-F_n(x)}{1-\Phi(x)} &= \exp\left\{\frac{x^3}{\sqrt{n}}\lambda\left(\frac{x}{\sqrt{n}}\right)\right\}\left[1+O\left(\frac{x+1}{\sqrt{n}}\right)\right], \\ \frac{F_n(-x)}{\Phi(-x)} &= \exp\left\{-\frac{x^3}{\sqrt{n}}\lambda\left(-\frac{x}{\sqrt{n}}\right)\right\}\left[1+O\left(\frac{x+1}{\sqrt{n}}\right)\right] \end{aligned} \tag{2}$$

where $\lambda(t) = \sum_{k=0}^{\infty} c_k t^k$ *is Cramér's series, a power series whose coefficients depend only on the cumulants of the random variable* X_1.

The right-hand side of Eq.(2) is known as the Cramér correction, describing the deviation from the Gaussian approximation. As the expression shows, by imposing the additional condition $x = o(n^{1/6})$—which restricts attention to an even narrower neighborhood around $x = 0$—one recovers Eq.(1). This means that the normal convergence guaranteed by the central limit theorem begins near $x = 0$ and expands gradually as n increases; however, outside the regions $x = o(n^{1/6})$ or $x = o(n^{1/2})$, normal convergence is not generally valid. This observation is crucial for our purposes, as our interest lies in the tail region, where x is large. In empirical settings, n is finite, and Pareto tails concern those large values of x where the Gaussian approximation may fail. As we show below, the core of our theory is an analysis of the tail behavior of $\widetilde{S}_n$ in those non-Gaussian regions.

Before turning to the general case, we consider two illustrative examples in which X_k follows a specific distribution and examine how the distribution of the sum $\widetilde{S}_n$ evolves as n increases. The

[10] See the next section for details on Cramér's condition. As discussed there, convergence to a Gaussian distribution near $x = 0$ holds even without Cramér's condition.



(a) Laplace distribution **(b)** Weibull-tail distribution

Figure 6: Density functions of $\widetilde{S}_n$ for $n = 1, \ldots, 12$. In both panels, the probability density functions of the sum $\widetilde{S}_n$ are shown for $n = 1, \ldots, 12$. When X_k follows a Laplace distribution, the density of $\widetilde{S}_n$ is given by (cf. Chapter 2.3.1 of Kotz et al. (2001)): $\mathbb{P}_{\widetilde{S}_n}(dx) = \frac{e^{-|x|}}{(n-1)!2^n} \sum_{j=0}^{n-1} \frac{(n-1+j)!}{(n-1-j)!j!} \frac{|x|^{n-1-j}}{2^j}$. Panel (a) plots these densities for $n = 1, \ldots, 12$. For comparison, a Gaussian density with the same standard deviation as $\widetilde{S}_{12}$ is also shown.

first case is when X_k follows a Laplace distribution with density

$$\mathbb{P}(dx) = \frac{1}{2} \exp(-|x|)dx.$$

The second case is when X_k has a Weibull tail given by

$$\mathbb{P}(X_k > x) = \frac{1}{2} \exp(-|x|^\alpha), \quad 0 < \alpha < 1.$$

In the former case, we can derive an explicit expression for the distribution of $\widetilde{S}_n$. In the latter case, although an explicit expression is not available, we generate samples by simulation and estimate the density function from these samples (with $\alpha = 0.7$).

Figure 6 displays, on a logarithmic scale, the density of $\widetilde{S}_n$ for both the Laplace and the Weibull-tail cases. In a neighborhood of $x = 0$, the central region of the density becomes flatter and approaches a bell shape as n increases. This behavior reflects the normal convergence discussed above: in the central region, the distribution approaches a Gaussian distribution as n grows. By contrast, in the region of large x, the deviation from the Gaussian distribution is pronounced. On the logarithmic scale, the tail of the distribution of $\widetilde{S}_n$ shifts upward in an approximately parallel fashion as n increases. This indicates that, even as n becomes large, the tail behavior of $\widetilde{S}_n$ is essentially determined by the tail probabilities of its components, $\mathbb{P}(X_k > x)$. A rigorous explanation of this phenomenon will be provided in the next section.

3.2 Three Zones of the Distribution of $\widetilde{S}_n$

This section discusses how the probability distribution of $\widetilde{S}_n$ is characterized as a function of n and x. As we show below, the behavior of $\widetilde{S}_n$ differs substantially depending on whether the tail of the distribution of its components X_k is heavier or lighter than an exponential distribution. More precisely, we call the distribution of X_k light-tailed if it satisfies the Cramér condition:[11]

$$\mathbb{E}e^{\lambda X_k} < \infty \qquad \text{for some } \lambda > 0$$

The Gaussian and Laplace distributions are examples of light-tailed distributions. If the distribution is not light-tailed (i.e., if $\mathbb{E}e^{\lambda X_k}$ is infinite for all $\lambda > 0$), it is referred to as heavy-tailed. Among heavy-tailed distributions, those satisfying certain minimal regularity conditions are known as subexponential distributions.[12] Well-known heavy-tailed distributions such as the Weibull and Pareto distributions belong to the subexponential class.

A key property of subexponential distributions is that when $X_1, \ldots, X_n$ are i.i.d., the tail probability of their sum satisfies

$$\mathbb{P}(\widetilde{S}_n > x) \sim n\mathbb{P}(X_k > x) \quad \text{as } x \to \infty$$

Note that the right-hand side, $n\mathbb{P}(X_k > x)$, is simply the probability that the maximum element $\max\{X_1, \ldots, X_n\}$ exceeds x.[13] Thus, as $x \to \infty$, the tail of the sum becomes asymptotically equivalent to the tail of the maximum. In other words, large deviations of the sum are generated by a single large component—a property known as the principle of a single big jump.

In our analysis, we assume that the growth rates $X_1, X_2, \ldots, X_n$ follow a common subex-

[11] Much of the existing literature assumes light-tailed growth rate distributions (see, e.g., Equation (2) in Beare and Toda (2022)). In contrast, we assume that growth rate distributions are heavy-tailed. We argue that the inconsistency between existing models and the data stems from the light-tailed assumption, and that incorporating heavy-tailedness resolves it.

[12] The regularity condition is as follows: Let F denote the distribution of the random variable $X_k^+ := \max\{0, X_k\}$ on the positive real half-line $\mathbb{R}^+$, and let $\overline{F}(x) := F[x, \infty)$. The distribution F is subexponential if the following limit exists: $\lim_{x\to\infty} \frac{\overline{F*F}(x)}{\overline{F}(x)}$, where $F * F(x)$ denotes the convolution of F. In empirical applications, commonly used heavy-tailed distributions—such as Weibull-type and Pareto-type tails—satisfy this condition. Hence, for practical purposes, subexponential and heavy-tailed distributions can be treated as essentially equivalent; see Chapter 3 of Foss et al. (2011).

[13] This can be shown as follows. Let F be the distribution function of X_k (i.e., $F(x) = \mathbb{P}(X_k \leq x)$). The tail probability of the maximum $\max\{X_1, \ldots, X_n\}$ can be expressed as:

$$\mathbb{P}(\max\{X_1, \ldots, X_n\} > x) = 1 - F^n(x) = (1 - F(x)) \sum_{k=0}^{n-1} F^k(x) \sim n(1 - F(x)), \quad \text{as } x \to \infty$$

ponential distribution. Subexponential distributions include several subclasses, among which two major ones are the Pareto-type and Weibull-type tails. We assume that the growth rate distribution belongs to the latter class (the empirical validity of this assumption is examined in Section 4.3).

Assumption 3.2. The distribution of X_k has a Weibull-type tail of the form:

$$\mathbb{P}(X_k > x) = e^{-\ell(x)}, \quad \ell(x) := x^{\alpha} L(x),\ 0 < \alpha < 1$$

where $L(x)$ is a slowly varying function at infinity.

A characteristic feature of a Weibull tail is that, for sufficiently large x, the curvature of the tail on the log scale (i.e., the second derivative of ℓ) becomes small. Consequently, over a wide range of the upper tail, the slope remains nearly constant, making the tail appear approximately linear on a log scale. We exploit this linear approximation of ℓ in Section 3.3, Section 5.1, and Section 5.2.

What shape does the distribution of the sum $\widetilde{S}_n$ take under the above assumptions? As noted in Section 3.1, by the central limit theorem the distribution of $\widetilde{S}_n$ is approximated by a Gaussian distribution in a neighborhood of $x = 0$, and the width of this region expands as n increases. In contrast, by the subexponential property, the tail of the distribution of $\widetilde{S}_n$ should be approximated by n times the tail probability of the growth rates X_k. Thus, the behavior of $\widetilde{S}_n$ depends on both n and x. A rigorous characterization is provided by the following result.

Theorem 3.2 (Theorem 5.4.1 in Borovkov and Borovkov (2008))**.** *For* $x \leq \sigma_1(n)$,

$$\mathbb{P}(\widetilde{S}_n \geqslant x) = \left[1 - \Phi\left(\frac{x}{\sqrt{n}}\right)\right] e^{-n\Lambda_{\kappa}^{0}(x/n)}(1 + o(1))$$

Here, $\Lambda_{\kappa}^{0}(x/n) := \Lambda_{\kappa}(x/n) - \frac{x^2}{2n^2}$*, where* $\Lambda_{\kappa}(x/n)$ *is the truncated Cramér series. For* $x \gg \sigma_1(n)$,

$$\mathbb{P}(\widetilde{S}_n \geqslant x) = ne^{-M(x,n)}(1 + \varepsilon(x, n))$$

In particular, for $x \gg \sigma_2(n)$,

$$\mathbb{P}(\widetilde{S}_n \geq x) = n\mathbb{P}(X_k > x)(1 + o(1))$$

Here, boundaries $\sigma_1(n)$ *and* $\sigma_2(n)$ *are given by* $\sigma_1(n) := n^{1/(2-\alpha)} L_1(n)$ *and* $\sigma_2(n) := n^{1/(2-2\alpha)} L_2(n)$ *with* L_1 *and* L_2 *being some slowly varying functions, respectively.*

According to this theorem, the distribution of $\widetilde{S}_n$ can be divided into three regions depending on x and n: (i) the Cramér approximation region, (ii) the intermediate deviation region, and (iii) the extreme deviation region. In region (i) (i.e., $x \leq \sigma_1(n)$), as in the discussion of Section 3.1, the Cramér approximation holds in a neighborhood of $x = 0$. In an even smaller neighborhood of $x = 0$, normal convergence applies, and the distribution can be approximated by a Gaussian distribution. By contrast, region (iii) corresponds to the domain of the principle of a single big jump, where the distribution is governed by the tail probability of the individual components X_k.

In the intermediate deviation region (ii), where $\sigma_1(n) \ll x \ll \sigma_2(n)$, the distribution of $\widetilde{S}_n$ exhibits characteristics that lie between those of the Cramér and extreme deviation regions, and in general it admits no simple closed-form representation. However, when we examine the distribution on a logarithmic scale (i.e., $\log \mathbb{P}(\widetilde{S}_n > x)$), we can use the relation $M = \ell(x)(1 + o(1))$ (cf. Equation (5.4.30) in Borovkov and Borovkov (2008)) to derive the following approximation: for $x \gg \sigma_1(n)$,

$$\log \mathbb{P}(\widetilde{S}_n > x) = (1 + o(1)) \log n\mathbb{P}(X_k > x)$$

Thus, if our object of interest is the shape of the distribution on a logarithmic scale—as in the analysis of Pareto tails—then for $x \gg \sigma_1(n)$, the tail behavior of $\widetilde{S}_n$ can be approximated by $\log n\mathbb{P}(X_k > x)$ in both the intermediate and extreme deviation regions. In the next section, we analyze the shape of the distribution of S_n, which incorporates the initial size S_0.

3.3 Initial Size and the Distribution of S_n

In this section, we examine the distribution of S_n when the distribution of the initial size S_0 is also taken into account, and analyze how its shape varies with x and n. We assume that S_0 follows a subexponential distribution (possibly different from that of X_k), in particular one with a Weibull-type tail:

$$\mathbb{P}(S_0 > x) = e^{-\ell_0(x)}, \quad \ell_0(x) := x^{\alpha_0} L_0(x),\ 0 < \alpha_0 < 1$$

where $L_0(x)$ is a slowly varying function at infinity.[14] An extension of Theorem 3.2 is presented below.

Theorem 3.3 (Theorem 11.3.1(iii) in Borovkov and Borovkov (2008))**.** *For $x \gg \sigma_1(n)$ and $x \gg \sigma_1^0(n)$,*

$$\mathbb{P}(S_n > x) \sim e^{-M_0(x,n)} + ne^{-M(x,n)}$$

In particular, for $x \gg \sigma_2(n)$ and $x \gg \sigma_2^0(n)$,

$$\mathbb{P}(S_n > x) \sim \mathbb{P}(S_0 > x) + n\mathbb{P}(X_k > x)$$

Here, $\sigma_1(n), \sigma_2(n)$, and $M(x,n)$ are the same as given in Theorem 3.2. The functions $\sigma_1^0(n)$ and $\sigma_2^0(n)$ take the form of $\sigma_1^0(n) = n^{1/(2-\alpha_0)} L_3(n)$ and $\sigma_2^0(n) = n^{1/(2-2\alpha_0)} L_4(n)$, where L_3 and L_4 are slowly varying functions. For $x \gg \sigma_1^0(n)$, $M_0(x,n)$ takes the form of $M_0 = \ell_0(x)(1 + o(1))$.

[14] The assumptions regarding S_0 are not as crucial as Assumption 3.1 or Assumption 3.2. If the tail of S_0 is much lighter than that of the growth rates X_k (or $\widetilde{S}_n$), its effect on S_n is negligible, and the distribution of S_n essentially coincides with that of $\widetilde{S}_n$. Conversely, if the tail of S_0 is much heavier, S_n is dominated by S_0, and its distribution coincides with that of S_0. Our interest therefore concerns the nontrivial case in which the tail behavior of S_0 is comparable to that of the growth rates. For this reason, we assume that S_0 has a tail of the form $e^{-\ell_0(x)}$.

Based on this theorem, we now examine the tail slope of the distribution of S_n on the logarithmic scale. In the extreme deviation region, the theorem shows that the tail behavior of S_n is determined by the tail probabilities of S_0 and X_k. Note that the tail slopes of S_0 and X_k are given by $-\ell_0'(x)$ and $-\ell'(x)$, respectively. Differentiating the logarithm of the right-hand side yields

$$\frac{d}{dx}\log \mathbb{P}(S_n > x) = -w_0\ell_0'(x) - w_n\ell'(x),$$
$$w_0 := \frac{\mathbb{P}(S_0 > x)}{\mathbb{P}(S_0 > x) + n\mathbb{P}(X_k > x)},\ w_n := \frac{n\mathbb{P}(X_k > x)}{\mathbb{P}(S_0 > x) + n\mathbb{P}(X_k > x)}$$

To study the tail exponent, consider sufficiently large x such that $\ell_0(x)$ and $\ell(x)$ can be approximated linearly as

$$\ell_0(x) \approx a_0 x + b_0, \quad \ell(x) \approx a_1 x + b_1$$

If the slopes coincide ($a_0 = a_1$), then the slope of $\log \mathbb{P}(S_n > x)$ also matches this common slope. As n increases, the slope remains unchanged, and only the intercept shifts upward. When the slopes differ ($a_0 \neq a_1$), the slope of $\log \mathbb{P}(S_n > x)$ becomes a weighted average of a_0 and a_1, with weights w_0 and w_n. In particular, because w_n increases with n, the slope gradually shifts from that of S_0 toward that of X_k.

A similar argument applies to the distribution of S_n in the intermediate deviation region. Using the fact that, in this region, the functions M_0 and M satisfy the approximations $\ell_0(x)(1+o(1))$ and $\ell(x)(1+o(1))$ (see Section 11.3 in Borovkov and Borovkov (2008)), the slope of $\log \mathbb{P}(S_n > x)$ is given by

$$\begin{aligned} &\frac{d}{dx}\log \mathbb{P}(S_n > x) = (1+o(1))(-w_0\ell_0'(x) - w_n\ell'(x)) \\ &w_0 := \frac{e^{-M_0(x,n)}}{e^{-M_0(x,n)} + ne^{-M(x,n)}}, w_n = \frac{ne^{-M(x,n)}}{e^{-M_0(x,n)} + ne^{-M(x,n)}} \end{aligned} \tag{3}$$

As in the extreme deviation region, the slope of $\log \mathbb{P}(S_n > x)$ is expressed as a weighted average of a_0 and a_1, with weights w_0 and w_n. Thus, the tail slope of S_n is directly determined by the tail slopes of S_0 and X_k.[15] In Section 5.1, we examine whether this relationship is borne out empirically.

[15]Here, we assume that the $o(1)$ terms in M_0 and M do not change rapidly, and that their derivatives with respect to x are sufficiently small. Alternatively, when evaluating the tail slope of the distribution using finite differences—rather than exact derivatives—these $o(1)$ terms become effectively smoothed out, so the assumption is justified when n is sufficiently large. In the empirical analysis below, we use Eq.(3) to examine the tail slope of the distribution.

3.4 Patterns Driving Large Deviations in $\widetilde{S}_n$

Here, we return to $\widetilde{S}_n$ and consider how large deviations (i.e., events in which $\widetilde{S}_n > x$ for large x) arise. As shown below, the pattern by which $\widetilde{S}_n > x$ occurs differs substantially depending on whether the growth rate distribution is light-tailed or heavy-tailed.

As an illustrative example, let X_k be a non-negative random variable with density function f (i.e., $F(dx) = f(x)dx$).[16] Suppose that $f(x)$ admits the representation

$$f(x) = e^{-h(x)}, \quad x \geq 0$$

For example, if $f(x)$ is the density of an exponential distribution, then $h(x) = x$. If n i.i.d. random variables follow the distribution F, the probability that their sum equals x is given by

$$\mathbb{P}(\widetilde{S}_n = x) = \int_{\widetilde{S}_n = x} \exp\left(-\sum_{k=1}^{n} h(X_k)\right) dX_1 \dots dX_n$$

Given that the sum is fixed at x, which configuration of $(X_1, \dots, X_n)$ is most likely to occur? This problem is equivalent to minimizing the quantity $\sum_{k=1}^{n} h(X_k)$ subject to the constraint $\widetilde{S}_n = x$.

Let us consider first the case in which h is a convex function (e.g., a Weibull distribution with $\alpha > 1$). Jensen's inequality implies that the sum $\sum_{k=1}^{n} h(X_k)$ attains its minimum at $X_1 = \cdots = X_n = x/n$.[17] In other words, the most likely configuration of $X_1, \dots, X_n$ that generates the sum $\widetilde{S}_n = x$ is the one in which all components take the common value of x/n. Thus, each component contributing equally to the sum is the most probable outcome. In contrast, when h is a concave function (e.g., a Weibull distribution with $\alpha < 1$), the way in which the components $X_1, \dots, X_n$ produce the sum $\widetilde{S}_n = x$ is qualitatively different. The quantity $\sum_{k=1}^{n} h(X_k)$ is minimized when $X_{k^*} = x$ for some $k = k^*$ and $X_k = 0$ for $k \neq k^*$.[18] This indicates that a single component dominates the sum, whereas the remaining components contribute almost nothing. The boundary case arises when h is linear, namely in the exponential distribution. Thus, large deviations of $\widetilde{S}_n$ occur in two distinct ways, depending on whether the distribution has a tail lighter or heavier than the exponential distribution.

---

[16] This example is taken from Chapter 3 of Sornette (2006).

[17] Indeed, let $\widehat{X}_k$ be the deviation from x/n, i.e., $\widehat{X}_k := X_k - x/n$. Jensen's inequality states that for a real convex function φ, $\varphi\left(\frac{\sum_k x_k}{n}\right) \leq \frac{\sum_k \varphi(x_k)}{n}$. Thus, $\sum_k h(X_k) = h\left(\frac{x}{n} + \widehat{X}_1\right) + \ldots + h\left(\frac{x}{n} + \widehat{X}_n\right) \geq nh\left(\frac{x}{n}\right)$, where we used $\sum_k \widehat{X}_k = 0$ by construction.

[18] This can be shown as follows: Suppose that the statement does not hold. Then, there exist at least two k such that $0 < X_k < x$. Take such two indices (denoted by k_1, k_2) with $X_{k_1} \geq X_{k_2}$. By the concavity of h, the value of $\sum_k h(X_k)$ can be reduced by replacing X_{k_1}, X_{k_2} with $X_{k_1} + \varepsilon, X_{k_2} - \varepsilon$ for a sufficiently small $\varepsilon > 0$. This yields a contradiction.

The fact that large deviations of the sum arise in two qualitatively different ways can be verified by examining the ratio of the contribution of each component to the sum.[19] Let $X_1, X_2 \geq 0$ be two independent random variables drawn from a Weibull distribution with parameter α. Consider the distribution of the ratio $X_1/(X_1 + X_2)$ conditional on the event that their sum equals x (i.e., $X_1 + X_2 = x$). The conditional probability density function of the ratio, denoted by $g_{\alpha,x}$, is given by

$$g_{\alpha,x}(r) = c(r(1-r))^{\alpha-1} e^{-x^\alpha (r^\alpha + (1-r)^\alpha)} \tag{4}$$

where c is a normalizing constant independent of r.[20]

Figure 7 displays the density $g_{\alpha,x}$ for three different values of α. The density is symmetric around $1/2$ for all cases, which follows from the fact that X_1 and X_2 are i.i.d. Let us examine the density more closely for each value of α. When $\alpha > 1$ (i.e., the light-tailed case), the density is unimodal with a peak at $1/2$. This indicates that the most probable event is that X_1 and X_2 take similar values (i.e., $X_1 = X_2 = x/2$). In particular, for large x, the density becomes increasingly concentrated around $1/2$. In contrast, when $\alpha < 1$ (i.e., the heavy-tailed case), the density peaks at 0 and 1, exhibiting a U-shaped pattern. This means that it is more probable that either X_1 or X_2—but not both—takes a large value and dominates the sum x. Moreover, as suggested by Eq.(4), the density concentrates at 0 and 1 as $x \to \infty$. For large x, it becomes highly unlikely that both X_1 and X_2 are large and contribute equally to the sum. Thus, how each component contributes to the sum is determined by whether the tail of the distribution is heavier or lighter than an exponential distribution.

The intuition provided above can be formalized more rigorously as follows. First, consider the case in which the distribution of the growth rate X_k is light-tailed. When the growth rate distribution is light-tailed, the moment generating function is well-defined (denoted by $\psi(\lambda) := \mathbb{E}e^{\lambda X_k}$). We then introduce the Cramér transform of the random variable X_k, denoted by X_k^β, which is defined as follows:

$$\mathbb{P}(X_k^\beta \in dx) = \frac{e^{\lambda(\beta)x}\mathbb{P}(X_k \in dx)}{\psi(\lambda(\beta))},$$

where $\lambda(\beta) := \arg\sup_\lambda(\beta\lambda - \log\psi(\lambda))$. When the sum $\widetilde{S}_n$ takes a large value, the conditional distribution of each growth rate X_k is given by the following.

[19] This example is taken from Chapter 1 of Foss et al. (2011).

[20] This can be proved as follows: Define the random variables ξ_1, ξ_2 by $\xi_1 := \frac{X_1}{X_1+X_2}, \xi_2 := X_1 + X_2$. Then, $\Pr(\xi_1 = r|\xi_2 = x) = \frac{\Pr(\xi_1=r,\xi_2=x)}{\Pr(\xi_2=x)} = \frac{\Pr(X_1=rx,X_2=x(1-r))}{\Pr(\xi_2=x)}$. The numerator is computed using the independence of X_1 and X_2. The denominator depends only on x and is independent of r, and can therefore be absorbed into the normalizing constant c.

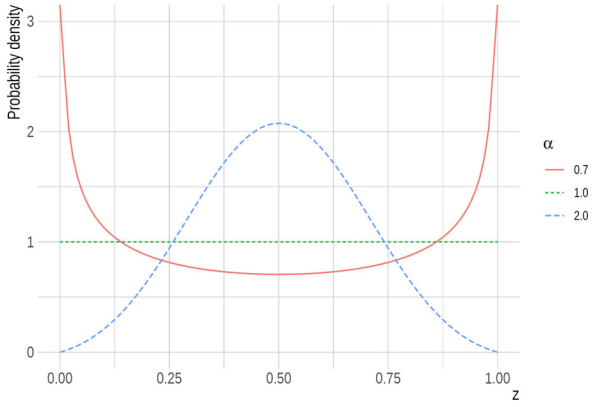


Figure 7: Plot of the density function $g_{\alpha,x}$. Three values of α are considered here: $\alpha = 0.7$, 1.0, and 2.0.

Theorem 3.4 (Corollary 3.1.2 in Borovkov (2020))**.** *Suppose that* $\beta = x/n \to \beta_0$ *as* $n \to \infty$. *Then, for any Borel sets* $B_1, \ldots, B_m$ *from* $\mathbb{R}$, *any* $k_1, \ldots, k_m$,

$$\prod_{i=1}^{m} \mathbb{P}\big(X_i^{\beta_0} \in B_i\big) = \lim_{n\to\infty} \mathbb{P}\big(X_{k_1} \in B_1, \ldots, X_{k_m} \in B_m \mid \widetilde{S}_n \in [x, x+\Delta)\big)$$

Note that Theorem 3.4 considers the case in which x grows on the order of n. When $\widetilde{S}_n$ attains a large value (of order n), the conditional distribution of the growth rates X_k is given by the distribution of their Cramér transform, X_k^{β}. As an example, let us consider the Gaussian distribution and its Cramér transform. If X_k follows a Gaussian distribution with mean μ and variance σ^2, its moment generating function is given by $\psi(\lambda) = e^{\mu\lambda+\sigma^2\lambda^2/2}$, and $\lambda(\beta) = \frac{\beta-\mu}{\sigma^2}$. Therefore, the Cramér transform of X_k is again Gaussian, with mean β and variance σ^2 (i.e., the Cramér transform simply shifts the distribution horizontally along the x-axis). For example, suppose that the unconditional distribution of the growth rate X_k has mean 0 (i.e., $\mu = 0$). Then, if we restrict attention to those samples that realize the large deviation $\widetilde{S}_n = x = n\beta$, the average growth rate among those samples equals $\beta = x/n$. In other words, the most typical way to achieve a large deviation of size $\widetilde{S}_n = x$ is through a path in which the growth rate increases gradually by x/n each period.

The behavior of the conditional distribution of the growth rates changes drastically when the growth rate distribution is subexponential. Rigorous results are provided by Armendáriz and Loulakis (2011).

Theorem 3.5 (Theorem 2 in Armendáriz and Loulakis (2011))**.** *Let* μ *be the probability measure of* X_k, *i.e.,* $\mu(A) := \mathbb{P}(X_k \in A)$. *Suppose that* μ *is subexponential. Then, the conditional probability*

$\mathbb{P}((X_1, \ldots, X_n) \in \cdot \mid \widetilde{S}_n > x)$ *converges in the total variance to a product of* $n-1$ *copies of* μ *and* ν_x*, where* $\nu_x(A) := \mathbb{P}(X_k \in A \mid X_k > x)$.

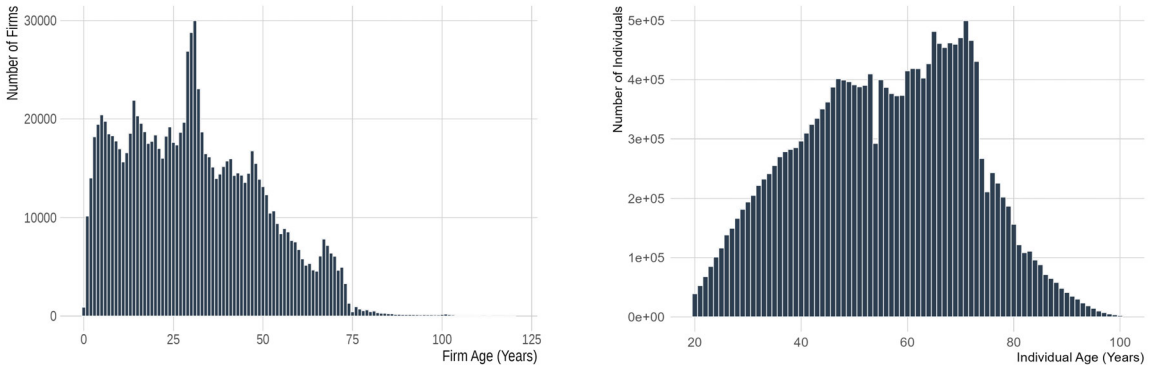
Theorem 3.5 states that when a large deviation event $\widetilde{S}_n > x$ occurs, the largest component (i.e., the jump) follows the distribution ν_x, while the remaining $n-1$ components follow the product measure μ_0^{n-1}. Note that μ_0 denotes the unconditional distribution of X_k. Thus, the large deviation of $\widetilde{S}_n$ is generated by the single largest growth rate, whereas the remaining growth rates are unaffected by the conditioning and retain their unconditional distribution. This property stands in sharp contrast to Theorem 3.4 and can be exploited for empirical validation. If the growth rate distribution is heavy-tailed, then the $n-1$ components of $X_1, \ldots, X_n$ excluding the jump should follow the same distribution as the unconditional growth rates. In contrast, if the distribution of X_k is light-tailed, the conditional distribution of X_k given $\widetilde{S}_n > x$ is instead given by the Cramér transform X_k^{β} (and, in the Gaussian case, corresponds to a shift by x/n). Therefore, by examining empirical data to determine which of these two patterns it more closely resembles, one can identify the typical mechanism that produces large deviations of $\widetilde{S}_n > x$.

4 Empirical Validation of the Two Assumptions

This section presents empirical evidence in support of our theory. Section 4.1 describes the data and reports summary statistics for the growth rates. Section 4.2 examines the empirical validity of the random walk assumption. Section 4.3 shows that the growth rate distribution is subexponential and exhibits a Weibull tail.

4.1 Data and Summary Statistics

We begin by defining the variables used in our empirical analysis. The initial size S_0 is defined as the logarithm of an agent's size measured at a fixed point after entry. For firm sales, S_0 is the logarithm of sales five years after incorporation; for individual income, it is the logarithm of income at age 25. We define S_0 in this way—rather than at entry—because growth patterns immediately after entry differ substantially from those in later periods. For firms, incorporation may occur as part of a corporate group reorganization, and the transfer of business units can generate apparent growth that is not comparable to subsequent growth. Measuring S_0 five years after incorporation mitigates such entry-related distortions. A similar issue arises for individual income: the sharp income increase associated with the transition from school to full-time employment is not comparable to



(a) Firm ages **(b)** Individual ages

Figure 8: Age distributions. Panel (a) shows the histogram of firm ages using the 2020 firm sales sample. Panel (b) shows the histogram of individual ages using the 2020 individual income sample.

later income growth. Accordingly, we define S_0 as income at age 25, where the random walk assumption provides a more reasonable approximation.

Firm sales

The firm-level sales data used in our analysis are compiled by Tokyo Shoko Research (TSR). TSR is a credit rating agency, and the dataset is based on surveys conducted for its corporate clients. It covers both listed and unlisted firms, with annual coverage exceeding one million firms. Nearly all large firms are known to be included in the dataset, and their information is updated at least once per year. Therefore, the upper tail of the firm size distribution is essentially fully covered by the dataset.

We impose several sample selection criteria when estimating the firm size distribution. Firm sales are measured using non-consolidated (stand-alone) financial statements, and firms with missing sales data are excluded. Because the TSR dataset contains firms that file financial statements more than once within a given year (i.e., firms whose fiscal period is shorter than 12 months), we restrict the sample to firms with a 12-month fiscal period.[21] We also exclude firms in the government and financial sectors. After applying these procedures, the sample size for 2020 is $1{,}169{,}440$ firms. The age distribution in 2020 is shown in **Figure 8**(a).

In our analysis, another key variable is the firm sales growth rate X_k. The sample used

[21] In our analysis, *year* refers to the fiscal year determined by the fiscal year-end date. For example, sales recorded for April 2010-March 2011 with a fiscal year-end of March 31, 2011 are treated as sales for year 2011.

Summary statistics of firm growth rates

Data: Tokyo Shoko Research from 2010 to 2020

Year	# of Samples	Mean	Std. Dev.	1st Quartile	Median	3rd Quartile
2010–11	711897	−0.022	0.288	−0.087	0.000	0.066
2011–12	712206	−0.008	0.282	−0.069	0.000	0.074
2012–13	718530	−0.010	0.272	−0.069	0.000	0.066
2013–14	717509	0.018	0.268	−0.043	0.000	0.096
2014–15	724215	−0.016	0.268	−0.075	0.000	0.061
2015–16	731788	−0.017	0.266	−0.073	0.000	0.059
2016–17	739739	−0.010	0.264	−0.061	0.000	0.060
2017–18	738764	−0.001	0.263	−0.049	0.000	0.070
2018–19	733870	−0.002	0.257	−0.050	0.000	0.064
2019–20	719949	−0.051	0.281	−0.118	−0.005	0.032

Table 1: Summary statistics for firm growth rates. The table reports summary statistics for one-year firm growth rates across different years.

for the analysis of growth rates is subject to the same selection criteria as the firm size sample described above, with two additional conditions. The first condition excludes extremely small firms. Specifically, we restrict the sample to firms whose size at the initial year used to compute the growth rate (e.g., sales in 2010 when calculating the growth rate from 2010 to 2015) is at least 40 million yen.[22] This condition is imposed because growth rates based on very small sales tend to be excessively volatile and deviate from the random walk assumption. The second condition concerns firm age. As explained in the definition of S_0, growth during the first five years after incorporation is included in S_0. Therefore, the analysis of firm growth rates X_k focuses on firms that are at least five years old.

Table 1 reports summary statistics for the one-year growth rates based on the sample that satisfies these conditions.[23] Note that the dispersion of growth rates shows little variation over the sample period.

[22] See the Supplemental Appendix for the choice of the 40 million yen threshold.

[23] In Assumption 3.1, the mean growth rate is set to zero for simplicity. Strictly speaking, it would be appropriate to estimate the mean growth rate for each year and conduct the analysis using demeaned growth rates. However, as shown in **Table 1** and **Table 2**, the sample mean and median of growth rates are small and close to zero. Moreover, the mean growth rate μ does not directly affect the analysis of tail shape or large deviations conducted below. Accordingly, we use the growth rates without demeaning in the empirical analysis.

Individual income

For the analysis of individual income, we use administrative tax return records provided by the National Tax College. In Japan, individuals file a tax return to report their annual income (from January 1 to December 31) to the tax authority, which is used to calculate their income tax and local residence taxes.[24] Although filing a tax return is not mandatory for all individuals, it becomes obligatory, for example, when annual salary income exceeds 20 million yen. Self-employed individuals (those with business income) are also required to file if their taxable income exceeds the basic exemption amount (480,000 yen as of 2020). In addition, individuals may voluntarily file a tax return to claim deductions, such as medical expense deductions, in order to reduce their tax liability. As a result, more than 20 million individuals file tax returns each year. Tax return data are particularly suitable for our analysis because they provide nearly complete coverage of high-income individuals in the upper tail of the income distribution. Moreover, the data include a unique identifier for each individual as well as age information, allowing us to construct a panel dataset.

In the analysis of individual income, we define income as the *gross income* (or gross receipts) derived from labor and business activities, which serves as the measure of an individual's economic scale. Temporary or non-recurring income sources (such as insurance payouts) and capital gains are excluded. Specifically, we construct gross income as the sum of business income, agricultural income, real estate income, salary income, and miscellaneous business income. We use gross income, rather than net income (which deducts expenses and tax deductions), to maintain conceptual consistency with our analysis of firm size. In the literature on Zipf's law and Gibrat's law for firms, firm size is typically measured by sales, which reflect the scale of economic activity. By contrast, profits—sales net of expenses—are not considered an appropriate measure of firm size, since even very large firms may report zero profits and thus not reflect their underlying economic scale. The same logic applies to individual income: gross receipts before deductions correspond to firm sales and reflect economic scale, whereas net income corresponds to firm profits. Accordingly, to ensure conceptual consistency between firm sales and individual income, we adopt gross income as the size variable for individuals.

Individuals with zero gross income are excluded from the sample. As a result, the sample size for the year 2020 is 19,680,907 individuals. The age distribution of the 2020 sample is presented in **Figure 8**(b). As the figure shows, the age distribution differs substantially between firm sales and

[24]For further details, see the National Tax Agency website: https://www.nta.go.jp/taxes/shiraberu/taxanswer/code/bunya-kakuteishinkoku.htm

Summary statistics of the growth rates of individual income

Data: Tax return data from National Tax College from 2014 to 2020

Year	# of Samples	Mean	Std. Dev.	1st Quartile	Median	3rd Quartile
2014–15	5055488	−0.020	0.330	−0.045	0.010	0.067
2015–16	5202343	−0.023	0.331	−0.049	0.009	0.065
2016–17	5365982	−0.020	0.323	−0.045	0.007	0.064
2017–18	5501765	−0.017	0.318	−0.042	0.010	0.066
2018–19	5567901	−0.017	0.321	−0.043	0.008	0.065
2019–20	5564641	−0.037	0.333	−0.074	0.000	0.059

Table 2: Summary statistics of individual income growth rates. The table reports summary statistics for one-year growth rates across different years.

individual income. The fact that Pareto tails are observed in both cases despite these differences suggests that, contrary to existing models, the age distribution is unlikely to be a driving force behind the emergence of Pareto tails.

The sample used for analyzing individual income growth rates is constructed by applying two additional conditions to the income sample described above. The first condition requires that an individual's income in the initial year used to compute the growth rate be at least 4 million yen. The second condition restricts the sample to individuals aged 25 to 60 in that year. **Table 2** presents summary statistics for the one-year growth rates for these samples. As in the case of firm sales, the dispersion of growth rates shows little variation over the sample period. Based on these samples, the common statistical properties of the growth rate distributions for firm sales and individual income are examined in the following section.

4.2 Random Walk Assumption

Methods

This section examines the independence assumption, that is, growth rates for a given firm are independent over time. Pearson's correlation coefficient is a widely used measure of dependence, but we do not employ it here because it is affected by marginal distributions.[25] Instead, we use Spearman's rank correlation coefficient, ρ_S, which depends solely on the dependence structure

[25] See the Supplemental Appendix for the limitations of Pearson's correlation coefficient and for a copula-based analysis of dependence.

Matrix of correlation coefficients Spearman's rank correlation coefficient										
Year	2019–20	2018–19	2017–18	2016–17	2015–16	2014–15	2013–14	2012–13	2011–12	2010–11
2019–20	NA	−0.056	0.052	0.054	0.061	0.044	0.050	0.046	0.032	0.021
2018–19	−0.056	NA	−0.073	0.059	0.055	0.055	0.059	0.046	0.049	0.043
2017–18	0.052	−0.073	NA	−0.078	0.046	0.060	0.071	0.038	0.042	0.058
2016–17	0.054	0.059	−0.078	NA	−0.090	0.046	0.063	0.050	0.039	0.040
2015–16	0.061	0.055	0.046	−0.090	NA	−0.076	0.046	0.046	0.043	0.036
2014–15	0.044	0.055	0.060	0.046	−0.076	NA	−0.080	0.046	0.050	0.049
2013–14	0.050	0.059	0.071	0.063	0.046	−0.080	NA	−0.066	0.060	0.066
2012–13	0.046	0.046	0.038	0.050	0.046	0.046	−0.066	NA	−0.076	0.035
2011–12	0.032	0.049	0.042	0.039	0.043	0.050	0.060	−0.076	NA	−0.075
2010–11	0.021	0.043	0.058	0.040	0.036	0.049	0.066	0.035	−0.075	NA

Table 3: Matrix of Spearman's rank correlation coefficients ρ_S for firm sales growth rates. The sample is the same as that used in **Table 1**.

between variables and is not affected by their marginal distributions. Spearman's ρ_S ranges from -1 to 1, and it equals 0 when the two variables are independent.

While Spearman's ρ_S provides important insights into dependence between growth rates, it is largely driven by regions where observations are dense, namely the central region of the distribution. Thus, it may fail to capture dependence in the tail region, where observations are sparse and extreme values occur. To address this concern, we employ the tail dependence measure. This measure represents the likelihood that one variable takes an extreme value given that the other also takes an extreme value. It is defined as:

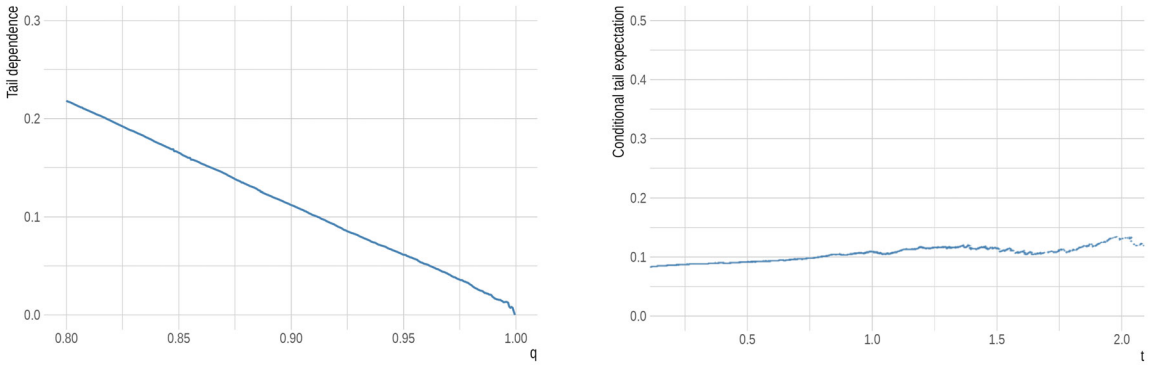
$$\lambda_U := \lim_{q \to 1} \mathbb{P}(X_2 > F_2^{-1}(q) \mid X_1 > F_1^{-1}(q))$$

Intuitively, λ_U can be interpreted as the probability that an extreme growth episode in the second period occurs given an extreme growth episode in the first period. When $\lambda_U > 0$ (i.e., the conditional probability above converges to a positive value), the two variables are said to exhibit tail dependence. When $\lambda_U = 0$, the variables are said to exhibit tail independence, meaning that large deviations do not occur consecutively.

As another measure of dependence in the tail region, we consider the conditional tail expectation:

$$\mathbb{E}[X_2 | X_1 > t]$$

We examine the behavior of this function as $t \to \infty$. When X_1 and X_2 represent growth rates



(a) Tail dependence measure

(b) Conditional tail expectation

Figure 9: Tail dependence measure and conditional tail expectation for firm sales growth rates. Let X_1 denote the growth rate for 2010-11 and X_2 the growth rate for 2011-12. Panel (a) plots the conditional probability $\mathbb{P}(X_2 > F_2^{-1}(q) \mid X_1 > F_1^{-1}(q))$ against q. Panel (b) plots the conditional tail expectation $\mathbb{E}[X_2 \mathbf{1}_{X_2>0} \mid X_1 \mathbf{1}_{X_1>0} > t]$ against t, using the nonnegative variable $X_k \mathbf{1}_{X_k>0}$.

in consecutive periods, it can be interpreted as the expected growth rate in the second period for agents that experienced exceptionally high growth in the first period. If the two random variables are independent, this expectation remains constant as a function of t. In contrast, if the variables exhibit tail dependence (i.e., $\lambda_U > 0$), then the conditional tail expectation grows linearly in t, that is, $\mathbb{E}[X_2|X_1 > t] \sim O(t)$ as $t \to \infty$ (see Section 2.20 in Joe (2014)).

Results

We apply the above methods to the growth rates of firm sales (the results for individual income are very similar and are briefly reported at the end of this section). The correlation matrix of growth rates across different years, computed using Spearman's ρ_S, is presented in **Table 3**. As expected, the coefficients approach zero as the time interval between the two growth rates increases. Moreover, even for consecutive periods, the absolute values of the coefficients remain below 0.1 and close to zero. In what follows, we restrict attention to consecutive periods and analyze the dependence structure in greater detail.

Figure 9(a) reports the tail dependence measure for the growth rates for 2010-11 and 2011-12 (more precisely, the conditional probability in the definition of λ_U) and shows how it behaves as $q \to 1$. As the figure illustrates, the tail dependence measure decreases and converges toward zero as $q \to 1$. This indicates that the dependence between growth rates in consecutive periods

Matrix of correlation coefficients Spearman's rank correlation coefficient						
Year	2019–20	2018–19	2017–18	2016–17	2015–16	2014–15
2019–20	NA	0.012	0.010	0.019	0.030	0.028
2018–19	0.012	NA	0.020	0.023	0.037	0.042
2017–18	0.010	0.020	NA	0.027	0.025	0.035
2016–17	0.019	0.023	0.027	NA	0.021	0.020
2015–16	0.030	0.037	0.025	0.021	NA	0.029
2014–15	0.028	0.042	0.035	0.020	0.029	NA

Table 4: Matrix of Spearman's rank correlation coefficients ρ_S for individual income growth rates. The sample is the same as that used in **Table 2**.

becomes increasingly weak in the upper tail, approaching tail independence as larger growth rates are considered.

The results for the conditional tail expectation are presented in **Figure 9**(b). Since growth rates can take both positive and negative values, we compute the conditional tail expectation for the non-negative variable $X_k \mathbf{1}_{X_k>0}$. As shown in the figure, the conditional tail expectation remains nearly constant in t. In other words, even if a firm experiences exceptionally high growth in one period, this does not raise the expected growth rate in the subsequent period. These findings are consistent with the result that $\lambda_U = 0$, indicating that dependence between consecutive-period growth rates in the upper tail region (i.e., cases where both periods exhibit high growth) is weak.

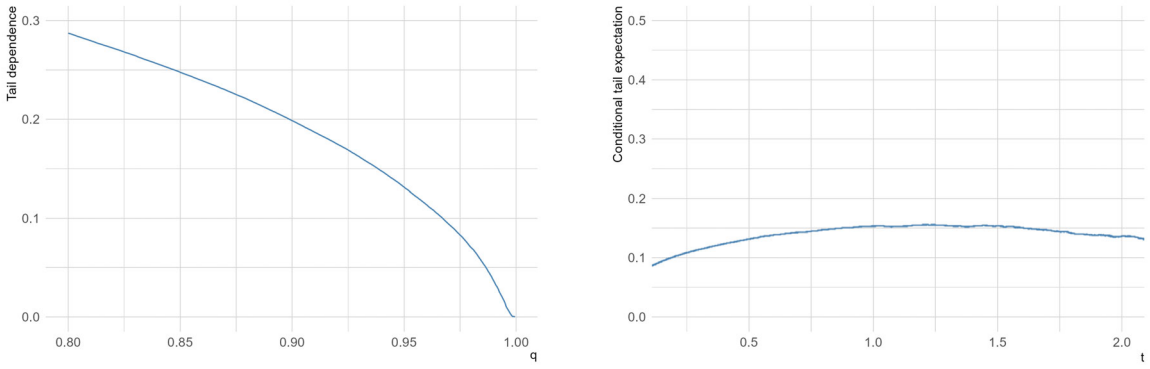
The same set of analyses is applied to individual income growth rates, and the results are reported in **Table 4** and **Figure 10**. The results for individual income closely mirror those for firm sales: in particular, dependence between growth rates is weak in the tail region. These results indicate that the growth rates in both cases can be well approximated by a random walk.

4.3 Subexponentiality and Weibull Tail of Growth Rates

Methods

This section examines whether growth rates follow a subexponential distribution, and in particular, whether they exhibit a Weibull tail. Our analysis employs kernel density estimation, the mean excess function, statistical tests for exponentiality, and tail-shape estimators proposed by Gardes et al. (2011) and El Methni et al. (2012). We describe these methods in turn below.

The mean excess function at a threshold u (denoted $e(u)$) is defined as the conditional expec-



(a) Tail dependence measure

(b) Conditional tail expectation

Figure 10: Tail dependence measure and conditional tail expectation for individual income growth rates. Let X_1 denote the growth rate for 2014-15 and X_2 the growth rate for 2015-16. Panel (a) plots the conditional probability $\mathbb{P}(X_2 > F_2^{-1}(q) \mid X_1 > F_1^{-1}(q))$ against q. Panel (b) plots the conditional tail expectation $\mathbb{E}[X_2 \mathbf{1}_{X_2>0} \mid X_1 \mathbf{1}_{X_1>0} > t]$ against t, using the nonnegative variable $X_k \mathbf{1}_{X_k>0}$.

tation of the overshoot $X - u$ given that $X > u$ (see, e.g., Embrechts et al. (1997)):

$$e(u) := \mathbb{E}[X - u \mid X > u] \quad \text{for } u > 0.$$

The mean excess function is useful because its shape reflects the heaviness of the tail of the distribution of X. In particular, our analysis focuses on the following three cases. First, when the distribution has an exponential tail, $e(u)$ remains constant in u. Second, when the distribution has a Pareto tail, $e(u)$ is a linearly increasing function of u. Third, when the distribution has a Weibull tail, its behavior lies between these two cases. Specifically, as u becomes large, it is known that

$$e(u) = \frac{u^{1-\alpha}}{c\alpha}(1 + o(1))$$

as $u \to \infty$ (cf. Beirlant et al. (1995)). By comparing the empirical mean excess function with these theoretical forms, we can characterize the tail heaviness of the growth rate distribution.

To statistically test that the distribution has a subexponential tail, we formulate the null hypothesis that the tail is exponential and test whether this hypothesis can be rejected. Numerous tests for exponentiality have been proposed in the literature (see, e.g., the surveys in Ascher (1990) and Henze and Meintanis (2005)). In particular, Ascher (1990) reports that the Cox-Oakes test, originally proposed by Cox and Oakes (1984), has the highest power among the various tests proposed in the literature. In our analysis, we apply the Cox-Oakes test to two subsamples of growth rates: observations with $X_k \geq 0.1$ and those with $X_k \geq 0.2$.

As discussed in Section 3.2, within the class of subexponential distributions, two major

subclasses are those with Pareto tails and those with Weibull tails. In our analysis, it is necessary to verify that the growth rate distribution is well approximated by a Weibull tail. As a graphical diagnostic, we use the following linear relationship, which holds when the underlying distribution has a Weibull tail. For $0 < u < v < 1$,

$$\log(-\log u) - \log(-\log v) \approx \alpha(\log x_u - \log x_v)$$

where x_u and x_v denote the u- and v-quantile values of the growth rate distribution, respectively. Let N be the sample size and let X_{N-i+1} denote the i-th largest observed growth rate. If the distribution follows a Weibull tail, then the points $(\log\log(N/i), \log X_{N-i+1})$ should lie approximately on a straight line. We exploit this property to assess whether the empirical distribution of growth rates is consistent with a Weibull tail.

To statistically assess whether the empirical distribution of growth rates is closer to a Pareto-type or a Weibull-type tail, we employ the methods proposed by Gardes et al. (2011) and El Methni et al. (2012). We consider the following family of distributions, which encompasses both Pareto-type and Weibull-type tails:

$$\overline{F}(x) = \exp(-K_\tau^{\leftarrow}(\log H(x))) \quad \text{for } x \geq x_* > 0, \text{ with } \tau \in [0, 1] \tag{5}$$

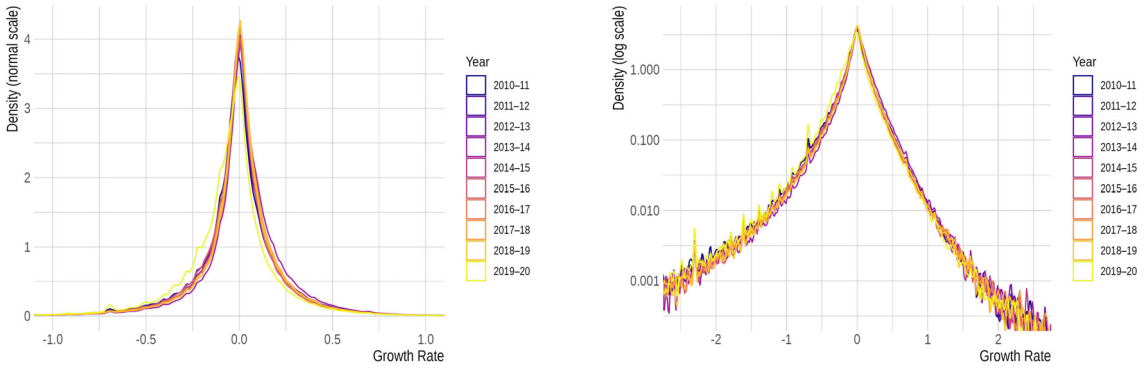
Here, H is a function whose inverse function satisfies $H^{\leftarrow}(t) = t^\theta L(t),\ \theta > 0$, where $L(t)$ is a slowly varying function, and $K_\tau(x) = \int_1^x u^{\tau-1} du$. The parameter τ characterizes the heaviness of the tail: larger values of τ correspond to heavier tails. In particular, $\tau = 0$ corresponds to a Weibull-type tail,[26] whereas $\tau = 1$ corresponds to a Pareto-type tail (the associated tail-shape parameter is $\alpha = \theta^{-1}$). Thus, by estimating τ, we can assess how close the empirical growth rate distribution is to a Weibull tail.

Results

Here, we apply the above methods to the growth rates of firm sales. The results of the density estimation for one-year growth rates are presented in **Figure 11**. Panel (a) shows that the distribution departs markedly from a Gaussian, with a sharp central peak and heavy tails. Panel (b) plots the same densities on a logarithmic scale, where the upper tail deviates upward from linearity. This curvature indicates that the tail is heavier than exponential, providing evidence that the growth rate distribution is subexponential.

We next examine the distribution of multi-period growth rates. **Figure 12** reports the estimated

[26] Note that when $\tau = 0$, we have $K_\tau(x) = \log x$, and hence $K_\tau^{\leftarrow}(x) = \exp(x)$. Therefore, $\overline{F}(x) = \exp\big(-\exp(\log H(x))\big) = \exp\big(-H(x)\big)$, which reduces to the Weibull-type tail case.



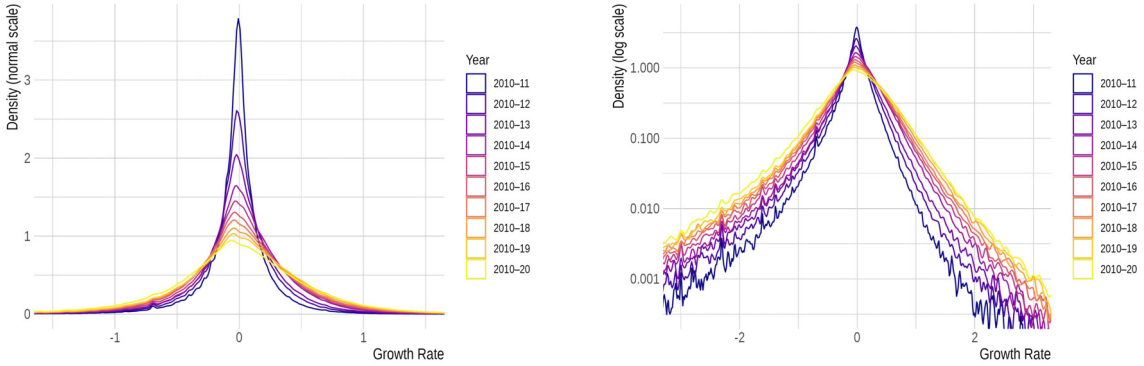
(a) Normal scale **(b)** Log scale

Figure 11: Density estimates of annual growth rates of firm sales. Here, we present the estimated density functions of the annual growth rates over the period 2010-2020. Panel (a) displays the densities on a normal scale, while Panel (b) displays them on a logarithmic scale.

densities for cumulative growth rates over $n = 1, 2, \ldots, 10$ years. As shown in Panel (a), the density around zero gradually approaches a bell-shaped curve as n increases. Panel (b) shows that, on a log scale, the right tail becomes approximately linear; moreover, the linear segment shifts upward as n increases. These features are consistent with the theoretical shape described in Section 3.2.

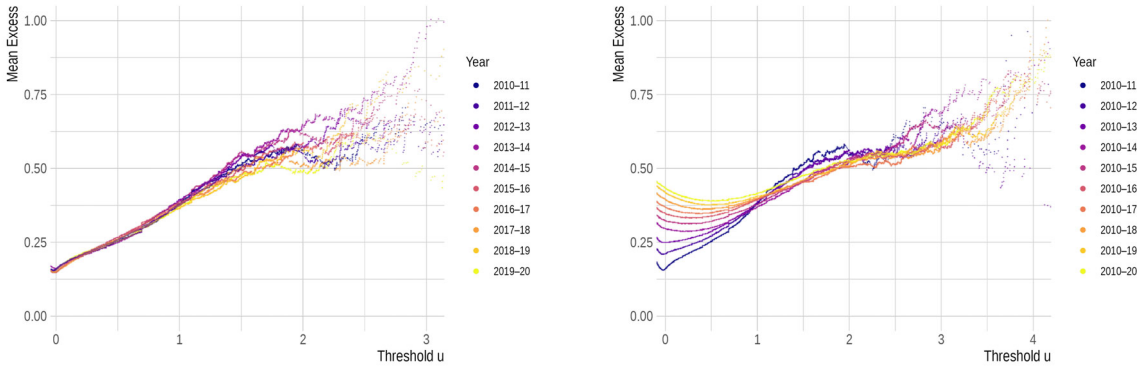
The mean excess function of the growth rates is reported in **Figure 13**. In all years, the estimated $e(u)$ is an increasing function of u, indicating that the distribution has tails heavier than exponential tails. Consistent with this, the Cox-Oakes test rejects exponentiality in all cases, with p-values less than 0.01. **Figure 13**(b) further shows that, in the tail region, the shape of $e(u)$ is largely independent of n. This is consistent with a property of subexponential distributions: the tail probabilities of n-period growth rates coincide with those of one-period growth rates, up to a multiplicative constant. This provides additional evidence that the growth rate distribution is subexponential. Moreover, in all panels, $e(u)$ increases with u, but its slope flattens as u grows. This pattern suggests that the growth rate distribution is closer to a Weibull tail than to a Pareto tail. In the following, we examine whether the distribution is more consistent with a Pareto or a Weibull tail.

We examine whether a Weibull tail provides a good approximation to the upper tail of the growth rate distribution, using the sample of one-year growth rates for 2019-20. **Figure 14**(a) plots $(\log\log(N/i), \log X_{N-i+1})$ using the top 5% of the growth-rate observations (i.e., $i = 1, \ldots, 0.05N$). The plotted points lie close to a straight line, indicating that the upper tail of the growth rate distribution is well approximated by a Weibull tail. The estimation results for the



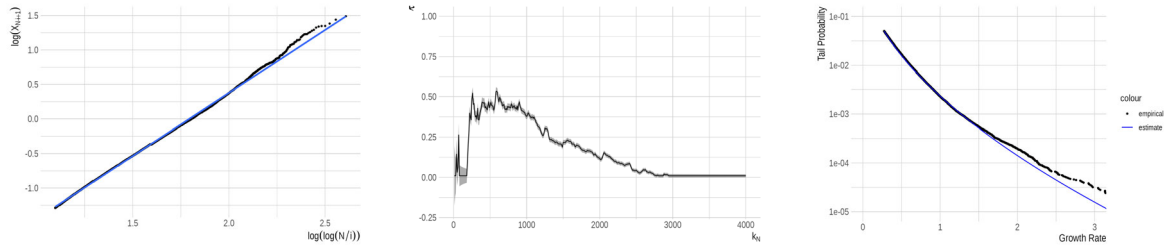
(a) Normal scale **(b)** Log scale

Figure 12: Density estimates of long-term growth rates of firm sales. This figure presents the estimated density functions of the n-year growth rates for $n = 1, 2, \ldots, 10$, using the data from the period 2010-2020. The initial year for all n-year growth rates is fixed at 2010. Panel (a) displays the densities on a normal scale, while Panel (b) displays them on a logarithmic scale.



(a) One-year growth rates **(b)** n-year growth rates

Figure 13: Mean excess function over threshold u for firm sales growth rates. Panel (a) presents the mean excess function of one-year growth rates, while Panel (b) reports the mean excess function of n-year growth rates ($n = 1, 2, \ldots, 10$), using data from 2010 to 2020. For all n, the initial year for the n-year growth rate is fixed at 2010. We conduct the Cox-Oakes test using the sample of growth rates for 2019-20. The standardized test statistic is -53.4 for observations with $X_k > 0.1$ and -33.7 for those with $X_k > 0.2$, and in both cases the null hypothesis of exponentiality is rejected at the 1% significance level.



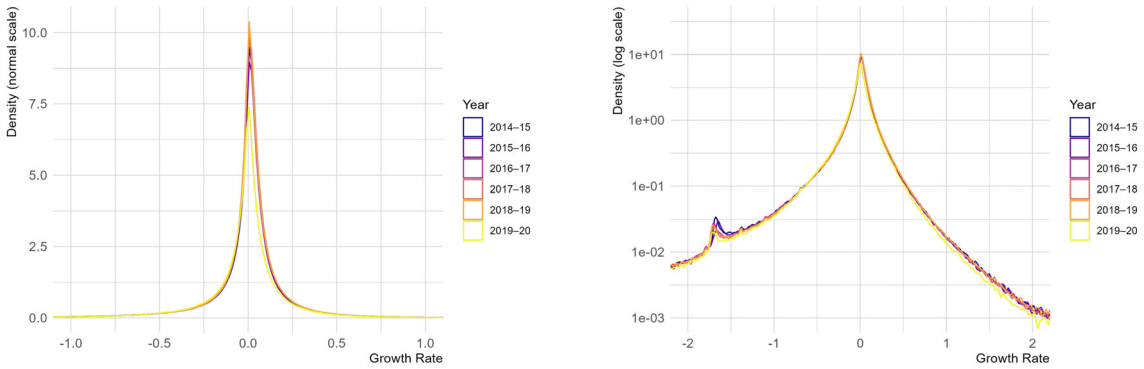
(a) Linear relation **(b)** Estimate of τ **(c)** Estimate of the tail probability

Figure 14: Parameter estimates of the tail probability of firm sales growth rates. This figure uses the one-year growth rates for 2019-20. In Panel (a), we plot $(\log\log(N/i), \log X_{N-i+1})$ using the top 5% of the growth-rate observations (i.e., $i = 1, \ldots, 0.05N$). Panel (b) reports the estimated values of τ together with the 99% confidence intervals. Since the estimation of τ is based only on the largest k_n observations, the x-axis shows the choice of k_n, while the y-axis shows the corresponding estimate of τ. Panel (c) presents the tail probabilities computed from Eq.(5), using the estimate of α obtained under the assumption $\tau = 0$. For comparison, the empirical complementary cumulative distribution function (i.e., $1 - F_n(x)$) of the growth rates is also plotted.

parameter τ in Eq.(5) are presented in **Figure 14**(b). The estimates are close to zero, implying that a Weibull-type tail provides a better approximation than a Pareto-type tail. Setting $\tau = 0$, we then estimate the remaining parameter α and, based on Eq.(5), compute the corresponding tail probabilities. **Figure 14**(c) compares these estimated tail probabilities with the empirical complementary cumulative distribution function (i.e., $1 - F_n(x)$, where F_n denotes the empirical distribution function). The estimated tail probabilities closely match those observed in the data. Taken together, and consistent with the findings from the mean excess function, these results support our assumption that the growth rate distribution follows a Weibull tail.

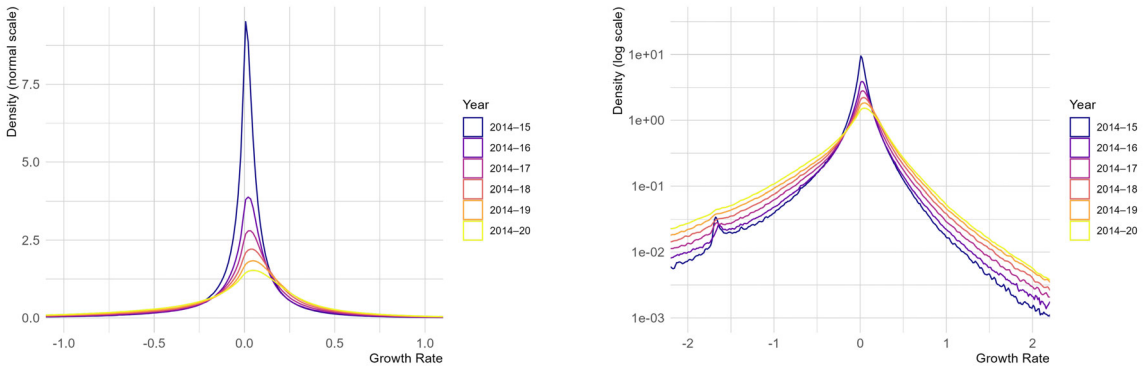
Applying the same set of analyses to individual income data yields **Figure 15**, **Figure 16**, **Figure 17**, and **Figure 18**. As in the case of firm sales, these results indicate that the distribution of individual income growth rates exhibits heavy tails, and in particular, is well approximated by a Weibull tail.

Thus, the dependence structure and distributional shape of growth rates are remarkably similar for firm sales and individual incomes. Since our theoretical framework in Section 3 relies solely on these statistical properties, this similarity implies that Pareto tails should emerge in both settings. This helps explain why Pareto tails are observed across a wide range of empirical domains. The next section shows that the predictions of our theory are indeed consistent with the empirical data.



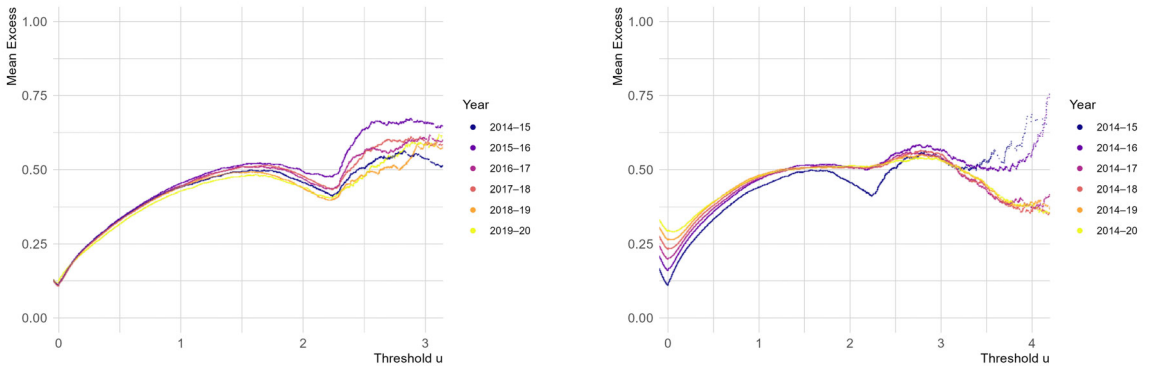
(a) Normal scale **(b)** Log scale

Figure 15: Density estimates of annual growth rates of individual income. Here, we present the estimated density functions of the annual growth rates over the period 2014-2020. Panel (a) displays the densities on a normal scale, while Panel (b) displays them on a logarithmic scale.



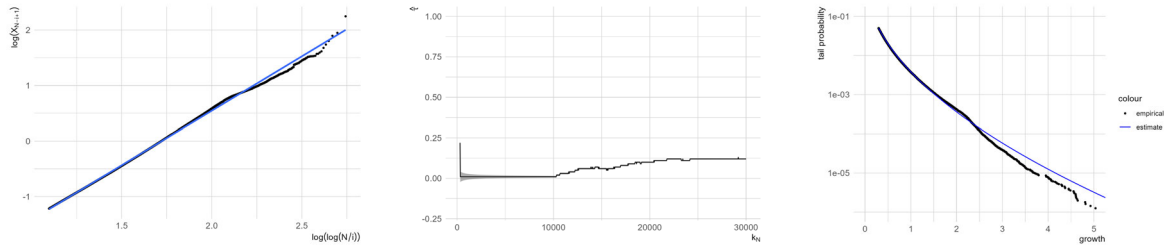
(a) Normal scale **(b)** Log scale

Figure 16: Density estimates of long-term growth rates of individual income. This figure presents the estimated density functions of the n-year growth rates for $n = 1, 2, \ldots, 10$, using the data from the period 2014–2020. The initial year for all n-year growth rates is fixed at 2014. Panel (a) displays the densities on a normal scale, while Panel (b) displays them on a logarithmic scale.



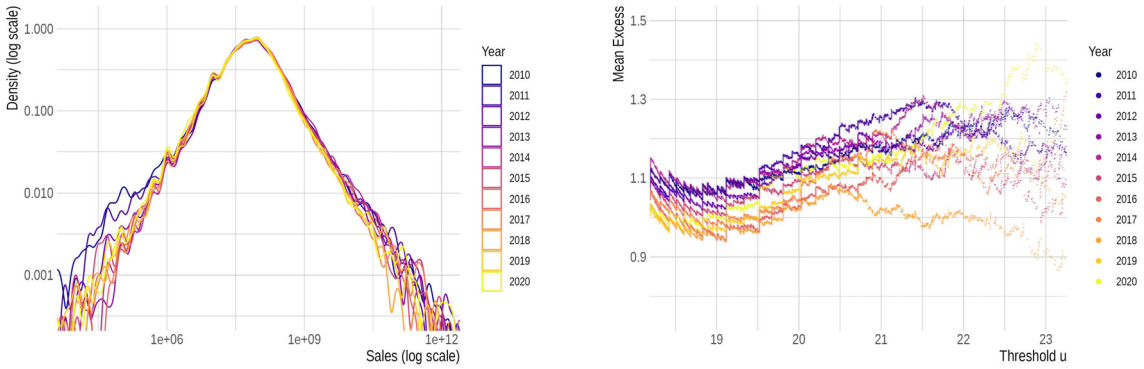
(a) One-year growth rates **(b)** n-year growth rates

Figure 17: Mean excess function over threshold u for individual income growth rates. Panel (a) presents the mean excess function of one-year growth rates, while Panel (b) reports the mean excess function of n-year growth rates ($n = 1, 2, \ldots, 6$), using data from 2014 to 2020. For all n, the initial year for the n-year growth rate is fixed at 2014. We conduct the Cox-Oakes test using the sample of growth rates for 2019-20. The standardized test statistic is -286.6 for observations with $X_k > 0.1$ and -172.7 for those with $X_k > 0.2$, and in both cases the null hypothesis of exponentiality is rejected at the 1% significance level.



(a) Linear relation **(b)** Estimate of τ **(c)** Estimate of the tail probability

Figure 18: Parameter estimates of the tail probability of individual income growth rates. This figure uses the one-year growth rates for 2019-20. In Panel (a), we plot $(\log\log(N/i), \log X_{N-i+1})$ using the top 5% of the growth-rate observations (i.e., $i = 1, \ldots, 0.05N$). Panel (b) reports the estimated values of τ together with the 99% confidence intervals. Since the estimation of τ is based only on the largest k_n observations, the x-axis shows the choice of k_n, while the y-axis shows the corresponding estimate of τ. Panel (c) presents the tail probabilities computed from Eq.(5), using the estimate of α obtained under the assumption $\tau = 0$. For comparison, the empirical complementary cumulative distribution function (i.e., $1 - F_n(x)$) of the growth rates is also plotted.



(a) Density estimates

(b) Mean excess functions

Figure 19: The distribution of the initial size S_0. We consider the logarithm of sales S_0 for firms that are five years old. Panel (a) shows the density estimate of S_0, with the vertical axis plotted on a log scale. Panel (b) presents the mean excess function of S_0.

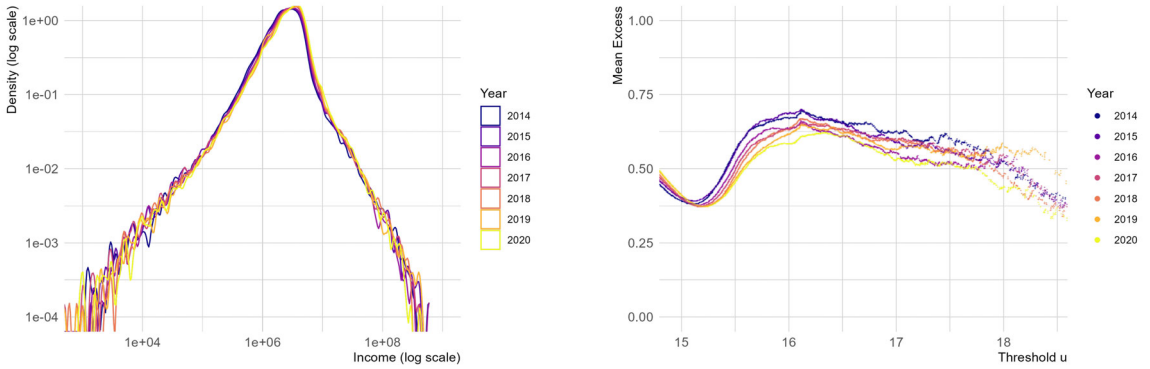
5 Empirical Validation of Theoretical Predictions

In this section, we examine whether the predictions of our theory are consistent with the data. Section 5.1 evaluates the tail exponent of the size distribution S_n. Section 5.2 examines the age-specific contribution to the tail of the aggregate size distribution. Section 5.3 examines whether jump-type growth patterns can be observed in the data.

5.1 Tail Exponent of the Size Distribution

As discussed in Section 3.3, the tail slope of the size distribution S_n on a log scale is determined by the tail slopes of the initial size distribution S_0 and the growth rate distribution. When the tail slopes of the growth rate distribution and the initial size distribution coincide, the size distribution shares this common slope, and an increase in n simply shifts the tail upward in parallel. By contrast, when the two slopes differ, its tail slope gradually converges to that of the growth rate distribution as n increases. Moreover, as n becomes large, the Cramér approximation region expands, and the size distribution approaches a Gaussian. In this section, using linear approximations of ℓ_0 and ℓ, we empirically estimate the tail slopes of the size distribution, the initial size distribution, and the growth rate distribution, and examine whether these theoretical predictions hold.

We begin with the case of individual income. Let S_0 denote the income of individuals who are 25 years old in 2020. The probability density function and the mean excess function of S_0 are



(a) Density estimates

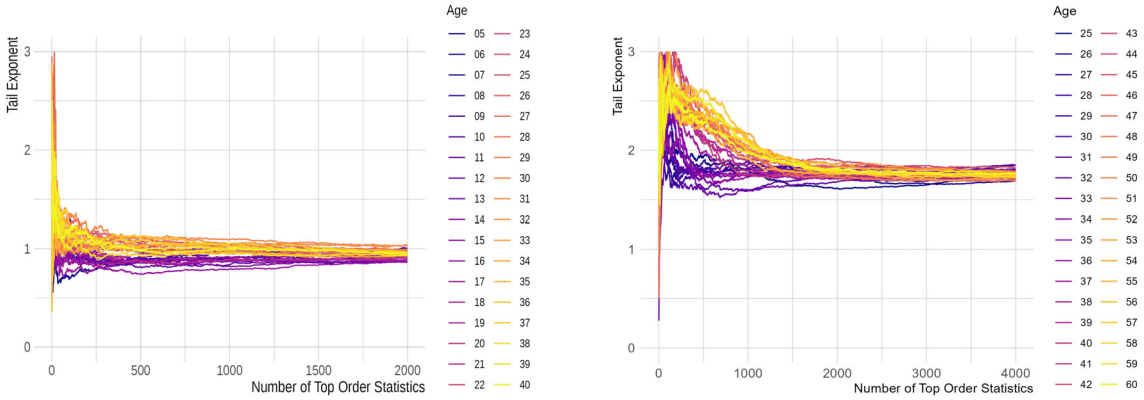
(b) Mean excess functions

Figure 20: The distribution of the initial size S_0. We consider the logarithm of income S_0 for individuals who are 25 years old. Panel (a) shows the density estimate of S_0, with the vertical axis plotted on a log scale. Panel (b) presents the mean excess function of S_0.

presented in **Figure 20**. Using the top 5% of the sample, the Hill estimate of the tail exponent of the distribution of S_0 is $a = 1.787$. Next, to estimate the tail exponent of the growth rate distribution, we use six-year growth rates $\widetilde{S}_6$ starting from 2014. Using the top 0.1% of the sample, the Hill estimate of the tail exponent of this distribution is $a = 1.890$. The difference between the tail exponents of S_0 and the growth rate distribution is small. Since these two exponents are close, the tail exponent of S_n should also take a similar value. Indeed, as shown in **Figure 21**(b), the tail exponent of the distribution of S_n remains around 1.8 and shows little variation with respect to n. These results are consistent with our prediction that the tail exponent of S_n is determined by the tail exponents of S_0 and the growth rate distribution.

Next, we turn to the case of firm sales. Here, we define S_0 as the sales of firms that are five years old in 2020. **Figure 19** presents the probability density function and the mean excess function of S_0. Using the top 5% of the sample, the Hill estimate of the tail exponent of the S_0 distribution is $a = 0.873$. To estimate the tail exponent of the growth rate distribution, we consider ten-year growth rates $\widetilde{S}_{10}$ starting from 2010. Using the top 0.1% of the sample, the Hill estimate of the tail exponent of this distribution is $a = 1.960$. Compared with the S_0 distribution, the growth rate distribution thus has a lighter tail.

According to our theory, the tail exponent of the size distribution S_n should lie between the tail exponents of S_0 and of the growth rate distribution, and it should move toward the latter as n increases. In other words, as n increases, the tail exponent of S_n is predicted to increase. Indeed, as shown in **Figure 21**(a), the tail exponent of S_n takes values between those of S_0 and the growth rate



(a) Firm sales **(b)** Individual incomes

Figure 21: Estimates of tail exponents. This figure reports the Hill estimates of the tail exponents of the distribution of S_n. The x-axis indicates the number of upper order statistics used in the estimation. Panel (a) shows the estimates for firm sales, using S_n for firms of age $n = 5, \ldots, 40$ as of 2020. Panel (b) shows the estimates for individual incomes, using S_n for individuals of age $n = 25, \ldots, 60$ as of 2020.

distribution, and it tends to be larger for older firms (although the increase is modest, reflecting the heavy-tailed nature of the S_0 distribution). Moreover, as n becomes large, the tail of the distribution exhibits downward curvature rather than forming a straight line, as seen in **Figure 2**(b). This is consistent with the prediction that, as the Cramér approximation region expands, the distribution becomes well approximated by a Gaussian distribution over a wide range. Taken together, these results are consistent with our theoretical prediction regarding the relationship among the tail exponents of S_0, the growth rate distribution, and S_n, as well as the resulting change in the shape of the size distribution as n increases.

5.2 Age-Specific Contribution to the Tail of the Aggregate Size Distribution

The proportion of the tail of the size distribution accounted for by different age groups is a key feature that distinguishes our theory from existing models. As in Section 5.1, suppose that the tail of the size distribution S_n for agents of age n can be approximated by

$$\log \mathbb{P}(S_n > x \mid \text{age} = n) \approx -a_n x + b_n.$$

In particular, assume that the slopes are common across ages, that is, $a_1 = a_2 = \cdots =: a$. Let $p_n := \mathbb{P}(\text{age} = n)$ denote the share of agents of age n in the population. Then the size distribution

S, aggregated over all age groups, satisfies

$$\log \mathbb{P}(S > x) = \log \sum_n p_n \mathbb{P}(S_n > x \mid \text{age} = n) \approx \log \left(\left(\sum_n p_n e^{b_n} \right) e^{-ax} \right) = -ax + b$$

where $b := \log \sum_n p_n e^{b_n}$. Thus, if the tail of each age-specific distribution has a common slope, then the tail of the aggregate size distribution also exhibits the same slope. This is why, in our theory, a Pareto tail emerges for the aggregate size distribution.

In this case, note that $\log \mathbb{P}(S_n > x \mid \text{age} = n) - \log \mathbb{P}(S > x) = \text{const.}$, which implies that the ratio $\mathbb{P}(S_n > x \mid \text{age} = n)/\mathbb{P}(S > x)$ is independent of x. Hence, the share of agents of age n in the tail of the aggregate size distribution is

$$\frac{\mathbb{P}(S_n > x, \text{age} = n)}{\mathbb{P}(S > x)} = p_n \frac{\mathbb{P}(S_n > x \mid \text{age} = n)}{\mathbb{P}(S > x)} = \text{const.}$$

In other words, the proportion of each age group in the tail depends only on n and remains constant with respect to x.

The above property can be confirmed using individual income data. As shown in **Figure 5**(b), the proportion of each age group within the tail probability remains stable across a wide range of x. This stands in contrast to the predictions of existing models, which imply that, as x increases, the tail of the size distribution becomes increasingly dominated by older agents. This result supports our theory: the emergence of Pareto tails is not the consequence of aggregating agents of different ages, but rather a property that already holds within each age group.

The corresponding results for firm sales are presented in **Figure 3**(a). As the figure shows, the proportion of younger firms increases with x. Strictly speaking, this pattern deviates from a Pareto tail; however, the nature of this deviation is consistent with the findings in Section 5.1. Specifically, for firm sales, the tail of the initial size distribution S_0 is heavier than that of the growth rate distribution. Consequently, when n is small, the tail of the size distribution is closer to that of S_0. As a result, the share of younger firms in the tail of the size distribution increases as x grows, in line with the predictions of our theory.

5.3 Jump-Type Growth Patterns in the Data

Here, following the discussion in Section 3.4, we examine whether a large deviation of $\widetilde{S}_n$ is driven not by average growth over n periods, but by a rapid growth occurring in a single period (i.e., a jump). Letting n be even, define the following ratio for each agent:

$$r = \frac{\sum_{k=1}^{n/2} X_k}{\sum_{k=1}^{n} X_k}.$$

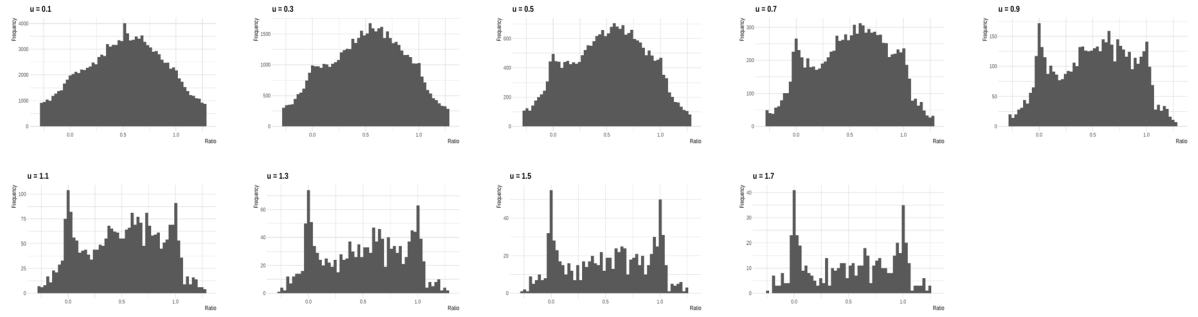


Figure 22: Histogram of r. Using the firm-level annual growth rates for 2010-11 and 2011-12, which we denote by X_{11} and X_{12}, we compute r with $n = 2$. We then restrict the subsample to firms satisfying the condition $X_{11} + X_{12} > u$ and plot the histogram of r. The value of u increases from 0.1 (top left) to 1.7 (bottom right) in increments of 0.2.

This ratio r represents the contribution of the first $n/2$ periods to the total growth over the n periods. For example, if the first half and the second half contribute equally to total growth, then $r = 1/2$. In what follows, we compute the histogram of r for agents who satisfy the condition $\sum_{k=1}^{n} X_k > u$, that is, agents who achieve high growth over n periods, and examine how its shape changes as u becomes large. According to the discussion in Section 3.4, when the growth rate distribution is subexponential, the event $r = 1/2$ should be among the least likely outcomes for large u. Below, we investigate whether such a pattern is indeed observed in the data.

Using firm sales growth rates for 2010-11 and 2011-12 (i.e., $n = 2$), we compute the ratio r and examine how the histogram of r changes shape as u increases. The value of u ranges from 0.1 to 1.7 in increments of 0.2. As shown in **Figure 22**, when u is small, the histogram of r exhibits a peak at $1/2$. This indicates that when the two-year growth rate is low, both years tend to contribute almost equally to the total two-year growth. In contrast, as u becomes large (e.g., $u \geq 0.9$), the histogram of r develops peaks near 0 and 1. This implies that high two-year growth is not typically achieved through equal contributions from both years, but rather through exceptionally rapid growth—a jump—occurring in a single year.

A similar pattern is even more clearly observed for individual income growth rates. Using the six-year growth rates from 2014 to 2020, we compute the ratio r, which measures the contribution of the first three years to the total six-year growth. We then restrict the sample to individuals whose total six-year growth exceeds a threshold u, and construct the histogram of r for each value of u. The value of u ranges from 0.1 to 2.8 in increments of 0.3. As shown in **Figure 23**, the same pattern as in the case of firm sales emerges. When u is small, the histogram of r exhibits a peak at $1/2$. This indicates that when the six-year growth rate is low, the first and second three-year

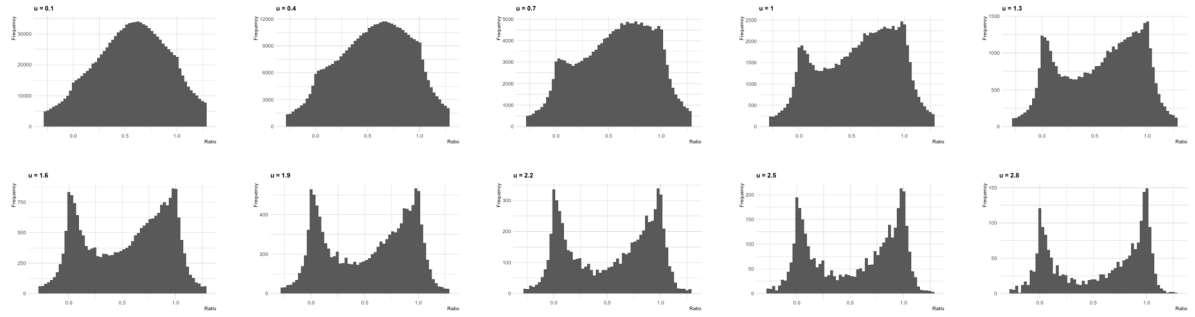
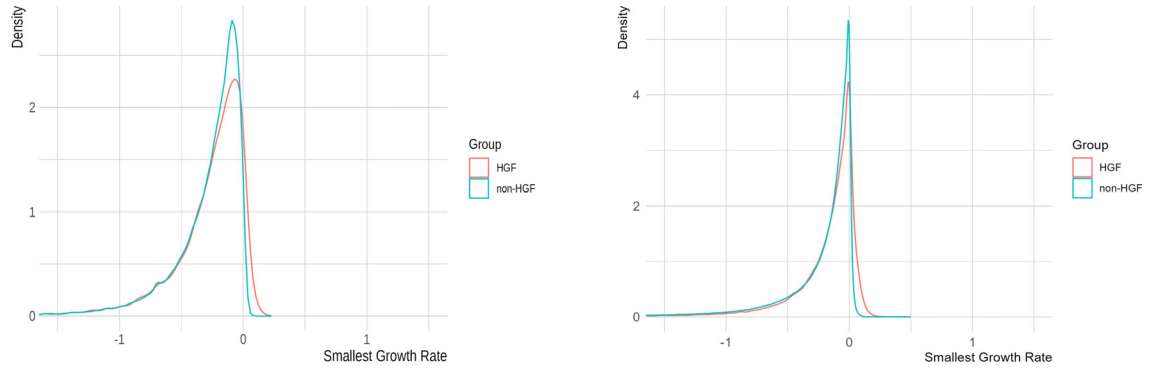


Figure 23: Histogram of r. Using the six-year individual income growth rates $X_{15}, \ldots, X_{20}$, we compute r with $n = 6$. We then restrict the subsample to individuals satisfying the condition $\sum_{k=15}^{20} X_k > u$ and plot the histogram of r. The value of u increases from 0.1 (top left) to 2.8 (bottom right) in increments of 0.3.

periods tend to contribute almost equally to total growth. In contrast, as u increases, the histogram develops peaks near 0 and 1, implying that high six-year growth rates are concentrated in either the first or the second half. Because the individual income dataset is much larger than the firm-level dataset, this change in the shape of the histogram of r can be observed clearly even for large values of u.

As an additional check for the jump-driven process, we examine the conditional distribution of growth rates under a large deviation event of $\widetilde{S}_n$. According to Section 3.4, if the growth rate distribution is subexponential, then high n-period growth (i.e., $\widetilde{S}_n > x$) is generated by rapid growth in a single period, that is, a jump. In this case, the growth rates in the remaining $n-1$ periods follow the unconditional distribution, and therefore the minimum growth rate among these periods should have the same distribution as the minimum drawn from the unconditional distribution. In contrast, if the growth rate distribution is light-tailed, the conditional distribution of growth rates given $\widetilde{S}_n > x$ differs from the unconditional one. For example, if the growth rates are Gaussian, then conditioning on $\widetilde{S}_n > x$ shifts the mean of the conditional distribution by x/n to the right. To distinguish between these two cases, we divide agents into those satisfying $\widetilde{S}_n > x$ and those who do not, and compute, for each agent, the minimum growth rate among the remaining $n-1$ periods. We then compare the resulting histograms for the two groups. If the histograms differ by x/n, the data are consistent with a light-tailed explanation; if they are indistinguishable, the data support the subexponential explanation.

We now consider the case of firm sales. Let $n = 10$, and use the 97th percentile of $\widetilde{S}_n$ as the threshold for a large deviation. From the data, this value is 1.11. If the growth rate distribution were Gaussian, then conditioning on $\widetilde{S}_n > 1.11$ would shift the mean of the conditional distribution by $1.11/10 = 0.111$ to the right. **Figure 24**(a) compares the histograms of the minimum growth



(a) Firm sales growth

(b) Individual income growth

Figure 24: Comparison of the distributions of the minimum growth rate within the sample period. We divide the sample into two groups: agents satisfying $\widetilde{S}_n > x$ and those who do not, where x is set to the 97th percentile of $\widetilde{S}_n$. For each agent, we compute the minimum growth rate within the sample periods and compare the histograms between the two groups.

rate among the remaining $n-1$ periods for the group satisfying $\widetilde{S}_n > 1.11$ and for the group that does not. The two distributions are extremely similar. The medians are -0.194 and -0.199, respectively, differing by only 0.005. Given that the Gaussian case predicts a rightward shift of 0.111, the difference observed in the data is negligibly small. This result is consistent with the behavior predicted by a subexponential growth rate distribution.

We conduct the same analysis for individual income growth. Here, with $n = 6$, the 97th percentile of the six-year growth rate $\widetilde{S}_n$ is 0.891. If the growth rate distribution were Gaussian, then conditioning on $\widetilde{S}_n > 0.891$ would shift the mean of the conditional distribution by $0.891/6 = 0.149$ to the right. **Figure 24**(b) compares the histograms of the minimum growth rate among the remaining $n-1$ periods between the group satisfying $\widetilde{S}_n > 0.891$ and the group that does not. The medians are -0.0942 and -0.110, respectively, differing by only 0.016. Compared with the expected rightward shift of 0.149 under a Gaussian distribution, the difference observed in the data is clearly small. This result is consistent with the subexponential nature of the growth rate distribution.

Taken together, the results on the ratio r and on the minimum growth rate within the sample period show that large deviations in $\widetilde{S}_n$ are generated by a jump occurring in a short period. This stands in sharp contrast to existing models, which assume that large deviations arise from the accumulation of moderately high growth over many periods. These empirical findings indicate that our theory is more consistent with the data than the existing models.

6 Conclusion

Pareto tails are a characteristic statistical regularity commonly observed in the upper tails of size distributions and have long attracted attention from economists. However, recent studies have identified several inconsistencies between existing models and empirical data, particularly regarding the time required for agents to reach the upper tail. To address these inconsistencies, we develop a new theoretical explanation for Pareto tails. Using Japanese firm-level sales and individual-income data, we demonstrate that the key predictions of our theory are borne out by empirical evidence, such as age-specific size distributions and jump-driven growth patterns.

The central idea of our theory is the link between the tail-heaviness of the growth rate distribution and the growth patterns through which agents reach the upper tail of the size distribution. When growth rates are light-tailed, reaching the upper tail requires accumulating moderate growth rates over long periods. Existing models implicitly assume this pattern and therefore imply an unrealistically long time for agents to reach the upper tail. By contrast, when growth rates are heavy-tailed, a short episode of exceptionally high growth—a jump—can propel an agent into the upper tail. Our theory is built on this growth pattern and thereby resolves the empirical inconsistencies of existing models.

These two types of processes have important implications for macroeconomic theories of growth and income inequality. The gradual process rests on the accumulation of small growth differences over long periods, a mechanism widely embedded in existing models. For example, Piketty (2014) explains the concentration of wealth as the compounded effect of a persistent gap between the return on capital and the growth rate. In contrast, the jump-driven process represents an alternative mechanism in which extreme outcomes arise from a single big jump. Understanding such jump-driven processes may therefore offer new insights into the origins of growth and income inequality.

References

Arata, Y. (2019). Firm growth and Laplace distribution: The importance of large jumps. *Journal of Economic Dynamics and Control*, 103:63–82.

Arata, Y., Miyakawa, D., and Mori, K. (2023). The U-shaped law of high-growth firms. *NTC Discussion Paper*, 230201-01HJ(2).

Armendáriz, I. and Loulakis, M. (2011). Conditional distribution of heavy tailed random variables on large deviations of their sum. *Stochastic processes and their applications*, 121(5):1138–1147.

Ascher, S. (1990). A survey of tests for exponentiality. *Communications in Statistics-Theory and Methods*, 19(5):1811–1825.

Axtell, R. L. (2001). Zipf distribution of US firm sizes. *science*, 293(5536):1818–1820.

Beare, B. K. and Toda, A. A. (2022). Determination of Pareto exponents in economic models driven by Markov multiplicative processes. *Econometrica*, 90(4):1811–1833.

Beirlant, J., Broniatowski, M., Teugels, J. L., and Vynckier, P. (1995). The mean residual life function at great age: Applications to tail estimation. *Journal of Statistical Planning and Inference*, 45(1-2):21–48.

Borovkov, A. A. (2020). *Asymptotic analysis of random walks: light-tailed distributions*, volume 176. Cambridge University Press.

Borovkov, A. A. and Borovkov, K. (2008). *Asymptotic analysis of random walks*, volume 118. Cambridge University Press.

Bottazzi, G., Coad, A., Jacoby, N., and Secchi, A. (2011). Corporate growth and industrial dynamics: Evidence from French manufacturing. *Applied Economics*, 43(1):103–116.

Bottazzi, G. and Secchi, A. (2006). Explaining the distribution of firm growth rates. *The RAND Journal of Economics*, 37(2):235–256.

Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, 63(250):318–351.

Coad, A. (2009). *The growth of firms: A survey of theories and empirical evidence*. Edward Elgar Publishing.

Coad, A., Daunfeldt, S.-O., Hölzl, W., Johansson, D., and Nightingale, P. (2014). High-growth firms: Introduction to the special section. *Industrial and Corporate Change*, 23(1):91–112.

Commonwealth of Australia (2017). *Australian Innovation System Report 2017*. Office of the Chief Economist, Department of Industry, Innovation and Science, Canberra.

Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*. CRC press.

Dosi, G., Grazzi, M., Moschella, D., Pisano, G., and Tamagni, F. (2020). Long-term firm growth: An empirical analysis of US manufacturers 1959–2015. *Industrial and Corporate Change*, 29(2):309–332.

Dosi, G., Pereira, M. C., and Virgillito, M. E. (2017). The footprint of evolutionary processes of learning and selection upon the statistical properties of industrial dynamics. *Industrial and Corporate Change*, 26(2):187–210.

El Methni, J., Gardes, L., Girard, S., and Guillou, A. (2012). Estimation of extreme quantiles from heavy and light tailed distributions. *Journal of Statistical Planning and Inference*, 142(10):2735–2747.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons, New York, 3 edition.

Foss, S., Korshunov, D., and Zachary, S. (2011). *An Introduction to Heavy-Tailed and Subexponential Distributions*.

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

Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ.*, 1(1):255–294.

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

Gardes, L., Girard, S., and Guillou, A. (2011). Weibull tail-distributions revisited: a new look at some tail estimators. *Journal of Statistical Planning and Inference*, 141(1):429–444.

Goswami, A. G., Medvedev, D., and Olafsen, E. (2019). *High-growth firms: Facts, fiction, and policy options for*

emerging economies. World Bank Publications.

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2021). What do data on millions of US workers reveal about lifecycle earnings dynamics? *Econometrica*, 89(5):2303–2339.

Guvenen, F., Pistaferri, L., and Violante, G. L. (2022). Global trends in income inequality and income dynamics: New insights from GRID. *Quantitative Economics*, 13(4):1321–1360.

Henze, N. and Meintanis, S. G. (2005). Recent and classical tests for exponentiality: a partial review with comparisons. *Metrika*, 61:29–45.

Ijiri, Y. and Simon, H. A. (1977). *Skew distributions and the sizes of business firms*, volume 24. North-Holland.

Joe, H. (2014). *Dependence modeling with copulas*. CRC press.

Kotz, S., Kozubowski, T., and Podgórski, K. (2001). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media.

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3):1103–1144.

Luttmer, E. G. (2010). Models of growth and firm heterogeneity. *Annu. Rev. Econ.*, 2(1):547–576.

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

Petrov, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Clarendon Press.

Piketty, T. (2014). *Capital in the twenty-first century*. Harvard University Press.

Reed, W. J. (2001). The Pareto, Zipf and other power laws. *Economics letters*, 74(1):15–19.

Simon, H. A. and Bonini, C. P. (1958). The size distribution of business firms. *The American economic review*, 48(4):607–617.

Sornette, D. (2006). *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media.

Stanley, M. H., Amaral, L. A., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., and Stanley, H. E. (1996). Scaling behaviour in the growth of companies. *Nature*, 379(6568):804–806.

Toda, A. A. and Walsh, K. (2015). The double power law in consumption and implications for testing Euler equations. *Journal of Political Economy*, 123(5):1177–1200.

Wold, H. O. and Whittle, P. (1957). A model explaining the Pareto distribution of wealth. *Econometrica*, pages 591–595.

Supplemental Appendix: Rapid Growth and the Emergence of Pareto Tails

Yoshiyuki Arata*
Research Institute of Economy, Trade and Industry
National Tax College

Hiroshi Yoshikawa
University of Tokyo

Shingo Okamoto
National Tax College

Date: February 9, 2026

1 Appendix

This appendix examines the robustness of our empirical findings. Section 1.1 investigates how growth rates depend on firm size and age. Section 1.2 analyzes the dependence between growth rates in consecutive periods using copula theory. Section 1.3 reports results obtained when alternative definitions of individual income are used.

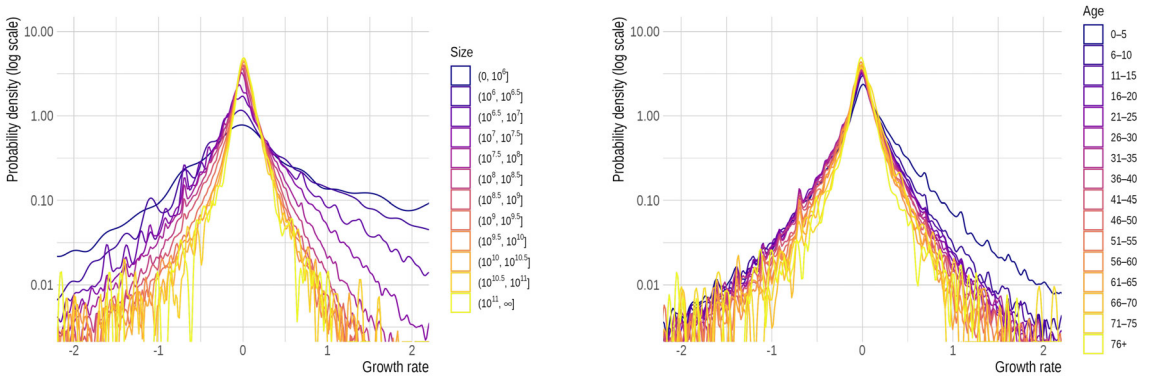
1.1 Dependence of Growth Rates on Size and Age

The dependence of growth rates on firm size and age has long been discussed in the literature on firm growth. In particular, it is well established that for firms that exceed a minimum size and are no longer in the initial start-up stage, the independence of growth rates (i.e., Gibrat's law) holds reasonably well (e.g., Lotti et al. (2009); Daunfeldt and Elert (2013)). In this section, we examine the distribution of growth rates across size and age categories to assess the ranges over which this approximation is valid.

Firm sales

We use firm sales growth rates for the period 2010-11. To examine the dependence of growth rates on firm size, we divide the sample into subsamples based on firms' sales in 2010, using size brackets defined by $10^{6.0}, 10^{6.5}, \ldots, 10^{11.0}$ yen. For each subsample, we estimate the probability

*Corresponding author: y.arata0325@gmail.com



(a) Size group **(b)** Age group

Figure 1: Density estimates of growth rates by size and age groups. We use firm sales growth rates for the period 2010-11. Panel (a) divides the sample into subsamples based on firms' sales in 2010, using size brackets of $10^{6.0}, 10^{6.5}, \ldots, 10^{11.0}$ yen. Panel (b) divides the sample into subsamples based on firms' age in 2010, using five-year intervals.

density function of the growth rate, and the results are presented in **Figure 1**(a). As the figure shows, smaller firms exhibit larger variance and a higher probability of large deviations. By contrast, once firm size becomes sufficiently large, this size dependence weakens, and the estimated density functions for firms with sales of at least $10^{7.5}$ yen almost coincide. This suggests that, for firms with sales above $10^{7.5}$ yen, the size dependence of growth rates is weak, and Gibrat's law provides a better approximation.

Similarly, to examine the dependence of growth rates on firm age, we divide the sample into five-year age groups based on the firms' age in 2010 and estimate the density function of the growth rate for each group. The results are shown in **Figure 1**(b). The growth rate distribution of newly established firms—those five years old or younger—differs markedly from that of the other age groups. The density functions for the remaining age groups are nearly identical. Thus, when the sample is restricted to firms older than five years, the age dependence of growth rates is essentially negligible. These results are consistent with previous studies and suggest that, once firms exceed a certain minimum size and are no longer in their initial start-up phase, growth rates can be regarded as approximately independent of both size and age.

Finally, we confirm that the heavy tails observed in the growth rate distributions—specifically, the fact that they are heavier than exponential—are not the result of aggregating firms of different sizes. In **Figure 2**, we restrict the sample to firms that exceed a certain minimum size and are not newly established (i.e., we exclude firms with sales below $10^{7.5}$ yen and firms five years old or

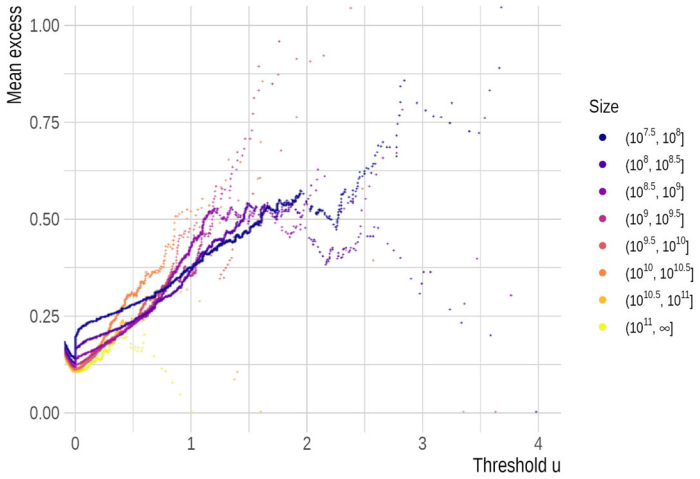


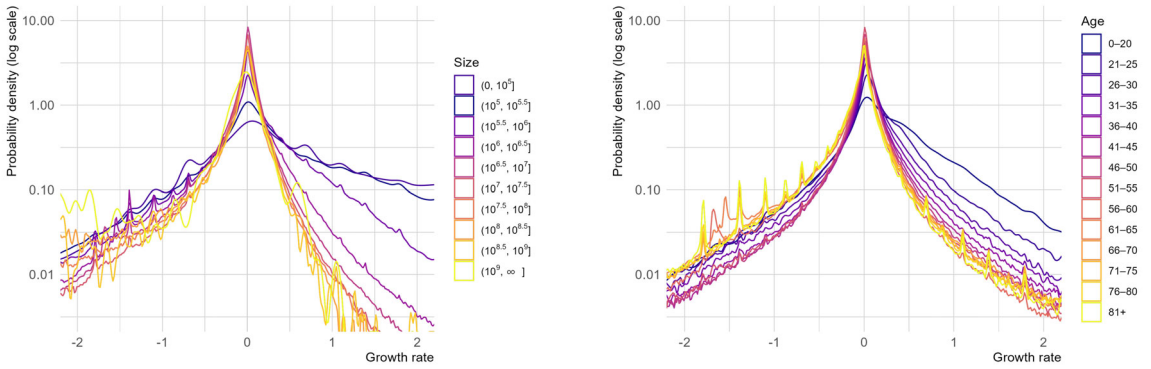
Figure 2: Mean excess function over threshold u by size group. The mean excess function of firms' sales growth rates for 2010-11 is computed by size group, restricting the sample to firms with sales of at least $10^{7.5}$ yen in 2010 and firms older than six years.

younger), and compute the mean excess function of growth rates for each size group. This figure shows that the mean excess function is increasing in the threshold u for all size groups. Thus, even when the sample is partitioned by firm size, the tails of the growth rate distribution remain heavier than exponential. These results indicate that the subexponential property is not driven by aggregation, but reflects an inherent and statistically robust characteristic of the growth rate distribution.

Individual incomes

We apply the same analysis to individual income growth rates for 2014–15. We divide the sample into subsamples based on individual income in 2014, using income brackets of $10^{5.0}, 10^{5.5}, \ldots, 10^{9.5}$ yen, and estimate the probability density function of income growth rates for each subsample. The results shown in **Figure 3**(a) indicate that the density functions for low-income groups exhibit greater variance and heavier tails than those for higher-income groups. However, this dependence weakens as income increases. In particular, for individuals with incomes above $10^{6.5}$ yen, the density functions almost overlap, suggesting that the dependence of the growth rate distribution on income level becomes negligible above this level.

To examine the dependence of the growth rate distribution on age, we divide the sample



(a) Income group **(b)** Age group

Figure 3: Density estimates of growth rates by size and age groups. We use individual income growth rates for the period 2014-15. Panel (a) divides the sample into subsamples based on individual income in 2014, using income brackets of $10^{5.0}, 10^{5.5}, \ldots, 10^{9.0}$ yen. Panel (b) divides the sample into subsamples based on individuals' age in 2014, using five-year intervals.

into five-year age groups based on individuals' age in 2014 and estimate the probability density function of income growth rates for each group. The results in **Figure 3**(b) show that the growth rate distribution for individuals aged 25 or younger differs substantially from that of other age groups, whereas the age dependence becomes significantly weaker for older individuals. Thus, once individuals aged 25 or younger are excluded from the sample, the age dependence of income growth rates is unlikely to pose a significant concern for our analysis. These results support the empirical validity of the independence assumption adopted in our main analysis, as long as the sample is restricted to individuals above certain thresholds of income and age.

Finally, we confirm that the subexponentiality of the growth rate distribution is not the result of aggregating individuals with different income levels. **Figure 4** shows the mean excess function of income growth rates by income group, after excluding individuals with income below $10^{6.5}$ yen and those aged 25 or younger from the sample. The mean excess function is increasing in the threshold u for each income group. That is, for all income groups, the tail of the growth rate distribution is heavier than exponential. These results indicate that the subexponentiality of the growth rate distribution is an inherent property of the distribution itself.

As shown above, in both the firm-sales and individual-income datasets, the distribution of growth rates can be reasonably approximated as independent of size and age, provided that very small or very young firms and individuals are excluded from the sample. Because the region in which this independence holds constitutes the majority of observations in both datasets, the

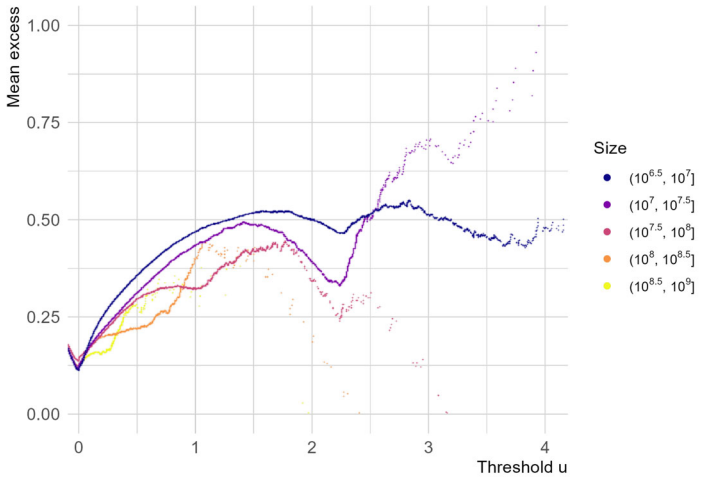


Figure 4: Mean excess function over threshold u by income group. The mean excess function of income growth rates for 2014-15 is computed by income group, restricting the sample to individuals aged between 26 and 60 and with income of at least $10^{6.5}$ yen in 2014.

assumption that growth rates are independent of size and age represents an empirically valid approximation for our analysis. Furthermore, the subexponential property of the growth rate distribution is consistently observed in both firm-sales and individual-income data and remains robust even when each dataset is further partitioned by size. This subexponentiality lies at the core of our theory, and its robustness suggests that the universality of Pareto tails arises from this distributional property.

1.2 Serial Dependence in Growth Rates: A Copula-Based Analysis

In this section, we analyze the dependence structure of growth rates in consecutive periods using copula theory. We begin with a brief introduction to copula theory and then provide empirical results based on copula analysis.

In the existing literature, the most common measure used to quantify the degree of dependence is Pearson's correlation coefficient. However, it is well known in the statistics literature that this coefficient does not purely reflect the dependence between two random variables (see, e.g., Embrechts et al. (2002)). This is because Pearson's correlation also depends on their marginal distributions. For example, if h_1 and h_2 are strictly increasing functions, the Pearson correlation between X_1 and X_2 does not, in general, coincide with the correlation between $h_1(X_1)$ and $h_2(X_2)$.

Such sensitivity to the marginals makes it difficult to determine whether the magnitude of the coefficient reflects genuine dependence or merely properties of the marginal distributions. This issue is particularly important in our setting, where the marginal distribution of growth rates differs substantially from a Gaussian distribution.

These issues can be addressed using copula theory (see Joe (2014) for details). The key idea of copula theory is that any bivariate distribution function F can be uniquely decomposed as

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)),$$

where C is the copula function, and F_1 and F_2 are the marginal distributions of X_1 and X_2, respectively. All information on the dependence between X_1 and X_2 is encapsulated in the copula C. The problem with Pearson's correlation coefficient arises from the fact that it is not determined solely by C, but also depends on the marginals F_1 and F_2. As a dependence measure determined only by C, Spearman's ρ_S is commonly used:

$$\rho_S := \mathrm{Corr}[F_1(X_1), F_2(X_2)]$$

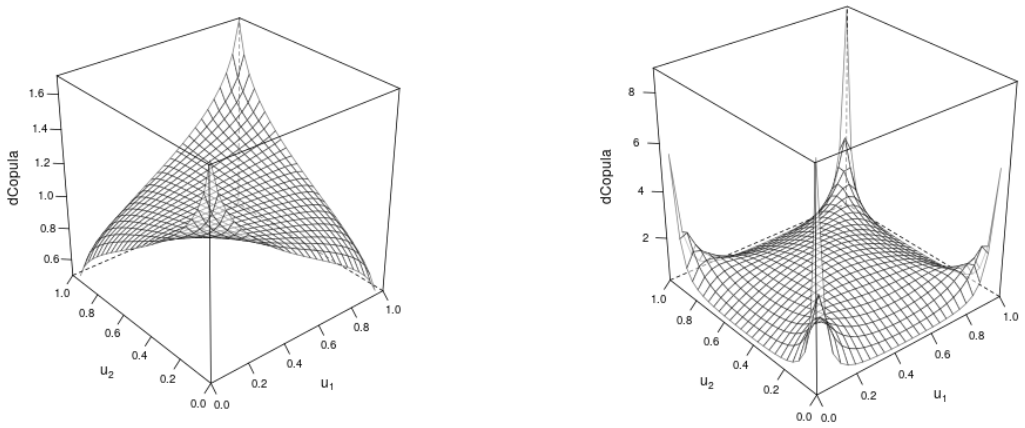
Since $F_1(X_1)$ and $F_2(X_2)$ indicate the rank positions of each observation within their respective marginal distributions, Spearman's ρ_S can be interpreted as a measure of rank concordance. For this reason, we use Spearman's ρ_S in the main text instead of Pearson's correlation coefficient.

To obtain a more detailed understanding of the dependence structure, we examine the functional form of C directly. In what follows, we describe two widely used copulas: the Gaussian copula and the Student-t copula. The Gaussian copula and the Student-t copula are derived from the bivariate Gaussian distribution and the bivariate Student-t distribution, respectively. Specifically, the Gaussian copula is defined as

$$C(u_1, u_2; \rho) = \Phi_2\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho\right)$$

where Φ_2 denotes the bivariate standard Gaussian distribution with correlation parameter ρ, and Φ denotes the univariate standard Gaussian distribution. In other words, the Gaussian copula extracts only the dependence structure of the bivariate Gaussian distribution by removing the effects of its marginal components. The Student-t copula is defined analogously, retaining only the dependence structure of the bivariate Student-t distribution.

The Gaussian copula and the Student-t copula each exhibit distinct characteristic shapes. The density functions of these two copulas are shown in **Figure 5**. The density of the Gaussian copula diverges at the corners $(0, 0)$ and $(1, 1)$. Note that even copulas such as the Gaussian copula—which exhibit weak tail dependence—can produce spikes at $(0, 0)$ and $(1, 1)$. Thus, spikes at these corners in a copula density plot do not necessarily indicate strong tail dependence. In contrast, the Student-t



(a) Gaussian copula

(b) Student-t copula

Figure 5: Density plots of the copula function C. For the Gaussian copula, the correlation parameter ρ is set to 0.1. For the Student-t copula, the degrees of freedom are set to $\nu = 1$, and the correlation parameter is set to $\rho = 0.1$.

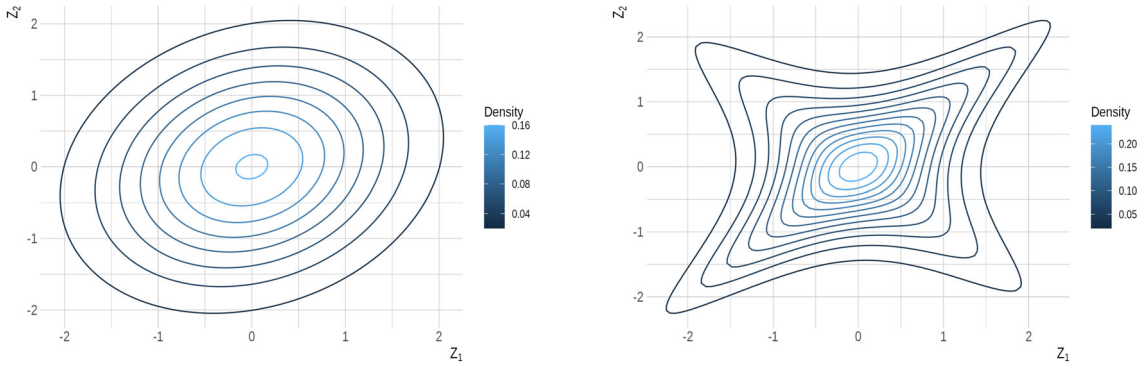
copula with $\rho > 0$ exhibits spikes not only near $(0, 0)$ and $(1, 1)$ but also near $(0, 1)$ and $(1, 0)$. This implies that, relative to the Gaussian copula, the Student-t copula exhibits a higher likelihood of joint extreme movements in opposite directions—even when the correlation parameter is positive.

While empirical copula plots are commonly used to examine the shape of a copula, the associated density functions can diverge near the corners $(0, 0)$ and $(1, 1)$. To address this issue, Joe (2014) recommends transforming the variables to normal scores. Specifically, each variable X_j is transformed as $Z_j = \Phi^{-1} \circ F_j(X_j)$, and we consider the resulting distribution F_N of (Z_1, Z_2), whose marginals are standard normal:

$$F_N(Z_1, Z_2) := C(\Phi(Z_1), \Phi(Z_2))$$

If the copula C is Gaussian, then F_N reduces to a bivariate Gaussian distribution. Therefore, if the plot of (Z_1, Z_2) departs from the elliptical shape of a bivariate Gaussian distribution, this deviation indicates a departure from the Gaussian copula. **Figure 6** shows the normal-score plots for the Gaussian and Student-t copulas. The Gaussian copula yields an elliptical pattern (i.e., the shape of the bivariate Gaussian distribution), whereas the Student-t copula produces a diamond-shaped pattern that extends toward the four corners, reflecting its stronger tail dependence.

Furthermore, we consider upper and lower semi-correlation coefficients as measures of de-



(a) Gaussian copula **(b)** Student-t copula

Figure 6: Density plots of F_N obtained via normal scores. The parameters used here are the same as those employed in **Figure 5**.

pendence based on the transformed variables Z_1 and Z_2:

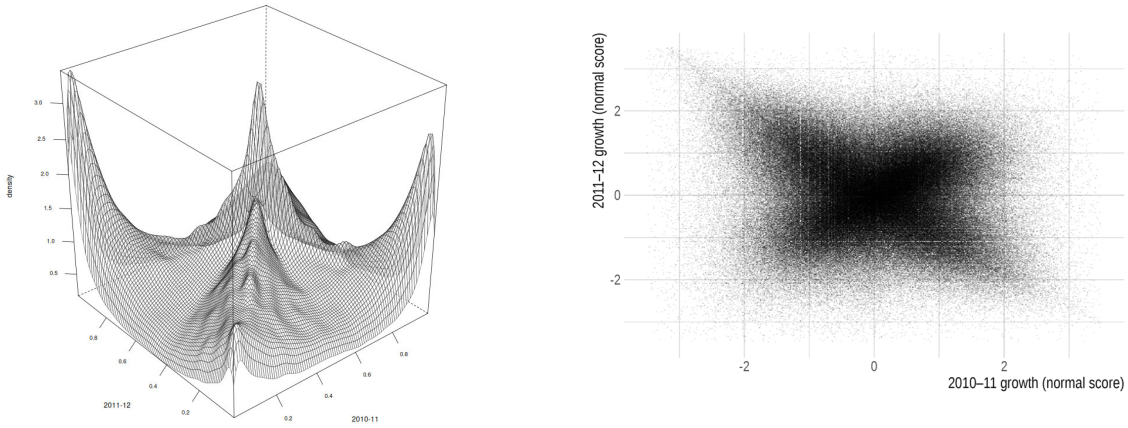
$$\rho_N^+ = \text{Corr}\left[Z_1, Z_2 \mid Z_1 > 0, Z_2 > 0\right],$$
$$\rho_N^- = \text{Corr}\left[Z_1, Z_2 \mid Z_1 < 0, Z_2 < 0\right]$$

These coefficients correspond to the correlation in the upper-right and lower-left quadrants of the scatter plot of the transformed variables. Since the theoretical values of ρ_N^+ and ρ_N^- under the Gaussian copula can be derived analytically, comparing them with the empirical values allows us to evaluate deviations from the Gaussian copula.

The parameters of the copula are estimated using the pseudo maximum likelihood method proposed by Genest et al. (1995). This method consists of two steps: (1) obtaining nonparametric estimates of the marginal distributions, and (2) estimating the parameters of the copula. To examine whether the Student-t copula provides a statistically significantly better fit to the data than the Gaussian copula, we employ Vuong's method (Vuong (1989)). This method compares the log-likelihoods of the two models to assess whether the difference in fit is statistically significant. In our analysis, we compare the Student-t copula with the Gaussian copula.

Firm sales

We apply the above methods to firm sales data using the growth rates for 2010-11 and 2011-12. **Figure 7** presents the empirical copula and the scatter plot of the normal scores. The empirical copula density exhibits spikes at the four corners—$(1,1)$, $(0,0)$, $(1,0)$, and $(0,1)$—which is a characteristic feature of the Student-t copula. Likewise, the normal-scores plot displays a diamond-



(a) Empirical copula

(b) Normal scores

Figure 7: Empirical copula and normal-score plots. Firm sales growth rates for 2010-11 and 2011-12 are used. Panel (a) presents the density estimate of the empirical copula. Panel (b) shows the growth rates transformed into normal scores. The semi-correlation coefficients are $\rho_N^+ = 0.239$ and $\rho_N^- = 0.306$.

shaped pattern rather than an elliptical one. These results suggest that the underlying copula is closer to a Student-t copula than to a Gaussian copula.

This tendency is also reflected in the correlation coefficients. While Spearman's ρ_S is close to zero or slightly negative, as shown in the main text, the semi-correlation coefficients are $\rho_N^+ = 0.239$ and $\rho_N^- = 0.306$. These values are substantially larger than those expected under the Gaussian copula ($\rho_N^+ = \rho_N^- = -0.042$). This indicates that, although there is no strong positive or negative correlation over the entire domain, relatively strong correlations emerge when attention is restricted to the first and third quadrants.

We compare the Gaussian copula and the Student-t copula to determine which provides a better fit to the data. The estimation results for each parametric copula are reported in **Table 1**. Based on the AIC, the Student-t copula with degrees of freedom $\nu = 2$ provides the best fit. We further assess whether the improvement in fit of the Student-t copula over the Gaussian copula is statistically significant using Vuong's method. The test rejects the null hypothesis in favor of the Student-t copula with a p-value of 0.01. This indicates that the Student-t copula provides a significantly better fit to the data than the Gaussian copula.

Finally, we directly examine how well the estimated Student-t copula fits the data. Contour plots of the empirical copula and the estimated copula are presented in **Figure 8**. The Student-t copula provides a close approximation to the empirical copula. We also assess whether the estimated Student-t copula can replicate the observed semi-correlation coefficients of the normal

Parameter estimates and model fit for parametric copulas

Data: Tokyo Shoko Research

copula	shape	maximum log-likelihood	Vuong's test statistic (s.e.)	semi-correlation coefficient
Gaussian	–0.127	4427	-	–0.042
Student-t with df 1	–0.028	-970	-	0.607
Student-t with df 2	–0.070	33496	0.0530(0.00042)	0.375
Student-t with df 3	–0.091	32625	-	0.256
Student-t with df 4	–0.103	29351	-	0.187
Student-t with df 5	–0.110	26392	-	0.142
Student-t with df 10	–0.122	17927	-	0.049

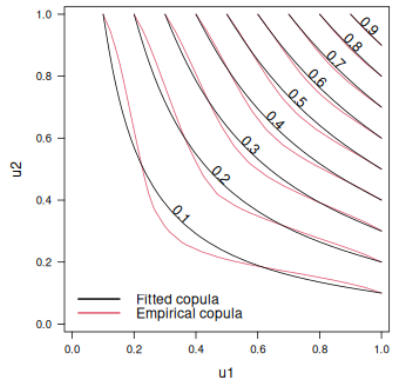
Table 1: Parametric estimates and model fit for parametric copulas. Firm sales growth rates for 2010–11 and 2011–12 are used. Since all copula families considered have the same number of parameters, model selection based on the AIC is equivalent to selection based on the maximum likelihood. The test statistic used in Vuong's method is defined as $\widehat{D} := \log(f^S(y_i;\widehat{\theta}^S)/f^G(y_i;\widehat{\theta}^G))$, where f^S and $\widehat{\theta}^S$ denote the likelihood function and the maximum likelihood estimate for the Student-t copula, and f^G and $\widehat{\theta}^G$ denote those for the Gaussian copula. The last column reports the values of ρ_N^+ and ρ_N^- implied by the estimated copulas.

scores. Under the estimated Student-t copula with degrees of freedom $\nu = 2$, the semi-correlation coefficients are $\rho_N^+ = \rho_N^- = 0.375$. These results further reinforce the finding that the Student-t copula captures the empirical pattern in which stronger dependence emerges when the sample is restricted to the first or third quadrant.

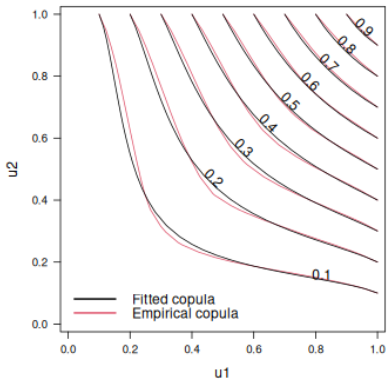
Individual incomes

We apply the above method to individual income growth rates for 2014-15 and 2015-16. **Figure 9** presents the estimated empirical copula and the scatter plot of the normal scores. The empirical copula density exhibits spikes at all four corners, which is characteristic of the Student-t copula. Likewise, the scatter plot of the normal scores forms a diamond shape rather than an ellipse, another feature indicative of the Student-t copula.

Furthermore, this characteristic feature of the Student-t copula is also reflected in the correlation coefficients. While the overall correlation coefficient over the entire domain is close to zero, the semi-correlation coefficients are $\rho_N^+ = 0.392$ and $\rho_N^- = 0.399$. This indicates that, compared with the Gaussian copula, the empirical data exhibit stronger dependence when attention is restricted to the first and third quadrants.

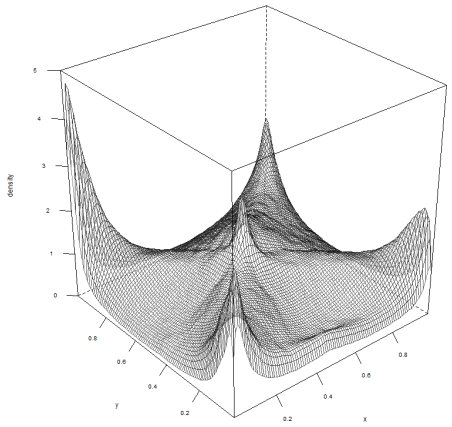


(a) Gaussian copula

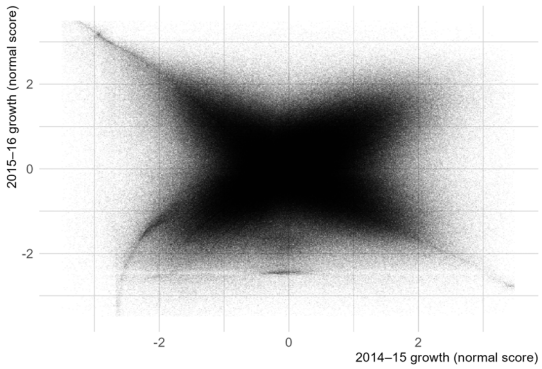


(b) Student-t copula with $\nu = 2$

Figure 8: Comparison between the empirical copula and model-estimated parametric copulas. Firm sales growth rates for 2010-11 and 2011-12 are used. Contour plots of the empirical copula (red lines) and the parametric copulas estimated from the data (black lines) are presented. Panel (a) uses the Gaussian copula, and Panel (b) uses the Student-t copula with $\nu = 2$.



(a) Empirical copula



(b) Normal scores

Figure 9: Empirical copula and plots of normal scores. Individual income growth rates for 2014-15 and 2015-16 are used. Panel (a) displays the density estimate of the empirical copula. Panel (b) presents the growth rates transformed into normal scores. The semi-correlation coefficients are $\rho_N^+ = 0.392$ and $\rho_N^- = 0.399$.

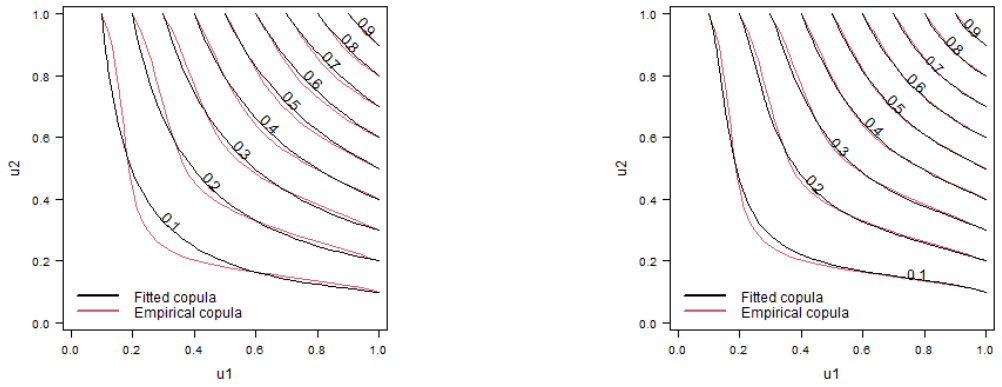
Estimates for parametric copulas Data: Tax return data from National Tax College from 2014 to 2020				
copula	shape	maximum log-likelihood	Vuong's test	semi-correlation coefficient
Gaussian	0.013	8	-	0.005
Student-t with df 1	0.051	4277	-	0.618
Student-t with df 2	0.057	8352	0.0834(0.0009)	0.401
Student-t with df 3	0.055	7366	-	0.295
Student-t with df 4	0.052	6321	-	0.232
Student-t with df 5	0.048	5496	-	0.191
Student-t with df 10	0.037	3317	-	0.102

Table 2: Parametric estimates and model fit for parametric copulas. Individual income growth rates for 2014-15 and 2015-16 are used. Due to computational constraints, we do not use the full dataset; instead, we randomly draw $100,000$ observations from the sample to conduct the estimation. For other details, refer to the description in **Table 1**.

The above characteristics of the copula can also be confirmed by the parameter estimates of the Gaussian copula and the Student-t copula. As shown by the estimation results in **Table 2**, the Student-t copula with degrees of freedom $\nu = 2$ provides the best fit to the data according to the AIC. Moreover, Vuong's method confirms that the difference between the Student-t copula with $\nu = 2$ and the Gaussian copula is statistically significant, with a p-value of 0.01.

Finally, to assess how well the estimated copulas fit the data, **Figure 10** presents contour plots of the empirical copula and the estimated copulas. The Student-t copula with degrees of freedom $\nu = 2$ provides a close approximation to the empirical copula. Moreover, under the estimated Student-t copula with $\nu = 2$, the semi-correlation coefficients are $\rho_N^+ = \rho_N^- = 0.401$. These values are close to those computed from the data, $\rho_N^+ = 0.392$ and $\rho_N^- = 0.399$. These results indicate that the Student-t copula captures the dependence structure between growth rates as reflected in the empirical copula.

The above analysis shows that the statistical properties of firm sales growth rates and individual income growth rates are remarkably similar. Specifically, not only are their marginal growth rate distributions and tail dependencies similar, but the copulas summarizing their dependence structures are also closely aligned. In particular, the results indicate that when the current-period growth rate is positive (denoted by $+$), the next period is more likely to exhibit either positive growth $(+,+)$ or negative growth $(+,-)$, whereas cases in which the next period shows approximately zero growth



(a) Gaussian copula

(b) Student-t copula with $\nu = 2$

Figure 10: Comparison between empirical and model-estimated copulas. Individual income growth rates for 2014-15 and 2015-16 are used. Contour plots of the empirical copula (red lines) and the parametric copula estimated from the data (black lines) are presented. Panel (a) uses the Gaussian copula, while Panel (b) uses the Student-t copula with degrees of freedom $\nu = 2$.

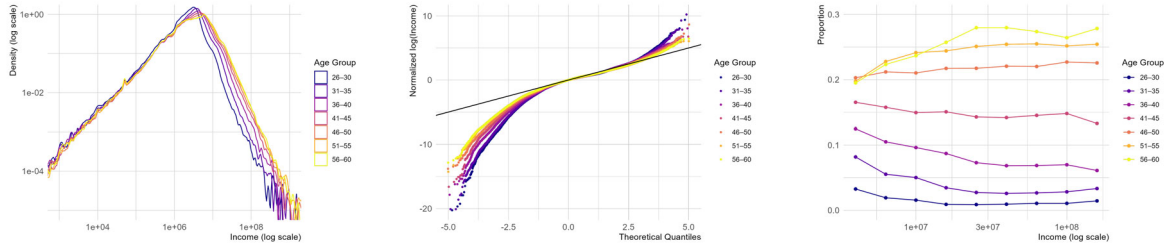
$(+, 0)$ are relatively rare. While a full characterization of the dependence structure lies beyond the scope of this paper, the similarity in the statistical properties of firm sales and individual income growth—despite differences in their underlying economic mechanisms—may help explain why macro-level regularities such as Pareto tails emerge in both settings.

1.3 Robustness to Alternative Income Definitions

Here, we verify that the main findings of our analysis remain robust even when alternative definitions of individual income are used. We repeat the same analysis using two alternative income definitions: net income (after deducting expenses and tax deductions) and wage and salary income. Using these alternative definitions, we find that our main results are replicated:

- Pareto tails emerge in the size distribution by age;
- The share of the tail probability remains nearly constant over a wide range of x, and this pattern holds across all age groups;
- The growth rate distribution exhibits heavier tails than an exponential function;
- Apart from jump episodes, the distribution of period-by-period growth rates, conditional on cumulative growth over n periods, differs little between individuals with high cumulative growth and those without.

We conduct the analysis by defining individual income as net income, i.e., income after



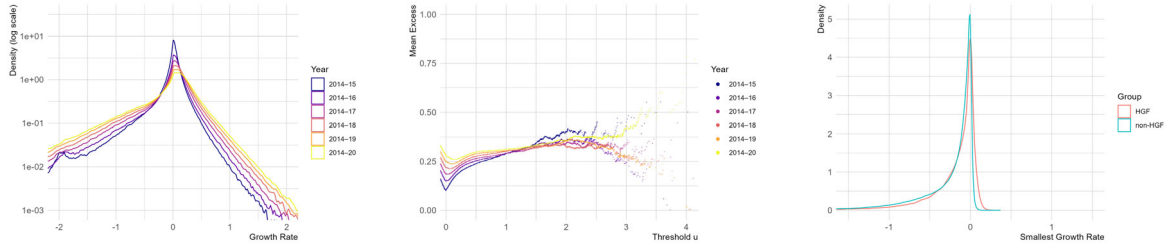
(a) Size distributions **(b)** QQ-plots **(c)** Age-group proportions

Figure 11: Size distributions by age group. We use net income as the measure of individual income. We restrict the sample to individuals aged 26 to 60 and divide them into five-year age groups. Panel (a) presents the estimated density functions of size S for each age group. Panel (b) shows the QQ-plots of the size distributions for each age group. Panel (c) reports the proportion of each age group in the tail probability $\mathbb{P}(S > x)$.

deducting expenses and tax deductions. Results using wage and salary income as an alternative definition are provided in **Figure 13** and **Figure 14**, and the interpretation is analogous to that for the net-income case. **Figure 11**(a) presents the estimated densities of the size distributions across age groups. All age groups exhibit a Pareto tail, and the tail slopes are similar across groups. **Figure 11**(b) shows the QQ-plots of these age-specific size distributions. Consistent with the density estimates, the distributions deviate from the Gaussian distribution, indicating heavy-tailed behavior. **Figure 11**(c) reports the share of each age group in the tail probability $\mathbb{P}(S > x)$. These shares remain stable as x increases, implying that the upper tail is not disproportionately driven by older age groups.

Next, we present the results for growth rates. **Figure 12**(a) shows the density estimates of the n-year growth rate distributions for $n = 1, 2, \ldots, 6$. In all cases, the distributions deviate clearly from the Gaussian distribution. **Figure 12**(b) presents the mean excess function $e(u)$ for the n-year growth rates. Consistent with the density estimates, $e(u)$ is an increasing function of u, indicating that the growth rate distributions have heavier tails than an exponential function. Moreover, the shape of $e(u)$ is largely invariant to n, exhibiting a similar pattern across all horizons, which implies that the distributions are subexponential.

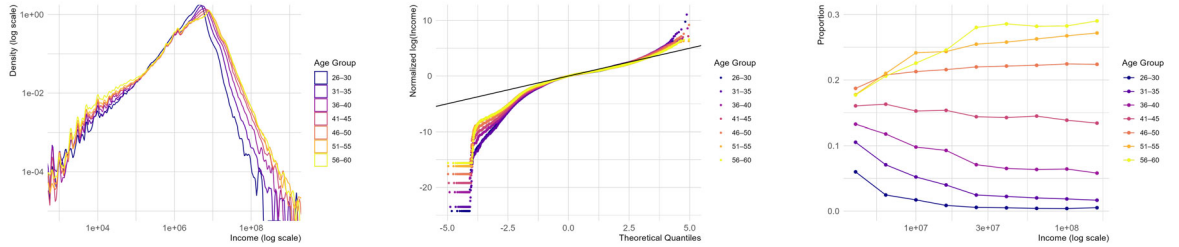
Finally, to examine the characteristic patterns of growth, we compare the histograms of the minimum growth rate within the n-year period for individuals who achieved $\widetilde{S}_n > x$ and those who did not. For $n = 6$, the 97th percentile of $\widetilde{S}_n$ is 0.814. If the distribution were Gaussian, the distribution of growth rates for those who achieved $\widetilde{S}_n > x$ (i.e., the conditional distribution) would be expected to shift to the right by approximately $0.814/6 = 0.136$. The results are shown



(a) Growth rate distribution **(b)** Mean excess function **(c)** Minimum growth rate

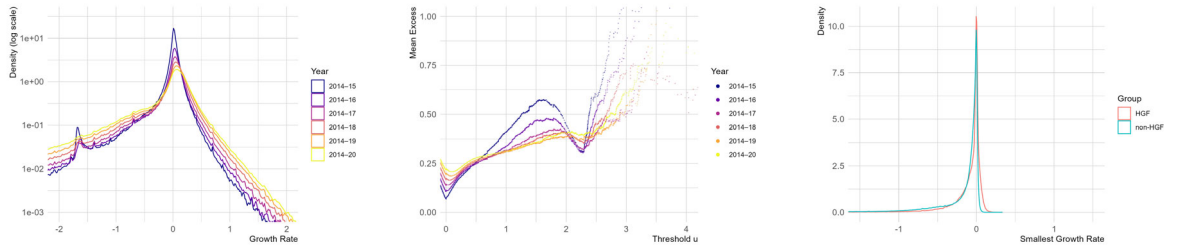
Figure 12: Results based on n-year growth rates We use net income as the measure of individual income. In computing growth rates, we restrict the sample to individuals aged 25 to 60. Panel (a) reports the density estimates of the n-year growth rates ($n = 1, 2, \ldots, 6$), computed using 2014 as the starting year. The vertical axis is shown on a logarithmic scale. Panel (b) presents the mean excess function for the n-year growth rates. Panel (c) compares the histograms of the minimum growth rate within the n-year period for individuals who achieved $\widetilde{S}_n > x$ and those who did not. The 97th percentile of $\widetilde{S}_n$ is 0.814. The medians of the two histograms are -0.0867 and -0.121, respectively.

in **Figure 12**(c). The two histograms are similar: their medians are -0.0867 and -0.121, with a difference of 0.0343. This difference is much smaller than the 0.136 gap expected under a Gaussian distribution, providing evidence in favor of jump-driven growth.



(a) Size distributions **(b)** QQ-plots **(c)** Age-group proportions

Figure 13: Size distributions by age group. We use wage and salary income as the measure of individual income. We restrict the sample to individuals aged 26 to 60 and divide them into five-year age groups. Panel (a) presents the estimated density functions of size S for each age group. Panel (b) shows the QQ-plots of the size distributions for each age group. Panel (c) reports the proportion of each age group in the tail probability $\mathbb{P}(S > x)$.



(a) Growth rate distribution **(b)** Mean excess function **(c)** Minimum growth rate

Figure 14: Results based on n-year growth rates We use wage and salary income as the measure of individual income. In computing growth rates, we restrict the sample to individuals aged 26 to 60. Panel (a) reports the density estimates of the n-year growth rates ($n = 1, 2, \ldots, 6$), computed using 2014 as the starting year. The vertical axis is shown on a logarithmic scale. Panel (b) presents the mean excess function for the n-year growth rates. Panel (c) compares the histograms of the minimum growth rate within the n-year period for individuals who achieved $\widetilde{S}_n > x$ and those who did not. The 97th percentile of $\widetilde{S}_n$ is 0.652. The medians of the two histograms are -0.0237 and -0.0559, respectively.

References

Daunfeldt, S.-O. and Elert, N. (2013). When is Gibrat's law a law? *Small Business Economics*, 41:133–147.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

Joe, H. (2014). *Dependence modeling with copulas*. CRC press.

Lotti, F., Santarelli, E., and Vivarelli, M. (2009). Defending Gibrat's Law as a long-run regularity. *Small Business Economics*, 32(1):31–44.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, pages 307–333.