

# A Schumpeterian Model of Top Income Inequality

Charles I. Jones

Stanford GSB and NBER

Jihee Kim \*

KAIST

June 8, 2017 – Version 3.1

*Forthcoming in the Journal of Political Economy*

## Abstract

Top income inequality rose sharply in the United States over the last 40 years but increased only slightly in economies like France and Japan. Why? This paper explores a model in which heterogeneous entrepreneurs, broadly interpreted, exert effort to generate exponential growth in their incomes. On its own, this force leads to rising inequality. Creative destruction by outside innovators restrains this expansion and induces top incomes to obey a Pareto distribution. The development of the world wide web and a reduction in top tax rates are examples of changes that raise the growth rate of entrepreneurial incomes and therefore increase Pareto inequality. In contrast, policies that stimulate creative destruction reduce top inequality. Examples include research subsidies or a decline in the extent to which incumbent firms can block new innovation. Differences in these considerations across countries and over time may explain the varied patterns of top income inequality that we see in the data.

---

\*For many helpful comments and suggestions, we are grateful to Daron Acemoglu, Philippe Aghion, Ufuk Akcigit, Jess Benhabib, Sebastian Di Tella, Xavier Gabaix, Mike Harrison, Pete Klenow, Ben Moll, Chris Tonetti, Alwyn Young, Gabriel Zucman; seminar participants at the AEA annual meetings, Brown, Chicago Booth, CREI, the Federal Reserve Bank of San Francisco, Groningen, HKUST, IIES Stockholm, Korea University, the NBER EFG group, the NBER Income Distribution group, Princeton, the SED 2015 meeting, Stanford, U.C. Santa Cruz, USC, U. Washington, Yale, Yonsei University, and Zurich; and especially to the editor and four referees.

## 1. Introduction

As documented extensively by Piketty and Saez (2003) and Atkinson, Piketty and Saez (2011), top income inequality — such as the share of income going to the top 1% or top 0.1% of earners — has risen sharply in the United States since around 1980. The pattern in other countries is different and heterogeneous. For example, top inequality rose only slightly in France and Japan. Why? What economic forces explain the varied patterns in top income inequality that we see around the world?

It is well-known that the upper tail of the income distribution follows a power law. One way of thinking about this is to note that income inequality is fractal in nature, as we document more carefully below. In particular, the following questions all have essentially the same answer: What fraction of the income going to the top 10% of earners accrues to the top 1%? What fraction of the income going to the top 1% of earners accrues to the top 0.1%? What fraction of the income going to the top 0.1% of earners accrues to the top 0.01%? The answer to each of these questions — which turns out to be around 40% in the United States today — is a simple function of the parameter that characterizes the power law. Therefore *changes* in top income inequality naturally involve changes in the power law parameter. This paper considers a range of economic explanations for such changes.

The model we develop uses the Pareto-generating mechanisms that researchers like Gabaix (1999) and Luttmer (2007) have used in other contexts. Gabaix studies why the distribution of city populations is Pareto with its key parameter equal to unity. Luttmer studies why the distribution of employment by firms has the same structure. It is worth noting that both cities and firm sizes exhibit substantially more inequality than top incomes (power law inequality for incomes is around 0.5, as we show below, versus around 1 for city populations and firm employment). Our approach therefore is slightly different: why are incomes Pareto and why is Pareto inequality changing over time, rather than why is a power law inequality measure so close to unity.<sup>1</sup>

The basic insight in this literature is that exponential growth, tweaked appropriately, can deliver a Pareto distribution for outcomes. The tweak is needed for the fol-

---

<sup>1</sup>These papers in turn build on a large literature on such mechanisms outside economics. For example, see Reed (2001), Mitzenmacher (2004), and Malevergne, Saichev and Sornette (2013). Gabaix (2009) and Luttmer (2010) have excellent surveys of these mechanisms, written for economists. Benhabib (2014) and Moll (2016) provide very helpful teaching notes.

lowing reason. Suppose that city populations (or incomes or employment by firms) grow exponentially at 2% per year plus some random normally-distributed shock. In this case, the log of population would follow a normal distribution with a variance that grows over time. To keep the distribution from spreading out forever, we need an additional force. For example, a constant probability of death will suffice to render the distribution stationary.

In the model we develop below, researchers create new ideas — new computer chips or manufacturing techniques, but also best-selling books, smartphone apps, financial products, surgical techniques, or even new ways of organizing a law firm. Ideas should be interpreted broadly in this model. The random growth process corresponds to the way entrepreneurs increase their productivity and build market share for their new products. The growth rate of this process is tied to entrepreneurial effort, and anything that raises this effort, resulting in faster growth in entrepreneurial income, will raise top income inequality. The “death rate” in our setup is naturally tied to creative destruction: researchers invent new ideas that make the previous state-of-the-art surgical technique or best-selling iPad application obsolete. A higher rate of creative destruction restrains entrepreneurial income growth and results in lower top income inequality. In this way, the interplay between existing entrepreneurs growing their profits and the creative destruction associated with new ideas determines top income inequality.

This paper proceeds as follows. Section 2 presents some basic facts of top income inequality, emphasizing that the rise in the United States is accurately characterized by a change in the power law parameter. Section 3 considers a brief toy model to illustrate the main mechanism in the paper. The next two sections then develop the model, first with an exogenous allocation of labor to research and then more fully with an endogenous allocation of labor. Section 6 uses the IRS public use panel of tax returns as well as data from the Social Security Administration to estimate several of the key parameters of the model, illustrating that the mechanism is economically significant. Section 7 highlights the important role played by transition dynamics in this framework.

### **1.1. The existing literature**

A number of other recent papers contribute to our understanding of the dynamics of top income inequality. Piketty, Saez and Stantcheva (2014) and Rothschild and Scheuer

(2016) explore the possibility that the decline in top tax rates has led to a rise in rent seeking, leading top inequality to increase. Philippon and Reshef (2012) focus explicitly on finance and the extent to which rising rents in that sector can explain rising inequality; see also Bell and Van Reenen (2014). Bakija, Cole and Heim (2010) and Kaplan and Rauh (2010) note that the rise in top inequality occurs across a range of occupations; it is not just focused in finance or among CEOs, for example, but includes doctors and lawyers and star athletes as well. Benabou and Tirole (2016) discuss how competition for the most talented workers can result in a “bonus culture” with excessive incentives for the highly skilled. Haskel, Lawrence, Leamer and Slaughter (2012) suggest that globalization may have raised the returns to superstars via a Rosen (1981) mechanism. Aghion, Akcigit, Bergeaud, Blundell and Hemous (2015) show that innovation and top income inequality are positively correlated within U.S. states and across U.S. commuting zones; we discuss how this finding might be reconciled with our framework in a later section. There is of course a much larger literature on changes in income inequality throughout the distribution. Katz and Autor (1999) provide a general overview, while Autor, Katz and Kearney (2006), Gordon and Dew-Becker (2008), and Acemoglu and Autor (2011) provide more recent updates. Banerjee and Newman (1993) and Galor and Zeira (1993) study the interactions between economic growth and income inequality.

Lucas and Moll (2014) explore a model of human capital and the sharing of ideas that gives rise to endogenous growth. Perla and Tonetti (2014) study a similar mechanism in the context of technology adoption by firms. These papers show that if the initial distribution of human capital or firm productivity has a Pareto upper tail, then the ergodic distribution also inherits this property and the model can lead to endogenous growth, a result reminiscent of Kortum (1997). The Pareto distribution, then, is more of an “input” in these models rather than an outcome.<sup>2</sup>

The most closely-related papers to this one are Levy (2003), Benhabib, Bisin and Zhu (2011), Nirei (2009), Moll (2012), Piketty and Saez (2013), Piketty and Zucman (2014), Toda (2014), Benhabib and Bisin (2016), Hubmer, Krusell and Smith (2016), and Nirei and Aoki (2016). These papers study economic mechanisms that generate endoge-

---

<sup>2</sup>Luttmer (2014) extends this line of work in an attempt to get endogenous growth without assuming a Pareto distribution and also considers implications for inequality. Koenig, Lorenz and Zilibotti (2016) derive a Zipf distribution in the upper tail for firm productivity in an endogenous growth setting.

nously a Pareto distribution for wealth, and therefore for capital income. The mechanism responsible for random growth in these papers is either the asset accumulation equation (which naturally follows a random walk when viewed in partial equilibrium) or the capital accumulation equation in a neoclassical growth model. Geerolf (2016) connects both top income inequality and firm size inequality in a Garicano (2000)-style model of hierarchies, building on the assignment model of Gabaix and Landier (2008).<sup>3</sup>

The present paper differs most directly from much of the previous literature by focusing explicitly on labor income and entrepreneurial income.<sup>4</sup> Since much of the rise in top income inequality in the United States is due to labor income — e.g. see Piketty and Saez (2003) — this focus is appropriate. Our paper also differs by embedding the discussion of Pareto inequality in a model with endogenous growth, allowing us to study the potential tradeoffs between growth and inequality.

Finally, Gabaix, Lasry, Lions and Moll (2016) show that the basic random growth model has trouble matching the transition dynamics of top income inequality. Building on Luttmer (2011), they suggest that a model with heterogeneous mean growth rates for top earners will be more successful, and we incorporate their valuable insights, as discussed further below.

## 2. Some Basic Facts

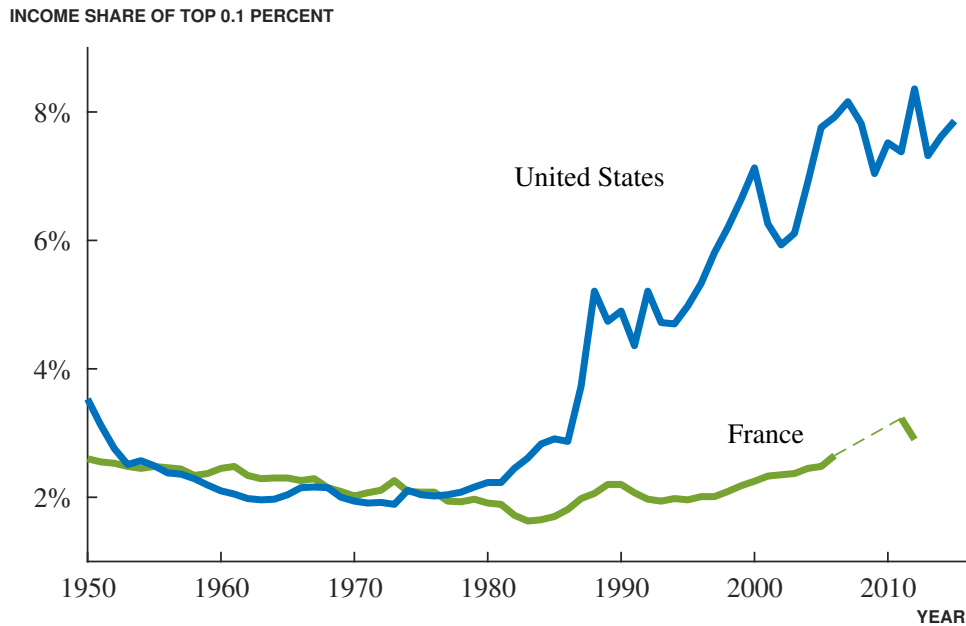
Figures 1 and 2 show some of the key facts about top income inequality that have been documented by Piketty and Saez (2003) and Atkinson, Piketty and Saez (2011). For example, the first figure shows the large increase in top inequality for the United States since 1980, compared to the relative stability of inequality in France.

Figure 2 shows the dynamics of top income inequality for a range of countries, illustrating that the U.S. and France are large countries close to the two extremes. The horizontal axis shows the share of aggregate income earned by the top 1%, averaged between 1980 and 1982, while the vertical axis shows the same share for 2006–2008.

---

<sup>3</sup>The mechanism by which Geerolf (2016) generates the Pareto distribution is different from the random growth mechanism in most of these other papers. Instead, Geerolf exploits the fact that power functions (like Cobb-Douglas production functions) are closely related to Pareto distributions and that the first-order Taylor expansion of a function with  $f(0) = 0$  around zero is itself a power function (a linear one).

<sup>4</sup>Classic papers on generating Pareto distributions for income include Champernowne (1953), Simon (1955), and Mandelbrot (1960).

**Figure 1: Top Income Inequality in the United States and France**

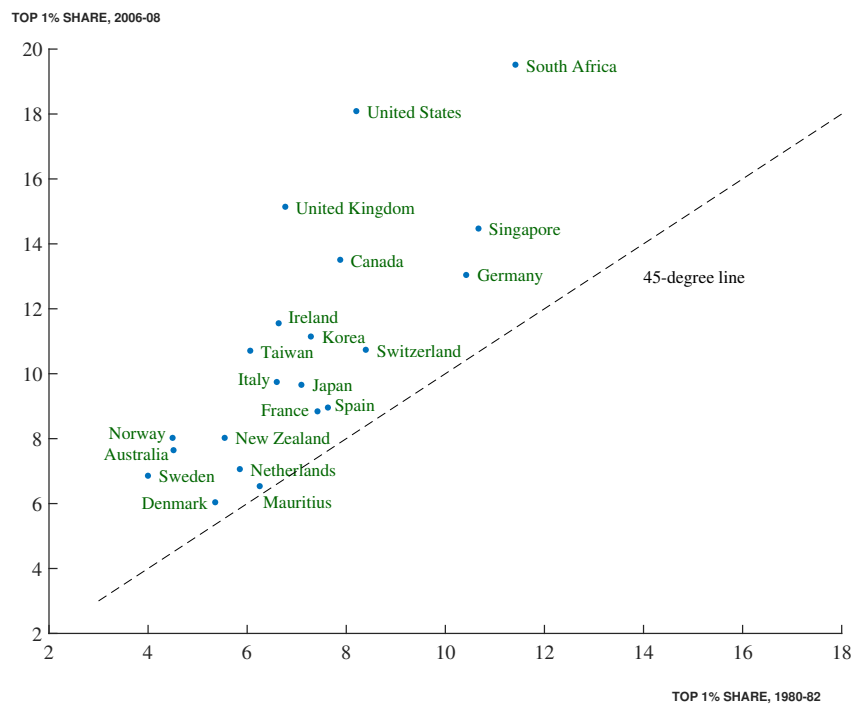
Source: Alvaredo, Atkinson, Piketty, Saez and Zucman (2016), World Wealth and Income Database. Includes interest and dividends but not capital gains.

All the economies for which we have data lie above the 45-degree line: that is, top income inequality has risen everywhere. The rise is the largest in the United States, South Africa, and Norway, but substantial increases are also seen elsewhere, such as in Ireland, Portugal, Singapore, Italy, and Sweden. Japan and France exhibit smaller but still noticeable increases. For example, the top 1% share in France rises from 7.4% to 9.0%.

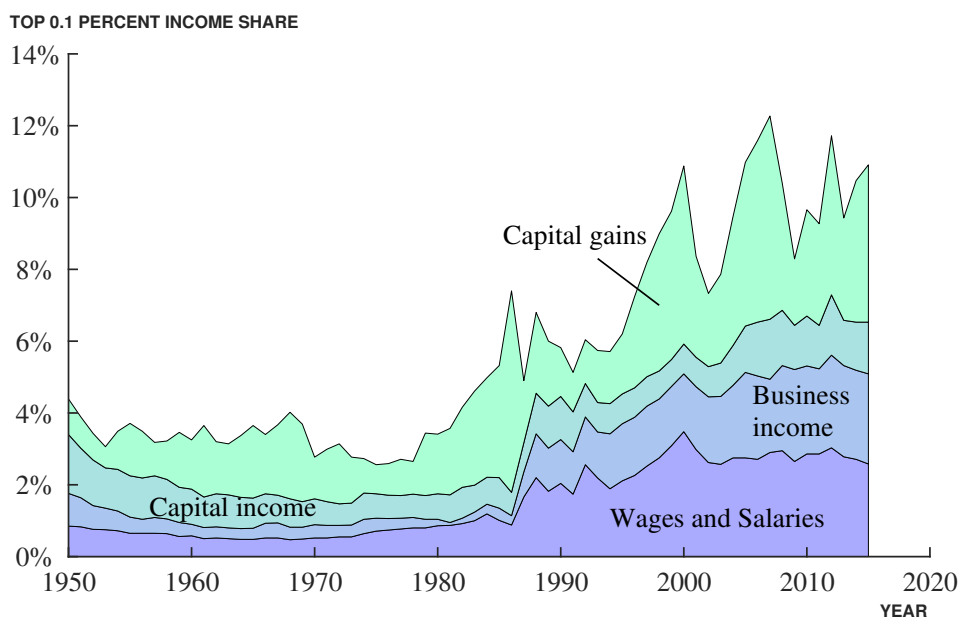
## 2.1. The Role of Labor Income

As discussed by Atkinson, Piketty and Saez (2011) and Piketty, Saez and Zucman (2016), a substantial part of the rise in U.S. top income inequality represents a rise in labor income inequality, particularly if one includes “business income” (i.e. profits from sole proprietorships, partnerships and S-corporations) in the labor income category. Given our focus on entrepreneurs, our ideal income measure would always include entrepreneurial income. From now on, when we speak of “labor income,” we will include entrepreneurial income as well. Figure 3 shows an updated version of a graph

Figure 2: Top Income Inequality around the World, 1980–82 and 2006–2008



Note: Top income inequality has increased since 1980 in every country for which we have data. The size of the increase varies substantially, however. Source: Alvaredo, Atkinson, Piketty, Saez and Zucman (2016), World Wealth and Income Database.

**Figure 3: The Composition of the Top 0.1 Percent Income Share**

Note: The figure shows the composition of the top 0.1 percent income share. Source: These data are taken from the “data-Fig4B” tab of the June 2016 update of the spreadsheet appendix to Piketty and Saez (2003).

from Piketty and Saez (2003) for the period since 1950, supporting the observation that much of the rise in top income inequality is associated with this broad concept of labor income.

Because the model in this paper is based on labor income as opposed to capital income, documenting the Pareto nature of labor income inequality in particular is also important. It is well known, dating back to Pareto (1896), that the top portion of the income distribution can be characterized by a power law. That is, at high levels, the income distribution is approximately Pareto. In particular, if  $Y$  is a random variable denoting incomes, then, at least above some high level (i.e. for  $Y \geq y_0$ )

$$\Pr[Y > y] = \left(\frac{y}{y_0}\right)^{-\xi}, \quad (1)$$

where  $\xi$  is called the “power law exponent.”

Saez (2001) shows that wage and salary income from U.S. income tax records in the early 1990s is well-described by a Pareto distribution. Figure 4 replicates his analysis



for 1980 and 2005 for a broader income concept that includes both wage and salary income as well as entrepreneurial income from businesses. In particular, the figures plot mean income above some threshold as a ratio to the threshold itself. If income obeys a Pareto distribution like that in (1), then this ratio should equal the constant  $\frac{\xi}{\xi-1}$ , *regardless of the threshold*. That is, as we move to higher and higher income thresholds, the ratio of average income above the threshold to the threshold itself should remain constant.<sup>5</sup> Figure 4 shows that this property holds reasonably well in 1980 and 2005, and also illustrates that the ratio has risen substantially over this period, reflecting the rise in top income inequality.

## 2.2. Fractal Inequality and the Pareto Distribution

There is a tight connection between Pareto distributions and the “top  $x$  percent” shares that are the focus of Piketty and Saez (2003) and others. To see this, let  $\tilde{S}(p)$  denote the share of income going to the top  $p$  percentiles. For the Pareto distribution defined in equation (1) above, this share is given by  $(p/100)^{1-1/\xi}$ . A larger power-law exponent,  $\xi$ , is associated with lower top income inequality. It is therefore convenient to define the “power-law inequality” exponent as

$$\eta \equiv \frac{1}{\xi} \quad (2)$$

so that

$$\tilde{S}(p) = \left(\frac{100}{p}\right)^{\eta-1}. \quad (3)$$

For example, if  $\eta = 1/2$ , then the share of income going to the top 1% is  $100^{-1/2} = .10$ . However, if  $\eta = 3/4$ , the share going to the top 1% rises sharply to  $100^{-1/4} \approx 0.32$ .

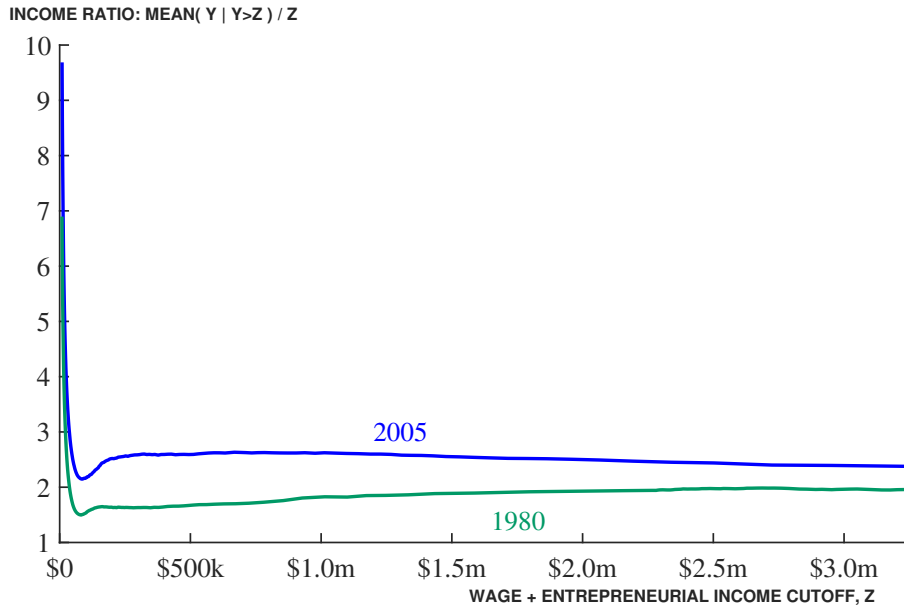
An important property of Pareto distributions is that they exhibit a fractal pattern of top inequality. To see this, let  $S(a) = \tilde{S}(a)/\tilde{S}(10a)$  denote the fraction of income earned by the top  $10 \times a$  percent of people that actually goes to the top  $a$  percent. For example,  $S(1)$  is the fraction of income going to the top 10% that actually accrues to the top 1%, and  $S(.1)$  is the fraction of income going to the top 1% that actually goes to the top 1 in 1000 earners. Under a Pareto distribution,

$$S(a) = 10^{\eta-1}. \quad (4)$$

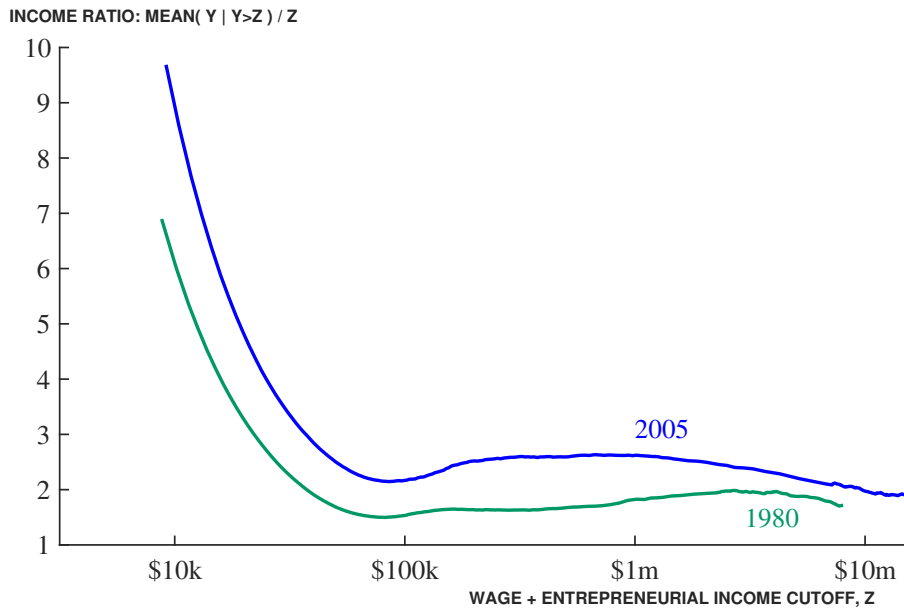
---

<sup>5</sup>This follows easily from the fact that the mean of a Pareto distribution is  $\frac{\xi y_0}{\xi-1}$  and that the conditional mean just scales up with the threshold.

Figure 4: The Pareto Nature of Labor Income (Broadly Defined)



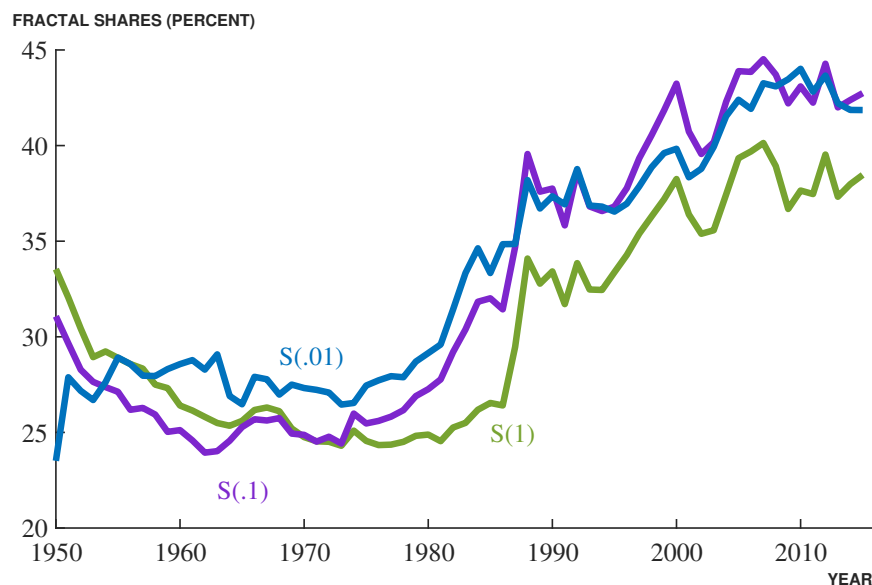
(a) Linear scale, up to \$3 million



(b) Log scale

Note: The figures plot the ratio of average wage plus entrepreneurial income above some threshold  $z$  to the threshold itself. For a Pareto distribution with Pareto inequality parameter  $\eta$ , this ratio equals  $1/(1 - \eta)$ . Saez (2001) produced similar graphs for 1992 and 1993 for wage and salary income using the IRS public use tax files available from the NBER at [www.nber.org/taxsim-notes.html](http://www.nber.org/taxsim-notes.html). The figures here replicate these results using the same data source and a broader income concept for 1980 and 2005.

Figure 5: Fractal Inequality of U.S. Income



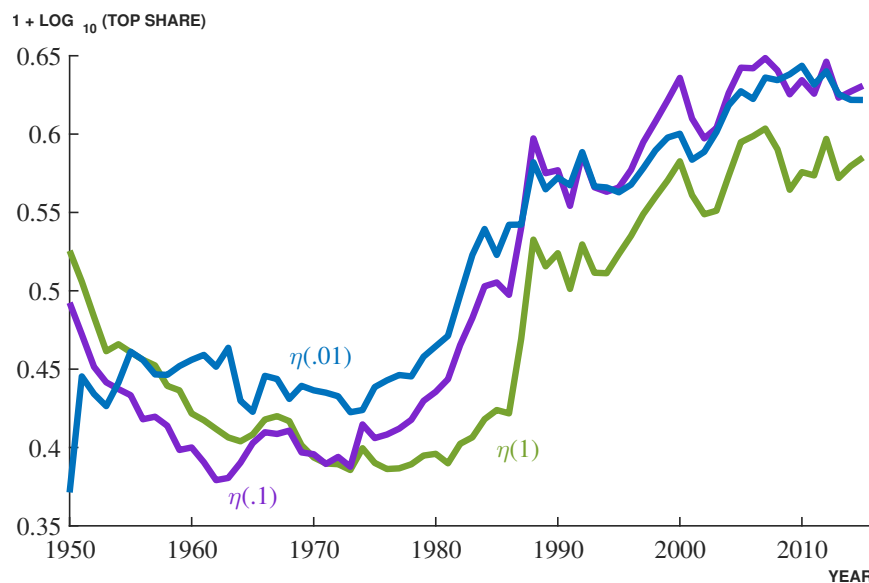
Note:  $S(a)$  denotes the fraction of income going to the top 10a percent of earners that actually goes to the top  $a$  percent. For example,  $S(1)$  is the share of the Top 10%'s income that accrues to the Top 1%. Source: Alvaredo, Atkinson, Piketty, Saez and Zucman (2016), World Wealth and Income Database. Includes interest and dividends but not capital gains.

Notice that this last result holds for all values of  $a$ , or at least for all values for which income follows a Pareto distribution. This means that top income inequality obeys a *fractal* pattern: the fraction of the Top 10 percent's income going to the Top 1 percent is the same as the fraction of the Top 1 percent's income going to the Top 0.1 percent, which is the same as the fraction of the Top 0.1 percent's income going to the Top 0.01 percent.

Not surprisingly, top income inequality is well-characterized by this fractal pattern, as shown in Figure 5.<sup>6</sup> At the very top, the fractal prediction holds remarkably well, and  $S(.01) \approx S(.1) \approx S(1)$ . Prior to 1980, the fractal shares are around 25 percent: one quarter of the Top  $X$  percent's income goes to the Top  $X/10$  percent. By the end of the sample in 2015, this fractal share is closer to 40 percent.

The rise in fractal inequality shown in Figure 5 can be related directly to the power-law inequality exponent using equation (4) and taking logs. The corresponding Pareto

<sup>6</sup>Others have noticed this before. For example, see Aluatiun.wordpress.com (2011).

Figure 6: The Power-Law Inequality Exponent  $\eta$ , United States

Note:  $\eta(a)$  is the power law inequality exponent obtained from the fractal inequality income shares in Figure 5 assuming a Pareto distribution. See equation (4) in the text.

inequality measures are shown in Figure 6. This figure gives us the quantitative guidance that we need for theory. The goal is to build a model that explains why top incomes are Pareto and that generates a Pareto exponent that rises from around 0.4 to around 0.6 for the United States but by much less in France and other countries.

### 2.3. Skill-Biased Technical Change?

Before moving on, it is worth pausing to consider a simple, familiar explanation in order to understand why it is incomplete: skill-biased technical change. For example, if the distribution of skill is Pareto and there is a rise in the return to skill, does this raise top inequality? The answer is no, and it is instructive to see why.

Suppose the economy consists of a large number of homogeneous low-skilled workers with fixed income  $\bar{y}$ . High-skilled people, in contrast, are heterogeneous: income for highly-skilled person  $i$  is  $y_i = \bar{w}x_i^\alpha$ , where  $x_i$  is person  $i$ 's skill and  $\bar{w}$  is the wage per unit of skill (ignore  $\alpha$  for now). If the distribution of skill across people is Pareto with inequality parameter  $\eta_x$ , then the income distribution at the top will be Pareto

with inequality parameter  $\eta_y = \alpha\eta_x$ . That is, if  $\Pr[x_i > x] = x^{-1/\eta_x}$ , then  $\Pr[y_i > y] = (\frac{y}{\bar{w}})^{-1/\eta_y}$ . An increase in  $\bar{w}$  — a skill-biased technical change that increases the return to skill — shifts the Pareto distribution right, increasing the gap between high-skilled and low-skilled workers. But it but does not change Pareto inequality  $\eta_y$ : a simple story of skill-biased technical change is not enough.

Notice that if the exponent  $\alpha$  were to rise over time, this would lead to a rise in Pareto inequality. But this requires something more than just a simple skill-biased technical change story. Moreover, even a rising  $\alpha$  would leave unexplained the question of why the underlying skill distribution is Pareto. The remainder of this paper can be seen as explaining why  $x$  is Pareto and what economic forces might cause  $\alpha$  to change over time or differ across countries.<sup>7</sup>

## 2.4. Summary

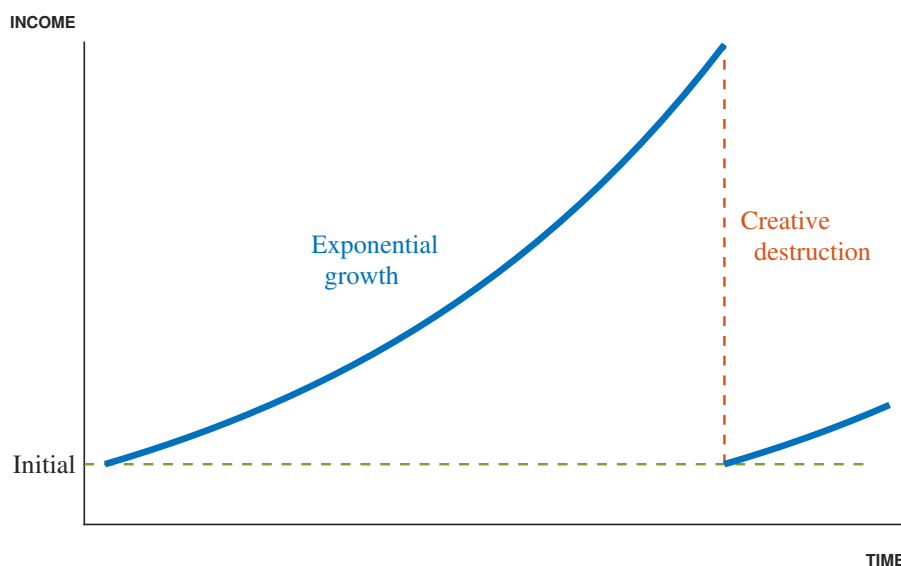
Here then are the basic facts related to top income inequality that we'd like to be able to explain. Between 1960 and 1980, top income inequality was relatively low and stable in both the United States and France. Since around 1980, however, top inequality has increased sharply in countries like the United States, Norway, and Portugal, while it has increased only slightly in others, including France and Japan. Finally, labor income is well-described by a Pareto distribution, and rising top income inequality is to a great extent associated with rising labor income inequality. Changing top income inequality corresponds to a change in the power-law inequality exponent, and the U.S. data suggest a rise from about 0.4 in the 1970s to about 0.6 by 2015. The remainder of this paper develops and analyzes a model to help us understand these facts.

## 3. A Simple Model of Top Income Inequality

It is well-known that exponential growth and Pareto distributions are tightly linked, and this link is at the heart of the main mechanism in this paper. To illustrate this point in the clearest way, we begin with a brief toy model, illustrated graphically in Figure 7.<sup>8</sup>

<sup>7</sup>Gabaix, Lasry, Lions and Moll (2016) explore an alternative approach they call “scale dependence” in an extension of Gabaix and Landier (2008), viewing a rise in  $\alpha$  as a convexification in the returns to skill.

<sup>8</sup>See Gabaix (2009) for a similar stylized model, which Gabaix attributes to Steindl (1965), applied to Zipf's Law for cities. Benhabib (2014) traces the history of Pareto-generating mechanisms and attributes the earliest instance of a simple model like that outlined here to Cantelli (1921).

Figure 7: Basic Mechanism: Exponential growth with death  $\Rightarrow$  Pareto

When a person first becomes a top earner (“entrepreneur”), she earns income  $y_0$ . As long as she remains a top earner, her income grows over time at rate  $\mu$ , so the income of a person who’s been a top earner for  $x$  years — think of  $x$  as “entrepreneurial experience” — is  $y(x) = y_0 e^{\mu x}$ .

People do not remain top earners forever. Instead, there is a constant probability  $\delta$  per unit of time (more formally, a Poisson process) that an existing entrepreneur is displaced. If this occurs, the existing entrepreneur drops out of the top, becoming a “normal” worker, and is replaced by a new entrepreneur who starts over at the bottom of the ladder and earns  $y_0$ .

What fraction of people in this economy have income greater than some level  $y$ ? The answer is simply the fraction of people who have been entrepreneurs for at least  $x(y)$  years, where

$$x(y) = \frac{1}{\mu} \log \left( \frac{y}{y_0} \right). \quad (5)$$

With a Poisson replacement process, it is well-known that the distribution of experience for a given individual follows an exponential distribution, i.e.  $\Pr[\text{Experience} > x] = e^{-\delta x}$ . Let’s take for granted that the stationary distribution of experience across a population of entrepreneurs is this same exponential distribution; this is shown more

formally in Appendix A. Then, the remainder of the argument is straightforward:

$$\begin{aligned} \Pr [\text{Income} > y] &= \Pr [\text{Experience} > x(y)] \\ &= e^{-\delta x(y)} \\ &= \left( \frac{y}{y_0} \right)^{-\frac{\delta}{\mu}} \end{aligned} \tag{6}$$

which is a Pareto distribution!

Pareto inequality in this model is then given by the inverse of the exponent above:

$$\eta_y = \frac{\mu}{\delta}. \tag{7}$$

Top income inequality can therefore change for two reasons. First, an increase in the growth rate of top earners,  $\mu$ , will widen the distribution: the higher is the growth rate, the higher is the ratio of top incomes to the income of a new entrepreneur. Second, an increase in the “death rate”  $\delta$  will reduce top inequality, as entrepreneurs have less time during which to build their advantage.

The logic of the simple model provides useful intuition about *why* the Pareto result emerges. First, in equation (5), the log of income is proportional to experience. This is a common and natural assumption. For example, in models where income grows exponentially over time, income and time are related in this way. Or in labor economics, log income and experience are linked in Mincer-style equations. Next, the distribution of experience is exponential. This is a property of a Poisson process with a constant arrival rate. Putting these two pieces together, log income has an exponential distribution. But this is just another way of saying that income has a Pareto distribution. More briefly, *exponential growth occurring over an exponentially-distributed amount of time delivers a Pareto distribution.*

What are the economic determinants of  $\mu$  and  $\delta$ , and why might they change over time or differ across countries? Answering these questions is one of the goals of the full model that we develop next.

## 4. A Schumpeterian Model of Top Income Inequality

The simple model illustrates in a reduced-form fashion the main mechanism at work in this paper. In our full model, we develop a theory in which the economic determinants of  $\mu$  and  $\delta$  are apparent, and we consider what changes in the economy could be responsible for the range of patterns we see in top income inequality across countries. Entrepreneurs undertake research to improve the productivity of their existing firms and increase their incomes. This process is assumed to be stochastic, which allows us to better match up the model with micro data on top incomes. At the same time, the death rate is endogenized by tying it to the process of creative destruction by outside research in a Schumpeterian growth model. The setup captures some of the key features of top incomes: the importance of entrepreneurial effort, the role of creative destruction, and the centrality of “luck” as some people succeed beyond their wildest dreams while others fail.

### 4.1. Entrepreneurs

An entrepreneur is a monopolist with the exclusive right to sell a particular variety, in competition with other varieties. We interpret this statement quite broadly. For example, think of a Silicon Valley startup, an author of a new book, a new rock band, an athlete just making it to the pro’s, or a doctor who has invented a new surgical technique. Moreover, we do not associate a single variety with a single firm — the entrepreneur could be a middle manager in a large company who has made some breakthrough and earned a promotion.

When a new variety is first introduced, it has a low quality/productivity, denoted by  $x$ , which can be thought of as the stock of the new incumbent’s innovation. The entrepreneur then expends effort (“incumbent research”) to improve  $x$ . We explain later how  $x$  affects firm production and profitability. For the moment, it is sufficient to assume that the entrepreneur’s income is proportional to  $x$ , as it will be in general equilibrium. Note that we are recycling notation: this  $x$  does not measure experience as it did in the simple model of Section 2 (though it is related).

Given an  $x$ , the entrepreneur maximizes the expected present discounted value of



flow utility,  $u(c, \ell) = \log c_t + \beta \log \ell_t$ , subject to the following constraints:

$$c_t = \psi_t x_t \quad (8)$$

$$e_t + \ell_t + \tau = 1 \quad (9)$$

$$dx_t = \mu(e_t)x_t dt + \sigma x_t dB_t \quad (10)$$

$$\mu(e) = \phi e \quad (11)$$

For simplicity, we do not allow entrepreneurs to smooth their consumption and instead assume that consumption equals income, which in turn is proportional to the entrepreneur's productivity  $x$ . The factor of proportionality,  $\psi_t$ , is exogenous to the individual's actions and is the same for all entrepreneurs; it is endogenized in general equilibrium shortly. The entrepreneur has one unit of time each period, which can be used for effort  $e$  or leisure  $\ell$  or it can be wasted, in amount  $\tau$ . This could correspond to time spent addressing government regulations and bureaucratic red tape, for example.

Equation (10) describes how effort improves the entrepreneur's productivity  $x$  through a geometric Brownian motion. The average growth rate of productivity is  $\mu(e) = \phi e$ , where  $\phi$  is a technological parameter converting effort into growth.  $dB_t$  denotes the standard normal increment to the Brownian motion. This equation can be viewed as a stochastic version of an Aghion-Howitt research equation for incumbents. Alternatively, it is also reminiscent of the human capital accumulation process in Lucas (1988). Interestingly, as we discuss below, the inherent linearity of this equation does *not* give rise to long-run growth. Instead the model will deliver a stationary distribution of  $x$  across heterogeneous entrepreneurs.

Finally, there is a Poisson creative destruction process by which the entrepreneur loses her monopoly position and is replaced by a new entrepreneur. This occurs at the (endogenized in general equilibrium) rate  $\delta$ . In addition, there is an exogenous piece to destruction as well, which occurs at a constant rate  $\bar{\delta}$ .

The Bellman equation for the entrepreneur is

$$\rho V(x_t, t) = \max_e \log \psi_t + \log x_t + \beta \log(\Omega - e_t) + \frac{\mathbb{E}[dV(x_t, t)]}{dt} + (\delta + \bar{\delta})(V^w(t) - V(x_t, t)) \quad (12)$$

subject to (10), where  $\Omega \equiv 1 - \tau$  and  $\frac{\mathbb{E}[dV(x_t, t)]}{dt}$  is short-hand for the Ito calculus terms, i.e.  $\frac{\mathbb{E}[dV(x_t, t)]}{dt} \equiv \mu(e_t)x_t V_x(x_t, t) + \frac{1}{2}\sigma^2 x_t^2 V_{xx}(x_t, t) + V_t(x_t, t)$ .  $V(x, t)$  is the expected utility of an entrepreneur with quality  $x$  and rate of time preference  $\rho$ . The flow of the value function depends on the “dividend” of utility from consumption and leisure, the “capital gain” associated with the expected change in the value function, and the possible loss associated with creative destruction, in which case the entrepreneur becomes a worker with expected utility  $V^w$ .

The first key result describes an existing entrepreneur’s choice of research effort. (Proofs of all propositions are given in Appendix D).

**Proposition 1 (Entrepreneurial Effort):** *Entrepreneurial effort solves the Bellman problem in equation (12) and along the balanced growth path is given by*

$$e^* = 1 - \tau - \frac{1}{\phi} \cdot \beta(\rho + \delta + \bar{\delta}). \quad (13)$$

This proposition implies that entrepreneurial effort is an increasing function of the technology parameter  $\phi$  but decreases whenever  $\tau$ ,  $\beta$ ,  $\rho$ ,  $\delta$ , or  $\bar{\delta}$  are higher.

## 4.2. The Stationary Distribution of Entrepreneurial Income

Assume there is a continuum of entrepreneurs of unit measure at any point in time. The initial distribution of entrepreneurial productivity  $x$  is given by  $f_0(x)$ , and the distribution evolves according to the geometric Brownian motion process given above. Entrepreneurs can be displaced in one of two ways. Endogenous creative destruction (the Poisson process at rate  $\delta$ ) leads to replacement by a new entrepreneur who inherits the existing quality  $x$ ; hence the distribution is not mechanically altered by this form of destruction. In large part, this is a simplifying assumption; otherwise one has to worry about the extent to which the step up the quality ladder by a new entrepreneur trades off with the higher  $x$  that the previous entrepreneur has accumulated. We treat the exogenous destruction at rate  $\bar{\delta}$  differently. In this case, existing entrepreneurs are replaced by new “young” entrepreneurs with a given initial productivity  $x_0$ . Exogenous destruction could correspond to the actual death or retirement of existing entrepreneurs, or it could stand in for policy actions by the government: one form of misallocation may be that the government appropriates the patent from an existing

entrepreneur and gives it to a new favored individual. Finally, it simplifies the analysis to assume that  $x_0$  is the minimum possible productivity: there is a “reflecting barrier” at  $x_0$ ; this assumption could be relaxed.

We’ve set up the stochastic process for  $x$  so that we can apply a well-known result in the literature for generating Pareto distributions.<sup>9</sup> If a variable follows a Brownian motion, like  $x$  above, the density of the distribution  $f(x, t)$  satisfies a Kolmogorov forward equation:<sup>10</sup>

$$\frac{\partial f(x, t)}{\partial t} = -\bar{\delta}f(x, t) - \frac{\partial}{\partial x} [\mu(e^*)xf(x, t)] + \frac{1}{2} \cdot \frac{\partial^2}{\partial x^2} [\sigma^2 x^2 f(x, t)] \quad (14)$$

If a stationary distribution,  $\lim_{t \rightarrow \infty} f(x, t) = f(x)$  exists, it therefore satisfies

$$0 = -\bar{\delta}f(x) - \frac{d}{dx} [\mu(e^*)xf(x)] + \frac{1}{2} \cdot \frac{d^2}{dx^2} [\sigma^2 x^2 f(x)] \quad (15)$$

Guessing that the Pareto form  $f(x) = Cx^{-\xi-1}$  solves this differential equation, one obtains the following result:

**Proposition 2 (The Pareto Income Distribution):** *The stationary distribution of (normalized) entrepreneurial income is given by*

$$F(x) = 1 - \left( \frac{x}{x_0} \right)^{-\xi^*} \quad (16)$$

where

$$\xi^* = -\frac{\tilde{\mu}^*}{\sigma^2} + \sqrt{\left( \frac{\tilde{\mu}^*}{\sigma^2} \right)^2 + \frac{2\bar{\delta}}{\sigma^2}} \quad (17)$$

and  $\tilde{\mu}^* \equiv \mu(e^*) - \frac{1}{2}\sigma^2 = \phi(1 - \tau) - \beta(\rho + \delta^* + \bar{\delta}) - \frac{1}{2}\sigma^2$ . Power-law inequality is therefore given by  $\eta^* \equiv 1/\xi^*$ .

The word “normalized” in the proposition refers to the fact that the income of an entrepreneur with productivity  $x$  is  $\psi_t x$ . Aggregate growth occurs via the  $\psi_t$  term, as discussed when we turn to general equilibrium, while the distribution of  $x$  is what is stationary. Finally, we put a “star” on  $\delta$  as a reminder that this value is determined in

<sup>9</sup>For more detailed discussion, see Reed (2001), Mitzenmacher (2004), Gabaix (2009), and Luttmer (2010). Malevergne, Saichev and Sornette (2013) is closest to the present setup.

<sup>10</sup>This is the stochastic generalization of an equation like (A2) in the appendix, related to the simple model at the start of the paper.

general equilibrium as well.

*Comparative statics:* Taking  $\delta^*$  as exogenous for the moment, the comparative static results are as follows: power-law inequality,  $\eta^*$ , increases if effort is more effective at growing entrepreneurial income (a higher  $\phi$ ), decreases if the time endowment is reduced by government policy (a higher  $\tau$ ), decreases if entrepreneurs place more weight on leisure (a higher  $\beta$ ), and decreases if either the endogenous or exogenous rates of creative destruction rise (a higher  $\delta^*$  or  $\bar{\delta}$ ).<sup>11</sup>

The analysis so far shows how one can endogenously obtain a Pareto-shaped income distribution. We've purposefully gotten to this result as quickly as possible while deferring our discussion of the general equilibrium in order to draw attention to the key economic forces that determine top income inequality.

### 4.3. Heterogeneous Mean Growth Rates

As pointed out by Luttmer (2011) and Gabaix, Lasry, Lions and Moll (2016), the basic random growth framework that forms the heart of the model so far has trouble explaining features of the data associated with transition dynamics. For example, in the firm dynamics studied by Luttmer (2011), Google and Microsoft become billion-dollar companies seemingly overnight, much faster (and more frequently) than occurs in plausibly-calibrated basic random growth models. Gabaix, Lasry, Lions and Moll (2016) also note that the speed of convergence to the stationary distribution is very slow in such models, making it hard for those models to match the rapid rise in top income inequality observed in the data.

Both papers suggest that a solution to these problems can be found by introducing heterogeneous mean growth rates: that is, it is possible for some entrepreneurs to grow extremely rapidly, at least for awhile (this is sometimes called the “Luttmer rocket”). This insight is consistent with recent empirical work: Guvenen, Karahan, Ozkan and Song (2016) show that growth rates for top earners are extremely heterogeneous, with the distribution of growth rates featuring a thick upper tail that even appears to be Pareto itself.

We follow the implementation by Gabaix, Lasry, Lions and Moll (2016) and augment

---

<sup>11</sup>The effect of  $\sigma^2$  on power-law inequality is more subtle. If  $\eta^* > \mu^*/\bar{\delta}$ , then a rise in  $\sigma^2$  increases  $\eta^*$ . Since  $\eta^* \rightarrow \mu^*/\bar{\delta}$  as  $\sigma^2 \rightarrow 0$ , this is the relevant case. Notice the similarity of this limit to the result in the simple model given at the start of the paper.

our basic setup to include two growth states for entrepreneurs.<sup>12</sup> When researchers discover a new idea, a fraction of them inherit the high growth  $\phi_H$  parameter. They then face a Poisson process with arrival rate  $\bar{p}$  for transitioning permanently down to the more normal  $\phi_L$  low growth parameter. In addition, we allow the variance of the shocks to also depend on the state, distinguishing  $\sigma_H$  and  $\sigma_L$ . This change is easily introduced and has a straightforward effect on the analysis we've done so far, as shown in the next proposition.

**Proposition 3** (Pareto Inequality with Heterogeneous Mean Growth Rates): *Extending the model to include high and low growth rates as in Luttmer (2011) and Gabaix, Lasry, Lions, and Moll (2015), for  $\phi_H$  sufficiently large, the stationary distribution of (normalized) entrepreneurial income has an upper tail with a Pareto inequality exponent  $\eta^* \equiv 1/\xi_H$ , where*

$$\xi_H = -\frac{\tilde{\mu}_H^*}{\sigma_H^2} + \sqrt{\left(\frac{\tilde{\mu}_H^*}{\sigma_H^2}\right)^2 + \frac{2(\bar{\delta} + \bar{p})}{\sigma_H^2}} \quad (18)$$

and  $\tilde{\mu}_H^* \equiv \mu_H(e^*) - \frac{1}{2}\sigma_H^2 = \phi_H(1 - \tau) - \beta(\rho + \delta^* + \bar{\delta}) - \frac{1}{2}\sigma_H^2$ .

That is, Pareto inequality is determined just as before, only with the key parameters replaced by those in the “high” growth case. The addition of  $\phi_H$  allows some entrepreneurs to grow very rapidly, addressing the Google/Microsoft problem. And the speed of convergence to steady state is governed by the Poisson “death rate.” Here, the relevant death rate includes  $\bar{p}$ , the rate at which entrepreneurs “die” out of the high growth state. We later estimate this rate to be very rapid, thereby substantially speeding up the transition to the stationary distribution.

#### 4.4. Production and General Equilibrium

Next, we flesh out the rest of the general equilibrium: how the entrepreneur's productivity  $x$  enters the model, how  $x$  affects entrepreneurial income (the proportionality factor  $\psi_t$ ), and how creative destruction  $\delta^*$  is determined.

The remainder of the setup is a relatively conventional model of endogenous growth with quality ladders and creative destruction, in the tradition of Aghion and Howitt

<sup>12</sup>Also, see Luttmer (2016). The logic of the proposition below suggests that the restriction to only two states instead of more is not especially important: the Pareto distribution will be dominated by the single state that delivers the thickest tail.

(1992) and Grossman and Helpman (1991). A fixed population of people choose to be basic laborers, outside researchers (searching for a new idea), or entrepreneurs (who have found an idea and are in the process of improving it).

A unit measure of varieties exist in the economy, and varieties combine to produce a single final output good:

$$Y = \left( \int_0^1 Y_i^\theta di \right)^{1/\theta}, \quad 0 < \theta < 1 \quad (19)$$

Each variety is produced by an entrepreneur using a production function that exhibits constant returns to basic labor  $L_i$ :

$$Y_i = \gamma^{n_t} x_i^\alpha L_i. \quad (20)$$

The productivity in variety  $i$ 's production function depends on two terms. The first captures aggregate productivity growth. The variable  $n_t$  measures how far up the quality ladder the variety is, and  $\gamma > 1$  is the step size. For simplicity, we assume that a researcher who moves a particular variety up the quality ladder generates spillovers that move all varieties up the quality ladder: in equilibrium, every variety is on the same rung of the ladder. (This just avoids us having to aggregate over varieties at different positions on the ladder.) The second term is the key place where the entrepreneur's idiosyncratic productivity enters: labor productivity depends on  $x_i^\alpha$ . As usual, variety  $i$ 's market share is increasing in  $x_i$ .

The main resource constraint in this environment involves labor:

$$L_t + R_t + 1 = \bar{N}, \quad L_t \equiv \int_0^1 L_{it} di \quad (21)$$

A fixed measure of people,  $\bar{N}$ , are available to the economy. People can work as the raw labor making varieties, or as outside researchers,  $R_t$ , or as entrepreneurs — of which there is always just a unit measure, though their identities can change. It is convenient to define  $\bar{L} \equiv \bar{N} - 1$ .

Outside researchers discover new ideas through a Poisson process with arrival rate  $\lambda$  per researcher. Research is undirected and a successful discovery, if implemented, increases the productivity of a randomly chosen variety by a proportion  $\gamma > 1$ . Once

the research is successful, the researcher becomes the entrepreneur of that variety, replacing the old entrepreneur by endogenous creative destruction. In addition, as explained above, the new idea generates spillovers that raise productivity in all other varieties as well. Existing entrepreneurs, however, may use the political process to block new ideas. We model this in a reduced form way: a fraction  $\bar{z}$  of new ideas are successfully blocked from implementation, preserving the monopoly (and productivity) of the existing entrepreneur.

The flow rate of innovation is therefore

$$\dot{n}_t = \lambda(1 - \bar{z})R_t \quad (22)$$

and this also gives the rate of creative destruction:

$$\delta_t = \dot{n}_t. \quad (23)$$

#### 4.5. The Allocation of Resources

There are 12 key endogenous variables in this economic environment:  $Y$ ,  $Y_i$ ,  $x_i$ ,  $L_i$ ,  $L$ ,  $R$ ,  $n$ ,  $\delta$ ,  $e_i$ ,  $c_i$ ,  $\ell_i$ ,  $\psi$ . (Table 2 in the appendix summarizes the notation used in the paper.) The entrepreneur's choice problem laid out earlier pins down  $c$ ,  $\ell$ , and  $e$  for each entrepreneur. Production functions and resource constraints determine  $Y$ ,  $Y_i$ ,  $L$ ,  $x_i$ ,  $n$ , and  $\delta$ . This leaves us needing to determine  $R$ ,  $L_i$ , and  $\psi$ .

It is easiest to do this in two stages. Conditional on a choice for  $R$ , standard equilibrium analysis can easily pin down the other variables, and the comparative statics can be calculated analytically. So to begin, we focus on a situation in which the fraction of people working as researchers is given exogenously:  $R/\bar{L} = \bar{s}$ . Later, we let markets determine this allocation as well and provide numerical results.

We follow a standard approach in decentralizing the allocation of resources. The final goods sector is perfectly competitive, while each entrepreneur engages in monopolistic competition in selling their varieties. Each entrepreneur is allowed by the patent system to act as a monopolist and charges a markup over marginal cost given by  $1/\theta$ . In equilibrium, then, wages and profits are given by the following proposition.

**Proposition 4 (Output, Wages, and Profits):** *Let  $w$  denote the wage per unit of raw*

labor, and let  $\pi_i$  denote the profit earned by the entrepreneur selling variety  $i$ . Assume now and for the rest of the paper that  $\alpha = (1 - \theta)/\theta$ .<sup>13</sup> The equilibrium with monopolistic competition leads to

$$Y_t = \gamma^{n_t} X_t^\alpha L_t \quad (24)$$

$$w_t = \theta \gamma^{n_t} X_t^\alpha \quad (25)$$

$$\pi_{it} = (1 - \theta) \gamma^{n_t} X_t^\alpha \left( \frac{x_{it}}{\bar{X}_t} \right) L_t \quad (26)$$

where  $X_t \equiv \int_0^1 x_{it} di$  is the mean of the  $x$  distribution across entrepreneurs.

According to the proposition, aggregate output is an increasing function of the mean of the idiosyncratic productivity distribution,  $X$ . In the baseline case with only a single state for  $\phi$ , the stationary distribution is Pareto throughout, and an important intuition is available. The mean of the  $x$  distribution is then  $X = \frac{x_0}{1-\eta}$ . More inequality (a higher  $\eta$ ) therefore has a long-run level effect in this economy, raising both output and wages.

We can now determine the value of  $\psi_t$ , the parameter that relates entrepreneurial income to  $x$ . Entrepreneurs earn the profits from their variety,  $\pi_{it}$ . In the entrepreneur's problem, we previously stated that the entrepreneur's income is  $\psi_t x_{it}$ , so these two equations define  $\psi_t$  as

$$\psi_t = (1 - \theta) \gamma^{n_t} X_t^{\alpha-1} L_t. \quad (27)$$

Finally, we can determine the overall growth rate of the economy along a balanced growth path. Once the stationary distribution of  $x$  has been reached,  $X$  is constant. Since  $L$  is also constant over time, the aggregate production function in equation (24) implies that growth in output per person is  $\dot{n}_t \log \gamma = \lambda(1 - \bar{z}) \bar{s} \bar{L} \log \gamma$  if the allocation of research is given by  $R/\bar{L} = \bar{s}$ . This insight pins down the key endogenous variables of the model, as shown in the next result.<sup>14</sup>

**Proposition 5 (Growth and inequality in the  $\bar{s}$  case):** *If the allocation of research is given exogenously by  $R/\bar{L} = \bar{s}$  with  $0 < \bar{s} < 1$ , then along a balanced growth path, the*

<sup>13</sup>This is merely a simplifying assumption that makes profits a linear function of  $x_i$ . It can be relaxed with a bit more algebra.

<sup>14</sup>At least one of the authors feels a painful twinge writing down a model in which the scale of the economy affects the long-run growth rate. This is certainly one target for valuable future work.



*growth of final output per person,  $g_y$ , and the rate of creative destruction are given by*

$$g_y^* = \lambda(1 - \bar{z})\bar{s}\bar{L} \log \gamma \quad (28)$$

$$\delta^* = \lambda(1 - \bar{z})\bar{s}\bar{L}. \quad (29)$$

*Power-law inequality is then given by [Proposition 2](#) or [Proposition 3](#) with this value of  $\delta^*$ .*

#### **4.6. Growth and Inequality: Comparative Statics**

In the setup with an exogenously-given allocation of research, the comparative static results are easy to see, and these comparative statics can be divided into those that affect top income inequality only, and those that also affect economic growth.

First, a technological change that increases  $\phi_H$  will increase top income inequality in the long run. This corresponds to anything that increases the effectiveness of entrepreneurs in building the market for their product. A canonical example of such a change might be the rise in the World Wide Web. For a given amount of effort, the rise of information technology and the internet allows successful entrepreneurs to grow their profits much more quickly than before, and we now see many examples of firms that go from being very small to very large quite quickly. Such a change is arguably not specific to any particular economy but rather common to the world. This change can be thought of as contributing to the overall rise in top income inequality throughout most economies, as was documented back in [Figure 2](#).

Interestingly, this technological change has no effect on the long-run growth rate of the economy, at least as long as  $\bar{s}$  is held fixed. The reason is instructive about how the model works. In the long run, there is a stationary distribution of entrepreneurial productivity  $x$ . Some varieties are extraordinarily successful, while most are not. Even though an increase in  $\phi_H$  increases the rate of growth of  $x$ , this only serves to widen the stationary distribution, as we showed back in [Section 4.2](#). There is a level effect on overall GDP (working through  $X$ ), but no growth effect. Long-run growth comes about only through the arrival of new external ideas from outside research, not through the productivity growth associated with improving an existing idea. In light of the endogenous growth literature, it is interesting that the log-linear differential equation

inherent in the geometric Brownian motion leads only to level effects in the model, rather than to growth effects. The ultimate reason underlying this fact is the “death rate”  $\delta$  and the decay rate  $\bar{p}$  that cause entrepreneurs to exit, generating a stationary distribution of  $x$ . This was the logic shown in the toy model back in Section 3.

The parameters  $\tau$  and  $\beta$  also affect top income inequality without affecting growth when  $\bar{s}$  is held constant. An increase in  $\tau$  corresponds to a reduction in the time endowment available to entrepreneurs — an example of such a policy might be the red tape and regulations associated with starting and maintaining a business. With less time available to devote to the productive aspects of running a business, the distribution of  $x$  and therefore the distribution of entrepreneurial income is narrowed and top income inequality declines.

The two key parameters in the model that affect both growth and top income inequality are  $\bar{s}$  and  $\bar{z}$ , and they work the same way. If a larger fraction of the labor works in research ( $\uparrow\bar{s}$ ) or if fewer innovations are blocked by incumbents ( $\downarrow\bar{z}$ ), the long-run growth rate will be higher — a traditional result in Schumpeterian growth models. Here, however, there will also be an effect on top income inequality. In particular, faster growth means more creative destruction — a higher  $\delta$ . This means that entrepreneurs have less time to build successful businesses, and this reduces top income inequality in the stationary distribution.

These are the basic comparative statics of top income inequality. Notice that a rise in top income inequality can be the result of either favorable changes in the economy — a new technology like the World Wide Web — or unfavorable changes — like policies that protect existing entrepreneurs from creative destruction.

## 5. Endogenizing R&D

We now endogenize the allocation of labor to research,  $s$ . This allocation is pinned down by the following condition: ex ante, people are indifferent between being a worker and being a researcher.

A worker earns a wage that grows at a constant rate and simply consumes this labor income. The worker’s value function is therefore

$$\rho V^w(t) = \log w_t + \frac{dV^w(t)}{dt} \tag{30}$$

A researcher searches for a new idea. If successful, the researcher becomes an entrepreneur. If unsuccessful, we assume the researcher still earns a wage  $\bar{m}w$ , where  $\bar{m}$  is a parameter measuring the amount of social insurance for unsuccessful research.

The value function for a researcher at time  $t$  is

$$\rho V^R(t) = \log(\bar{m}w_t) + \frac{dV^R(t)}{dt} + \lambda(1 - \bar{z}) (\mathbb{E}[V(x, t)] - V_t^R) + \bar{\delta}_R (\mathbb{E}[V(x_0, t)] - V^R(t)). \quad (31)$$

The first two terms on the right-hand side capture the basic consumption of an unsuccessful entrepreneur and the capital gain associated with wage growth. The last two terms capture the successful transition a researcher makes to being an entrepreneur when a new idea is discovered. This can happen in two ways. First, with Poisson flow rate  $\lambda(1 - \bar{z})$  the researcher innovates, pushing the research frontier forward by the factor  $\gamma$ , and replaces some randomly-selected existing entrepreneur. Alternatively, the researcher may benefit from the exogenous process: at rate  $\bar{\delta}_R \equiv \bar{\delta}/R$ , the researcher replaces a randomly-chosen variety and becomes a new entrepreneur with productivity  $x_0$ .

Finally the indifference condition  $V^w(t) = V^R(t)$  determines the allocation of labor as summarized in the following proposition.

**Proposition 6 (Allocation of Labor):** *In the stationary general equilibrium, the allocation of labor to research,  $s$ , is determined by the condition that  $V^w(t) = V^R(t)$ , where expressions for these value functions are given by equations (30) and (31).*

The key equations that describe the stationary general equilibrium are then shown in Table 1. However, it is not easy to discuss comparative statics as there is no closed-form solution for  $s^*$ . Instead, in the next section we show numerically how each parameter affects growth and inequality. The appendix explains in detail how the model is solved.

### 5.1. Steady-State Comparative Statics

Figure 8 shows the effect of various parameters on steady-state growth and inequality when  $s \equiv R/\bar{L}$  is endogenously determined. The effects on Pareto inequality are similar to those from the exogenous  $\bar{s}$  case. Now, however, we can also study the effects on

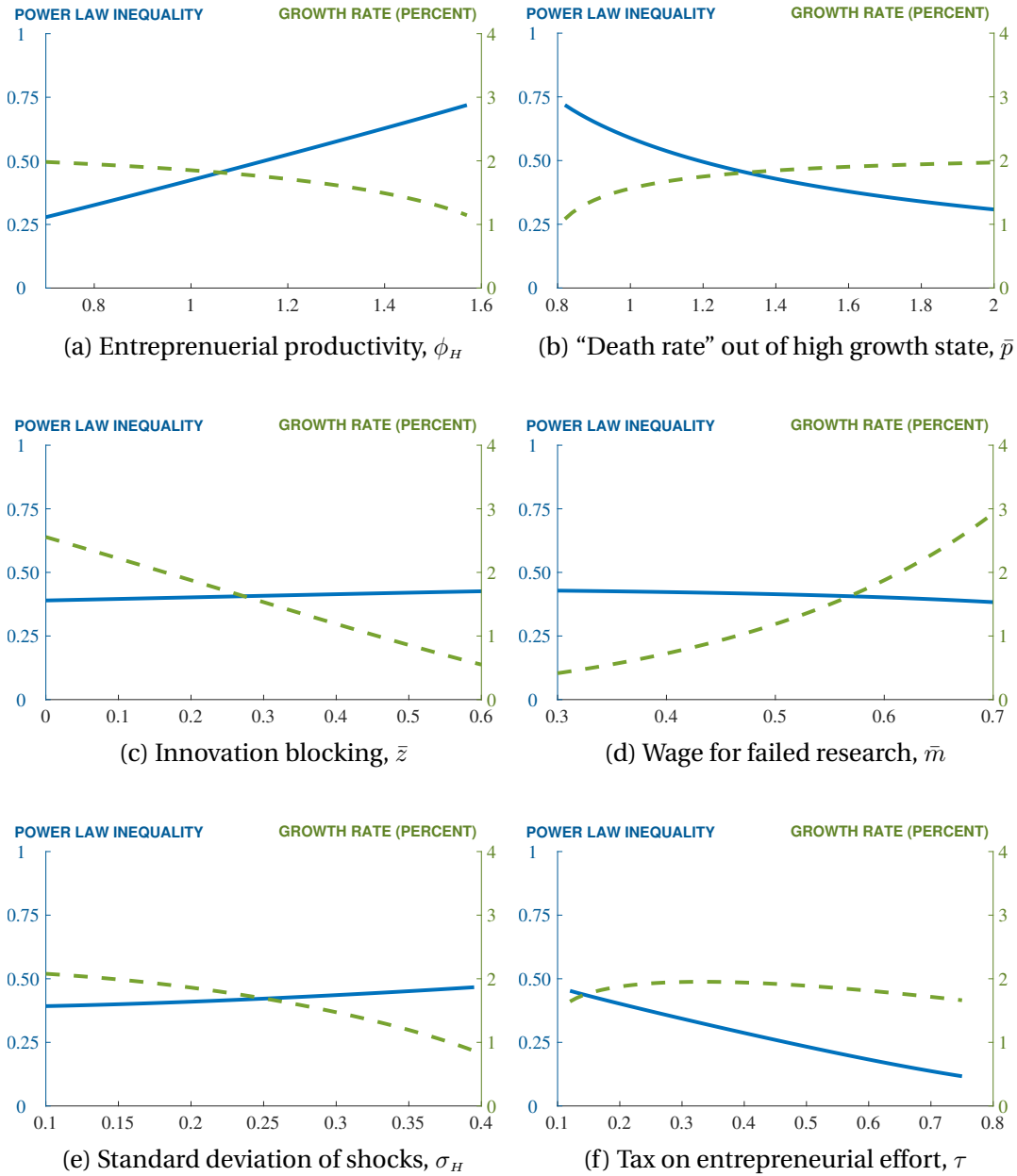
**Table 1:** Key Equations Characterizing the Stationary General Equilibrium

<b>Drift of log x</b>	$\tilde{\mu}_H^* = \phi_H(1 - \tau) - \beta(\rho + \delta^* + \bar{\delta}) - \frac{1}{2}\sigma_H^2$
<b>Pareto inequality</b>	$\eta^* = 1/\xi^*, \quad \xi^* = -\frac{\tilde{\mu}_H^*}{\sigma_H^2} + \sqrt{\left(\frac{\tilde{\mu}_H^*}{\sigma_H^2}\right)^2 + \frac{2(\bar{\delta} + \bar{p})}{\sigma_H^2}}$
<b>Creative destruction</b>	$\delta^* = \lambda(1 - \bar{z})s^*\bar{L}$
<b>Growth</b>	$g^* = \delta^* \log \gamma$
<b>Research allocation</b>	$V^w(s^*) = V^R(s^*)$

economic growth. For example, consider the effect of an increase in the technology parameter  $\phi_H$ , shown in Figure 8a: an increase in  $\phi_H$  raises Pareto inequality, as discussed earlier, but — perhaps surprisingly — causes a decline in the long-run growth rate of GDP per person. Similar results occur throughout Figure 8: parameter changes that increase Pareto inequality tend to reduce economic growth.

To understand this result, recall that the growth rate of the economy is determined by the fraction of people who decide to enter the research process, prospecting for the possibility of becoming successful entrepreneurs. On the one hand, an increase in  $\phi_H$  makes it easier for entrepreneurs to grow their profits, which tends to make research more attractive. However, from the standpoint of a researcher who has not yet discovered a new idea, another effect dominates. The positive technological improvement from a rising  $\phi_H$  raises average wages in the economy through  $X$ , both for workers and for unsuccessful researchers. The mean effect on the level of wages and profits is therefore neutral with respect to the allocation of labor. However, it also increases the inequality among successful researchers, making the research process itself more risky. Our researchers are risk-averse individuals with log utility, and the result of this risk aversion is that a rise in  $\phi_H$  results in a smaller fraction of people becoming researchers, which lowers the long-run growth rate in this endogenous growth model.

Figure 8: Steady-State Comparative Statics



Note: The figures show the steady-state values of Pareto inequality (solid line) and long-run growth (dashed) when a single parameter changes away from its baseline value. The baseline values are  $\rho = .01$ ,  $\theta = 2/3$ ,  $\gamma = 1.4$ ,  $\lambda = .02$ ,  $\phi_H = 0.955$ ,  $\beta = 1$ ,  $\sigma_H = 0.15$ ,  $\bar{\delta} = .08$ ,  $\bar{m} = .6$ ,  $\bar{z} = 0.2$ ,  $\tau = 0.2$ ,  $\bar{L} = 30$ ,  $\bar{p} = 1.5$ , and  $\bar{q} = .9875$ . These values will be discussed in more detail in Section 7.

One can, of course, imagine writing down the model in a different way. For example, if research is undertaken by risk-neutral firms, then this effect would not be present. Ultimately, this question must be decided by empirical work. Our model, however, makes it clear that this additional force is present, so that increases in Pareto inequality that result from positive technological changes need not increase the rate of growth.

The model generally features a negative relationship between long-run growth and top income inequality for two reasons. First is the reason just given: higher inequality tends to reduce growth by making research riskier. The second completes the cycle of feedback: faster growth leads to more creative destruction, which lowers inequality. Along a transition path, however, this negative effect on long-run growth is temporarily offset by a positive level effect (e.g. associated with the improved technology). Section 7. below shows that this effect can be large, even for periods as long as twenty or thirty years.

## 6. Micro Evidence

To what extent is our model consistent with empirical evidence? The first point to make is that the basic stochastic process for incomes assumed in our model — a geometric random walk with positive drift — is the canonical data-generating process estimated in an extensive empirical literature on income dynamics. Meghir and Pistaferri (2011) survey this literature, highlighting prominent examples such as MaCurdy (1982), Abowd and Card (1989), Topel and Ward (1992), Baker and Solon (2003), and Meghir and Pistaferri (2004). There are of course exceptions and some papers prefer alternative specifications, with the main one being the “heterogeneous income profiles” which allow for individual-specific means and returns to experience — consistent with the extended model with heterogeneous mean growth rates — but often find a persistence parameter less than one; for example, see Lillard and Willis (1978), Baker (1997), and Guvenen (2007, 2009). While debate continues within this literature, it is fair to say that a fundamental benchmark is that the log of income features a random walk component. In that sense, the basic data generating process we assume in this paper has solid micro-econometric foundations.

With unlimited access to micro data, our model makes some clear predictions that

could be tested. In particular, one could estimate the stochastic process for incomes around the top of the income distribution. In addition to the geometric random walk with heterogeneous drifts, one could estimate the creative destruction parameters — to what extent do high income earners see their incomes drop by a large amount in a short time? Guvenen, Ozkan and Song (2014b) provide evidence for precisely this effect, stating “[I]ndividuals in higher earnings percentiles face persistent shocks that are more negatively skewed than those faced by individuals that are ranked lower, consistent with the idea that the higher an individual’s earnings are, the more room he has to fall” (p. 20).

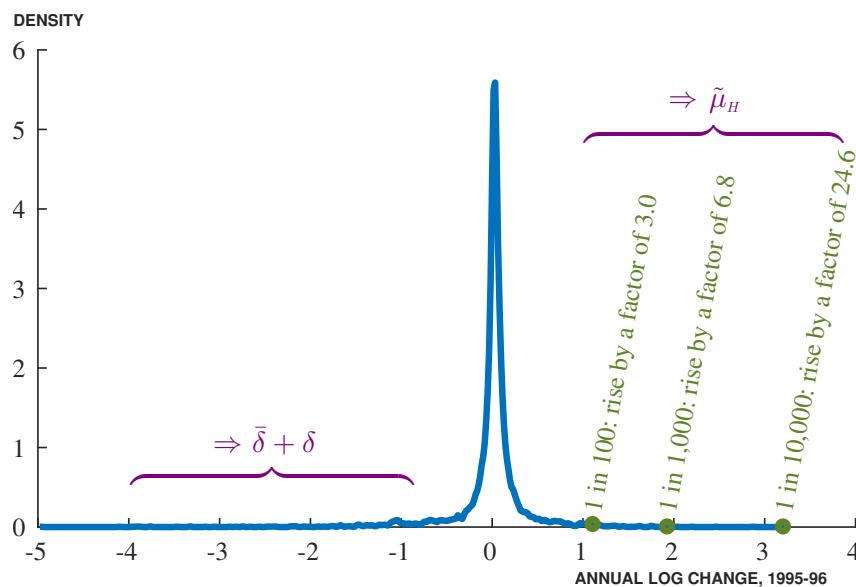
Beyond estimating this stochastic process, one could also see how the process differs before and after 1980 in the United States and how it differs between the United States and other countries. For example, one would expect the positive drift of the random walk to be higher for top incomes after 1980 than before. And one would expect this drift to be higher in the United States in the 2000s than in France; there could also be differences in the creative destruction parameters or the decay rate out of the high-growth state between countries and over time that could be estimated.

In the remainder of this section, we present estimates of the determinants of  $\eta$  and how they have changed over time. We have two sources for these estimates. First, we use the rich set of moments for wage and salary (W2) income based on the Social Security Administration data, as reported in the extensive data appendix of Guvenen, Karahan, Ozkan and Song (2016); we refer to this as the “SSA data.” These results have the advantage of being based on a large random sample of more than a million workers, with moments available annually between 1981 and 2011. Second, we use the U.S. Internal Revenue Service public use tax model panel files created by the Statistics of Income Division from 1979 to 1990, hosted by the NBER (the “IRS data”).<sup>15</sup> This is a random sample of taxpayers who can be followed over time between 1979 and 1990. The disadvantage is that the sample sizes are small and the time frame is relatively short. The advantage is that it allows us to examine entrepreneurial income in addition to wages and salaries.

---

<sup>15</sup>See <http://www.nber.org/taxsim-notes.html>.

Figure 9: The Distribution of Top Income Growth Rates, 1995–1996 (SSA Data)



Note: This figure shows the density of annual log income changes in 1995–96 for the 90th percentile of top earners aged 45-50, obtained from the spreadsheet data appendix of Guvenen, Karahan, Ozkan and Song (2016) for wage and salary income.

### 6.1. The Distribution of Top Income Growth Rates

Guvenen, Karahan, Ozkan and Song (2016) provide evidence for thick tails on both sides of the growth rate distribution for wage and salary income, and some of their evidence is shown in Figure 9; similar facts can be documented in the IRS panel data, but the sample sizes are much smaller. According to our model, the distribution of income growth rates for top earners should display thick tails at both the top and the bottom, as we see in the figure. At the bottom, the destruction shocks result in a potentially large downward shift in incomes, causing the growth rate distribution to be left-skewed. Quantitatively, the left-skewness of the distribution of growth rates helps us to identify  $\delta_t^e \equiv \bar{\delta}_t + \delta_t$ .

At the top, the presence of a “high growth” group leads to a mixture of normal distributions that thickens the right tail; this helps us to identify  $\tilde{\mu}_H$ . For example, in Figure 9, 1 in 1000 top earners see their incomes rise by a factor of 6.8 over the course of a year, and 1 in 10,000 see an increase by a factor of nearly 25!<sup>16</sup>

<sup>16</sup>Their evidence suggests very thick Pareto-like tails for the growth rate distribution, a fact that our



## 6.2. Empirical Results Based on Social Security Data

Our data and estimation are discussed in more detail in Appendix B. In brief, the parameters are estimated from the distribution of growth rates for top earners, as suggested in a stylized way in Figure 9. The upper tail of the growth rate distribution is used to estimate  $\tilde{\mu}_H$ ,  $\sigma_H$ , and  $\bar{p}$ . We estimate  $\tilde{\mu}_H$  as the median of growth rates above the 95th percentile of the growth rate distribution, i.e. as the growth rate at the 97.5th percentile. We estimate  $\sigma_H$  as the standard deviation of growth rates above the 95th percentile of the growth rate distribution.<sup>17</sup> If top incomes fall by more than 40 percent, we consider this a destruction event. The fraction of growth rates below this cutoff is an estimate of  $\delta^e$ , which is estimated to be around 13% in the Social Security data.

To estimate  $\bar{p}$ , we again use moments provided by Guvenen, Karahan, Ozkan and Song (2016). For each year and each percentile of the income distribution and for each percentile of the change in log earnings in year  $t$ , they report the fraction of people in a cell who have a growth rate above the 95th percentile of the growth rate distribution in year  $t + 1$ . Call this “ProbStayHigh.” In our model, this probability equals  $e^{-(\bar{p} + \delta^e)}$ , which is used to recover an estimate of  $\bar{p}$  in each year. Empirically, ProbStayHigh is relatively small: only around 5 percent of top earners with growth rates above the 95th percentile in year  $t$  continue in that high growth state in year  $t + 1$ . This yields estimates of  $\bar{p}$  that average about 2.8. From this value, one can see how transition dynamics are much faster when based on  $\bar{p}$  than when involving only  $\bar{\delta}$ , a key point made by Gabaix, Lasry, Lions and Moll (2016). In fact, the implied half life for the mean of high-growth incomes is less than half a year with this value.<sup>18</sup>

The heavy solid black line in Figure 10 shows the implied steady-state measure of Pareto inequality, based on (smoothed) values of the parameter estimates in each year

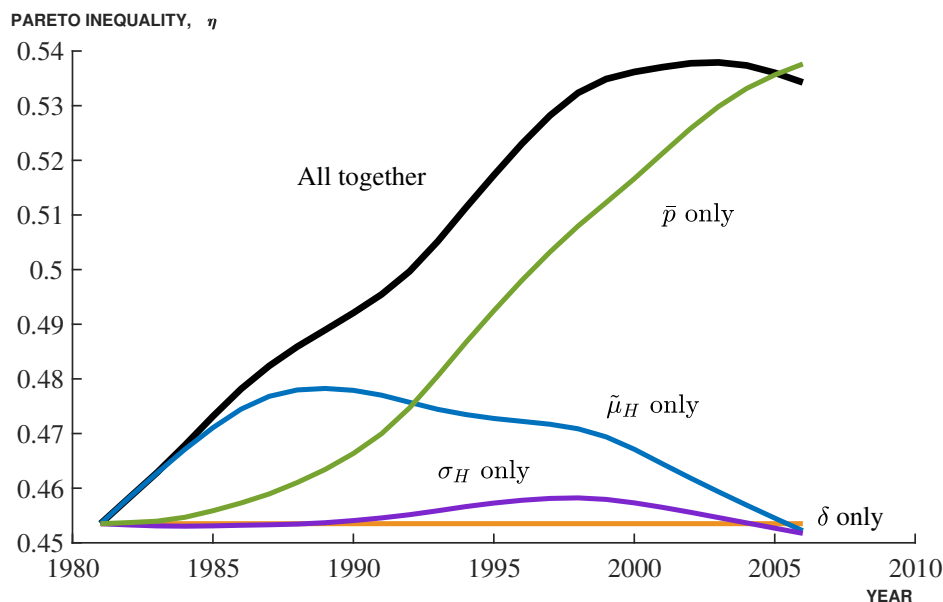
---

simple model cannot match. Our model predicts that growth rates would exhibit two overlapping normal distributions, together with a thick left tail associated with  $\delta + \bar{\delta}$  exits. We suspect that a model with more states for the heterogenous growth rates could do a better job of matching this data.

<sup>17</sup>Motivated by the empirical income dynamics literature, we make an adjustment for the presence of temporary income shocks, which are absent from our theory. Calculations from Blundell, Pistaferri and Preston (2008) and Heathcote, Perri and Violante (2010) suggest that the variance of the random walk innovation accounts for only about 1/6 to 1/3 of the variance of income growth rates. It is unclear how this applies to top incomes. Hence we make the following correction: we calculate  $\sigma^2$ , the variance of the random walk innovation, to be 1/4 the variance calculated from the highest decile of the growth rate distribution. Because our estimate of  $\eta$  is relatively insensitive to  $\sigma$ , this adjustment does not play a significant role.

<sup>18</sup>Using equation (29) from Gabaix, Lasry, Lions and Moll (2016), the speed of convergence for incomes in the high-growth state is  $\lambda_H(-1) = -\tilde{\mu}_H - \frac{1}{2} \cdot \sigma_H^2 + \bar{p} + \bar{\delta}$ , and the half-life is  $\ln(2)$  divided by this value.

Figure 10: Decomposing Pareto Inequality: Social Security Data



Note: Estimates based on the wage and salary data obtained from Guvenen, Karahan, Ozkan and Song (2016)

using the formula in equation (18).<sup>19</sup> The thinner colored lines in the graph show how the steady-state  $\eta$  would evolve if only one parameter changed, with the others held constant at their 1981 values. Of course this ignores transition dynamics, which we turn to in Section 7.

Several features of Figure 10 stand out. First, the overall level of  $\eta$  is roughly consistent with what we see in the data: the initial value is a little higher than what we observed in the early 1980s (at 0.45 instead of 0.40), and the value in the year 2000 is lower than in the data (0.53 instead of 0.63). Put differently, the model and the moments from the wage and salary data are consistent with about half the observed increase in Pareto inequality between 1980 and 2000 in the steady-state analysis.

Next, the colored lines in Figure 10 provide a decomposition of the overall movements in  $\eta$ . For example, the initial rise in  $\eta$  is driven by increases in  $\tilde{\mu}_H$  in the 1980s and changes in  $\bar{p}$  after that: the largest increases in  $\eta$  come from an increase in ProbStayHigh — the high growth state appears to have gotten more persistent in these data (i.e. a

<sup>19</sup>We smooth the parameter estimates using an HP filter with the smoothing parameter equal to 100, using the data from 1981 until 2006 to avoid the financial crisis having an undue influence on the smoothed values.

decline in the decay rate  $\bar{p}$ ). Interestingly, there is little evidence in the SSA wage and salary data for a sustained increase in  $\tilde{\mu}_H$ . Also, changes in the death rate  $\delta$  or in the idiosyncratic variance  $\sigma_H$  play a much smaller role, according to this decomposition. For example, reducing  $\sigma_H$  all the way to zero at the 2000 parameter values only lowers  $\eta^*$  from 0.536 to 0.500. Luck matters in our calibration, but it is luck in the form of becoming and remaining a high-growth entrepreneur that is most crucial.

### 6.3. Empirical Results Based on IRS Data

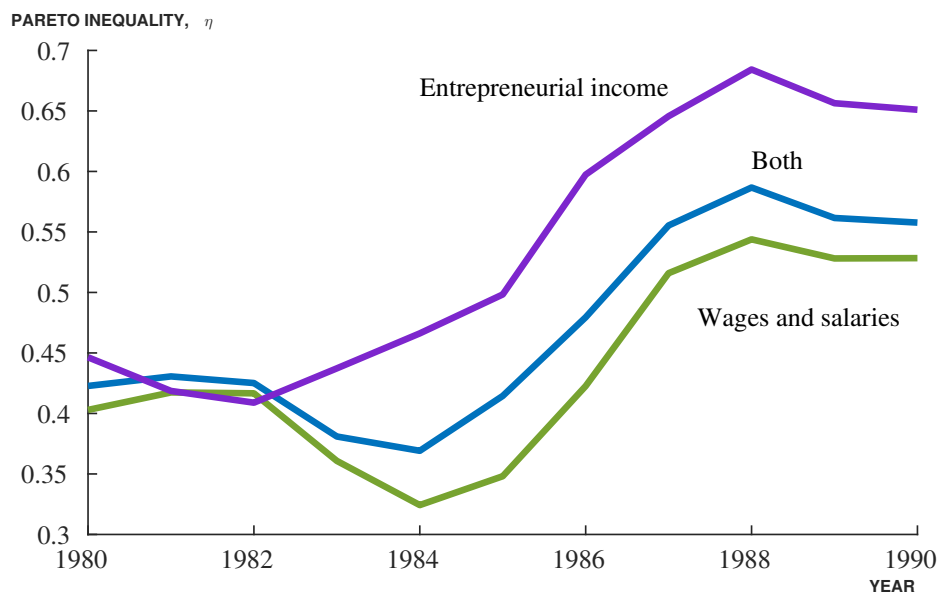
Similar calculations are possible using the IRS public use micro data, as explained in more detail in Appendix C. Unfortunately, the sample sizes are smaller (we can follow between 200 and 1100 earners in the Top 5% across a given two-year period; hence we focus on the Top 5% in the IRS data rather than the Top 1% which was possible in the SSA data) and the time frame is shorter (1979–1990). But these data have the advantage that we can see beyond wage and salary income to include business and entrepreneurial income. We follow Piketty and Saez (2003) and define entrepreneurial income to be the sum of income from Schedule C, partnerships, S corporations, and farm income.<sup>20</sup>

In the IRS panel data, we consider the Top 5% of earners in a given year. We then estimate  $\tilde{\mu}_H$  as the median growth rate above the 90th percentile of the growth rate distribution, i.e. as the growth rate at the 95th percentile. We estimate  $\sigma_H$  as the standard deviation of growth rates above the 90th percentile of the growth rate distribution. The fraction of top earners experiencing a decline in earnings of more than 40 percent is used to estimate  $\delta^e$ , and we assume  $\delta = 0.02$  and recover  $\bar{\delta}$  as the difference between  $\delta^e$  and  $\delta$ . We impose a constant value of 0.8 for  $\bar{p}$ . Appendix C reports further details on estimation.<sup>21</sup>

<sup>20</sup>The shares of wage income and entrepreneurial income in top incomes have seen some changes over time (see Table A7 of the June 2016 update of the spreadsheet appendix to Piketty and Saez (2003)). For example, the share of entrepreneurial income in top incomes increased after the Tax Reform Act of 1986 as it became advantageous to file as a partnership or S corporation to avoid the corporate-level tax.

<sup>21</sup>The main place our approach deviates from what we did for the Social Security data is in the choice of  $\bar{p}$ . The IRS panel data are not sufficiently rich to provide an estimate of  $\bar{p}$  because the panel dimension is too short for individual earners. If we use the value obtained from the moments in the Guvenen, Karahan, Ozkan and Song (2016) data for the top 1% of earners, 2.8, the implied value of  $\eta$  averages about 0.2. The reason is that the growth rates for the top 1% are substantially higher than the growth rates for the top 5%, so this high decay rate paired with lower  $\tilde{\mu}_H$  leads to the low  $\eta$  estimates. By choosing  $\bar{p} = 0.8$ , the level of  $\eta$  moves up to roughly match the data. The overall trend we obtain in  $\eta$  is unaffected by the choice of  $\bar{p}$ .

Figure 11: Pareto Inequality: IRS Data



Note: Steady-state Pareto inequality is calculated from equation (18) using the IRS public use micro data, as explained in the text and in Appendix C.

Figure 11 shows the implied steady-state Pareto inequality from the parameter estimates for each year (pair) between 1980 and 1990, using equation (18). As we saw earlier using the moments from Guvenen, Karahan, Ozkan and Song (2016), Pareto inequality rises in the 1980s for wage and salary income.<sup>22</sup> The figure shows that the upward trend is even more substantial if one focuses on entrepreneurial income.

We see several main take-aways from Figures 10 and 11. First, the basic moments emphasized by the model lead to Pareto inequality that is of the right magnitude and generates an upward trend in the 1980s and 1990s. Some of the trend results from an increase in the growth rate of top incomes, especially in the IRS data where we can see entrepreneurial income. A significant part of the trend in the wage and salary data comes from a decline in the “decay rate” out of the high-growth state. Finally, there is a perhaps surprising decline in the growth rate of top incomes in the wage and salary data after 1990. Obviously, exploring these features in the restricted-access administrative data is a valuable direction for future research.

<sup>22</sup>The timing is surprisingly different between Figure 10 and 11. The underlying moments of the data look somewhat different, partly due to the fact that the larger sample size in the SSA data lets us look higher up in the distribution and potentially in part due to the sampling error in the IRS data.

## 7. Transition Dynamics

We now explore the role of transition dynamics in this Schumpeterian model. This is important for two reasons. First, Gabaix, Lasry, Lions and Moll (2016) suggest that transition dynamics in models like this one can be very slow, making the steady-state calculations of the preceding section potentially misleading. Second, “level effects” in GDP per person may be important, as opposed to just the long-run growth rate. The first set of examples below explores one-time shocks to  $\phi_H$ ,  $\bar{p}$ , and  $\tau$ , while the second set feeds in the shocks recovered from the IRS and SSA data in the previous section. We trace through the transition dynamics associated with the Kolmogorov forward equation to study how top inequality and GDP per person evolve in response to these shocks.

These remain examples, however, for three main reasons. First and most importantly, the micro data we have access to is imperfect. The detailed SSA data allows us to look at the top of the income distribution and follow top earners over time, but it is based only on wages and salaries, rather than entrepreneurial income. The IRS panel allows us to see entrepreneurial income, but is only available during the 1980s, has small sample sizes, and has a limited panel dimension. The numerical examples we report below, then, are merely intended to verify that changes like what we’ve seen in the micro data have the potential when filtered through our model to explain the changes in top inequality that we see.

Second, the “reduced-form” empirical evidence is insufficient to identify the underlying structural parameters of the model. As one simple example, movements in both  $\phi_H$  and  $\tau$  can deliver changes in  $\tilde{\mu}_H$  over time; without additional data, it is hard to know which has changed.

Finally, we solve for only an approximation to the true transition dynamics of our model. It is notoriously difficult to solve for transition dynamics in heterogeneous agent models, in part because of the large state space they imply.<sup>23</sup> The simplification we use to make our problem computationally tractable is to assume control variables

---

<sup>23</sup>To be more precise, the HJB and KFE equations are coupled, with one running forward in time and one running backward. A full solution requires solving for a fixed point in the time path of two general equilibrium objects (the “wage” of the entrepreneurs  $\psi_t$  and the rate of creative destruction), which is at the frontier of existing methods. We are grateful to Ben Moll for advice on the solution technique we undertake.

(i.e. the effort choice by entrepreneurs and the research allocation) jump immediately to their steady-state values, while state variables evolve according to their laws of motion. The transition dynamics then come only from the Kolmogorov equation as the distribution of  $x$  evolves slowly over time. Importantly, this allows us to study transition dynamics in GDP per person as well, since the mean of the  $x$  distribution will also evolve slowly. These transition dynamics turn out to be quite important.

### 7.1. One-Time Shocks to $\phi_H$ , $\bar{p}$ , and $\tau$

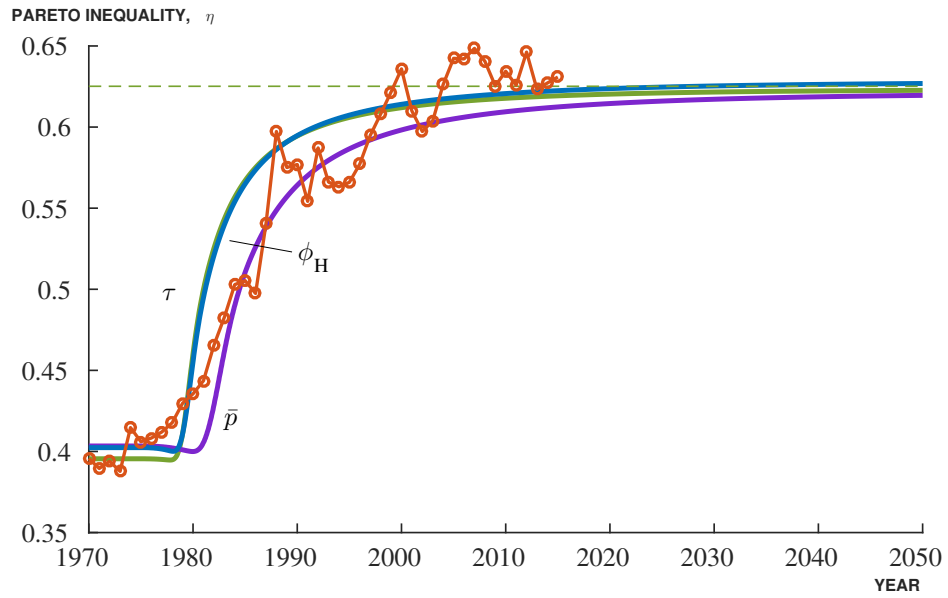
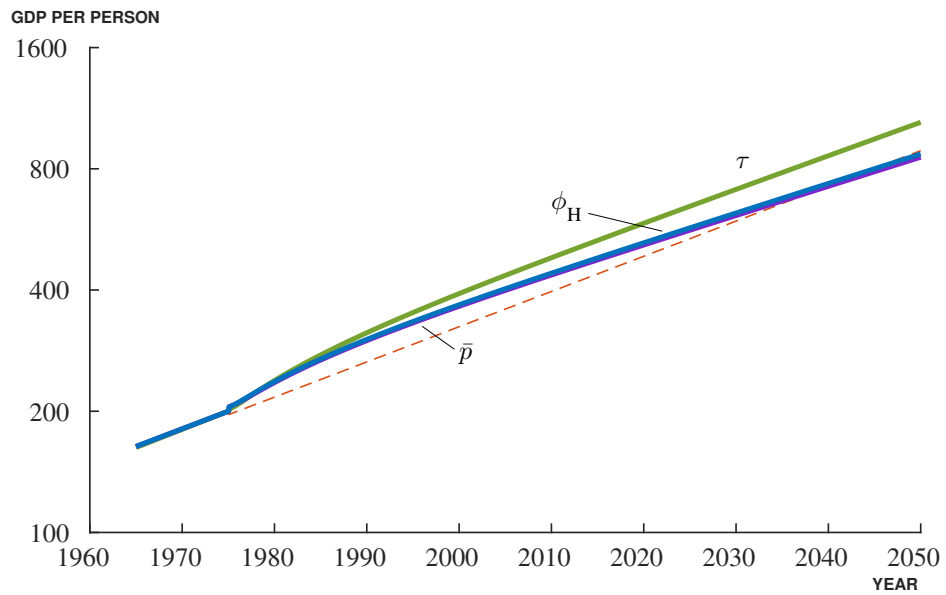
Our first set of examples consider one-time shocks to  $\phi_H$ ,  $\bar{p}$ , and  $\tau$ . We start with a set of baseline parameters that match U.S. Pareto inequality in 1975. Where possible, these are chosen to be consistent with the empirical estimates of the previous section. For example, we assume that  $\sigma_H$  for the United States is constant and equal to 0.15, broadly consistent with the evidence in Section 6. We assume  $\bar{\delta} = 0.08$  and  $\gamma = 1.4$ , so that  $\delta \approx 0.06$  when the economy's growth rate is 2 percent, and therefore  $\bar{\delta} + \delta \approx 0.14$ , similar to what we estimated using the SSA data in the previous section.<sup>24</sup>

Starting the economy off in steady state, we then shock one of our parameter values in the year 1980 by an amount that raises the new steady-state Pareto inequality to its average value at the end of our sample, 0.63. We feed in the new steady-state values of entrepreneurial effort and the research share and discretize the state space to solve the Kolmogorov Forward Equation using a finite difference method discussed by Moll (2016). The results are shown in Figure 12.<sup>25</sup>

Several findings stand out. First, consider Pareto inequality. One-time shocks to all three parameter values can generate rising Pareto inequality that roughly matches the U.S. experience. Some of the subtleties are interesting. For example, while the shock occurs in 1975, Pareto inequality does not start rising immediately and in fact declines at first. The reason is that top incomes take time to accumulate, and “slightly less than

<sup>24</sup>Our complete set of values are  $\rho = 0.01$ ,  $\bar{L} = 30$ ,  $\tau = 0.2$ ,  $\theta = 2/3$ ,  $\beta = 1$ ,  $\lambda = 0.02$ ,  $\bar{z} = 0.2$ ,  $\sigma_H = 0.15$ ,  $\bar{\delta} = 0.08$ ,  $\phi_H = 0.955$ ,  $\bar{p} = 1.5$ ,  $\bar{m} = 0.6$ , and  $\bar{q} = .9875$ . The value of  $\bar{q}$  is chosen so that the fraction of high-growth entrepreneurs in the stationary distribution is 5 percent.

<sup>25</sup>To solve the KFE, we discretize the state space into 2000 grid points and consider a time interval of  $dt = 1/25$  periods (about two weeks). We then apply the finite difference method described in Moll (2016) and used by Gabaix, Lasry, Lions and Moll (2016), for example in their “fig4b.m” and “fig5b.m” programs. In particular, the discretized KFE becomes a (sparse) Markov transition problem once we discretize the state space, and this equation can be solved in Matlab with standard techniques. For the second case in Figure 13 where we consider the sequence of shocks recovered from the IRS and SSA data, the transition matrix changes over time, but otherwise the same solution method applies.

Figure 12: One-Time Shocks to  $\phi_H$ ,  $\bar{p}$ , and  $\tau$ (a) Pareto inequality,  $\eta$ 

(b) GDP per person, log scale

Note: The figures show the U.S. data on top inequality (the orange circles) as well as the dynamic response of Pareto inequality and GDP per person to a one-time shock to  $\phi_H$ ,  $\bar{p}$  and  $\tau$  that occurs in 1975, using the approximation method described in the text. The shocks are sufficient to move Pareto inequality from an initial steady-state value of 0.39 to a new steady-state value of 0.63. All other parameters are held constant.

top” incomes accumulate first. In addition, the transition dynamics are remarkably slow, even given the very high decay rates out of the high growth state. As explained by Gabaix, Lasry, Lions and Moll (2016), this is because convergence rates at the top are much slower than convergence rates for the mean.

Next, consider the results for GDP per person shown in the bottom panel of Figure 12. We saw earlier that each of these shocks *lowers* the growth rate of GDP per person in the long run. However, we see here that the long-run is very far away! In particular, the “level effect” associated with the gradual increase in the mean of the distribution of entrepreneurial productivity dominates for at least 30 years. This shows that the model can reproduce the positive correlation between top inequality and growth rates that has been documented elsewhere, as discussed more in the concluding section.

This exercise suggests that something like these shocks — or perhaps some combination of them or even a sequence of such shocks — can potentially explain the patterns that we see in top income inequality across countries.<sup>26</sup>

## 7.2. Transition Dynamics and the IRS/SSA-Inspired Shocks

An alternative set of examples is shown in Figure 13. Here, we take the (smoothed) time paths of  $\tilde{\mu}_H$ ,  $\bar{p}$ ,  $\bar{\delta}$ , and  $\sigma_H$  that we observe in the IRS and SSA data, feed them into the model, and show the transition path of Pareto inequality and GDP per person. As in our previous exercises, we cannot solve the full model for transition dynamics. In this case, we are additionally limited by a basic identification problem: we do not know what structural changes in the deep parameters of the model led to the changes in  $\tilde{\mu}_H$ , for example. However, in our previous exercises, we found that the effects on  $s$  and  $g$  were not especially large. So for this set of exercises, we hold those macro-level variables constant and once again solve the discretized Kolmogorov equation. We assume the parameters that are impacted by shocks remain constant at the final values that we observe.

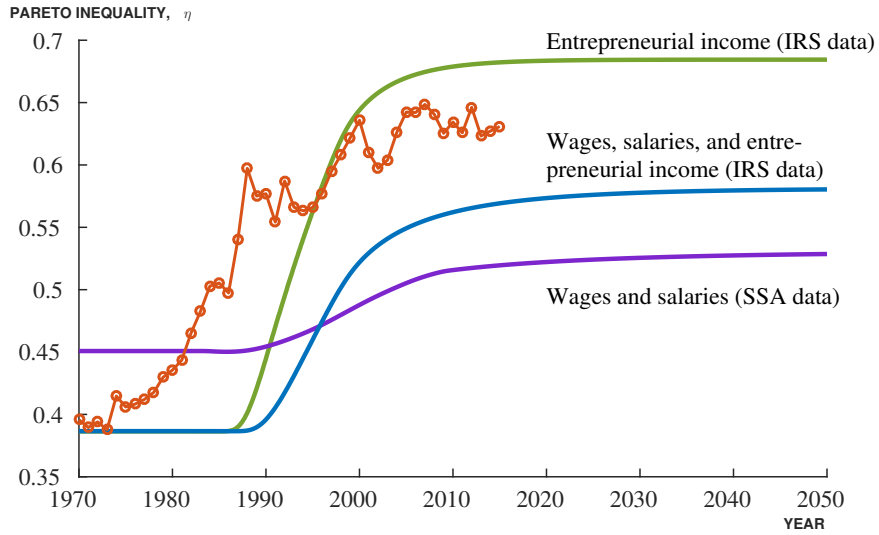
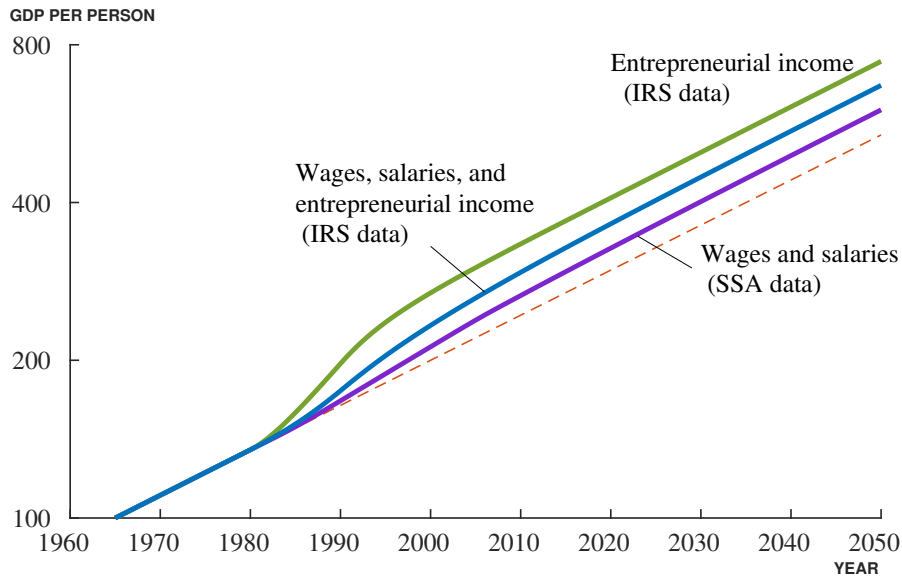
The results for GDP per person confirm what we saw earlier: there are substantial

---

<sup>26</sup>The exercise also confirms that the “Luttmer rocket” with high decay rates can successfully address the problems in simple random growth models of slow transition dynamics and “very old” people at the top of the distribution. For example, in the steady state of these simulations, the average length of time that someone in the top 1% has been an entrepreneur is around 20 years, and the average for someone in the top 1% who is also in the high growth state is 3.7 years.



Figure 13: The Dynamic Response to IRS/SSA-Inspired Shocks

(a) Pareto inequality,  $\eta$ 

(b) GDP per person, log scale

Note: The figures show the dynamic response of Pareto inequality and GDP per person to the HP-filtered shocks to  $\tilde{\mu}_H$ ,  $\bar{p}$ ,  $\bar{\delta}$ , and  $\sigma_H$  estimated in Section 6, using the approximation method described in the text. The shocks start in 1981 and continue until 2006 for the SSA data and until 1990 for the other two cases, which use the IRS data. For the IRS cases, we hold  $\bar{p}$  constant at the value that delivers an initial value of  $\eta$  of 0.39. All other parameters are held constant.

“level effects” that emerge over 25 years or more that can easily mask any long-run growth effect during that time.

The results for Pareto inequality are notable in several ways. First, there is remarkably little action in the wage and salary data from the Social Security Administration, consistent with what we saw in the steady-state analysis of Figure 10. In part, this reflects the fact that the rise in Pareto inequality after a shock occurs with a delay (as shown in the previous figure), and  $\tilde{\mu}_H$  exhibits a hump-shaped pattern in the SSA data.

There is substantial action in the entrepreneurial income data from the IRS, enough that these changes can generate a rise in Pareto inequality that matches what we see in the data. The timing is off somewhat, but this could potentially be explained if there were shocks occurring before 1981 that we cannot see. The IRS series that combines both wages and salaries with entrepreneurial income, not surprisingly, generates results that are intermediate.

This analysis is only suggestive, in part because the IRS results are based on small samples only during the 1980s. However, it clearly points toward the value of obtaining administrative data on entrepreneurial incomes via the tax records. The SSA data results are a bit puzzling in that we know that wage and salary data also display a large increase in Pareto inequality. Clearly something beyond our basic model must be going on. For example, it could be that there is a differential growth rate for wage and salary income for people who have experienced rapid growth in entrepreneurial income in the past, something we cannot see with just the SSA data alone.

## 8. Conclusion

A model in which entrepreneurs expend effort to increase the profits from their existing ideas while researchers seek new ideas to replace incumbents in a process of creative destruction generates a Pareto distribution for top incomes. Moreover, it suggests economic forces that change top income inequality. Forces that increase the effort of fast-growing entrepreneurs in improving their products — or that increase the productivity of their effort — can increase top inequality. Forces that enhance creative destruction or that raise the rate at which high-growth entrepreneurs lose that status can decrease top inequality.

Globalization is a general economic phenomenon that could be driving these changes. Greater globalization allows entrepreneurs to grow their profits more rapidly for a given amount of effort, increasing  $\phi_H$  and raising inequality. On the other hand, as countries open their domestic markets to more competition via globalization, rates of creative destruction (including  $\bar{p}$ ) go up, reducing inequality. Changes in these impacts over time or differences in their strength across countries can potentially explain the patterns of top income inequality that we see in the data.

A theme that emerges clearly from our analysis is that there are rich connections between models of top income inequality and the underlying micro data on income dynamics. Work connecting these two literatures — including incorporating micro data from other countries — is likely to be quite fruitful in coming years. Aghion, Akcigit, Hyytinen and Toivanen (2016) and Bell, Chetty, Jaravel, Petkova and Van Reenen (2016) document that children of low-income parents are much less likely to become inventors. Guvenen, Kaplan and Song (2014a) study the role of gender differences in the rise in top earnings inequality. These same authors (in progress) are working to estimate a rich model of micro income dynamics and tie it more closely to the rise in top income inequality. Related work using the administrative data from the IRS on the dynamics of entrepreneurial income is likely to be a productive area for future research.

Aghion, Akcigit, Bergeaud, Blundell and Hemous (2015) and Akcigit, Grigsby and Nicholas (2016) document that innovation and top income inequality are positively correlated across U.S. states and commuting zones. On the surface, there might be some tension between their results and ours (innovation raising inequality empirically versus the creative destruction effect just discussed), but we instead see the results as complementary. The creative destruction force works to reduce inequality in our model, and that effect has a solid economic foundation. Empirical estimates of the correlation between innovation and inequality, however, surely reflect more than just creative destruction. For example, the incumbent innovation by existing entrepreneurs (raising  $x$ ) is also captured by the patent data, and this force tends to raise inequality in our model. And as we saw earlier, the “level effect” associated with rising inequality can dominate the long-run “growth effect” for many decades. Isolating these distinct forces empirically is an important direction for future research.

**Appendix:** Table 2 summarizes the notation used in the paper.

Table 2: Guide to Notation for Full Model

$\eta$	Pareto inequality measure (inverse of Pareto exponent)
$x$	Idiosyncratic productivity of an entrepreneur's variety
$f(x, t)$	Distribution of idiosyncratic productivity across entrepreneurs
$e$	Entrepreneur's effort
$\tilde{\mu}$	$\tilde{\mu} \equiv \phi e - \frac{1}{2}\sigma^2$ . Drift of $\log x$ .
$\ell$	Entrepreneur's leisure
$\tau$	"Tax" on the entrepreneur's time endowment
$\psi$	Entrepreneur's "wage" per unit of $x$
$\phi$	Technology parameter: how effort translates into growth of $x$
$\sigma$	Variance of the idiosyncratic shocks to $x$
$\Omega$	$\Omega \equiv 1 - \tau$
$\beta$	Weight on (log) leisure in utility
$\rho$	Rate of time preference
$\delta$	Endogenous rate of creative destruction
$\bar{\delta}$	Exogenous destruction of entrepreneurs
$\bar{p}$	Rate at which high-growth entrepreneurs decay to low growth
$\alpha$	Exponent on $x_i$ in production of variety $i$
$\theta$	CES curvature parameter in final goods production
$n_t$	"Height" up the quality ladder. Productivity is $\gamma^{n_t}$
$\gamma$	Step size for the quality ladder. $\gamma > 1$
$\bar{N}$	Aggregate labor endowment
$L_t$	Aggregate allocation of labor to goods production
$R_t$	Aggregate allocation of labor to idea production
$\bar{L}$	$\bar{L} \equiv \bar{N} - 1$ . Labor endowment net of entrepreneurs.
$\lambda$	Research productivity
$\bar{z}$	Fraction of innovations that are exogenously blocked
$V, V^R, V^w$	Expected lifetime utility for entrepreneurs, researchers, and workers
$\bar{m}$	Fraction of equilibrium wage paid to failed researchers
$X$	Mean of the distribution of idiosyncratic productivity, $x$
$w_t$	Wage of labor in producing goods
$\pi_{it}$	Flow profit in variety $i$
$s_t$	$s_t \equiv R/\bar{L}$ . Fraction of labor engaged in research
$g_y$	Growth rate of GDP per person ( $y$ )

## References

- Abowd, John M. and David Card, "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, March 1989, 57 (2), 411–45.
- Acemoglu, Daron and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, June 2011, chapter 12, pp. 1043–1171.
- Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, March 1992, 60 (2), 323–351.
- , Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hemous, "Innovation and Top Income Inequality," Working Paper 21247, National Bureau of Economic Research June 2015.
- , —, Ari Hyytinen, and Otto Toivanen, "Living the American Dream in Finland: The Social Mobility of Innovators," Technical Report, Harvard University unpublished manuscript 2016.
- Akcigit, Ufuk, John Grigsby, and Tom Nicholas, "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age," 2016. University of Chicago manuscript.
- Aluation.wordpress.com, "Fractals, fractiles and inequality," Technical Report August 2011.
- Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, "The World Wealth and Income Database," 2016. Accessed on October 28, 2016.
- Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez, "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 2011, 49 (1), 3–71.
- Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "The Polarization of the U.S. Labor Market," *American Economic Review*, May 2006, 96 (2), 189–194.
- Baker, Michael, "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, April 1997, 15 (2), 338–75.
- and Gary Solon, "Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records," *Journal of Labor Economics*, April 2003, 21 (2), 267–288.
- Bakija, Jon, Adam Cole, and Bradley Heim, "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data," working paper, Indiana University November 2010.

- Banerjee, Abhijit V. and Andrew F. Newman, "Occupational Choice and the Process of Development," *Journal of Political Economy*, April 1993, 101 (2), 274–98.
- Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "The Lifecycle of Inventors," 2016. Stanford University unpublished manuscript.
- Bell, Brian and John Van Reenen, "Bankers and Their Bonuses," *The Economic Journal*, 2014, 124 (574), F1–F21.
- Benabou, Roland and Jean Tirole, "Bonus Culture: Competitive Pay, Screening, and Multitasking," *Journal of Political Economy*, 2016, 124 (2), 305–370.
- Benhabib, Jess, "Wealth Distribution Overview," 2014. NYU teaching slides <http://www.econ.nyu.edu/user/benhabib/wealth%20distribution%20theories%20overview3.pdf>.
- , Alberto Bisin, and Shenghao Zhu, "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents," *Econometrica*, 01 2011, 79 (1), 123–157.
- and —, "Skewed Wealth Distributions: Theory and Empirics," 2016. NYU unpublished manuscript.
- Blundell, Richard, Luigi Pistaferri, and Ian Preston, "Consumption Inequality and Partial Insurance," *American Economic Review*, December 2008, 98 (5), 1887–1921.
- Cantelli, F.P., "Sulle applicazioni del calcolo delle probabilita alla fisica molecolare," *Metron*, 1921, 1 (3), 83–91.
- Champernowne, David G., "A Model of Income Distribution," *The Economic Journal*, 1953, 63 (250), 318–351.
- Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, August 1999, 114 (3), 739–767.
- , "Power Laws in Economics and Finance," *Annual Review of Economics*, 2009, 1 (1), 255–294.
- and Augustin Landier, "Why Has CEO Pay Increased So Much?," *Quarterly Journal of Economics*, 2008, 123 (1), 49–100.
- , Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "The Dynamics of Inequality," *Econometrica*, November 2016, 84 (6), 2071–2111.
- Galor, Oded and Joseph Zeira, "Income Distribution and Macroeconomics," *Review of Economic Studies*, January 1993, 60 (1), 35–52.

- Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, October 2000, 108 (5), 874–904.
- Geerolf, Francois, "A Theory of Pareto Distributions," *Unpublished manuscript, UCLA*, October 2016.
- Gordon, Robert J. and Ian Dew-Becker, "Controversies about the Rise of American Inequality: A Survey," NBER Working Papers 13982, National Bureau of Economic Research, Inc May 2008.
- Grossman, Gene M. and Elhanan Helpman, *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press, 1991.
- Guvenen, Fatih, "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *American Economic Review*, June 2007, 97 (3), 687–712.
- , "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, January 2009, 12 (1), 58–79.
- , Fatih Karahan, Serdar Ozkan, and Jae Song, "What do data on millions of U.S. workers reveal about life-cycle earnings risk?," August 2016. University of Minnesota, unpublished manuscript.
- , Greg Kaplan, and Jae Song, "The Glass Ceiling and The Paper Floor: Gender Differences among Top Earners, 1981-2012," Working Paper 20560, National Bureau of Economic Research October 2014.
- , Serdar Ozkan, and Jae Song, "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 2014, 122 (3), 621 – 660.
- Haskel, Jonathan, Robert Z. Lawrence, Edward E. Leamer, and Matthew J. Slaughter, "Globalization and U.S. Wages: Modifying Classic Theory to Explain Recent Facts," *Journal of Economic Perspectives*, Spring 2012, 26 (2), 119–40.
- Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante, "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006," *Review of Economic Dynamics*, January 2010, 13 (1), 15–51.
- Hubmer, Joachim, Per Krusell, and Anthony A. Smith, "The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation," Working Paper 23011, National Bureau of Economic Research December 2016.
- Kaplan, Steven N. and Joshua Rauh, "Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?," *Review of Financial Studies*, 2010, 23 (3), 1004–1050.

- Katz, Lawrence F. and David H. Autor, "Changes in the wage structure and earnings inequality," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3 of *Handbook of Labor Economics*, Elsevier, June 1999, chapter 26, pp. 1463–1555.
- Koenig, Michael, Jan Lorenz, and Fabrizio Zilibotti, "Innovation vs. imitation and the evolution of productivity distributions," *Theoretical Economics*, September 2016, 11 (3), 1053–1102.
- Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, 65 (6), 1389–1419.
- Levy, Moshe, "Are rich people smarter?," *Journal of Economic Theory*, 2003, 110 (1), 42 – 64.
- Lillard, Lee A. and Robert J. Willis, "Dynamic Aspects of Earning Mobility," *Econometrica*, September 1978, 46 (5), 985–1012.
- Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 1988, 22 (1), 3–42.
- and Benjamin Moll, "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, February 2014, 122 (1), 1–51.
- Luttmer, Erzo G.J., "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, 08 2007, 122 (3), 1103–1144.
- , "Models of Growth and Firm Heterogeneity," *Annual Review Economics*, 2010, 2 (1), 547–576.
- , "On the Mechanics of Firm Growth," *Review of Economic Studies*, 2011, 78 (3), 1042–1068.
- , "An Assignment Model of Knowledge Diffusion and Income Inequality," Working Papers 715, Federal Reserve Bank of Minneapolis September 2014.
- , "Further Notes on Micro Heterogeneity and Macro Slow Convergence," 2016. University of Minnesota unpublished manuscript.
- MaCurdy, Thomas E., "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 1982, 18 (1), 83–114.
- Malevergne, Y., A. Saichev, and D. Sornette, "Zipf's law and maximum sustainable growth," *Journal of Economic Dynamics and Control*, 2013, 37 (6), 1195 – 1212.
- Mandelbrot, Benoit, "The Pareto-Levy Law and the Distribution of Income," *International Economic Review*, 1960, 1 (2), 79–106.



- Meghir, Costas and Luigi Pistaferri, "Income Variance Dynamics and Heterogeneity," *Econometrica*, 01 2004, 72 (1), 1–32.
- and —, "Earnings, Consumption and Life Cycle Choices," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics, Volume 4*, Elsevier, 2011, chapter 9, pp. 773–854.
- Mitzenmacher, Michael, "A Brief History of Generative Models for Power Law and Lognormal Distributions," *Internet Mathematics*, 2004, 1 (2).
- Moll, Benjamin, "Inequality and Financial Development: A Power-Law Kuznets Curve," 2012. Princeton University working paper.
- , "Lectures 6 and 7: Theories of Top Inequality," Fall 2016. Princeton teaching slides [http://www.princeton.edu/~moll/ECO521\\_2016/Lectures6\\_7\\_ECO521.pdf](http://www.princeton.edu/~moll/ECO521_2016/Lectures6_7_ECO521.pdf).
- Nirei, Makoto, "Pareto Distributions in Economic Growth Models," IIR Working Paper 09-05, Institute of Innovation Research, Hitotsubashi University July 2009.
- and Shuhei Aoki, "Pareto distribution of income in neoclassical growth models," *Review of Economic Dynamics*, 2016, 20, 25 – 42.
- Pareto, Vilfredo, *Cours d'Economie Politique*, Geneva: Droz, 1896.
- Perla, Jesse and Christopher Tonetti, "Equilibrium Imitation and Growth," *Journal of Political Economy*, February 2014, 122 (1), 52–76.
- Philippon, Thomas and Ariell Reshef, "Wages and Human Capital in the U.S. Finance Industry: 1909–2006," *The Quarterly Journal of Economics*, 2012, 127 (4), 1551–1609.
- Piketty, Thomas and Emmanuel Saez, "Income Inequality In The United States, 1913–1998," *Quarterly Journal of Economics*, February 2003, 118 (1), 1–39.
- and —, "A Theory of Optimal Inheritance Taxation," *Econometrica*, 09 2013, 81 (5), 1851–1886.
- and Gabriel Zucman, "Wealth and Inheritance in the Long Run," April 2014. forthcoming in the Handbook of Income Distribution.
- , Emmanuel Saez, and Gabriel Zucman, "Distributional National Accounts: Methods and Estimates for the United States," Working Paper 22945, National Bureau of Economic Research December 2016.
- , —, and Stefanie Stantcheva, "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, February 2014, 6 (1), 230–71.

- Reed, William J., "The Pareto, Zipf and other power laws," *Economics Letters*, 2001, 74 (1), 15–19.
- Rosen, Sherwin, "The Economics of Superstars," *American Economic Review*, December 1981, 71 (5), 845–58.
- Rothschild, Casey and Florian Scheuer, "Optimal Taxation with Rent-Seeking," *Review of Economic Studies*, 2016, 83 (3), 1225–1262.
- Saez, Emmanuel, "Using Elasticities to Derive Optimal Tax Rates," *Review of Economic Studies*, 2001, 68, 205–229.
- Simon, Herbert A., "On a Class of Skew Distribution Functions," *Biometrika*, 1955, 42 (3/4), 425–440.
- Steindl, Joseph, *Random Processes and the Growth of Firms*, New York: Hafner, 1965.
- Toda, Alexis Akira, "Incomplete market dynamics and cross-sectional distributions," *Journal of Economic Theory*, 2014, 154, 310–348.
- Topel, Robert H and Michael P Ward, "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, May 1992, 107 (2), 439–79.

# Online Appendix to “A Schumpeterian Model of Top Income Inequality”

(Not for publication)

Charles I. Jones and Jihee Kim

June 8, 2017

## A Appendix: The Stationary Distribution of Experience

It is helpful to show the argument that the stationary distribution of experience in the simple model of Section 3 is exponential. The reason is that this illustrates a basic version of the Kolmogorov forward equation, which is used later in solving for inequality in the full stochastic model.

Let  $F(x, t)$  denote the distribution of experience at time  $t$ , and consider how this distribution evolves over some discrete time interval  $\Delta t$ :<sup>27</sup>

$$F(x, t + \Delta t) - F(x, t) = \underbrace{\delta \Delta t (1 - F(x, t))}_{\text{inflow from above } x} - \underbrace{[F(x, t) - F(x - \Delta x, t)]}_{\text{outflow as top folks age}}$$

Dividing both sides by  $\Delta t = \Delta x$  and taking the limit as the time interval goes to zero yields:

$$\frac{\partial F(x, t)}{\partial t} = \delta(1 - F(x, t)) - \frac{\partial F(x, t)}{\partial x} \tag{A1}$$

One can continue with this equation directly. But for what comes later, it is more useful to take the derivative of both sides of this equation with respect to  $x$ . Letting  $f(x, t) := \frac{\partial F(x, t)}{\partial x}$  denote the pdf,

$$\frac{\partial f(x, t)}{\partial t} = -\delta f(x, t) - \frac{\partial f(x, t)}{\partial x}. \tag{A2}$$

This equation is the non-stochastic version of the Kolmogorov forward equation that we will see later. The intuition underlying this equation is easiest to see in the version involving the cdf, equation (A1), which just involves the inflows and outflows mentioned earlier.

Finally, to solve for the stationary distribution, we seek  $f(x)$  such that  $\frac{\partial f(x, t)}{\partial t} = 0$ .

---

<sup>27</sup>This equation drops a term involving both  $\Delta t$  and  $\Delta x$ , as it goes to zero later when we take limits.

Therefore

$$0 = -\delta f(x) - \frac{df(x)}{dx}.$$

Integrating this equation twice yields the result that the stationary distribution is exponential:  $F(x) = 1 - e^{-\delta x}$ .

## **B Appendix: Estimating the U.S. Stochastic Income Process using the Moments from Guvenen et al (2016)**

Our results with the SSA data are based on the extensive data appendix provided by Guvenen, Karahan, Ozkan and Song (2016). Because their sample sizes are extremely large, it is possible to look in the extreme upper tails of the income and growth rate distributions. In particular, the moments below are all based on the top 1% of earners.

The upper tail of the growth rate distribution is used to estimate  $\tilde{\mu}_H$ ,  $\sigma_H$ , and  $\bar{p}$ . We estimate  $\tilde{\mu}_H$  as the median of growth rates above the 95th percentile of the growth rate distribution, i.e. as the growth rate at the 97.5th percentile, for the sample of Top 1% earners. We estimate  $\sigma_H$  as the standard deviation of growth rates above the 95th percentile of the growth rate distribution. If top incomes fall by more than 40 percent, we consider this a destruction event. The fraction of growth rates below this cutoff is an estimate of  $\delta^e$ .

To estimate  $\bar{p}$ , we again use moments provided by Guvenen, Karahan, Ozkan and Song (2016). For each year and each percentile of the income distribution and for each percentile of the change in log earnings in year  $t$ , they report the fraction of people in a cell who have a growth rate above the 95th percentile of the growth rate distribution in year  $t+1$ . Call this “ProbStayHigh.” In our model, this probability equals  $e^{-(\bar{p}+\delta^e)}$ , which is used to recover an estimate of  $\bar{p}$  in each year. Empirically, ProbStayHigh is relatively small: only around 5 percent of top earners with growth rates above the 95th percentile in year  $t$  continue in that high growth state in year  $t+1$ .

The moments that we obtain in this fashion are reported in Table B1, together with the steady-state Pareto inequality that is implied by the parameters in each year.

Table B1: SSA Data Parameter Estimates (smoothed)

Year	$\eta$	$\tilde{\mu}_H$	$\sigma_H$	$\bar{p}$	$\delta^e$
1981	0.4535	1.2885	0.3024	3.0078	0.1157
1982	0.4582	1.3039	0.3013	3.0006	0.1215
1983	0.4628	1.3188	0.3004	2.9929	0.1274
1984	0.4679	1.3331	0.3001	2.9820	0.1331
1985	0.4732	1.3461	0.3004	2.9679	0.1386
1986	0.4782	1.3571	0.3010	2.9531	0.1433
1987	0.4824	1.3648	0.3014	2.9376	0.1471
1988	0.4859	1.3686	0.3021	2.9204	0.1496
1989	0.4890	1.3695	0.3033	2.9026	0.1506
1990	0.4921	1.3684	0.3052	2.8826	0.1506
1991	0.4954	1.3655	0.3076	2.8593	0.1495
1992	0.4997	1.3612	0.3104	2.8294	0.1479
1993	0.5051	1.3570	0.3139	2.7936	0.1459
1994	0.5113	1.3539	0.3176	2.7570	0.1439
1995	0.5173	1.3516	0.3206	2.7226	0.1424
1996	0.5230	1.3499	0.3230	2.6898	0.1417
1997	0.5282	1.3481	0.3246	2.6592	0.1419
1998	0.5323	1.3454	0.3251	2.6312	0.1425
1999	0.5349	1.3406	0.3238	2.6062	0.1432
2000	0.5362	1.3331	0.3208	2.5821	0.1433
2001	0.5370	1.3244	0.3169	2.5580	0.1422
2002	0.5378	1.3159	0.3126	2.5361	0.1398
2003	0.5379	1.3076	0.3079	2.5184	0.1364
2004	0.5374	1.2996	0.3031	2.5051	0.1325
2005	0.5360	1.2918	0.2984	2.4965	0.1284
2006	0.5343	1.2842	0.2936	2.4905	0.1243

Note: Estimates of steady-state Pareto inequality and the income process parameters in the SSA data based on moments provided by Guvenen, Karahan, Ozkan and Song (2016). Parameters are smoothed using and HP-filter with smoothing parameter 100. The estimates of  $\sigma_H$  do not include the correction for permanent versus transitory shocks. This table comes from the program "GuvenenEta.m."

## C Appendix: Estimating the U.S. Stochastic Income Process using the IRS Data

### C1. Data

The data we use are the U.S. Internal Revenue Service public use tax model panel files created by the Statistics of Income Division from 1979 to 1990, hosted by the NBER. See <http://www.nber.org/taxsim-notes.html>. We restrict our sample to tax units that involved married taxpayers filing jointly in two consecutive years and convert nominal values to 2012 constant dollars using the consumer price index. In our model, the Pareto distribution applies to “normalized” incomes, i.e. netting out the effect of aggregate growth. For this reason, we divide our micro income observations by average taxable income in each year, excluding capital gains. We use the series from Table A0 in the updated spreadsheet for Piketty and Saez (2003). For each pair of consecutive years, we record initial income and the change in log income for each tax unit. This constitutes our main data used in the estimation. In the main case, we use only tax units in the Top 5 percent of our income measure; small sample sizes make it hard to consider narrower measures.

### C2. Estimation

For each set of consecutive years, we use our cross-section of income levels and growth rates to estimate the key parameters governing the stochastic process for top incomes in our model. Hence, our key parameters are indexed by time. We assume that any growth rates that reduce income by more than  $\Delta$  percent are due to the destruction shocks. For our benchmark estimates, we assume  $\Delta = 40$ . Our estimate of  $\delta_t^c$  is therefore the fraction of growth rates that reduce incomes by more than  $\Delta$  percent.

To estimate  $\tilde{\mu}_H$  and  $\sigma_H$ , we look at growth rates for the Top 5% of earners. From this distribution of growth rates, we focus on the upper tail. In particular, we estimate  $\tilde{\mu}_H$  as the median of growth rates above the 90th percentile of the growth rate distribution, i.e. as the growth rate at the 95th percentile. We estimate  $\sigma_H$  as the standard deviation of growth rates above the 90th percentile of the growth rate distribution. In an earlier version of this paper, with different estimation, we computed confidence intervals using bootstraps. These were generally wide, and we expect the same thing is true here

Table C1: IRS Data Parameter Estimates (Wages, Salaries, and Entrepreneurial Income)

Year	$\tilde{\mu}_H$	$\sigma_H$	$\delta^e$	$N$
1980	0.3369	0.1357	0.0978	1063
1981	0.3380	0.1540	0.0951	1041
1982	0.4273	0.1437	0.0773	220
1983	0.2762	0.1181	0.1075	214
1984	0.2405	0.1168	0.0762	210
1985	0.3647	0.1353	0.0897	223
1986	0.3472	0.1294	0.0667	225
1987	0.4978	0.2273	0.1233	227
1988	0.5584	0.2228	0.0996	452
1989	0.3969	0.1540	0.1341	440
1990	0.4768	0.1765	0.0897	457

Note: Estimates of the income process parameters when income includes both wages and salaries as well as entrepreneurial income in the IRS panel data. See text for details. The sample size  $N$  in the last column is the number of observations in our panel for which we can compute a growth rate for a Top 5% earner. The 95th percentile of the growth rate distribution will then be the  $.05 \cdot N$ th highest growth rate, typically around 106th or 46th from the top for the key 1979–81 and 1988–90 periods. The estimates of  $\sigma_H$  include the correction for permanent versus transitory shocks. These results are reported by the program “PanelMuH.m.”

given the small numbers of observations.

Table C1 shows the details of our baseline estimates, year by year using the panel data, where the income measure includes both wages and salaries as well as entrepreneurial income. Table C2 shows the results when we restrict our income measure to entrepreneurial income only. Table C3 shows the implied steady-state Pareto inequality for various income measures and cutoffs to illustrate robustness.

Table C2: Parameter Estimates (Entrepreneurial Income only)

Year	$\tilde{\mu}_H$	$\sigma_H$	$\delta^e$	$N$
1980	0.4436	0.1592	0.2338	1018
1981	0.3963	0.1657	0.2403	1003
1982	0.3085	0.1265	0.1628	215
1983	0.4057	0.2213	0.1770	209
1984	0.4877	0.1280	0.1748	206
1985	0.2825	0.1178	0.1800	200
1986	0.6407	0.2414	0.1185	211
1987	0.5221	0.1683	0.2201	209
1988	0.7672	0.2010	0.1737	380
1989	0.5035	0.2074	0.2163	393
1990	0.6734	0.1846	0.2575	400

Note: Estimates of the income process parameters for entrepreneurial income only in the IRS panel data. See notes to Table C1.



Table C3: Implied Pareto Inequality from IRS Panel Data

Income concept	Top income threshold	$\delta$ threshold	Pareto Inequality, $\eta$	
			1979–81	1988–90
Wages/Salaries + Entr. Income	95	40	0.41	0.52
Wages and Salaries	95	40	0.39	0.50
Entrepreneurial Income only	95	40	0.44	0.61
Wages/Salaries + Entr. Income	90	40	0.37	0.46
Wages/Salaries + Entr. Income	97	40	0.45	0.56
Wages/Salaries + Entr. Income	95	50	0.42	0.54

Note: The last two columns of the table report implied steady-state Pareto inequality based on parameter estimates from the IRS panel data, computed under various assumptions. The first column shows the different income concepts used. The second column reports the percentile threshold used to define “top earner.” The third column of the table reports the percentile cutoff of the growth rate distribution used to estimate  $\bar{\delta}$ . The values for  $\bar{p}$  and  $\bar{\delta}$  are held fixed at 1.0 and 0.02 respectively.

## D Appendix: Proofs of the Propositions

This appendix contains outlines of the proofs of the propositions reported in the paper.

### Proof of Proposition 1. Entrepreneurial Effort

The first order condition for the Bellman equation (12) yields

$$\frac{\beta}{\Omega - e^*} = \phi V_x(x_t, t)x_t, \quad (\text{D1})$$

where  $e^*$  denotes the optimal level of entrepreneurial effort.

Next, we conjecture that the value function takes the form of  $V(x_t, t) = \zeta_0 + \zeta_1 t + \zeta_2 \log x_t$  for some constants  $\zeta_0$ ,  $\zeta_1$ , and  $\zeta_2$ . We then rewrite (D1) as

$$\frac{\beta}{\Omega - e^*} = \phi \zeta_2. \quad (\text{D2})$$

Now substituting (D2) into (12), we have

$$\begin{aligned} & (\rho + \delta + \bar{\delta})(\zeta_0 + \zeta_1 t + \zeta_2 \log x_t) \\ &= \beta \log \left( \frac{\beta}{\phi \zeta_2} \right) + \zeta_2 \phi e^* - \frac{1}{2} \zeta_2 \sigma^2 + \zeta_1 + (\delta + \bar{\delta})V^w(t) + \log \psi_t + \log x_t \end{aligned} \quad (\text{D3})$$

Equating the coefficients on  $\log x_t$  yields

$$\zeta_2 = \frac{1}{\rho + \delta + \bar{\delta}}.$$

We then substitute  $\zeta_2$  into (D2) to obtain

$$e^* = \Omega - \frac{1}{\phi} \beta (\rho + \delta + \bar{\delta}).$$

To complete the proof, we next outline how to solve for  $\zeta_0$  and  $\zeta_1$  by showing that the right-hand side of (D3) has the same form as our conjecture. As we later show in the proof of Proposition 5,  $V^w(t) = \frac{1}{\rho} \log w_t + \frac{g}{\rho^2}$ , where  $g$  is some constant. Moreover, (25) and (27) imply that both  $\log w_t$  and  $\log \psi_t$  are linear functions of  $n_t$ . Since  $\dot{n}_t$  is constant in the stationary equilibrium,  $\log w_t$  and  $\log \psi_t$  are linear in  $t$ . Therefore, the right-hand side of (D3) will have the same form as our conjecture, and we obtain  $\zeta_0$  and  $\zeta_1$  by

equating the coefficients and constants in (D3). QED.

### Proof of Proposition 2. The Pareto Income Distribution

Substituting our guess  $f(x) = Cx^{-\xi-1}$  to (15), we obtain

$$0 = -\bar{\delta}f(x) + \xi\mu f(x) + \frac{1}{2}\xi(\xi - 1)\sigma^2 f(x).$$

To make this equation hold for every  $x$ , we require

$$\frac{1}{2}\xi(\xi - 1)\sigma^2 + \mu\xi - \bar{\delta} = 0.$$

Solving this equation for  $\xi$ , we obtain the positive root in (17). QED.

### Proof of Proposition 3. Pareto Inequality with Heterogeneous Mean Growth Rates

The heterogeneous random growth model differs from the baseline model by having entrepreneurs heterogeneous not only in productivity  $x$  but also in the growth state, which the growth parameter  $\phi$  and the variance of the shocks  $\sigma^2$  depend on. We explain in detail here how the growth state of an entrepreneur is determined.

When an entrepreneur is replaced by exogenous destruction (the Poisson process at rate  $\bar{\delta}$ ), the new replacing entrepreneur, who starts with an initial productivity  $x_0$ , inherits the high growth state  $(\phi_H, \sigma_H)$  with probability  $\bar{q}$  and the low growth state  $(\phi_L, \sigma_L)$  with probability  $1 - \bar{q}$ . On the other hand, in the case of replacement by endogenous creative destruction (the Poisson process at rate  $\delta$ ), the new entrepreneur inherits the growth state as well as the quality  $x$  of the replaced entrepreneur. Therefore, the distribution of high-growth and low-growth entrepreneurs as well as the distribution of  $x$  are not mechanically altered by creative destruction. In addition, we assume that high-growth entrepreneurs transition to the low-growth state following a Poisson process with arrival rate  $\bar{p}$ .

We summarize the Poisson arrival rates of the state transition events in the following

table.

(From \ To)	High growth	Low growth	Workers
High growth		$\bar{p}$	$\delta + \bar{\delta}$
Low growth	0		$\delta + \bar{\delta}$
Researchers	$\bar{q}\bar{\delta}_R + \pi_H\delta_R$	$(1 - \bar{q})\bar{\delta}_R + (1 - \pi_H)\delta_R$	

where  $\pi_H$  is the measure of the high-growth entrepreneurs and  $\delta_R = \delta/R$ ,  $\bar{\delta}_R = \bar{\delta}/R$ .

In addition, unlike in the baseline model, we do not assume a minimum possible productivity level, or a reflecting barrier,  $x_0$  to simplify the analysis.

In this proof, we first solve for the optimal entrepreneurial effort to specify the geometric Brownian motion process that productivity  $x$  follows for each state. We then solve for the stationary distribution of  $x$  to determine Pareto inequality.

We start from the optimal effort of entrepreneurs in the low-growth state. The Bellman equation for an entrepreneur in the low-growth state is given by

$$\rho V^L(x_t, t) = \max_e \log \psi_t + \log x_t + \beta \log(\Omega - e_t) + \frac{\mathbb{E}[dV^L(x_t, t)]}{dt} + (\delta + \bar{\delta})(V^w(t) - V^L(x_t, t)). \quad (\text{D4})$$

Notice that the Bellman equations (D4) and (12) differ only in the subscripts for the state-specific variables. Therefore, from the proof of Proposition 1, we find that (1) the value function takes the form of  $V^L(x_t, t) = \zeta_0 + \zeta_1 t + \zeta_2 \log x_t$  for some constants  $\zeta_0, \zeta_1$ , and  $\zeta_2 = \frac{1}{\rho + \delta + \bar{\delta}}$ , and (2)  $e_L^* = (1 - \tau) - \frac{1}{\phi_L} \beta(\rho + \delta + \bar{\delta})$ .

Next, the Bellman equation for an entrepreneur in the high-growth state is given by

$$\rho V^H(x_t, t) = \max_e \log \psi_t + \log x_t + \beta \log(\Omega - e_t) + \frac{\mathbb{E}[dV^H(x_t, t)]}{dt} + \bar{p}(V^L(x_t, t) - V^H(x_t, t)) + (\delta + \bar{\delta})(V^w(t) - V^H(x_t, t)). \quad (\text{D5})$$

Applying the same form of conjecture on  $V^H(x_t, t)$  and substituting in  $V^L(x_t, t)$ , we find that the optimal effort of entrepreneurs in the high-growth state is  $e_H^* = (1 - \tau) - \frac{1}{\phi_H} \cdot \beta(\rho + \delta + \bar{\delta})$ .

The optimal entrepreneurial efforts  $e_H^*, e_L^*$  pin down the mean growth rates in the geometric Brownian motion process (10) to have  $\mu_H^* = \mu(e_H^*) = \phi_H e_H^* = \phi_H(1 - \tau) - \beta(\rho + \delta + \bar{\delta})$  and  $\mu_L^* = \mu(e_L^*) = \phi_L e_L^* = \phi_L(1 - \tau) - \beta(\rho + \delta + \bar{\delta})$ .

Now we've set up the stochastic processes for  $x$  in each state so that we can move

on to study the density of the distribution  $f(x, t)$ . The remainder of the proof closely follows the heterogeneous mean growth model in Gabaix, Lasry, Lions and Moll (2016).

We change the variable of interest to “ $y \equiv \log x$ ”, which simplifies the mathematical analysis that follows. By applying Ito’s formula to  $dx_t = \mu(e_t)x_t dt + \sigma x_t dB_t$ , we obtain the following stochastic process for  $y$ :

$$dy = \tilde{\mu} dt + \sigma dB_t, \text{ where } \tilde{\mu} = \mu(e_t) - \frac{\sigma^2}{2}. \quad (\text{D6})$$

Let the density  $g(y, t) = g_H(y, t) + g_L(y, t)$ , where  $g_H(y, t)$  and  $g_L(y, t)$  are the densities of entrepreneurs in the high and low growth states, respectively. Then the densities satisfy the following Kolmogorov forward equations:

$$\frac{\partial g_H(y, t)}{\partial t} = -(\bar{\delta} + \bar{p})g_H(y, t) - \frac{\partial}{\partial y} [\tilde{\mu}_H g_H(y, t)] + \frac{1}{2} \cdot \frac{\partial^2}{\partial y^2} [\sigma^2 g_H(y, t)] + \bar{q}\bar{\delta}\delta_0(y) \quad (\text{D7})$$

$$\frac{\partial g_L(y, t)}{\partial t} = -\bar{\delta}g_L(y, t) + \bar{p}g_H(y, t) - \frac{\partial}{\partial y} [\tilde{\mu}_L g_L(y, t)] + \frac{1}{2} \cdot \frac{\partial^2}{\partial y^2} [\sigma^2 g_L(y, t)] + (1 - \bar{q})\bar{\delta}\delta_0(y), \quad (\text{D8})$$

where  $\delta_0$  is a Dirac delta function (a point mass function at  $y = 0$ , or equivalently at  $x = 1$ , which can be interpreted as the case when  $x_0$  is normalized to be 1).

By applying the Laplace transform  $\hat{g}(s, t) = E[e^{-sy}]$  to equations (D7) and (D8), we obtain the new system of Kolmogorov forward equations in the following.

$$\begin{aligned} \frac{\partial \hat{g}_H(s, t)}{\partial t} &= -(\bar{\delta} + \bar{p})\hat{g}_H(s, t) - [s\tilde{\mu}_H \hat{g}_H(s, t)] + \frac{1}{2} \cdot [s^2 \sigma^2 \hat{g}_H(s, t)] + \bar{q}\bar{\delta}, \\ \frac{\partial \hat{g}_L(s, t)}{\partial t} &= -\bar{\delta}\hat{g}_L(s, t) + \bar{p}\hat{g}_H(s, t) - [s\tilde{\mu}_L \hat{g}_L(s, t)] + \frac{1}{2} [s^2 \sigma^2 \hat{g}_L(s, t)] + (1 - \bar{q})\bar{\delta}, \end{aligned}$$

which is a system of ODE. This can be rewritten in a matrix form as  $\hat{g}_t = A\hat{g} + \beta$ , where

$$A = \begin{pmatrix} -(\bar{\delta} + \bar{p}) - s\tilde{\mu}_H + s^2 \frac{\sigma_H^2}{2} & 0 \\ \bar{p} & -\bar{\delta} - s\tilde{\mu}_L + s^2 \frac{\sigma_L^2}{2} \end{pmatrix} \equiv \begin{pmatrix} -\lambda_H(s) & 0 \\ \bar{p} & -\lambda_L(s) \end{pmatrix}, \beta = \begin{pmatrix} \bar{q}\bar{\delta} \\ (1 - \bar{q})\bar{\delta} \end{pmatrix}.$$

If a stationary distribution exists,  $\lim_{t \rightarrow \infty} \hat{g}_H(s, t) = \hat{g}_H(s)$  and  $\lim_{t \rightarrow \infty} \hat{g}_L(s, t) = \hat{g}_L(s)$  exist. Therefore, we obtain  $\hat{g}_H(s)$  and  $\hat{g}_L(s)$  by solving  $\hat{g}_t = A\hat{g} + \beta = 0$ . The

corresponding stationary distribution  $\hat{g}(s) = \hat{g}_H(s) + \hat{g}_L(s)$  is then given by

$$\begin{pmatrix} \hat{g}_H(s) \\ \hat{g}_L(s) \end{pmatrix} = \begin{pmatrix} \frac{\bar{q}\bar{\delta}}{\lambda_H(s)} \\ \frac{\bar{p}\bar{q}\bar{\delta}}{\lambda_H(s)\lambda_L(s)} + \frac{(1-\bar{q})\bar{\delta}}{\lambda_L(s)} \end{pmatrix}. \quad (\text{D9})$$

Note that the Laplace transform  $\hat{g}(s)$  is ‘ $-s$ ’-moment of the stationary distribution of income  $f(x)$  as  $\hat{g}(s) = E[e^{-sy}] = E[(e^y)^{-s}] = E[x^{-s}]$ . Moreover, for any distribution with a Pareto tail  $\xi$ , higher moments than  $\xi$  is infinite. Using this moment condition, we obtain the Pareto tail index  $\xi$  of the income distribution  $f(x)$ ,  $\xi = \min\{\xi_H, \xi_L\}$ , where  $\xi_H = -\frac{\tilde{\mu}_H^*}{\sigma_H^2} + \sqrt{\left(\frac{\tilde{\mu}_H^*}{\sigma_H^2}\right)^2 + \frac{2(\bar{\delta}+\bar{p})}{\sigma_H^2}}$  is a positive root of  $\lambda_H(-\xi) = -(\bar{\delta} + \bar{p}) + \xi\tilde{\mu}_H + \xi^2\frac{\sigma_H^2}{2} = 0$  and  $\xi_L = -\frac{\tilde{\mu}_L^*}{\sigma_L^2} + \sqrt{\left(\frac{\tilde{\mu}_L^*}{\sigma_L^2}\right)^2 + \frac{2\bar{\delta}}{\sigma_L^2}}$  is a positive root of  $\lambda_L(-\xi) = -\bar{\delta} + \xi\tilde{\mu}_L + \xi^2\frac{\sigma_L^2}{2} = 0$ .

It is easy to see that  $\xi = \min\{\xi_H, \xi_L\} = \xi_H$  when  $\phi_H$  is sufficiently large to have  $\frac{\tilde{\mu}_H^*}{\sigma_H^2} \gg \frac{\tilde{\mu}_L^*}{\sigma_L^2}$ . That is, Pareto inequality is determined by the distribution of high-growth entrepreneurs when there is a big gap in the mean growth rate (after adjusting for the variance),  $\frac{\tilde{\mu}^*}{\sigma^2}$ , between the two states.

We list here two relatively simple and intuitive example cases where high-growth entrepreneurs pin down the tail parameter. First, probably the most intuitive case we can think of where the tail of high-growth entrepreneurs is thicker than that of low-growth entrepreneurs is the case when the high growth is accompanied with slower arrival of death. Recall that the Pareto inequality is pinned down by the growth and death rates in the simple model of Section 3. It is actually easy to see that  $\xi_H$  is always smaller than  $\xi_L$  if  $\frac{\tilde{\mu}_H^*}{\sigma_H^2} > \frac{\tilde{\mu}_L^*}{\sigma_L^2}$  and  $\frac{2(\bar{\delta}+\bar{p})}{\sigma_H^2} < \frac{2\bar{\delta}}{\sigma_L^2}$ . Second, if  $\tilde{\mu}_L^* < 0 < \tilde{\mu}_H^*$  ( $\phi_L$  is small enough to make  $\tilde{\mu}_L^*$  negative), even when death arrives faster in the high-growth state, more precisely when  $\frac{2(\bar{\delta}+\bar{p})}{\sigma_H^2} > \frac{2\bar{\delta}}{\sigma_L^2}$ ,  $\xi_H$  is always smaller than  $\xi_L$  if  $\frac{\tilde{\mu}_H^*}{\sigma_H^2} > \left(\frac{\bar{\delta}+\bar{p}}{\sigma_H^2} - \frac{\bar{\delta}}{\sigma_L^2}\right) \sqrt{\frac{\sigma_L^2}{2\bar{\delta}}}$  ( $\phi_H$  is large enough to offset the faster arrival of death in the high-growth state) or  $\frac{\tilde{\mu}_L^*}{\sigma_L^2} < -\left(\frac{\bar{\delta}+\bar{p}}{\sigma_H^2} - \frac{\bar{\delta}}{\sigma_L^2}\right) \sqrt{\frac{\sigma_H^2}{2(\bar{\delta}+\bar{p})}}$  ( $\phi_L$  is small enough to offset the slower arrival of death in the low growth state). Note that  $\xi = \min\{\xi_H, \xi_L\} = \xi_H$  holds under more general parameter conditions than the ones just discussed here. **QED.**

#### **Proof of Proposition 4. Output, Wages, and Profits**

Note that we omit the time subscripts for convenience since the final goods sector’s

problem and the entrepreneurs' monopoly decisions are temporal.

We begin by solving the final goods sector's problem. A perfectly competitive final goods sector combines the varieties  $i$  of price  $p_i$  to produce the final good  $Y$ . This representative firm solves

$$\max_{Y_i, \forall i \in [0,1]} \left( \int_0^\infty Y_i^\theta di \right)^{\frac{1}{\theta}} - \int_0^\infty p_i Y_i di.$$

The demand equations for each variety  $i$  that follow from the first order conditions are

$$\left( \frac{Y}{Y_i} \right)^{1-\theta} = p_i. \quad (\text{D10})$$

Each variety  $i$  is produced by a monopolistic entrepreneur, who solves

$$\max_{Y_i} p_i(Y_i)Y_i - wL_i = Y^{1-\theta}Y_i^\theta - \frac{w}{\gamma^n x_i^\alpha} Y_i. \quad (\text{D11})$$

The solution involves a usual monopoly markup  $\frac{1}{\theta}$  over marginal cost and is given by

$$Y_i = \left( \frac{1}{\theta} \frac{w}{\gamma^n x_i^\alpha} \right)^{\frac{1}{\theta-1}} Y. \quad (\text{D12})$$

By plugging (D12) in the final goods production function, we obtain the equilibrium wage equation

$$w = \theta \gamma^n \left( \int_0^1 x_i^{\alpha \frac{\theta}{1-\theta}} \right)^{\frac{1-\theta}{\theta}} \equiv \theta \gamma^n X^\alpha, \quad (\text{D13})$$

where we assume  $\alpha \frac{\theta}{1-\theta} = 1$  and  $X \equiv \int_0^1 x_i di$ . Using this equation we can rewrite (D12) as

$$Y_i = \left( \frac{x_i}{X} \right)^{\frac{1}{\theta}} Y. \quad (\text{D14})$$

Next, combining (D14) and (20) to get an expression for  $L_i$  and substituting this into the labor market clearing condition  $\int_0^1 L_i di = L$  yields the following equation for the final output  $Y$ :

$$Y = \gamma^n X^\alpha L. \quad (\text{D15})$$

Lastly, the profit  $\pi_i$  is calculated from plugging the optimal solution (D10), (D12), and (D14) into the monopoly problem (D11). QED.

**Proof of Proposition 5.** Growth and inequality in the  $\bar{s}$  case

The proof is provided in the main text. QED.

**Proof of Proposition 6.** Allocation of Labor

**Allocation of labor in the baseline model**

To solve the indifference equation  $V^w(t) = V^R(t)$ , we begin by studying the value of being a worker  $V^w(t)$ . The value function given in (30) can be rewritten as

$$V^w(t) = \int_t^\infty \exp^{-\rho(\tau-t)} \log w_\tau d\tau = \frac{1}{\rho} \log w_t + \frac{g}{\rho^2}, \quad (\text{D16})$$

where the last equality comes from the fact that  $w_t$  given in Proposition 4 grows at the constant rate of growth  $g \equiv n_t \log \gamma$  in the stationary general equilibrium. Note that  $dV^w(t)/dt = \frac{g}{\rho}$ .

We next derive the value of being a researcher  $V^R(t)$ .  $V^R(t)$  given in (31) suggests that we start from studying the value function for an entrepreneur to get  $\mathbb{E}[V(x_t, t)]$  and  $V(x_0, t)$ . Recall that the value function for an entrepreneur with quality  $x_t$  is given in (12) and (D3). We rewrite (D3) as

$$(\rho + \delta + \bar{\delta})V(x_t, t) = \log \psi_t + \log x_t + C + (\delta + \bar{\delta})V^w(t) + \frac{dV(x, t)}{dt}, \quad (\text{D17})$$

where  $C = \beta \log(\Omega - e^*) + \frac{\phi e^* - \frac{1}{2}\sigma^2}{\rho + \delta + \bar{\delta}}$  contains constant terms. Differentiating (D17) with respect to time, we obtain

$$\frac{dV(x, t)}{dt} = \frac{1}{\rho + \delta + \bar{\delta}} \left( \frac{\dot{\psi}_t}{\psi_t} + (\delta + \bar{\delta}) \frac{dV^w}{dt} \right) = \frac{g}{\rho}. \quad (\text{D18})$$

The last equality comes from the fact that  $\psi_t = \frac{1-\theta}{\theta} \frac{w_t L_t}{X_t}$  and  $X_t = \int_0^1 x_{it} di = \mathbb{E}[x_t] = \frac{x_0}{1-\eta}$ . Substituting (D18) and (D16) into (D17) yields

$$(\rho + \delta + \bar{\delta})V(x, t) = \log \frac{1-\theta}{\theta} + \log L_t - \log x_0 + \log(1-\eta) + \log x_t + C + (\rho + \delta + \bar{\delta})V^w(t). \quad (\text{D19})$$



Now taking expectations on (D19) and rearranging, we get

$$\mathbb{E}[V(x_t, t)] = \frac{1}{\rho + \delta + \bar{\delta}} \left( \log \frac{1-\theta}{\theta} + \log L_t + \log(1-\eta) + \eta + C \right) + V^w(t), \quad (\text{D20})$$

where we set the minimum value  $x_0 = 1$  and use  $\mathbb{E}[\log x] = \eta$  if  $x$  follows a Pareto distribution with the inequality parameter  $\eta$ . Furthermore, we know from (D19) that

$$V(x_0, t) = \frac{1}{\rho + \delta + \bar{\delta}} \left( \log \frac{1-\theta}{\theta} + \log L_t + \log(1-\eta) + C \right) + V^w(t). \quad (\text{D21})$$

We then rewrite (D20) as

$$\mathbb{E}[V(x_t, t)] = V(x_0, t) + \frac{\eta}{\rho + \delta + \bar{\delta}}. \quad (\text{D22})$$

Next substituting (D22) into (31) and rearranging, we obtain

$$(\rho + \lambda(1-\bar{z}) + \bar{\delta}_R)V^R(t) = \underbrace{\log \bar{m} + \log w_t + \frac{g}{\rho}}_{=\rho V^w(t)} + (\lambda(1-\bar{z}) + \bar{\delta}_R)V(x_0, t) + \frac{\lambda(1-\bar{z})\eta}{\rho + \delta + \bar{\delta}}. \quad (\text{D23})$$

Lastly we substitute (D21) into (D23) and apply the indifference equation  $V^w(t) = V^R(t)$  to (D23) to get

$$0 = (\rho + \delta + \bar{\delta}) \log \bar{m} + (\lambda(1-\bar{z}) + \bar{\delta}_R) \left( \log \frac{1-\theta}{\theta} + \log L_t + \log(1-\eta) + C \right) + \lambda(1-\bar{z})\eta.$$

Solving the last equation for  $\log L_t$ , we finally obtain the allocation of labor to research

$s^* = 1 - \frac{L^*}{L}$ , where

$$\log L^* = \log \frac{\theta}{1-\theta} - \frac{(\rho + \delta^* + \bar{\delta}) \log \bar{m} + \lambda(1-\bar{z})\eta^*}{\lambda(1-\bar{z}) + \bar{\delta}_R} - \beta \log(1 - \tau - e^*) - \frac{\phi e^* - \frac{1}{2}\sigma^2}{\rho + \delta^* + \bar{\delta}} - \log(1 - \eta^*).$$

### Allocation of labor with heterogeneous mean growth rates

The value function for a worker,  $V^w(t)$ , does not change with heterogeneous mean growth rates, thus we have  $V^w(t) = \frac{1}{\rho} \log w_t + \frac{g}{\rho^2}$  as given in the equation (D16). On the other hand, the value function for a researcher at time  $t$ ,  $V^R(t)$ , changes from equa-

tion (31) to consider the two different growth states, which is shown below.

$$\begin{aligned} \rho V^R(t) = \log(\bar{m}w_t) + \frac{dV^R(t)}{dt} + \lambda(1 - \bar{z}) (\pi_H \mathbb{E}[V^H(x, t)] + \pi_L \mathbb{E}[V^L(x, t)] - V^R(t)) \\ + \bar{\delta}_R (\bar{q}V^H(x_0, t) + (1 - \bar{q})V^L(x_0, t) - V^R(t)), \end{aligned} \quad (\text{D24})$$

where  $\pi = (\pi_H, \pi_L)$  is the distribution of high-growth and low-growth entrepreneurs. In the remainder of the proof, we will solve for each term in (D24) to explicitly solve for  $L_t$ .

Starting from the first two terms, we have

$$\log(\bar{m}w_t) + \frac{dV^R(t)}{dt} = \log \bar{m} + \log w_t + \frac{dV^w(t)}{dt} = \log \bar{m} + \rho V^w(t). \quad (\text{D25})$$

We next calculate the stationary distribution  $\pi^*$ . It can be obtained by describing the state transitions among entrepreneurs as a continuous time Markov chain, where we consider the state space {'high growth', 'low-growth'} and ignore replacements by destructions.<sup>28</sup> Specifically, the rate matrix  $Q$  of this Markov chain is given by

$$Q = \begin{array}{cc} & \begin{array}{cc} \text{High growth} & \text{Low growth} \end{array} \\ \begin{array}{c} \text{High growth} \\ \text{Low growth} \end{array} & \begin{pmatrix} -(\bar{p} + (1 - \bar{q})\bar{\delta}) & \bar{p} + (1 - \bar{q})\bar{\delta} \\ \bar{q}\bar{\delta} & -\bar{q}\bar{\delta} \end{pmatrix}. \end{array}$$

We then obtain the stationary distribution  $\pi^*$  by solving  $\pi Q = 0$  as below:

$$\pi^* = (\pi_H^*, \pi_L^*) = \left( \frac{\bar{\delta}\bar{q}}{\bar{p} + \bar{\delta}}, \frac{\bar{p}\bar{q}}{\bar{p} + \bar{\delta}} + (1 - \bar{q}) \right).$$

Moving on to the next terms  $\mathbb{E}[V^H(x, t)]$  and  $\mathbb{E}[V^L(x, t)]$ , recall that the value functions for an entrepreneur with quality  $x_t$  is given in (D5) and (D4). We rewrite them as

$$\begin{aligned} (\rho + \delta + \bar{\delta})V^H(x_t, t) = \log \psi_t + \log x_t + C_H + \bar{p}(V^L(x_t, t) - V^H(x_t, t)) \\ + (\delta + \bar{\delta})V^w(t) + \frac{dV^H}{dt}, \end{aligned} \quad (\text{D26})$$

$$(\rho + \delta + \bar{\delta})V^L(x_t, t) = \log \psi_t + \log x_t + C_L + (\delta + \bar{\delta})V^w(t) + \frac{dV^L}{dt}, \quad (\text{D27})$$

<sup>28</sup>For example, if a high-growth entrepreneur is replaced by another high-growth entrepreneur due to an exogenous destruction event, we do not consider this event as a state transition because it does not change the distribution of entrepreneurs' growth states.

where  $C_H \equiv \beta \log(\Omega - e_H^*) + \frac{\phi_H e_H^* - \frac{1}{2}\sigma_H^2}{\rho + \delta + \bar{\delta}}$  and  $C_L \equiv \beta \log(\Omega - e_L^*) + \frac{\phi_L e_L^* - \frac{1}{2}\sigma_L^2}{\rho + \delta + \bar{\delta}}$  contain constant terms. Moreover, recall that  $\psi_t = \frac{1-\theta}{\theta} \frac{w_t L_t}{X_t}$  and  $X_t = \int_0^1 x_{it} di = \mathbb{E}[x_t]$ . Differentiating (D26) and (D27) with respect to  $t$  as in (D18), we get

$$\frac{dV^H(x, t)}{dt} = \frac{dV^L(x, t)}{dt} = \frac{g}{\rho}. \quad (\text{D28})$$

We can also replace  $V^L - V^H$  in (D26) by subtracting (D26) from (D27) to get

$$V^L(x_t, t) - V^H(x_t, t) = \frac{C_L - C_H}{\rho + \delta + \bar{\delta} + \bar{p}}. \quad (\text{D29})$$

Substituting (D28), (D29), and (D16) into (D26) and (D27) yields

$$V^H(x, t) = V^w(t) + \frac{1}{\rho + \delta + \bar{\delta}} \left( \log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x_t] + \log x_t + C_H + \frac{\bar{p}(C_L - C_H)}{\rho + \delta + \bar{\delta} + \bar{p}} \right) \quad (\text{D30})$$

$$V^L(x, t) = V^w(t) + \frac{1}{\rho + \delta + \bar{\delta}} \left( \log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x_t] + \log x_t + C_L \right). \quad (\text{D31})$$

Now taking expectations and rearranging, we get

$$\begin{aligned} \pi_H \mathbb{E}[V^H(x, t)] + \pi_L \mathbb{E}[V^L(x, t)] = \\ V^w(t) + \frac{1}{\rho + \delta + \bar{\delta}} \left( \log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x_t] + \mathbb{E}[\log x_t] + D(\pi_H, \pi_L) \right), \end{aligned} \quad (\text{D32})$$

where  $D(\pi_H, \pi_L) \equiv \pi_H \frac{(\rho + \delta + \bar{\delta})C_H + \bar{p}C_L}{\rho + \delta + \bar{\delta} + \bar{p}} + \pi_L C_L$ .

Similarly, we also obtain

$$\begin{aligned} \bar{q} V^H(x_0, t) + (1 - \bar{q}) V^L(x_0, t) = \\ V^w(t) + \frac{1}{\rho + \delta + \bar{\delta}} \left( \log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x_t] + D(\bar{q}, 1 - \bar{q}) \right), \end{aligned} \quad (\text{D33})$$

where we normalized  $x_0$  to be 1.

Substituting (D25), (D32), (D33), and the indifference equation  $V^w(t) = V^R(t)$  into (D24), we get

$$\begin{aligned} 0 = (\rho + \delta + \bar{\delta}) \log \bar{m} + (\lambda(1 - \bar{z}) + \bar{\delta}_R) \left( \log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x] \right) + \\ \lambda(1 - \bar{z}) \left( \mathbb{E}[\log x] + D(\pi_H, \pi_L) \right) + \bar{\delta}_R D(\bar{q}, 1 - \bar{q}). \end{aligned} \quad (\text{D34})$$

Solving (D34) for  $\log L_t$ , we finally obtain the allocation of labor to research with het-

erogeneous mean growth rates  $s^* = 1 - \frac{L^*}{L}$ , where

$$\log L^* = \log \frac{\theta}{1-\theta} + \log \mathbb{E}[x] - \frac{(\rho + \delta^* + \bar{\delta}) \log \bar{m} + \lambda(1 - \bar{z})(\mathbb{E}[\log x] + D(\pi_H, \pi_L)) + \bar{\delta}_R D(\bar{q}, 1 - \bar{q})}{\lambda(1 - \bar{z}) + \bar{\delta}_R}.$$

Note that we can further solve for  $\mathbb{E}[\log x]$  and  $\mathbb{E}[x]$  by applying the Laplace transform  $\hat{g}(s) = E[e^{-sy}]$  of which the explicit functional form is given in (D9). Specifically, they are given by

$$\begin{aligned} \mathbb{E}[x] &= \mathbb{E}[e^y] = \hat{g}(-1) = \frac{\bar{\delta}\bar{q}}{\bar{\delta} + \bar{p} - \mu_H} + \frac{\bar{\delta}(1-\bar{q})}{\bar{\delta} - \mu_L} + \frac{\bar{\delta}\bar{q}\bar{p}}{(\bar{\delta} + \bar{p} - \mu_H)(\bar{\delta} - \mu_L)}, \\ \mathbb{E}[\log x] &= \int (\log x) f(x) dx = \int y e^{-sy} g(y) dy|_{s=0} = -\hat{g}'(0) = \frac{\bar{\delta}\bar{q}\bar{\mu}_H + (\bar{p} + \bar{\delta}(1-\bar{q}))\bar{\mu}_L}{\bar{\delta}(\bar{\delta} + \bar{p})}. \end{aligned}$$

**QED.**

In the numerical solutions that we carry out in the paper to solve for the optimal allocation of labor, there can be two values of  $s^*$  that solve the condition  $V^R(s^*) = V^w(s^*)$ . The presence of two equilibria results from the “risk-averse researchers” force that we discuss in the paper, which provides positive feedback: a decrease in  $s$  will lower growth, reduce creative destruction, and raise inequality — which in turn can reduce  $s$  further because researchers are risk averse. Of the two equilibria, one is a “low research, low growth, high inequality” steady state, while the other is “high research, high growth, low inequality.” We focus on the high growth/low inequality steady state in our comparative statics, as this one clearly Pareto dominates the low growth/high inequality steady state.