

A Schumpeterian Model of Top Income Inequality

Charles I. Jones

Stanford GSB and NBER

Jihee Kim *

KAIST

July 2, 2015 – Version 1.5

Abstract

Top income inequality rose sharply in the United States over the last 35 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, perhaps associated with globalization, may explain the varied patterns of top income inequality that we see in the data.

*We are grateful to Philippe Aghion, Jess Benhabib, Sebastian Di Tella, Xavier Gabaix, Mike Harrison, Pete Klenow, Ben Moll, Chris Tonetti, Alwyn Young, Gabriel Zucman and 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 Income Distribution group, Princeton, the SED 2015 meetings, Stanford, U.C. Santa Cruz, USC, U. Washington, Yale, Yonsei University, and Zurich for helpful comments.

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 studied 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.¹

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-

¹These 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 (2010). Gabaix (2009) and Luttmer (2010) have excellent surveys of these mechanisms, written for economists. Benhabib (2014) and Moll (2012b) 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 some kind of “tweak.” 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, best-selling books, smartphone apps, financial products, surgical techniques, or even new ways of organizing a law firm. 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 simple 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 to estimate several of the key parameters of the model, illustrating that the mechanism is economically significant. A final section studies numerical examples to highlight several additional quantitative possibilities of the 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 (2011) 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 (2009) focus explicitly on finance and the extent to which rising rents in that sector can explain rising inequality; see also Bell and Van Reenen (2010). 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 (2013) 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 the concluding 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.²

The most closely-related papers to this one are Benhabib, Bisin and Zhu (2011), Nirei (2009), Aoki and Nirei (2013), Moll (2012a), Piketty and Saez (2012), Piketty and Zucman (2014), and Toda (2014). These papers study economic mechanisms that generate endogenously a Pareto distribution for wealth, and therefore for capital income. The mechanism responsible for random growth in these papers is either the asset ac-

²Luttmer (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 (2012) derive a Zipf distribution in the upper tail for firm productivity in an endogenous growth setting.

cumulation equation (which naturally follows a random walk when viewed in partial equilibrium) or the capital accumulation equation in a neoclassical growth model. The present paper differs most directly by focusing on labor income rather than wealth. 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. Geerolf (2014) connects both top income inequality and firm size inequality in a Garicano (2000)-style model of hierarchies.

Finally, Gabaix, Lasry, Lions and Moll (2015) 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.

2. Some Basic Facts

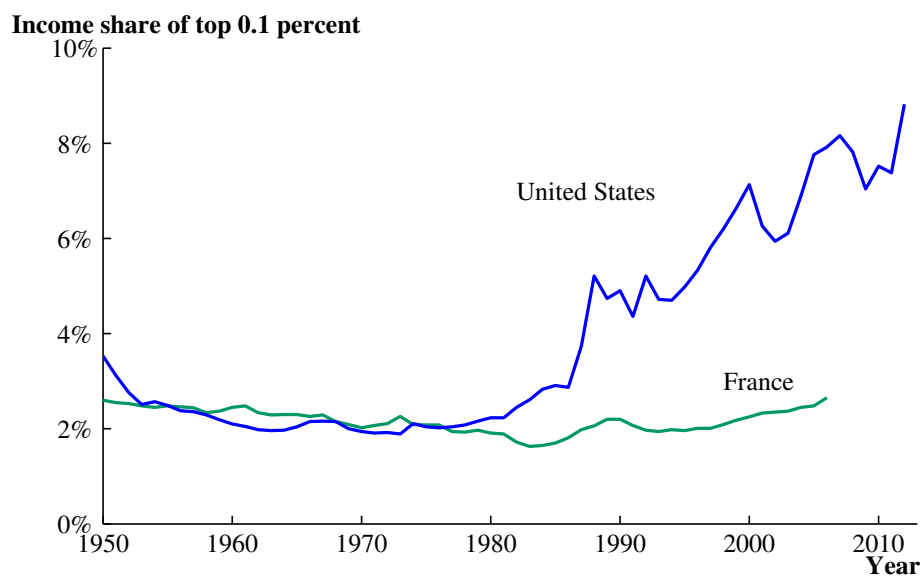
Figures 1 and 2 show some of the key facts about top income inequality that have been documented by Piketty and Saez (2003) and Atkinson, Piketty and Saez (2011). For example, the first figure shows the large increase in top inequality for the United States since 1980, compared to the relative stability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

Figure 2 shows the dynamics of top income inequality for a range of countries. The horizontal axis shows the share of aggregate income going to the top 1% of earners, averaged between 1980 and 1982, while the vertical axis shows the same share for 2006–2008. 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 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), 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-

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



Source: Alvaredo, Atkinson, Piketty and Saez (2013), World Top Incomes Database.

corporations) in the labor income category. Figure 3 shows an updated version of their graph for the period since 1950, supporting this observation.

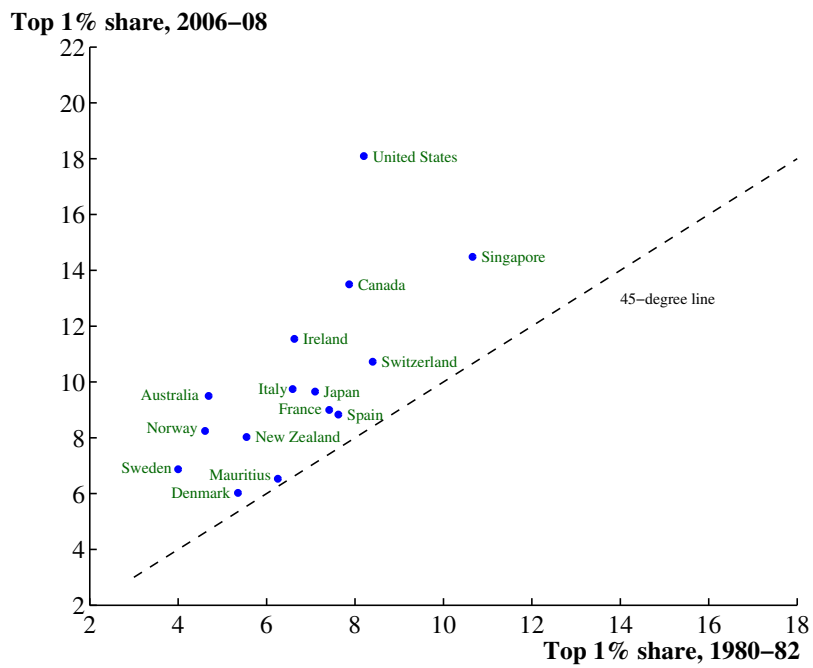
Because the model in this paper is based on labor income as opposed to capital income, documenting the Pareto nature of labor income inequality in particular is also important. It is well known, dating back to Pareto (1896), that the top portion of the income distribution can be characterized by a power 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 ξ is called the “power law exponent.”

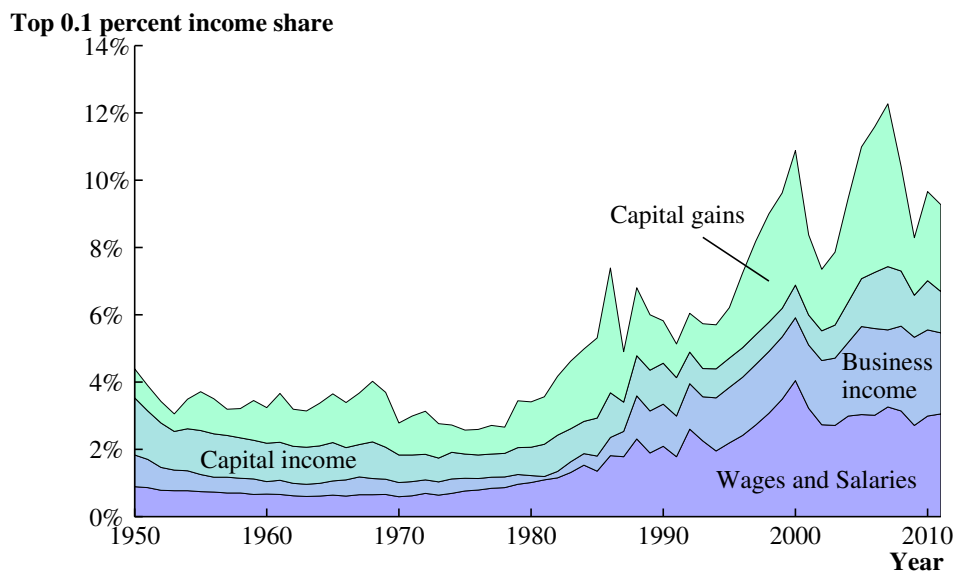
Saez (2001) shows that wage and salary income from U.S. income tax records in the early 1990s is well-described by a Pareto distribution. Figure 4 replicates his analysis for 1980 and 2005. In particular, the figures graph mean wage income above some threshold as a ratio to the threshold itself. If wage income obeys a Pareto distribution

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



Note: Top income inequality has increased since 1980 in most countries for which we have data. The size of the increase varies substantially, however. Source: World Top Incomes 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 September 2013 update of the spreadsheet appendix to Piketty and Saez (2003).

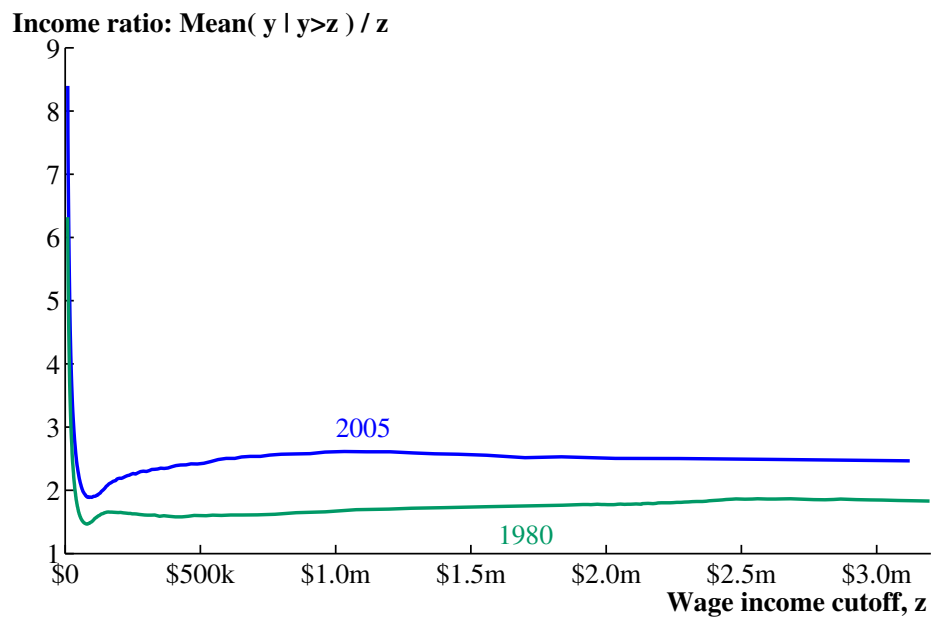
like that in (1), then this ratio should equal the constant $\frac{\xi}{\xi-1}$, *regardless of the threshold*. That is, as we move to higher and higher income thresholds, the ratio of average income above the threshold to the threshold itself should remain constant.³ Figure 4 shows that this property holds reasonably well in 1980 and 2005, and also illustrates that the ratio has risen substantially over this period, reflecting the rise in top wage income inequality.

2.2. Fractal Inequality and the Pareto Distribution

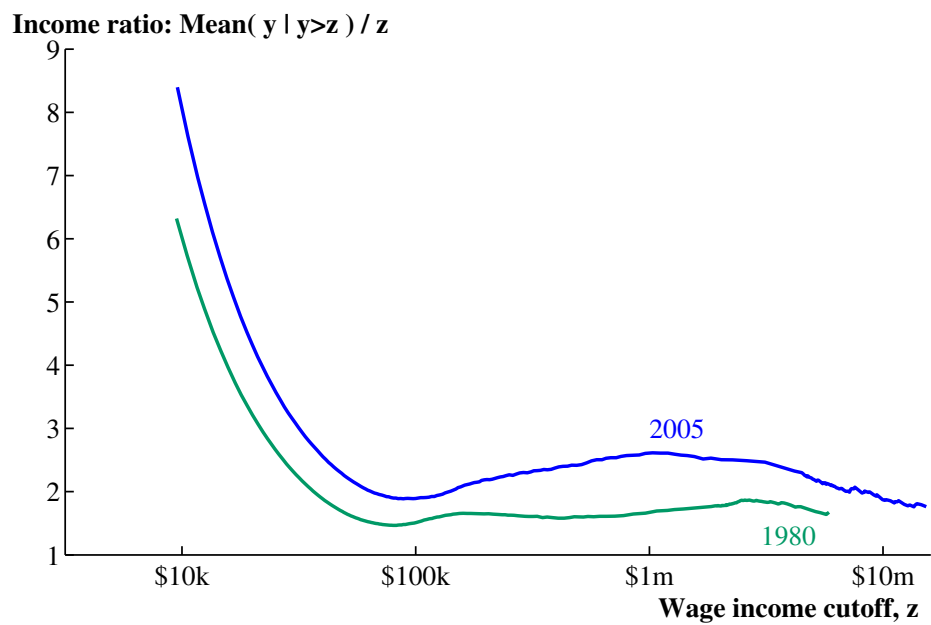
An important property of Pareto distributions is that they exhibit a fractal pattern of top inequality. To see this, let $\tilde{S}(p)$ denote the share of income going to the top p percentiles. For the Pareto distribution defined in equation (1) above, this share is given by $(p/100)^{1-1/\xi}$. A larger power-law exponent, ξ , is associated with lower top income

³This 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 Wage Income



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



(b) Log scale

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

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 $100^{-1/4} \approx 0.32$.

To see the fractal structure of top inequality under a Pareto distribution, 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,

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

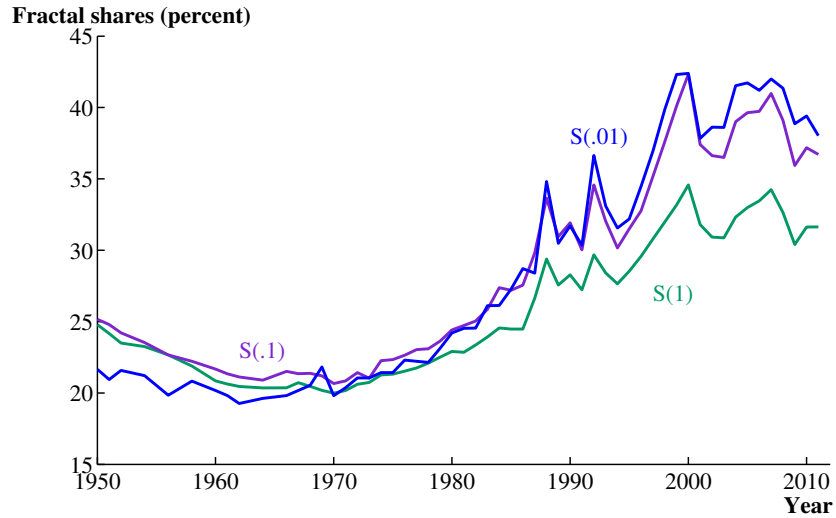
Notice that this last result holds for all values of a , or at least for all values for which income follows a Pareto distribution. This means that top income inequality obeys a *fractal* pattern: the fraction of the Top 10 percent’s income going to the Top 1 percent is the same as the fraction of the 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.⁴ Figure 5 shows the $S(a)$ shares directly. At the very top, the fractal prediction holds remarkably well, and $S(.01) \approx S(.1) \approx S(1)$. This is another indication that wage income is well-described as a Pareto distribution. 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 2010, this fractal share is closer to 40 percent.

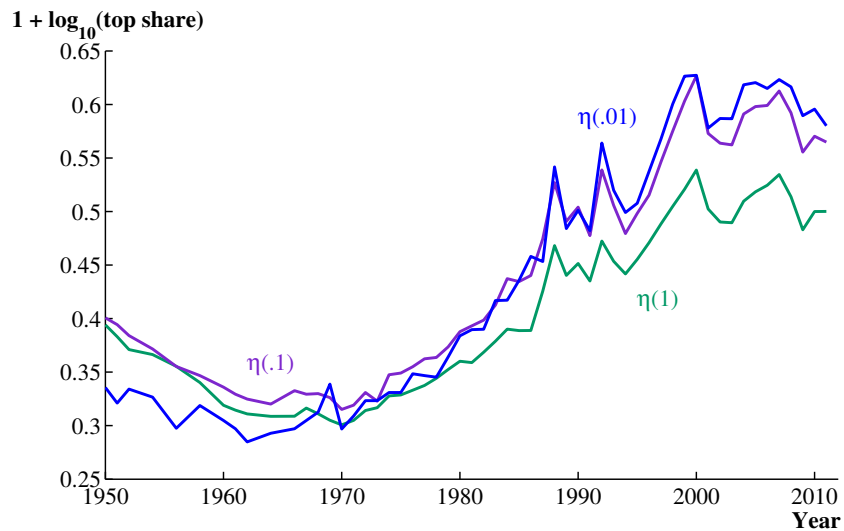
The rise in fractal inequality shown in Figure 5 can be related directly to the power-law inequality exponent using equation (4). Or, put another way, the *change* in the

⁴Others have noticed this before. For example, see Aluatiun.wordpress.com (2011).

Figure 5: Fractal Inequality of U.S. Wage 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: Underlying wage income shares are from the September 2013 update of the Piketty and Saez (2003) spreadsheet appendix.

Figure 6: The Power-Law Inequality Exponent η , United States

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

fractal shares is precisely equal to the change in the power law inequality exponent:

$$\Delta \log_{10} S(a) = \Delta \eta. \quad (5)$$

The corresponding Pareto 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 generates a Pareto exponent that rises from around 0.33 to around 0.55 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 α for now). If the distribution of skill across people is Pareto with inequality parameter η_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 η_y : a simple story of skill-biased technical change is not enough.

Notice that if the exponent α 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 α 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 α to change over time or differ across countries.

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 0.33 in the 1970s to about 0.55 by 2010. The remainder of this paper develops and analyzes a model to help us understand these facts.

3. A Simple Model of Top Income Inequality

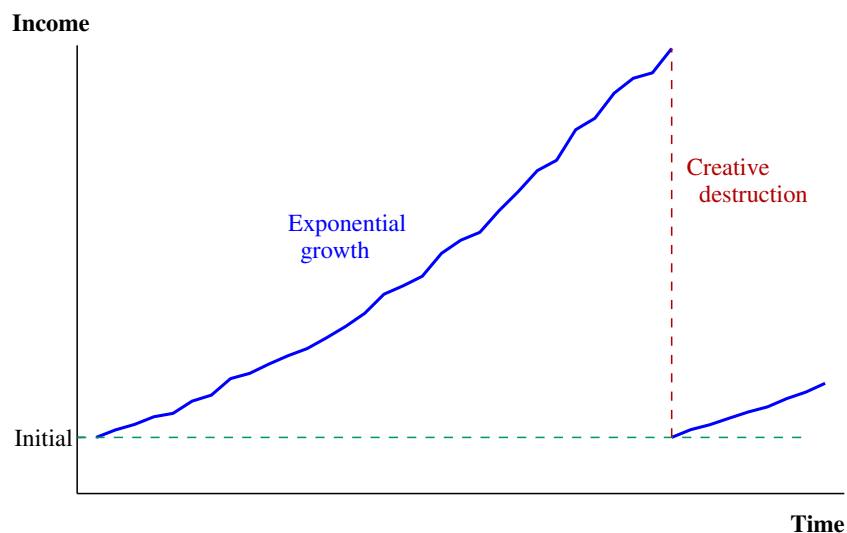
It is well-known that exponential growth and Pareto distributions are tightly linked, and this link is at the heart of the main mechanism in this paper. To illustrate this point in the clearest way, we begin with a simple, stylized model, illustrated graphically in Figure 7.⁵

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 μ , 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 δ 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

⁵See 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

$x(y)$ years, where

$$x(y) = \frac{1}{\mu} \log \left(\frac{y}{y_0} \right). \quad (6)$$

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 for the moment that the stationary distribution of experience across a population of entrepreneurs will be this same exponential distribution; this is shown more formally in Appendix A. Then, the remainder of the argument is straightforward:

$$\begin{aligned} \Pr[\text{Income} > y] &= \Pr[\text{Experience} > x(y)] \\ &= e^{-\delta x(y)} \\ &= \left(\frac{y}{y_0} \right)^{-\frac{\delta}{\mu}} \end{aligned} \quad (7)$$

which is a Pareto distribution!

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

$$\eta_y = \frac{\mu}{\delta}. \quad (8)$$

Top income inequality can therefore change for two reasons. First, an increase in the growth rate of top earners, μ , 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” δ will reduce top inequality, as entrepreneurs have less time during which to build their advantage.

What are the economic determinants of μ and δ , 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 below.

3.1. Intuition

The logic of the simple model provides useful intuition about *why* the Pareto result emerges. First, in equation (6), 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.*

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 μ and δ 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. In the full model, effort by entrepreneurs influences the growth of their incomes. In addition, this process is assumed to be stochastic, which allows us to better match up the model with micro data on top incomes. Finally, the death rate is made to be endogenous by tying it to the process of creative destruction in a Schumpeterian endogenous growth model. The setup seems to capture 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 entrepreneur’s human capital. The entrepreneur then expends effort to improve x . We explain later how x affects firm productivity 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:

$$c_t = \psi_t x_t \tag{9}$$

$$e_t + \ell_t + \tau = 1 \tag{10}$$

$$dx_t = \mu(e_t)x_t dt + \sigma x_t dB_t \tag{11}$$

$$\mu(e) = \phi e \tag{12}$$

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 human capital x . The factor of proportionality, ψ_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 ℓ or it can be wasted, in amount τ . This could correspond to time spent addressing government regulations and bureaucratic red tape, for example.

Equation (11) 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 ϕ is a technological parameter converting effort into growth. dB_t denotes the standard normal increment to the Brownian motion. This equation is a stochastic version of the human capital accumulation process in Lucas (1988).

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 δ . 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

$$\rho V(x_t, t) = \max_e \log \psi_t + \log x_t + \beta \log(\Omega - e_t) + \frac{\mathbb{E}[dV(x_t, t)]}{dt} + (\delta + \bar{\delta})(V^w(t) - V(x_t, t)) \quad (13)$$

subject to (11), where $\Omega \equiv 1 - \tau$ and $\frac{\mathbb{E}[dV(x_t, t)]}{dt}$ is short-hand for the Ito calculus terms, i.e. $\frac{\mathbb{E}[dV(x_t, t)]}{dt} = \mu(e_t)x_t V_x(x_t, t) + \frac{1}{2}\sigma^2 x_t^2 V_{xx}(x_t, t) + V_t(x_t, t)$. $V(x, t)$ is the expected utility of an entrepreneur with quality x and rate of time preference ρ . 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 the entrepreneur's choice of effort. (Proofs of all propositions are given in the [appendix](#)).

Proposition 1 (Entrepreneurial Effort): *Entrepreneurial effort solves the Bellman problem in equation (13) and along the balanced growth path is given by*

$$e^* = 1 - \tau - \frac{1}{\phi} \cdot \beta(\rho + \delta + \bar{\delta}). \quad (14)$$

This proposition implies that entrepreneurial effort is an increasing function of the technology parameter ϕ but decreases whenever τ , β , ρ , δ , 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 human capital 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 δ) 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 human capital 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.⁶ If a variable follows a Brownian motion, like x above, the density of the distribution $f(x, t)$ satisfies a Kolmogorov forward equation. This is the stochastic generalization of an equation like (A2) that we saw in the simple model at the start of the paper. In particular, the density satisfies

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

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

$$0 = -\bar{\delta}f(x) - \frac{d}{dx} [\mu(e^*)xf(x)] + \frac{1}{2} \cdot \frac{d^2}{dx^2} [\sigma^2 x^2 f(x)] \quad (16)$$

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

⁶For more detailed discussion, see Reed (2001), Mitzenmacher (2004), Gabaix (2009), and Luttmer (2010). Malevergne, Saichev and Sornette (2010) is closest to the present setup.

obtains the following result:

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

$$F(x) = 1 - \left(\frac{x}{x_0} \right)^{-\xi^*} \quad (17)$$

where

$$\xi^* = -\frac{\tilde{\mu}^*}{\sigma^2} + \sqrt{\left(\frac{\tilde{\mu}^*}{\sigma^2} \right)^2 + \frac{2\bar{\delta}}{\sigma^2}} \quad (18)$$

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 ψ_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 δ as a reminder that this value is determined in general equilibrium as well.

Comparative statics: Taking δ^* as exogenous for the moment, the comparative static results are as follows: power-law inequality, η^* , increases if effort is more effective at growing entrepreneurial income (a higher ϕ), decreases if the time endowment is reduced by government policy (a higher τ), decreases if entrepreneurs place more weight on leisure (a higher β), and decreases if either the endogenous or exogenous rates of creative destruction rise (a higher δ^* or $\bar{\delta}$).⁷

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 (2015), the basic random growth framework that forms the heart of the model so far has trouble

⁷The effect of σ^2 on power-law inequality is more subtle. If $\eta^* > \mu^*/\bar{\delta}$, then a rise in σ^2 increases η^* . 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.

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 (2015), partly motivated by a previous draft of our paper, 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. Fortunately, this insight is strongly supported by recent empirical work. Guvenen, Karahan, Ozkan and Song (2015) show that growth rates for top earners are quite 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 (2015) and augment our basic setup to include two growth states for entrepreneurs.⁸ When researchers discover a new idea, a fraction of them inherit the high growth ϕ_H parameter. They then face a Poisson process with arrival rate \bar{p} for transitioning permanently down to the more normal ϕ_L low growth parameter. In addition, we allow the variance of the shocks to also depend on the state, distinguishing σ_H and σ_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 ϕ_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*

$$\xi_H = -\frac{\tilde{\mu}_H^*}{\sigma_H^2} + \sqrt{\left(\frac{\tilde{\mu}_H^*}{\sigma_H^2}\right)^2 + \frac{2(\bar{\delta} + \bar{p})}{\sigma_H^2}} \quad (19)$$

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

⁸The 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.

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 ϕ_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 human capital x enters the model, how x affects entrepreneurial income (the proportionality factor ψ_t), and how creative destruction δ^* 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 (1992) and Grossman and Helpman (1991). A fixed population of people choose to be basic laborers, 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:

$$Y = \left(\int_0^1 Y_i^\theta di \right)^{1/\theta}. \quad (20)$$

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

$$Y_i = \gamma^{n_t} x_i^\alpha L_i. \quad (21)$$

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

human capital enters: labor productivity depends on x_i^α . As usual, variety i 's market share is increasing in x_i .

The main resource constraint in this environment involves labor:

$$L_t + R_t + 1 = \bar{N}, \quad L_t \equiv \int_0^1 L_{it} di \quad (22)$$

A fixed measure of people, \bar{N} , are available to the economy. People can work as the raw labor making varieties, or as 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 $\bar{L} \equiv \bar{N} - 1$.

Researchers discover new ideas through a Poisson process with arrival rate λ 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

$$\dot{n}_t = \lambda(1 - \bar{z})R_t \quad (23)$$

and this also gives the rate of creative destruction:

$$\delta_t = \dot{n}_t. \quad (24)$$

4.5. The Allocation of Resources

There are 12 key endogenous variables in this economic environment: $Y, Y_i, x_i, L_i, L, R, n, \delta, e_i, c_i, \ell_i, \psi$. The entrepreneur's choice problem laid out earlier pins down c, ℓ , and e for each entrepreneur. Production functions and resource constraints determine Y, Y_i, L, x_i, n , and δ . This leaves us needing to determine R, L_i , and ψ .

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 to 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 π_i denote the profit earned by the entrepreneur selling variety i . Assume now and for the rest of the paper that $\alpha = (1 - \theta)/\theta$.⁹ The equilibrium with monopolistic competition leads to*

$$Y_t = \gamma^{n_t} X_t^\alpha L_t \quad (25)$$

$$w_t = \theta \gamma^{n_t} X_t^\alpha \quad (26)$$

$$\pi_{it} = (1 - \theta) \gamma^{n_t} X_t^\alpha \left(\frac{x_{it}}{X_t} \right) L_t \quad (27)$$

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 ϕ , 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 η) therefore has a long-run level effect in this economy, raising both output and wages.

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

$$\psi_t = (1 - \theta) \gamma^{n_t} X_t^{\alpha-1} L_t. \quad (28)$$

⁹This is merely a simplifying assumption that makes profits a linear function of x_i . It can be relaxed with a bit more algebra.

Finally, we can now 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 (25) 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.¹⁰

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

$$g_y^* = \lambda(1 - \bar{z})\bar{s}\bar{L} \log \gamma \quad (29)$$

$$\delta^* = \lambda(1 - \bar{z})\bar{s}\bar{L}. \quad (30)$$

Power-law inequality is then given by Proposition 2 or Proposition 3 with this value of δ^ .*

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 ϕ 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 now 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

¹⁰At 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. The rationalization is that this allows us to focus on steady states and avoid transition dynamics, which would require difficult numerical solutions. This is certainly one target for valuable future work.

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 human capital x . Some varieties are extraordinarily successful, while most are not. Even though an increase in ϕ increases the rate of growth of x for all entrepreneurs, this only serves to widen the stationary distribution. 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 ideas, not through the productivity growth associated with enhancing an existing idea. Loosely speaking, the model features Lucas-style growth at the micro level, but long-run macro growth is entirely Romer/Aghion-Howitt/Grossman-Helpman.

The parameters τ and β also affect top income inequality without affecting growth when \bar{s} is held constant. An increase in τ 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. A similar result obtains if two economies differ with respect to β . An economy where preferences are such that entrepreneurs put more weight on leisure will spend less time building businesses and feature lower top income inequality in the long run.

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 δ . 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

$$\rho V^w(t) = \log w_t + \frac{dV^w(t)}{dt} \quad (31)$$

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

$$\rho V^R(t) = \log(\bar{m}w_t) + \frac{dV^R(t)}{dt} + \lambda(1 - \bar{z}) (\mathbb{E}[V(x, t)] - V_t^R) + \bar{\delta}_R (\mathbb{E}[V(x_0, t)] - V^R(t)). \quad (32)$$

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 γ , and replaces some randomly-selected existing entrepreneur. Alternatively, the researcher may benefit from the exogenous misallocation 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.

Table 1: Key Equations Characterizing the Stationary General Equilibrium

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

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 (31) and (32).*

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.

The model features transition dynamics away from steady state. The key reason for this is that the initial distribution of x can be anything, and the evolution of this distribution affects aggregate variables through X_t and δ_t . All numerical results that follow will therefore have the form of comparing steady states. Gabaix, Lasry, Lions and Moll (2015) conduct a careful analysis of transition dynamics in a heterogeneous random growth model and show that in the context of the present paper, the rate of convergence is $\bar{\delta} + \bar{p}$. This rate takes the value of approximately 0.8 (i.e. 80 percent, not 0.8 percent) in our baseline calibration. Transition dynamics are therefore very rapid, making the comparison of steady states a reasonable approach.

5.1. Comparative Statics

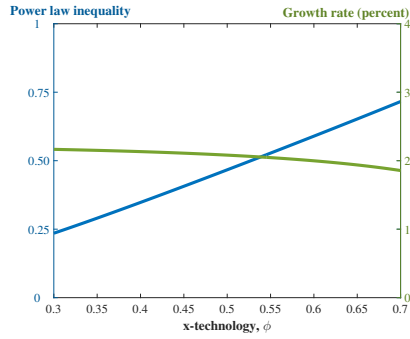
Figure 8 shows numerical results when $s \equiv R/\bar{L}$ is endogenously determined. The effects on Pareto inequality are similar to those from the exogenous \bar{s} case. Now, however, we can also study the effects on economic growth. For example, consider the effect of an increase in the technology parameter ϕ , shown in Figure 8a: an increase in ϕ 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 ϕ 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 ϕ 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 ϕ results in a smaller fraction of people becoming researchers, which lowers the long-run growth rate in this endogenous growth model.

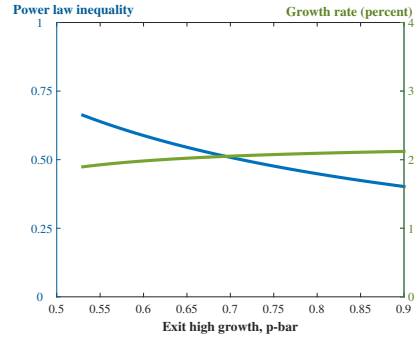
One can, of course, imagine writing down the model in a different way. For example, if research is undertaken by risk-neutral firms, then this effect would not be present. Ultimately, this question must be decided by empirical work. Our model, however, makes it clear that this additional force is present, 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 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. The implication for empirical work is complicated by level effects that would show up along any transition path, for example as discussed above with respect to a rise in ϕ .

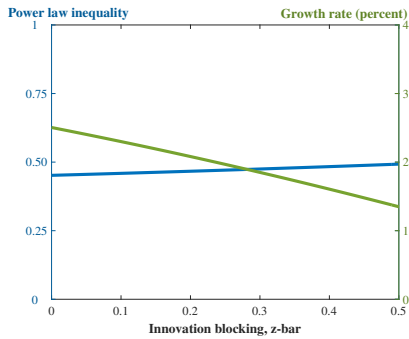
Figure 8: Numerical Examples: Endogenous s



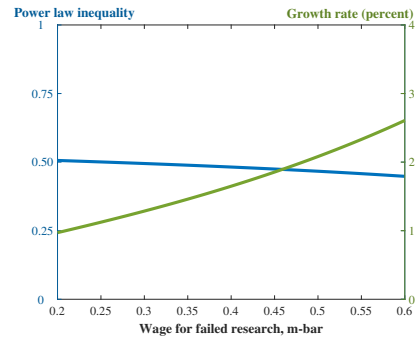
(a) Varying ϕ



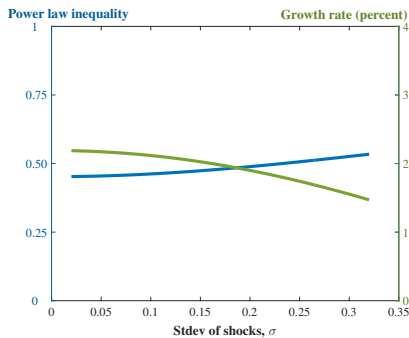
(b) Varying \bar{p}



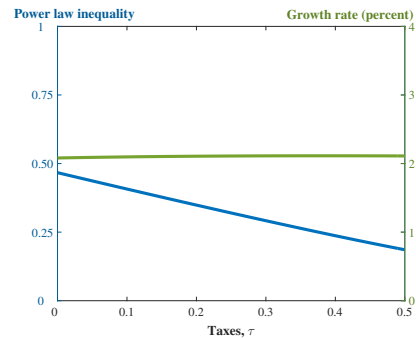
(c) Varying \bar{z}



(d) Varying \bar{m}



(e) Varying σ



(f) Varying τ

Note: The baseline parameter values in these examples are $\rho = .03$, $\bar{L} = 15$, $\theta = 2/3$, $\gamma = 1.4$, $\lambda = .027$, $\phi = .5$, $\beta = 1$, $\sigma = 0.122$, $\bar{\delta} = .04$, $\bar{m} = .5$, $\bar{z} = 0.2$, $\tau = 0$, $\bar{p} = .767$, and $\bar{q} = .504$. These values will be discussed in more detail in Section 7.

6. Micro Evidence

To what extent is our model consistent with empirical evidence? There are several ways to answer this question. Some relate to previous empirical work on income dynamics. Some relate to evidence we provide using public use micro data from the U.S. Internal Revenue Service. And some relate to evidence that we hope others with better access to administrative data can conceivably provide in the future.

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) surveys 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 between countries and over time that could be estimated.

Access to panel micro data on top income earners in the U.S. over time and in other countries is typically very hard to gain. However, we can make some progress by combining some of the moments made available by Guvenen, Karahan, Ozkan and Song (2015) with 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.¹¹

6.1. The Distribution of Top Income Growth Rates

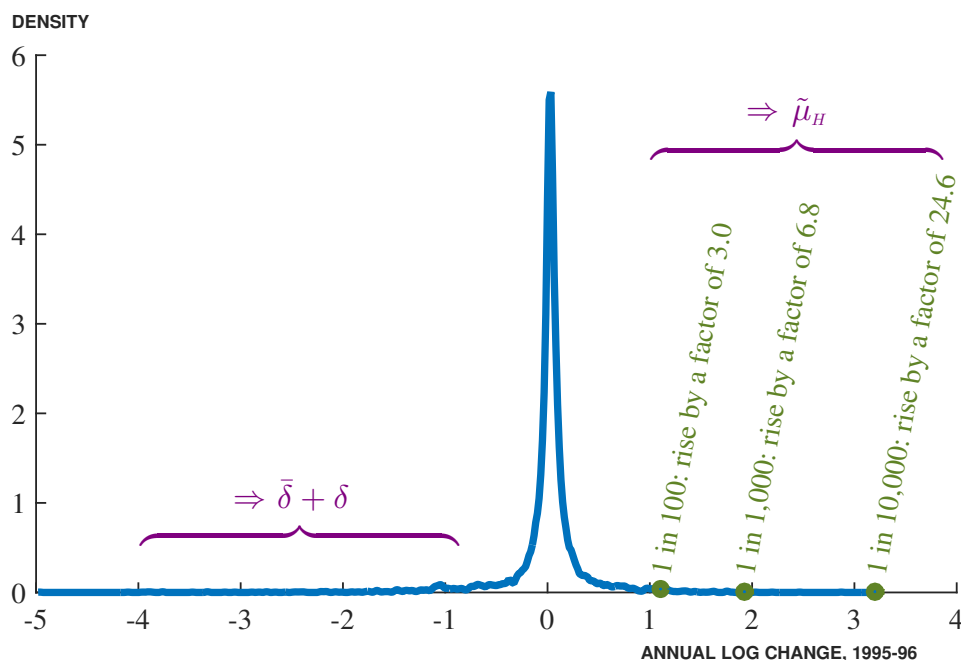
According to our model, the distribution of income growth rates for top earners should display thick tails at both the top and the bottom. At the bottom, the destruction shock results in a potentially large downward shift in incomes, causing the growth rate distribution to be left-skewed. At the top, the presence of a “high growth” group leads to a mixture of normal distributions that thickens the right tail. As noted earlier, Guvenen, Karahan, Ozkan and Song (2015) provide evidence for thick tails on both sides of the growth rate distribution, and some of their evidence is shown in Figure 9.¹²

We can see the same thing in our data from the IRS public use panel. Restricting our sample to tax units that involve married taxpayers filing jointly in two consecutive years, an example of this growth rate distribution for tax units starting in the 95th percentile and above in 1988 is shown in Figure 10. Quantitatively, the left-skewness of the distribution of growth rates helps us to identify $\delta_t^e \equiv \bar{\delta}_t + \delta_t$. A thick right tail is also evident in both the IRS and Guvenen et al data, and this right tail 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! The same phenomenon can be seen in the IRS data in Figure 10, but the relatively small sample size makes this tail less dramatic.

¹¹See <http://www.nber.org/taxsim-notes.html>.

¹²Their evidence suggests very thick Pareto-like tails for the growth rate distribution, a fact that our simple model cannot match.

Figure 9: The Distribution of Top Income Growth Rates, 1995–1996 (Güvenen et al.)



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 Güvenen, Karahan, Ozkan and Song (2015).

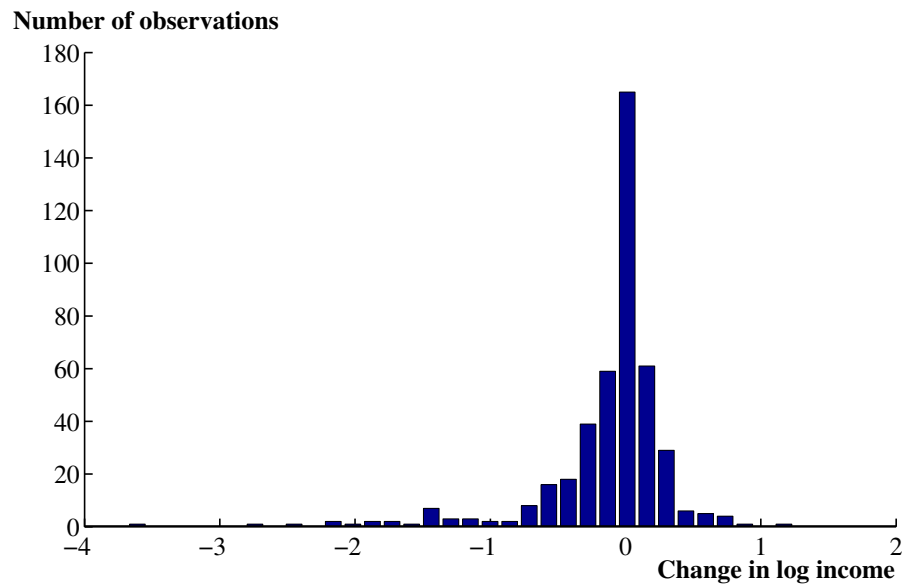
6.2. Empirical Results

Our data and estimation are discussed in more detail in Appendix B, and the results are summarized in Table 2. In brief, the parameters are estimated by specifying a cutoff value for growth rates, as suggested in a stylized way in Figure 9. For example, 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 δ^e , which is estimated to be around 6% in the IRS data.

To estimate $\tilde{\mu}_H$, σ_H , and \bar{p} using the IRS data, we look at growth rates for the Top 10% 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 σ_H as the standard deviation of growth rates above the 90th percentile of the growth rate distribution.

Next, and motivated by the empirical income dynamics literature, we make an ad-

Figure 10: The Distribution of Top Income Growth Rates 1988 and 1989 (IRS)



Note: This figure shows a histogram of the change in log wage and salary income for tax units consisting of married taxpayers filing jointly that are in the Top 5 percent of tax units in 1988 and for which we also observe income in 1989. The income levels are normalized by mean taxable income excluding capital gains.

Table 2: Empirical Estimates

Parameter	IRS 1979–81	IRS 1988–90	Guvenen et al. 1995–96
δ^e		0.06	...
σ_H		0.122	...
\bar{p}		0.767	...
$\tilde{\mu}_H$	0.244	0.303	0.435
Model: η^*	0.330	0.398	0.556
Data: η	0.33	0.48	0.55

Note: To estimate $\tilde{\mu}_H$, we look at growth rates for the Top 10% 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. The number from Guvenen, Karahan, Ozkan and Song (2015) for 1995–96 is the 95th percentile of the distribution of growth rates for the 90th percentile of top earners aged 45-50, obtained from the spreadsheet data appendix for that paper. We estimate σ_H as the standard deviation of growth rates above the 90th percentile of the growth rate distribution, which averages to 0.122. Finally, we choose \bar{p} to fit the value of η^* in 1980 of 0.33, which gives $\bar{p} = 0.767$. η^* in later years is computed from the other parameter values using equation (19) assuming $\bar{\delta} = .04$. The last row of the table reports empirical values for η in the United States taken in a stylized way from Figure 6 to capture “pre-1980,” 1988–90, and 1995–2010.

justment 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 σ^2 , the variance of the random walk innovation, to be 1/3 the variance calculated from the highest decile of the growth rate distribution.¹³

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. Instead, we choose the value of \bar{p} so that the model exactly matches Pareto inequality in 1979–81, using the formula in

¹³As shown later, this leads to estimates of σ of around 0.122 in 1979–1990, corresponding to a variance of the random walk innovations of about 0.015. This is in the right ballpark based on the empirical literature.

equation (19). We set the value of $\bar{\delta}$ to 0.04, a fairly conventional value, which leaves an estimate of $\delta = \delta^e - \bar{\delta} \approx 0.02$. Matching the Pareto inequality estimate in 1979–81 of 0.33 then leads to $\bar{p} = 0.767$. Over the course of a year, this implies that a high growth entrepreneur has a probability of approximately $1 - e^{-\bar{p}} \approx 0.54$ of transitioning to the low-growth state.¹⁴ 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 (2015).

We are now ready to put these numbers together to see what they imply about the evolution of top income inequality over time. For this purpose, we assume all the parameters other than $\tilde{\mu}_H$ are constant. (In our empirical work, there was no obvious trend in the other parameter values, though estimates often bounced around; see the appendix for more details.) We obtain estimates for $\tilde{\mu}_H$ for 1979–81 and 1988–90 from the IRS data. From Guvenen, Karahan, Ozkan and Song (2015), we obtain an estimate of $\tilde{\mu}_H$ for 1995–96, as the 95th percentile of the distribution of growth rates for the 90th percentile of top earners, obtained from the spreadsheet data appendix for that paper. As discussed in our appendix, we have done our best to make this number comparable to the IRS numbers, though some questions about comparability certainly remain. We anticipate using data from Guvenen et al in the near future when they make available some of the moments from the time series of their growth rate distributions.

The last two rows of Table 2 compare the estimated value of η^* from our model to the values observed in the U.S. data from Figure 6. As already mentioned, we fit the 1979–81 value by construction. The rise in $\tilde{\mu}_H$ through the 1980s then implies a value of η^* of 0.398 by 1988–90, compared to a value of around 0.48 in the data. The estimate of $\tilde{\mu}_H$ from Guvenen, Karahan, Ozkan and Song (2015) then rises sharply by 1995–96, implying a value of $\eta^* = 0.556$, very close to the empirically observed value of 0.55 from the 1995–2010 period.

Limited sample sizes in the IRS data and the comparability of the Guvenen, Karahan, Ozkan and Song (2015) data mean that this exercise should at best be viewed as a suggestive illustration. The basic model used in this paper is capable of explaining a substantial rise in U.S. top income inequality in a manner that is broadly consistent with the micro data. Access to additional moments from administrative data via the

¹⁴More precisely, the probability is $\bar{p}/(\bar{p} + \delta + \bar{\delta}) \cdot (1 - e^{-(\bar{p} + \delta + \bar{\delta})}) \approx 0.52$, because we have to consider the probability that high to low transition happens before arrival of death.

IRS and the Social Security Administration and from administrative authorities in other countries could be used to make additional progress and is a clear priority for future research.

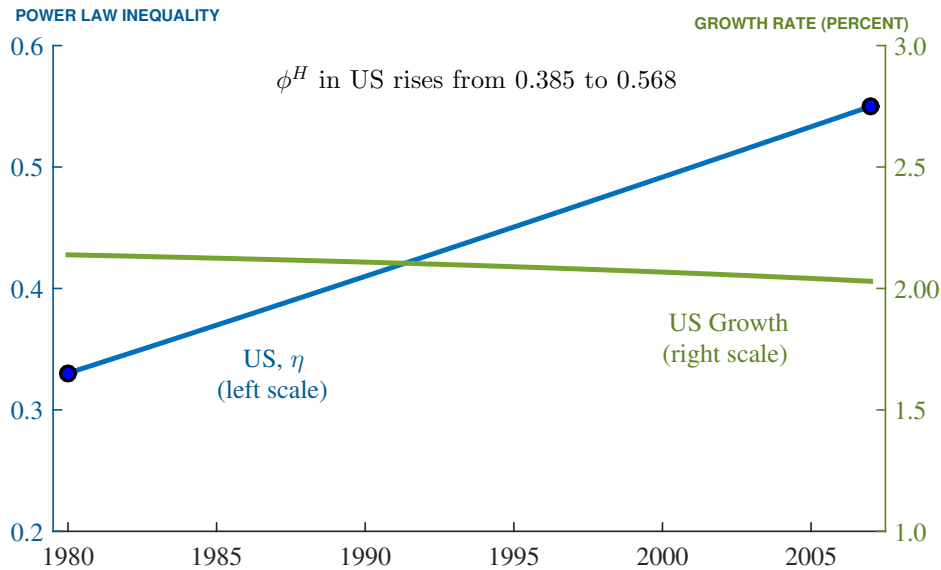
Based on these estimates, one can also explore the extent to which different parameters are responsible for the overall level of inequality. For example, Table 2 already shows the key role played by the drift parameter $\tilde{\mu}_H$. The rate at which high-growth entrepreneurs transit to the low-growth state, \bar{p} , also plays a crucial empirical role: raising this parameter by 10 percent essentially lowers η^* by around 10 percent. But two other parameters one might have thought to be important play a much smaller role. First, the exogenous destruction rate $\bar{\delta}$ is very small relative to \bar{p} and therefore plays a correspondingly small role. More interestingly, the standard deviation of the random walk process, σ_H , is relatively unimportant in our calibration: reducing this parameter all the way to zero at the 1995–96 parameter values only lowers η^* from 0.556 to 0.539; this small impact can also be seen earlier back in Figure 8. Luck matters in our calibration, but it is luck in the form of becoming and remaining a high-growth entrepreneur that is most crucial.

7. Numerical Examples

We now provide three numerical examples to illustrate the ability of the Schumpeterian model to match the levels and changes of top income inequality in the United States and France. Where possible, our baseline parameter values are chosen to be consistent with the empirical estimates of the previous section. For example, we assume that σ_H and σ_L for the United States are constant and equal at 0.122, broadly consistent with the evidence in Section 6. We assume $\bar{\delta} = 0.04$ and $\gamma = 1.4$, so that $\delta \approx 0.06$ when the economy's growth rate is 2 percent, and therefore $\bar{\delta} + \delta \approx 0.10$, similar to what we estimated in the previous section.

These remain examples, however, for two main reasons. First, the “reduced-form” empirical evidence is insufficient to identify the underlying structural parameters of the model. As one simple example, movements in both ϕ and τ can deliver changes in μ over time, as we show below. Second, for reasons discussed earlier, our numerical exercises do not consider transition dynamics and instead report a sequence of steady

Figure 11: Numerical Example: Matching U.S. Inequality



Note: Baseline parameter values are given in the text. Over time in this simulation, ϕ_H rises linearly from 0.385 to 0.568. All other parameters are held constant.

states. Nevertheless, the aim is to show that the Schumpeterian model can deliver changes in top income inequality consistent with both the overall inequality data and with the underlying micro data on top incomes.

7.1. Matching U.S. Inequality

Figure 11 shows a numerical example that illustrates one way in which the model can match the time series behavior of top income inequality in the United States. We start with a set of baseline parameters — $\rho = 0.03$, $\bar{L} = 15$, $\tau = 0$, $\theta = 2/3$, $\beta = 1$, $\lambda = 0.027$, $\bar{z} = 0.2$, $\sigma = 0.122$, $\bar{\delta} = 0.04$, $\phi_H = 0.5$, $\bar{p} = 0.767$, $\bar{m} = 0.5$, and $\bar{q} = .504$ — which match U.S. Pareto inequality in 1980.¹⁵ Next, we assume that ϕ_H , the technology parameter converting entrepreneurial effort into growth in a variety's productivity, rises over time. We calibrate the change in ϕ_H to match the rise in the U.S. top inequality between 1980 and 2007. The values of ϕ_H we recover produce values of $\tilde{\mu}_H$ ranging from 0.25 to 0.43, broadly consistent with the micro evidence presented earlier.

¹⁵The value of \bar{q} is chosen so that the fraction of high-growth entrepreneurs in the stationary distribution is 2.5 percent.

Although the increases in ϕ_H matches the changes in inequality, they have a negative impact on long-run growth. An increase in ϕ_H leads to a rise in entrepreneurial effort and to a rise in inequality. Because prospective researchers are risk averse, more inequality among entrepreneurs reduces the number of researchers, thereby limiting overall growth. This effect is relatively small in Figure 11; it could be offset in the short run by transition dynamics or by changes in other parameter values.

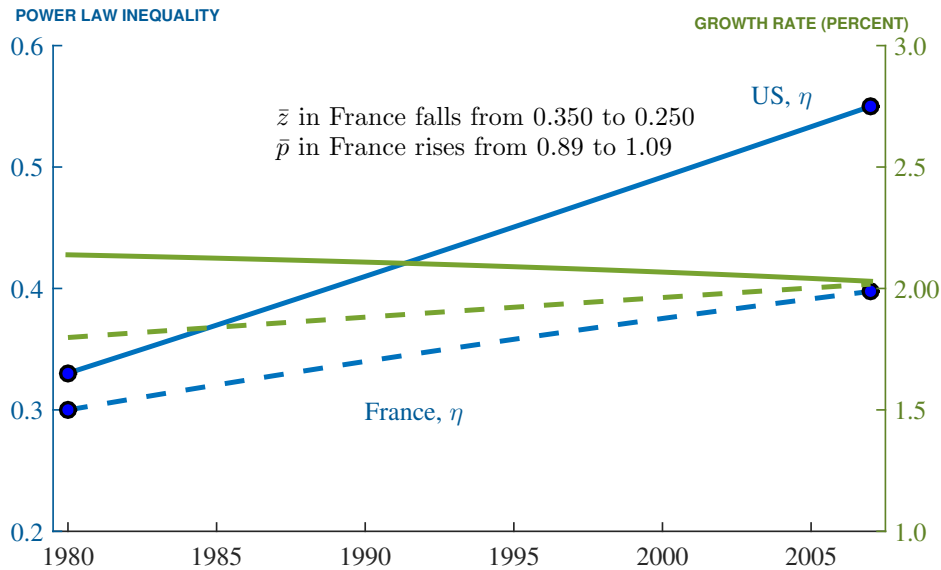
7.2. Matching Inequality in France

Next, we illustrate the ability of the model to simultaneously match the patterns of top income inequality in both the U.S. and France. For the United States, we assume the same parameter values just described. For France, we assume the same rise in ϕ_H applies, for example capturing the fact that the World Wide Web is a worldwide phenomenon. We know from our U.S. example that this would produce a large rise in top inequality, other things equal, so some other change is needed to offset the rise in France. For this example, we consider changes in \bar{p} and \bar{z} in France. First, we choose a higher value of \bar{p} in France to match their lower top inequality. This example then matches the modest increase in inequality in France by assuming a combination of a decline in innovation blocking, \bar{z} , and a rise in the rate at which high-growth entrepreneurs transition to normal entrepreneurs. These changes allow us to match the evidence on French inequality, as shown in Figure 12.

7.3. Taxes

As already discussed, these exercises are simply numerical examples, and there are other ways to match the change in top inequality in the model in ways that are broadly consistent with the data. In this example, we consider changes in the tax parameter τ . A literal interpretation of this parameter is that it is a tax on the time endowment of entrepreneurs — e.g. lost time due to “red tape.” It is tempting to wonder about the effects of marginal labor income tax rates. However, because consumption enters in log form in the utility function (to obtain analytic tractability), the income and substitution effects from a labor income tax cancel exactly and a labor income tax leaves equilibrium effort — and therefore top income inequality — unchanged in this setup. The time tax

Figure 12: Numerical Example: U.S. and France



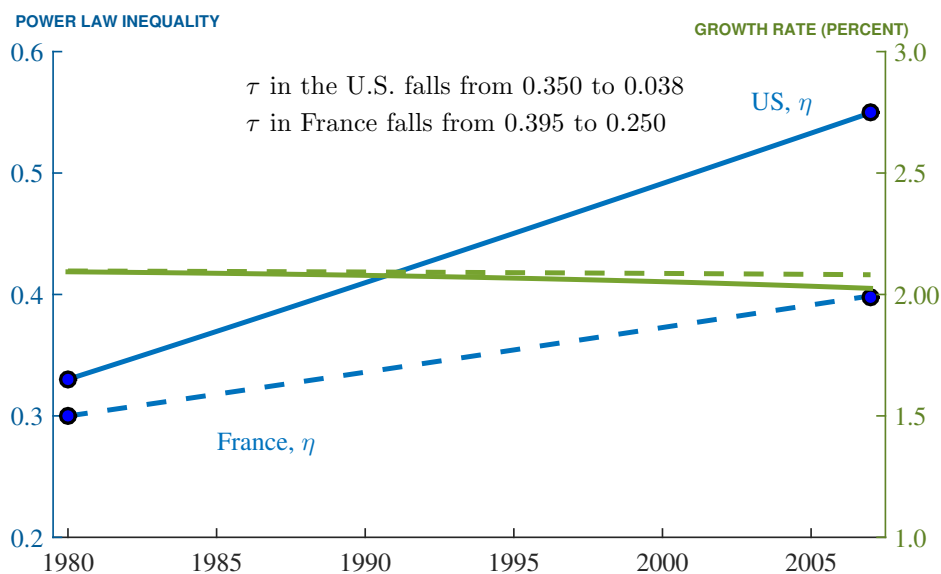
Note: The baseline parameter values described in the text are used; for example, the U.S. parameter values are the same as in Figure 11. In France, we assume ϕ rises in exactly the same way. To offset the large increase in inequality that would be implied, we assume \bar{z} and \bar{p} change, as shown in the figure.

τ is suggestive of what labor income taxes might imply in a richer framework.¹⁶

To what extent can changes in τ account for the differential patterns of top income inequality that we see in the United States and France? This question is considered in Figure 13. The baseline parameter values used in this example are identical to those used in Figure 12, with two exceptions. First, we raise the (constant) level of ϕ_H to 0.59 to match the level of U.S. inequality in 1980, when the initial tax rate is positive. Second, we consider linear declines in τ for both the U.S. and France. Having τ in the United States fall from 35 percent to 3.8 percent generates the U.S. inequality numbers. And a more modest decline from 39.5 percent to 25 percent can generate the French inequality facts. While it is possible to make the tax story work in our paper, this does

¹⁶Kim (2013) considers labor income taxes in a model in which goods are used to accumulate human capital, so that labor income taxes can affect top income inequality, and finds that tax changes can account for a modest portion of the changes in inequality over time and across countries. Kindermann and Krueger (2014) study optimal taxation in a model in which the coefficient of relative risk aversion is larger than one so that the income effect dominates the substitution effect and find that this supports very high optimal tax rates on top incomes. In our setting, this would imply that higher income taxes *increase* the effort of entrepreneurs and increase top (pre-tax) inequality. The decline in U.S. top marginal rates would not then be able to explain the rise in Pareto inequality.

Figure 13: Numerical Example: Taxes and Inequality



Note: The figure shows the results of linear declines in the tax parameter τ on top income inequality and on economy-wide average income growth. ϕ_H is held constant over time at 0.59.

require extraordinary efforts — viewing taxes as reducing the time endowment and requiring an implausibly large decline in τ for the United States. We therefore infer that the technology-based story surrounding ϕ_H is likely to be an important part of the explanation of rising top inequality.

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 ϕ and raising inequality. On the other hand, as countries open their domestic markets to more competition via globalization, rates of creative destruction 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 is likely to be quite fruitful in coming years. For example, 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. In a future version of this paper, we plan to use moments they release to compute estimates of Pareto inequality over time that are perhaps even more comparable than what we already show in Table 2.

Finally, in a recent paper, Aghion, Akcigit, Bergeaud, Blundell and Hémous (2015) document that innovation and top income inequality are positively correlated within U.S. states and across U.S. commuting zones. On the surface, there might be some tension between their results and ours (innovation raising inequality empirically versus the creative destruction effect in our model). There are at least two possible ways to reconcile these results. First, some of the x accumulation by existing entrepreneurs in our model might show up as patents in the data on innovation. Second, the empirical correlation may not estimate the creative destruction causal effect: perhaps the rise of information technology leads to many new patents but also causes the effect working through ϕ in our paper, leading to a positive empirical correlation. More work is needed to untangle this interesting empirical phenomenon.

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 Δt :¹⁷

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

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

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

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,

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

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 $\frac{\partial f(x, t)}{\partial t} = 0$. 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

B1. 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

¹⁷This equation drops a term involving both Δt and Δx , as it goes to zero later when we take limits.

involved married taxpayers filing jointly in two consecutive years and convert nominal values to 2012 constant dollars using the consumer price index. Our income measure is field 11 of the data, corresponding to wages and salaries. 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. Two sets of estimates are made, based on different cutoffs for the meaning of “top incomes.” In the main case, we use only tax units in the Top 10 percent of our income measure, which facilitates comparability to the Guvenen, Karahan, Ozkan and Song (2015) data. As a robustness check, we also consider a Top 5 percent cutoff.

B2. 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 Δ percent are due to the destruction shocks. For our benchmark estimates, we assume $\Delta = 40$. Our estimate of δ_t^e is therefore the fraction of growth rates that reduce incomes by more than Δ percent.

To estimate $\tilde{\mu}_H$ and σ_H , we look at growth rates for the Top 10% 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 σ_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 given the small numbers of observations. These will be provided in a future draft.

Table A1 shows the details of our baseline estimates for the Top 10% of earners, year by year using the panel data. For the results reported in the main text, we consider the observations based on the largest sample sizes, averaging the results for 1979–81 (the

Table A1: Parameter Estimates (Top 10 percentile cutoff)

Year	$\tilde{\mu}_H$	σ	δ^e	N
1980	0.2407	0.1906	0.0542	2123
1981	0.2470	0.2322	0.0465	2109
1982	0.3249	0.1588	0.0586	444
1983	0.2691	0.1177	0.0610	426
1984	0.2058	0.1483	0.0433	439
1985	0.2384	0.1885	0.0556	450
1986	0.2310	0.2602	0.0569	457
1987	0.3492	0.2052	0.0671	447
1988	0.3193	0.3090	0.0701	899
1989	0.2911	0.1761	0.0789	900
1990	0.3142	0.3335	0.0615	911

Note: 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 10% 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 σ do not include the $\sqrt{1/3}$ correction for permanent versus transitory shocks.

first two observations) and 1988–90 (the last two observations). We do not consider any trend in σ_H , in part because the data are very noisy, in part to focus on $\tilde{\mu}_H$, and in part because the effect of rises in σ in our model are very small (see Figure 8). Our estimate of σ_H is based on the average value over the entire decade, equal to 0.2109 before we make the permanent versus transitory adjustment.

Tables A2 shows the results for the Top 5% of earners, probably a better sample given our model. We do not make this the baseline, however, simply so we can compare our estimates to the 1995–96 estimates from Guvenen, Karahan, Ozkan and Song (2015), which are based on earners at the 90th percentile.

The estimate of $\tilde{\mu}_H$ reported in Table 2 from Guvenen, Karahan, Ozkan and Song (2015) for 1995–96 is the 95th percentile of the distribution of 1-year log changes for

Table A2: Parameter Estimates (Top 5 percentile cutoff)

Year	$\tilde{\mu}_H$	σ	δ^e	N
1980	0.2893	0.2321	0.0609	1067
1981	0.2843	0.2790	0.0535	1046
1982	0.4145	0.1544	0.0811	222
1983	0.2545	0.1070	0.0892	213
1984	0.2058	0.1203	0.0648	216
1985	0.2580	0.2034	0.0575	226
1986	0.2859	0.3228	0.0652	230
1987	0.4884	0.2132	0.1018	226
1988	0.4379	0.2714	0.0976	451
1989	0.3596	0.2049	0.1091	440
1990	0.3607	0.4301	0.0615	455

Note: 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 53rd or 24th from the top for the key 1979–81 and 1988–90 periods. The estimates of σ do not include the $\sqrt{1/3}$ correction for permanent versus transitory shocks.

the 90th percentile of top earners aged 45-50, obtained from the “Figure 3” tab in the spreadsheet data appendix for that paper (available [here](#)). They also report the growth rate distribution for earners aged 25-34; this estimate is even higher and would imply an even larger increase in top inequality.

In the future, we understand that Guvenen et al will provide time series data on the moments of the growth rate distribution, so we hope to be able to provide more comparable estimates over time in this fashion.

C 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 (13) yields

$$\frac{\beta}{\Omega - e^*} = \phi V_x(x_t, t)x_t, \quad (\text{C1})$$

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 ζ_0 , ζ_1 , and ζ_2 . We then rewrite (C1) as

$$\frac{\beta}{\Omega - e^*} = \phi \zeta_2. \quad (\text{C2})$$

Now substituting (C2) into (13), we have

$$\begin{aligned} & (\rho + \delta + \bar{\delta})(\zeta_0 + \zeta_1 t + \zeta_2 \log x_t) \\ & = \beta \log \left(\frac{\beta}{\phi \zeta_2} \right) + \zeta_2 \phi e^* - \frac{1}{2} \zeta_2 \sigma^2 + \zeta_1 + (\delta + \bar{\delta}) V^w(t) + \log \psi_t + \log x_t \end{aligned} \quad (\text{C3})$$

Equating the coefficients on $\log x_t$ yields

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

We then substitute ζ_2 into (C2) to obtain

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

To complete the proof, we next outline how to solve for ζ_0 and ζ_1 by showing that the right-hand side of (C3) 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, (26) and (28) imply that both $\log w_t$ and $\log \psi_t$ are linear functions of n_t . Since 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 (C3) will have the same form as our conjecture, and we obtain ζ_0 and ζ_1 by equating the coefficients and constants in (C3). QED.

Proof of Proposition 2. The Pareto Income Distribution

Substituting our guess $f(x) = Cx^{-\xi-1}$ to (16), 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 ξ , we obtain the positive root in (18). 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 ϕ and the variance of the shocks σ^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 human capital x_0 , inherits the high growth state (ϕ_H, σ_H) with probability \bar{q} and the low growth state (ϕ_L, σ_L) with probability $1 - \bar{q}$. On the other hand, in the case of replacement by endoge-

nous creative destruction (the Poisson process at rate δ), 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. Moreover, 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	$-(\bar{p} + \delta + \bar{\delta})$	\bar{p}	$\delta + \bar{\delta}$
Low growth	0	$-(\delta + \bar{\delta})$	$\delta + \bar{\delta}$
Researchers	$\bar{q}\bar{\delta}_R + \pi_H\delta_R$	$(1 - \bar{q})\bar{\delta}_R + (1 - \pi_H)\delta_R$	$-(\delta_R + \bar{\delta}_R)$

where π_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 that there is some 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

$$\rho V^L(x_t, t) = \max_e \log \psi_t + \log x_t + \beta \log(\Omega - e_t) + \frac{\mathbb{E}[dV^L(x_t, t)]}{dt} + (\delta + \bar{\delta})(V^w(t) - V^L(x_t, t)). \quad (\text{C4})$$

Notice that the Bellman equations (C4) and (13) 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 ζ_0, ζ_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

$$\rho V^H(x_t, t) = \max_e \log \psi_t + \log x_t + \beta \log(\Omega - e_t) + \frac{\mathbb{E}[dV^H(x_t, t)]}{dt} + \bar{p}(V^L(x_t, t) - V^H(x_t, t)) + (\delta + \bar{\delta})(V^w(t) - V^H(x_t, t)). \quad (\text{C5})$$

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 $e_H^* = (1 - \tau) - \frac{1}{\phi_H}\beta(\rho + \delta + \bar{\delta})$.

The optimal entrepreneurial efforts e_H^*, e_L^* pin down the mean growth rates in the geometric Brownian motion process (11) 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 (2015).

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 :

$$dy = \tilde{\mu} dt + \sigma dB_t, \text{ where } \tilde{\mu} = \mu(e_t) - \frac{\sigma^2}{2}. \quad (\text{C6})$$

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:

$$\frac{\partial g_H(y, t)}{\partial t} = -(\bar{\delta} + \bar{p})g_H(y, t) - \frac{\partial}{\partial y} [\tilde{\mu}_H g_H(y, t)] + \frac{1}{2} \cdot \frac{\partial^2}{\partial y^2} [\sigma^2 g_H(y, t)] + \bar{q}\bar{\delta}\delta_0(y) \quad (\text{C7})$$

$$\frac{\partial g_L(y, t)}{\partial t} = -\bar{\delta}g_L(y, t) + \bar{p}g_H(y, t) - \frac{\partial}{\partial y} [\tilde{\mu}_L g_L(y, t)] + \frac{1}{2} \cdot \frac{\partial^2}{\partial y^2} [\sigma^2 g_L(y, t)] + (1 - \bar{q})\bar{\delta}\delta_0(y), \quad (\text{C8})$$

where δ_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 (C7) and (C8), 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

$$\begin{pmatrix} \hat{g}_H(s) \\ \hat{g}_L(s) \end{pmatrix} = \begin{pmatrix} \frac{\bar{q}\bar{\delta}}{\lambda_H(s)} \\ \frac{\bar{p}\bar{q}\bar{\delta}}{\lambda_H(s)\lambda_L(s)} + \frac{(1-\bar{q})\bar{\delta}}{\lambda_L(s)} \end{pmatrix}. \quad (C9)$$

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 ξ , higher moments than ξ is infinite. Using this moment condition, we obtain the Pareto tail index ξ 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 ϕ_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 dominate the tail distribution. First, 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 where 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 also easy to prove that ξ_H is always smaller than ξ_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 $\tilde{\mu}_L^* < 0 < \tilde{\mu}_H^*$ (ϕ_L is small enough to make $\tilde{m}u_L^*$ negative), even when death arrives faster in the high-growth state, more precisely when $\frac{2(\bar{\delta} + \bar{p})}{\sigma_H^2} > \frac{2\bar{\delta}}{\sigma_L^2}$, ξ_H is always smaller than ξ_L if $\frac{\tilde{\mu}_H^*}{\sigma_H^2} > \left(\frac{\bar{\delta} + \bar{p}}{\sigma_H^2} - \frac{\bar{\delta}}{\sigma_L^2}\right) \sqrt{\frac{\sigma_L^2}{2\bar{\delta}}}$ (ϕ_H is large enough to

offset the faster arrival of death in the high-growth state) or $\frac{\tilde{\mu}_L^*}{\sigma_L^2} < -\left(\frac{\bar{\delta}+\bar{p}}{\sigma_H^2} - \frac{\bar{\delta}}{\sigma_L^2}\right) \sqrt{\frac{\sigma_H^2}{2(\bar{\delta}+\bar{p})}}$ (ϕ_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. **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

$$\left(\frac{Y}{Y_i} \right)^{1-\theta} = p_i. \quad (\text{C10})$$

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

$$\max_{Y_i} p_i(Y_i)Y_i - wL_i = Y^{1-\theta}Y_i^\theta - \frac{w}{\gamma^n x_i^\alpha} Y_i. \quad (\text{C11})$$

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

$$Y_i = \left(\frac{1}{\theta} \frac{w}{\gamma^n x_i^\alpha} \right)^{\frac{1}{\theta-1}} Y. \quad (\text{C12})$$

By plugging (C12) in the final goods production function, we obtain the equilibrium wage equation

$$w = \theta \gamma^n \left(\int_0^1 x_i^\alpha \frac{1}{1-\theta} \right)^{\frac{1-\theta}{\theta}} \equiv \theta \gamma^n X^\alpha, \quad (\text{C13})$$

where we assume $\alpha \frac{\theta}{1-\theta} = 1$ and $X \equiv \int_0^1 x_i di$. Using this equation we can rewrite (C12) as

$$Y_i = \left(\frac{x_i}{X} \right)^{\frac{1}{\theta}} Y. \quad (\text{C14})$$

Next, combining (C14) and (21) 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 :

$$Y = \gamma^n X^\alpha L. \quad (\text{C15})$$

Lastly, the profit π_i is calculated from plugging the optimal solution (C10), (C12), and (C14) into the monopoly problem (C11). 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 (31) can be rewritten as

$$V^w(t) = \int_t^\infty \exp^{-\rho(\tau-t)} \log w_\tau d\tau = \frac{1}{\rho} \log w_t + \frac{g}{\rho^2}, \quad (\text{C16})$$

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 (32) 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 (13) and (C3). We rewrite (C3) as

$$(\rho + \delta + \bar{\delta})V(x_t, t) = \log \psi_t + \log x_t + C + (\delta + \bar{\delta})V^w(t) + \frac{dV(x, t)}{dt}, \quad (\text{C17})$$

where $C = \beta \log(\Omega - e^*) + \frac{\phi e^* - \frac{1}{2}\sigma^2}{\rho + \delta + \bar{\delta}}$ contains constant terms. Differentiating (C17) with respect to time, we obtain

$$\frac{dV(x, t)}{dt} = \frac{1}{\rho + \delta + \bar{\delta}} \left(\frac{\dot{\psi}_t}{\psi_t} + (\delta + \bar{\delta}) \frac{dV^w}{dt} \right) = \frac{g}{\rho}. \quad (\text{C18})$$

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 (C18) and (C16) into (C17) yields

$$(\rho + \delta + \bar{\delta})V(x, t) = \log \frac{1-\theta}{\theta} + \log L_t - \log x_0 + \log(1-\eta) + \log x_t + C + (\rho + \delta + \bar{\delta})V^w(t). \quad (\text{C19})$$

Now taking expectations on (C19) and rearranging, we get

$$\mathbb{E}[V(x_t, t)] = \frac{1}{\rho + \delta + \bar{\delta}} (\log \frac{1-\theta}{\theta} + \log L_t + \log(1-\eta) + \eta + C) + V^w(t), \quad (\text{C20})$$

where we use $\mathbb{E}[\log x] = \log x_0 + \eta$ if x follows a Pareto distribution with the inequality parameter η and the minimum value x_0 . Furthermore, we know from (C19) that

$$V(x_0, t) = \frac{1}{\rho + \delta + \bar{\delta}} (\log \frac{1-\theta}{\theta} + \log L_t + \log(1-\eta) + C) + V^w(t). \quad (\text{C21})$$

We then rewrite (C20) as

$$\mathbb{E}[V(x_t, t)] = V(x_0, t) + \frac{\eta}{\rho + \delta + \bar{\delta}}. \quad (\text{C22})$$

Next substituting (C22) into (32) and rearranging, we obtain

$$(\rho + \lambda(1-\bar{z}) + \bar{\delta}_R)V^R(t) = \log \bar{m} + \log w_t + \underbrace{\frac{g}{\rho}}_{=\rho V^w(t)} + (\lambda(1-\bar{z}) + \bar{\delta}_R)V(x_0, t) + \frac{\lambda(1-\bar{z})\eta}{\rho + \delta + \bar{\delta}}. \quad (\text{C23})$$

Lastly we substitute (C21) into (C23) and apply the indifference equation $V^w(t) = V^R(t)$ to (C23) to get

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

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 heterogenous mean growth rates

The value function for a worker, $V^w(t)$, does not change with heterogenous growth

rates, thus we have $V^w(t) = \frac{1}{\rho} \log w_t + \frac{g}{\rho^2}$ as given in the equation (C16). On the other hand, the value function for a researcher at time t , $V^R(t)$, changes from equation (32) to consider the two different growth states that a researcher can take when they become a new entrepreneur as shown below.

$$\begin{aligned} \rho V^R(t) = \log(\bar{m}w_t) + \frac{dV^R(t)}{dt} + \lambda(1 - \bar{z}) (\pi_H \mathbb{E}[V^H(x, t)] + \pi_L \mathbb{E}[V^L(x, t)] - V^R(t)) \\ + \bar{\delta}_R (\bar{q}V^H(x_0, t) + (1 - \bar{q})V^L(x_0, t) - V^R(t)), \end{aligned} \quad (\text{C24})$$

where $\pi^* = (\pi_H, \pi_L)$ is the stationary distribution of high-growth and low-growth entrepreneurs. In the remainder of the proof, we will solve for each term in (C24) to explicitly solve for L_t .

Starting from the first two terms, we have

$$\log(\bar{m}w_t) + \frac{dV^R(t)}{dt} = \log \bar{m} + \log w_t + \frac{dV^w(t)}{dt} = \log \bar{m} + \rho V^w(t). \quad (\text{C25})$$

We next calculate the stationary distribution π^* . 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.¹⁸ 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 π^* 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

¹⁸For 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' states.

for an entrepreneur with quality x_t is given in (C5) and (C4). We rewrite them as

$$(\rho + \delta + \bar{\delta})V^H(x_t, t) = \log \psi_t + \log x_t + C_H + \bar{p}(V^L(x_t, t) - V^H(x_t, t)) \\ + (\delta + \bar{\delta})V^w(t) + \frac{dV^H}{dt}, \quad (\text{C26})$$

$$(\rho + \delta + \bar{\delta})V^L(x_t, t) = \log \psi_t + \log x_t + C_L + (\delta + \bar{\delta})V^w(t) + \frac{dV^L}{dt}, \quad (\text{C27})$$

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 (C26) and (C27) with respect to t as in (C18), we get

$$\frac{dV^H(x, t)}{dt} = \frac{dV^L(x, t)}{dt} = \frac{g}{\rho}. \quad (\text{C28})$$

We can also replace $V^L - V^H$ in (C26) by subtracting (C26) from (C27) to get

$$V^L(x_t, t) - V^H(x_t, t) = \frac{C_L - C_H}{\rho + \delta + \bar{\delta} + \bar{p}}. \quad (\text{C29})$$

Substituting (C28), (C29), and (C16) into (C26) and (C27) yields

$$V^H(x, t) = V^w(t) + \frac{1}{\rho + \delta + \bar{\delta}} \left(\log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x_t] + \log x_t + C_H + \frac{\bar{p}(C_L - C_H)}{\rho + \delta + \bar{\delta} + \bar{p}} \right) \quad (\text{C30})$$

$$V^L(x, t) = V^w(t) + \frac{1}{\rho + \delta + \bar{\delta}} \left(\log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x_t] + \log x_t + C_L \right). \quad (\text{C31})$$

Now taking expectations and rearranging, we get

$$\pi_H \mathbb{E}[V^H(x, t)] + \pi_L \mathbb{E}[V^L(x, t)] = \\ V^w(t) + \frac{1}{\rho + \delta + \bar{\delta}} \left(\log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x_t] + \mathbb{E}[\log x_t] + D(\pi_H, \pi_L) \right), \quad (\text{C32})$$

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

$$\bar{q}V^H(x_0, t) + (1 - \bar{q})V^L(x_0, t) = \\ V^w(t) + \frac{1}{\rho + \delta + \bar{\delta}} \left(\log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x_t] + D(\bar{q}, 1 - \bar{q}) \right), \quad (\text{C33})$$

where we normalized x_0 to be 1.

Substituting (C25), (C32), (C33), and the indifference equation $V^w(t) = V^R(t)$ into (C24), we get

$$0 = (\rho + \delta + \bar{\delta}) \log \bar{m} + (\lambda(1 - \bar{z}) + \bar{\delta}_R)(\log \frac{1-\theta}{\theta} + \log L_t - \log \mathbb{E}[x]) + \lambda(1 - \bar{z}) (\mathbb{E}[\log x] + D(\pi_H, \pi_L)) + \bar{\delta}_R D(\bar{q}, 1 - \bar{q}). \quad (\text{C34})$$

Solving (C34) for $\log L_t$, we finally obtain the allocation of labor to research with heterogeneous 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 $\mathbb{E}[x]$ with help of the Laplace transform $\hat{g}(s) = E[e^{-sy}]$ of which the explicit functional form is given in (C9). Specifically, they are given by

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

QED.

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.
- Aluation.wordpress.com, "Fractals, fractiles and inequality," Technical Report August 2011.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez, "The World Top Incomes Database," 2013. Accessed on October 15, 2013.

Aoki, Shuhei and Makoto Nirei, "Pareto Distributions and the Evolution of Top Incomes in the U.S.," MPRA Paper 47967, University Library of Munich, Germany July 2013.

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, Brian and John Van Reenen, "Bankers' pay and extreme wage inequality in the UK," Open Access publications from London School of Economics and Political Science <http://eprints.lse.ac.uk/>, London School of Economics and Political Science 2010.

Benabou, Roland and Jean Tirole, "Bonus Culture: Competitive Pay, Screening, and Multitasking," Working Paper 18936, National Bureau of Economic Research April 2013.

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.

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.

- 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.
- , Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "The Dynamics of Inequality," June 2015. Princeton University manuscript.
- 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 Static and Microfounded Theory of Zipf's Law for Firms and of the Top Labor Income Distribution," *Unpublished manuscript, UCLA, Los Angeles*, December 2014.
- 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?," Staff Reports 710, Federal Reserve Bank of New York February 2015.
- , 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.
- 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.
- Kim, Jihee, "The Effect of the Top Marginal Tax Rate on Top Income Inequality," 2013. KAIST, unpublished paper.
- Kindermann, Fabian and Dirk Krueger, "High Marginal Tax Rates on the Top 1Paper 20601, National Bureau of Economic Research October 2014.
- Koenig, Michael, Jan Lorenz, and Fabrizio Zilibotti, "Innovation vs. Imitation and the Evolution of Productivity Distributions," Discussion Papers, Stanford Institute for Economic Policy Research 11-008, Stanford Institute for Economic Policy Research February 2012.
- Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, 65 (6), 1389–1419.
- 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.

- 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,” Quantitative Finance Papers 1012.0199, arXiv.org December 2010.
- 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.
- , “Lecture 6: Income and Wealth Distribution,” 2012. Princeton teaching slides http://www.princeton.edu/~moll/ECO521Web/Lecture6_ECO521_web.pdf.
- Nirei, Makoto, “Pareto Distributions in Economic Growth Models,” IIR Working Paper 09-05, Institute of Innovation Research, Hitotsubashi University July 2009.
- 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. Financial Industry: 1909-2006,” NBER Working Papers 14644, National Bureau of Economic Research, Inc January 2009.
- 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 Capital Taxation,” NBER Working Papers 17989, National Bureau of Economic Research, Inc April 2012.
- and Gabriel Zucman, “Wealth and Inheritance in the Long Run,” April 2014. forthcoming in the Handbook of Income Distribution.

- , Emmanuel Saez, 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,” NBER Working Papers 17035, National Bureau of Economic Research, Inc May 2011.
- Saez, Emmanuel, “Using Elasticities to Derive Optimal Tax Rates,” *Review of Economic Studies*, 2001, 68, 205–229.
- 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.