A Schumpeterian Model of Top Income Inequality

Charles I. Jones — Stanford GSB and NBER

Jihee Kim * — KAIST

October 20, 2014 – Version 0.95

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, a reduction in top tax rates, and a decline in misallocation 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, Xavier Gabaix, Mike Harrison, Pete Klenow, Ben Moll, Chris Tonetti, Alwyn Young, Gabriel Zucman and seminar participants at the AEA annual meetings, Chicago Booth, CREI, the Federal Reserve Bank of San Francisco, Groningen, HKUST, IIES Stockholm, Korea University, the NBER Income Distribution group, Princeton, Stanford, USC, 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 0.6, 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 following reason. Suppose that

[1]These papers in turn build on a large literature on such mechanisms outside economics. For example, see Reed (2001), Mitzenmacher (2004), and Malevergne, Saichev and Sornette (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.

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 songs, 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 entreprenueurs 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 of idiosyncratic productivity, 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

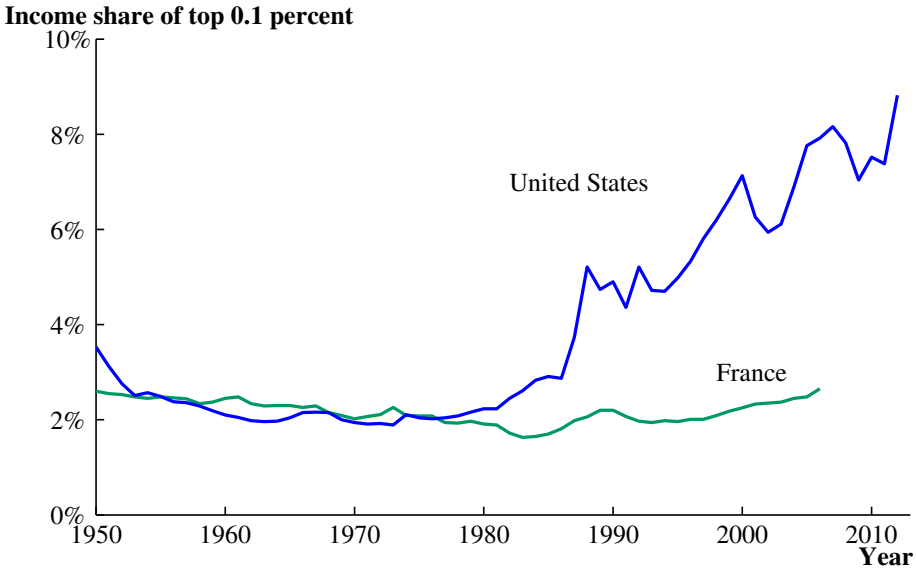
on finance and the extent to which rising rents in that sector can explain rising inequality; see also Bell and Van Reenen (2010). Bakija, Cole and Heim (2010) and Kaplan and Rauh (2010) note that the rise in top inequality occurs across a range of occupations; it is not just focused in finance or among 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. 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.

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), and Piketty and Zucman (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 accumulation 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.

[2]Luttmer (2014) extends this line of work in an attempt to get endogenous growth without assuming a Pareto distribution and also considers implications for inequality. Koenig, Lorenz and Zilibotti (2012) derive a Zipf distribution in the upper tail for firm productivity in an endogenous growth setting.

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



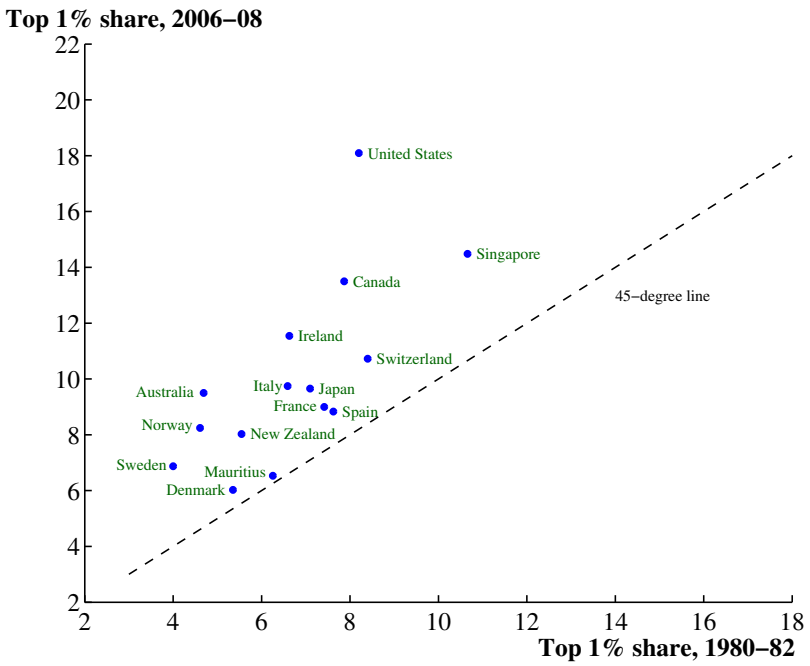
Source: World Top Incomes Database.

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.

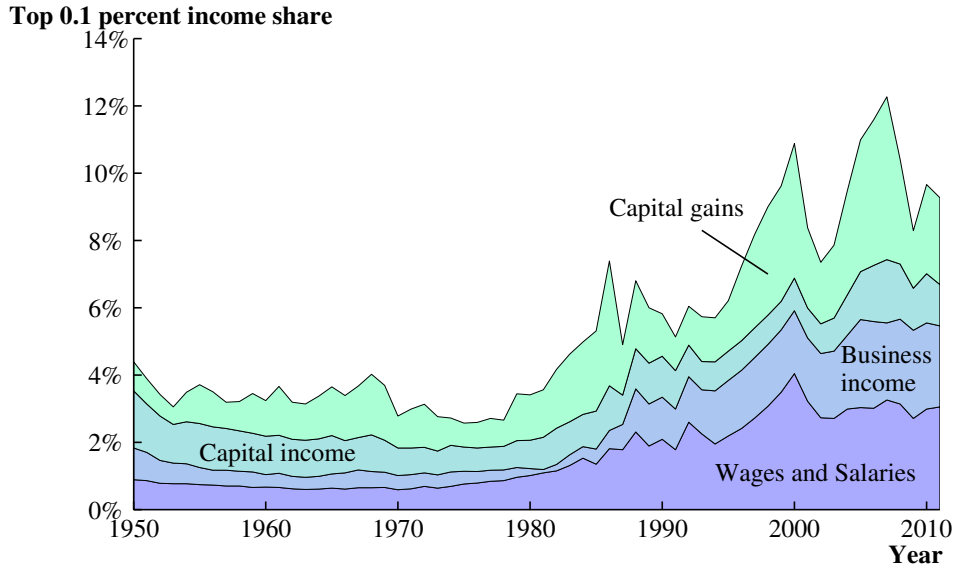
Figure 2 shows the dynamics of top income inequality for a range of countries. The horizontal axis shows the share of aggregate income going to the top 1% of earners, averaged between 1980 and 1982, while the vertical axis shows the same share for 2006–2008. All the economies for which we have data lie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

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).

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-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 $Y \geq y_0$)

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

where ξ is called the "power law exponent."

Saez (2001) documents 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 show mean wage income above some threshhold as a ratio to the threshhold itself. If wage income obeys a Pareto distribution like that in (1), then this ratio should equal the constant $\frac{\xi}{\xi-1}$, *regardless of the threshhold*. That is, as we move to higher and higher income threshholds, the ratio of average income above the threshhold to the threshhold 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 inequality inequality. It is therefore convenient to define the "power-law inequality" exponent as

$$\eta \equiv \frac{1}{\xi} \tag{2}$$

so that

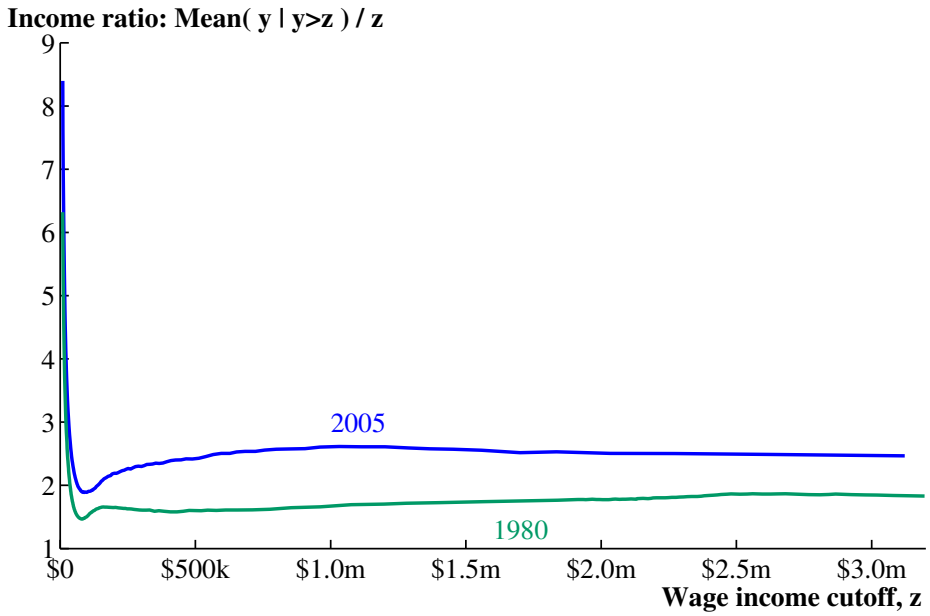
$$\tilde{S}(p) = \left(\frac{100}{p}\right)^{\eta-1}. \tag{3}$$

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

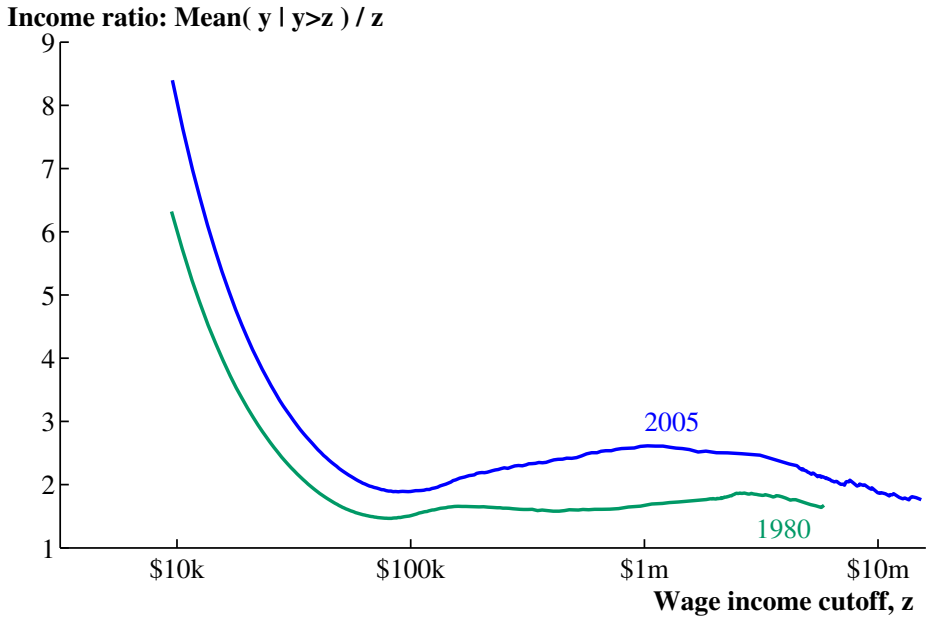
To see the fractal structure of top inequality under a Pareto distribution, let $S(a) = \tilde{S}(a)/\tilde{S}(10a)$ denote the fraction of income earned by the top $10 \times a$ percent of people that actually goes to the top a percent. For example, $S(1)$ is the fraction of income going to the top 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

[3]This follows easily from the fact that the mean of a Pareto distribution is $\frac{\xi y_0}{\xi-1}$ and that the conditional mean just scales up with the threshhold.

Figure 4: The Pareto Nature of Wage Income



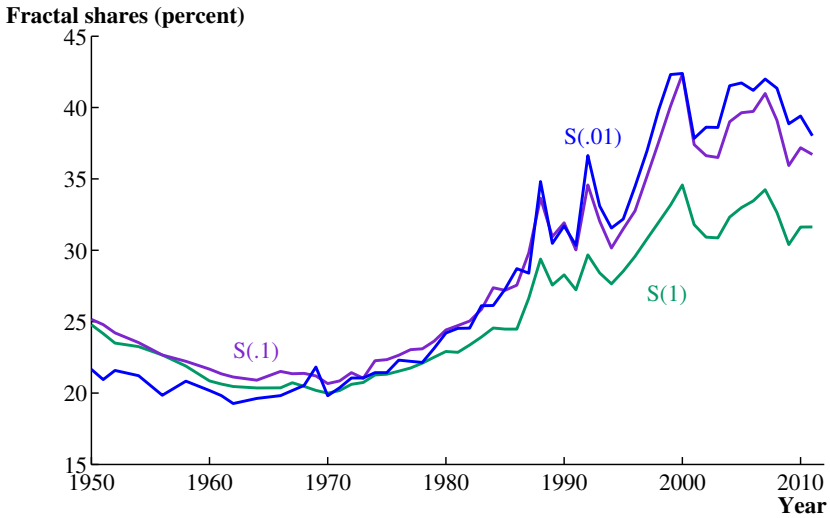
(a) Linear scale, up to $3 million



(b) Log scale

Note: The figures plot the ratio of average wage income above some threshhold z to the threshhold 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.

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.

distribution,

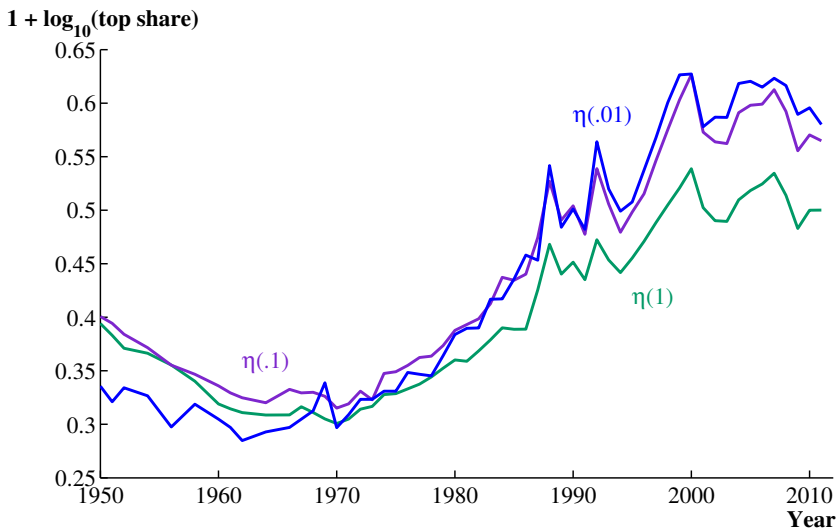
$$S(a) = 10^{\eta-1}. \tag{4}$$

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.

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

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



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

This rise in top income inequality shown in Figure 5 can be related directly to the power-law income inequality exponent using equation (4). Or, put another way, the *change* in the fractal shares is precisely equal to the change in the power law inequality exponent:

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

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.

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 work-

ers 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 $\Pr[y_i > y] = \left(\frac{y}{\bar{w}}\right)^{-1/\eta_y}$. An increase in $\bar{w}$ — a skill-biased technical change that increases the return to skill — shifts the entire 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.

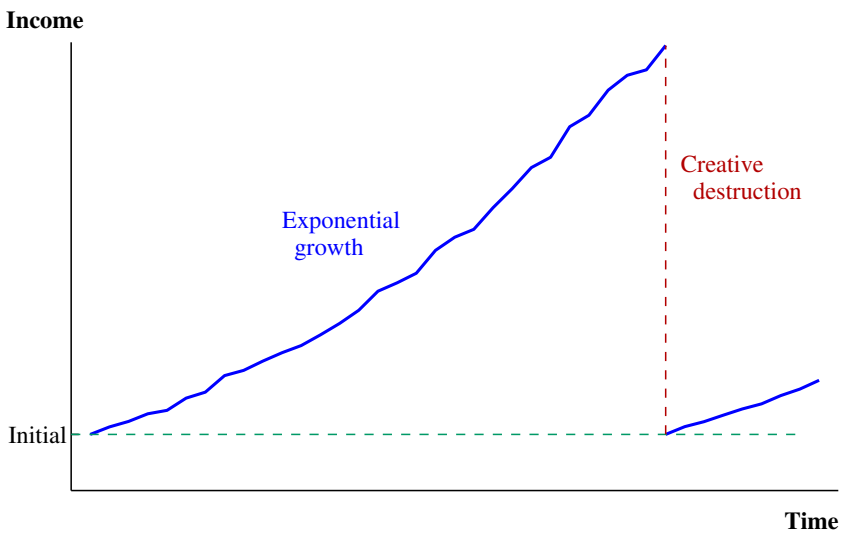
2.4. Summary

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

Figure 7: Basic Mechanism: Exponential growth with death ⇒ Pareto



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 $x(y)$ years, where

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

With a Poisson replacement process, it is well-known that the distribution of expe-

[5]See Gabaix (2009) for a similar stylized model 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).

rience 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; we'll return to consider this step in more detail shortly. Then, the remainder of the argument is straightforward:

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

which is a Pareto distribution!

The power law exponent for income in this model is then $\xi_y = \delta/\mu$, and power law inequality is given by

$$\eta_y = \frac{\mu}{\delta}. \tag{8}$$

In this model, top income inequality can 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.

3.2. The Stationary Distribution of Experience

It is helpful to show the argument that the stationary distribution of experience is exponential. The reason is that this illustrates a simple 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:[6]

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

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

$$\frac{\partial F(x,t)}{\partial t} = \delta(1-F(x,t)) - \frac{\partial F(x,t)}{\partial x} \tag{9}$$

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

$$\frac{\partial f(x,t)}{\partial t} = -\delta f(x,t) - \frac{\partial f(x,t)}{\partial x}. \tag{10}$$

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 (9), 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 expo-

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

nential. In particular,

$$F(x) = 1 - e^{-\delta x}.$$

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 and Market Share

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

$$c_t = \psi_t x_t \tag{11}$$

$$e_t + \ell_t + \tau = 1 \tag{12}$$

$$dx_t = \mu(e_t)x_t dt + \sigma x_t dB_t \tag{13}$$

$$\mu(e) = \phi e \tag{14}$$

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 (13) 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

$$\rho V(x_t,t) = \max_e \ \log \psi_t + \log x_t + \beta \log(\Omega - e_t) + \frac{\mathbb{E}[dV(x_t,t)]}{dt} + (\delta + \bar{\delta})(V^w(t) - V(x_t,t)) \tag{15}$$

subject to (13), 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. 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* (15) *and along the balanced growth path is given by*

$$e^* = 1 - \tau - \frac{1}{\phi} \cdot \beta(\rho + \delta + \bar{\delta}). \tag{16}$$

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 higher baseline quality of the 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 some base quality level 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 entreprenuer 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.[7] 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 (10) that we saw in the simple model at the start of the paper. In particular, the density satisfies

$$\frac{\partial f(x,t)}{\partial t} = -\bar{\delta} f(x,t) - \frac{\partial}{\partial x}\left[\mu(e^*) x f(x,t)\right] + \frac{1}{2}\cdot\frac{\partial^2}{\partial x^2}\left[\sigma^2 x^2 f(x,t)\right] \tag{17}$$

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

$$0 = -\bar{\delta} f(x) - \frac{d}{dx}\left[\mu(e^*) x f(x)\right] + \frac{1}{2}\cdot\frac{d^2}{dx^2}\left[\sigma^2 x^2 f(x)\right] \tag{18}$$

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

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

$$F(x) = 1 - \left(\frac{x}{x_0}\right)^{-\xi^*} \tag{19}$$

where

$$\xi^* = -\frac{\tilde{\mu}^*}{\sigma^2} + \sqrt{\left(\frac{\tilde{\mu}^*}{\sigma^2}\right)^2 + \frac{2\,\bar{\delta}}{\sigma^2}} \tag{20}$$

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

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

growing market share (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 $\bar{\delta}$).[8]

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 equilbrium in order to draw attention to the key economic forces that determine top income inequality.

Next, however, we flesh out the rest of the general equilibrium: how productivity x enters the model, how x affects entrepreneurial income (the proportionality factor ψ_t), and how creative destruction δ^* is determined.

4.3. Production and General Equilibrium

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). 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:

$$Y = \left(\int_0^1 Y_i^\theta di \right)^{1/\theta}. \tag{21}$$

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

$$Y_i = \gamma^{n_t} x_i^\alpha L_i. \tag{22}$$

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

[8]The 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 $\eta^* \to \mu^*/\bar{\delta}$ as $\sigma^2 \to 0$, 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.

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. In this sense, x_i most accurately captures the entrepreneur's productivity in making variety i. It could equivalently be specified as quality, and, as usual, variety i's market share is increasing in x_i.

The main resource constraint in this environment involves labor:

$$L_t + R_t + 1 = \bar{N}, \;\; L_t \equiv \int_0^1 L_{it} di \tag{23}$$

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

$$\dot{n}_t = \lambda(1 - \bar{z})R_t \tag{24}$$

and this also gives the rate of creative destruction:

$$\delta_t = \dot{n}_t. \tag{25}$$

4.4. The Allocation of Resources

There are 12 key endogenous variables in this economic environment: Y, Y_i, x_i, L_i, L, R, n, δ, e_i, c_i, ℓ_i, ψ. 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 3 (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*

$$Y_t = \gamma^{n_t} X_t^{\alpha} L_t \tag{26}$$

$$w_t = \theta \gamma^{n_t} X_t^{\alpha} \tag{27}$$

$$\pi_{it} = (1-\theta)\gamma^{n_t} X_t^{\alpha} \left(\frac{x_{it}}{X_t}\right) L_t \tag{28}$$

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. More inequality (a higher η) therefore has a level effect on output in this economy. Notice that this benefits workers as well by increasing the wage for raw labor. The entrepreneur's profits are linear in idiosyncratic productivity, x_i.

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

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

$$\psi_t = (1-\theta)\gamma^{n_t} X_t^{\alpha-1} L_t. \tag{29}$$

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, $X = \frac{x_0}{1-\eta}$ is constant. Since L is also constant over time, the aggregate production function in equation (26) 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 4 (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*

$$g_y^* = \lambda(1-\bar{z})\bar{s}\bar{L}\log\gamma \tag{30}$$

$$\delta^* = \lambda(1-\bar{z})\bar{s}\bar{L}. \tag{31}$$

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

As a reminder, from Proposition 2, power-law inequality is

$$\eta^* \equiv 1/\xi^*, \text{ where } \xi^* = -\frac{\tilde{\mu}^*}{\sigma^2} + \sqrt{\left(\frac{\tilde{\mu}^*}{\sigma^2}\right)^2 + \frac{2\,\bar{\delta}}{\sigma^2}} \tag{32}$$

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

4.5. 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

[10] At least one of the authors feels a painful twinge writing down a model in which the scale of the economy affects the long-run growth rate. 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.

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 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 affect 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.

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 in-

equality 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 funtion is therefore

$$\rho V^w(t) = \log w_t + \frac{dV^w(t)}{dt} \tag{33}$$

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

$$\rho V^R(t) = \log(\bar{m}w_t) + \frac{dV^R(t)}{dt} + \lambda(1-\bar{z})\left(\mathbb{E}[V(x,t)] - V_t^R\right) + \bar{\delta}_R\left(V(x_0,t) - V^R(t)\right). \tag{34}$$

The first two terms on the right-hand side capture the basic consumption of an unsuccesful 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.

Proposition 5 (Allocation of Labor): *In the stationary general equilibrium, the allocation of labor to research, s, is determined by the following two equations:*

$$s^* = 1 - \frac{L^*}{\bar{L}} \tag{35}$$

$$\log L^* = \log \frac{\theta}{1-\theta} - \beta \log(\Omega - e^*) - \frac{\mu(e^*) - \frac{1}{2}\sigma^2}{\rho + \delta^* + \bar{\delta}} - \frac{(\rho + \delta^* + \bar{\delta}) \log \bar{m} + \lambda(1-\bar{z})\eta^*}{\lambda(1-\bar{z}) + \bar{\delta}_R} - \log(1-\eta^*) \tag{36}$$

where e^* and η^* are given in Proposition 1 and Proposition 2, respectively.

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 in the stationary general equilibrium.

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. It would be valuable to solve the transition dynamics of this model, but this brings in many well-known complications (e.g. Krusell and Smith (1998)) and is beyond the scope of the present paper.[11]

[11] In models that only rely on Brownian motion and not on a Poisson death rate, transition dynamics can be very slow. However, calculations along the lines of Gabaix, Moll, Lasry and Lions (2014) indicate that the speed of convergence to steady state in this model is driven by the Poisson death rate. Later in the paper, we estimate values as high as 10% per year for this number, suggesting that transition dynamics would be relatively fast and helping to justify the focus on steady states. Achdou, Lasry, Lions and Moll (2014) currently explore how to solve for transition dynamics in continuous time models with heterogeneous agents.

Table 1: Equations Characterizing the Stationary General Equilibrium

Drift of log x	$\tilde{\mu}^* = \phi(1-\tau) - \beta(\rho + \delta^* + \bar{\delta}) - \frac{1}{2}\sigma^2$
Pareto inequality	$\eta^* = 1/\xi^*, \;\; \xi^* = -\frac{\tilde{\mu}^*}{\sigma^2} + \sqrt{\left(\frac{\tilde{\mu}^*}{\sigma^2}\right)^2 + \frac{2\bar{\delta}}{\sigma^2}}$
Creative destruction	$\delta^* = \lambda(1-\bar{z})s^*\bar{L}$
Growth	$g^* = \delta^* \log\gamma$
Research share	$s^* = 1 - \frac{L^*}{\bar{L}}$
Allocation of labor	$\log L^* = \log\frac{\theta}{1-\theta} - \beta\log(1-\tau-e^*) - \frac{\tilde{\mu}^*}{\rho+\delta^*+\bar{\delta}}$ $-\frac{(\rho+\delta^*+\bar{\delta})\log\bar{m}+\lambda(1-\bar{z})\eta^*}{\lambda(1-\bar{z})+\delta_R} - \log(1-\eta^*)$

5.1. Comparative Statics

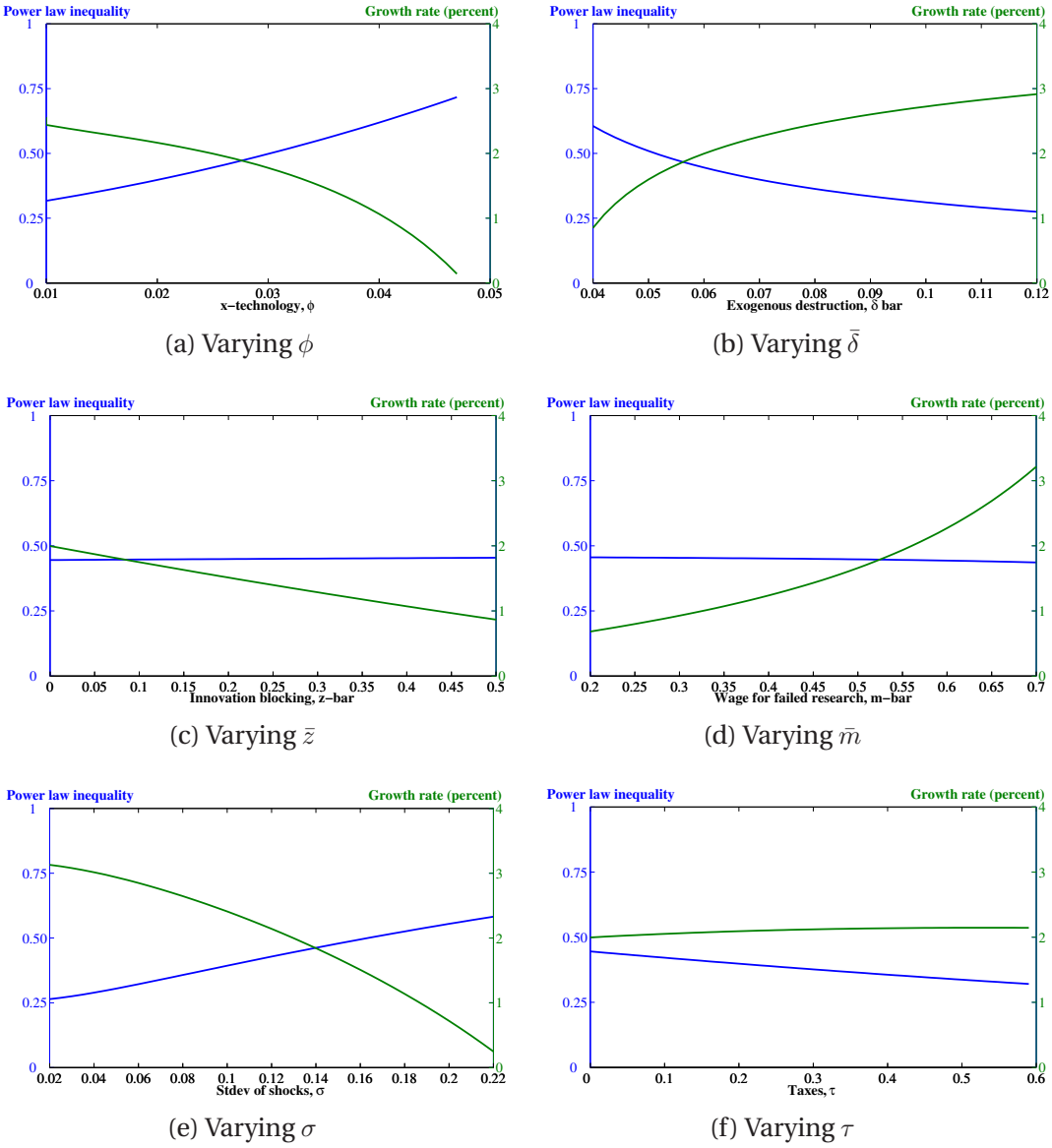
Figure 8 shows numerical results when $s \equiv R/\bar{L}$ is endogenously determined. The effects on Pareto inequality are similar to those from the exogenous $\bar{s}$ case. Now, however, we can also study the effects on economic growth, and those are quite interesting. For example, consider the effect 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 become 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, both for workers and for unsuccessful researchers. However, it also increases the inequality among successful researchers, making the research process itself more risky. Our researchers are risk-averse individuals with log uility, 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 $\bar{\delta}$

(c) Varying $\bar{z}$

(d) Varying $\bar{m}$

(e) Varying σ

(f) Varying τ

Note: The baseline parameter values in these examples are $\rho = .03$, $\bar{L} = 6$, $\theta = 2/3$, $\log\gamma = 1$, $\lambda = .032$, $\phi = .025$, $\beta = .08$, $\sigma = 0.13$, $\bar{\delta} = .06$, $\bar{m} = .56$, $\bar{z} = 0$, and $\tau = 0$. 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 and 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 drift, 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 (2013) 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 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.[12]

6.1. Estimating the Top Income Process for the United States

Motivated by our theory, consider the following empirical model for (normalized) top incomes:

$$\log y_t = \begin{cases} \log y_{t-1} + \tilde{\mu}_t + \sigma_t \epsilon_t & \text{with probability } 1 - \delta_t^e \\ \nu_t & \text{with probability } \delta_t^e \end{cases} \tag{37}$$

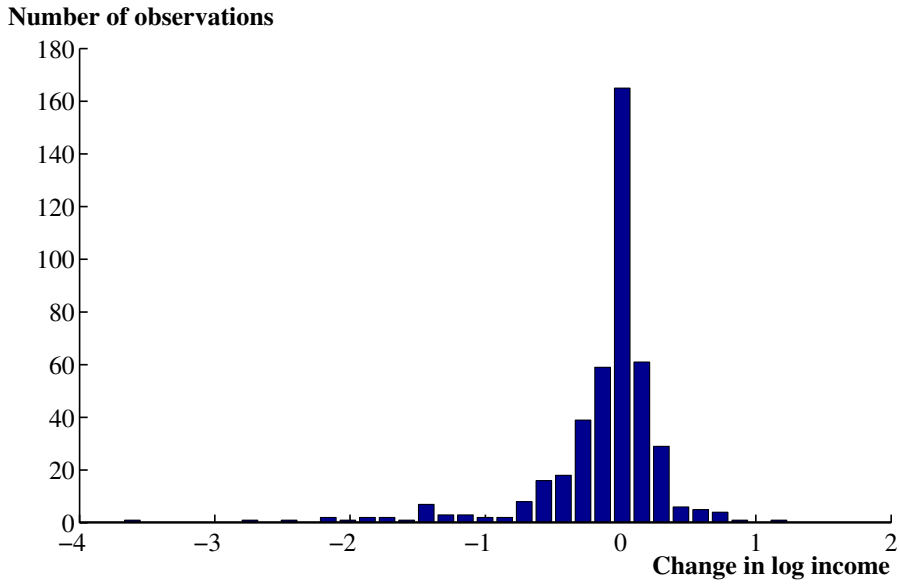
where $\epsilon_t \sim N(0,1)$. That is, with probability $1 - \delta_t^e$, the top income follows a geometric random walk with drift $\tilde{\mu}_t$. Alternatively, with probability δ_t^e, the top income earner reverts to being a "normal" worker and draws a new income from some unspecified distribution. (The "e" superscript in δ^e denotes the "empirical" model; we discuss how δ^e relates to the creative destruction parameter below.) According to this model, if the individual remains a top earner, then the distribution of income growth rates is normal. However, the destruction shock results in a potentially large downward shift in incomes, causing the growth rate distribution to be left-skewed. As noted above, Guvenen, Ozkan and Song (2013) provide evidence of this skewness.

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 9. Quantitatively, the left-skewness of the distribution of growth rates helps us to identify δ_t^e, while the center part of the distribution helps us to identify $\tilde{\mu}$ and σ.

Our data and estimation are discussed in more detail in Appendix A. In brief, the empirical model in equation (37) is estimated by specifying a cutoff value for growth rates: for example, in our benchmark case, if top incomes fall by more than 40 percent,

[12] See http://www.nber.org/taxsim-notes.html.

Figure 9: The Distribution of Top Income Growth Rates between 1988 and 1989



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.

we consider this a destruction event. The fraction of growth rates below this cutoff is an estimate of δ^e, and the remaining growth rates are used to estimate the mean and variance parameters. Finally, and motivated by the empirical income dynamics literature, we make an adjustment for the presence of temporary income shocks, which are absent from our theory. Calculations from Blundell, Pistaferri and Preston (2008) and Heathcote, Perri and Violante (2010) suggest that the variance of the random walk innovation accounts for only about 1/6 to 1/3 of the variance of income growth rates. It is unclear how this applies to top incomes. Hence we make the following correction: we have already reduced the variance of growth rates to some extent by cutting off the thick left tail through our destruction shocks. In addition, we calculate σ^2, the variance of the random walk innovation, to be 1/3 the variance of the remaining growth rates.[13]

For each year between 1980 and 1990, we use the cross section of growth rates from a two-year panel of tax returns to estimate the 3 time-varying parameters of this empirical model. We compute 95 percent confidence intervals using 5000 bootstrap random samples. Finally, we do this for two cut-offs of "top" incomes, corresponding to incomes in the Top 5 percent and incomes in the Top 10 percent, respectively.[14]

6.2. Results

Figures 10 through 12 shows the point estimates and 95 percent confidence intervals for various key parameters of the model. In each figure, the blue line and the grey bands correspond to estimates using the Top 5 percent of incomes, while the green line and the tan bands correspond to estimates from the Top 10 percent.

Overall, the key takeaways from this estimation are

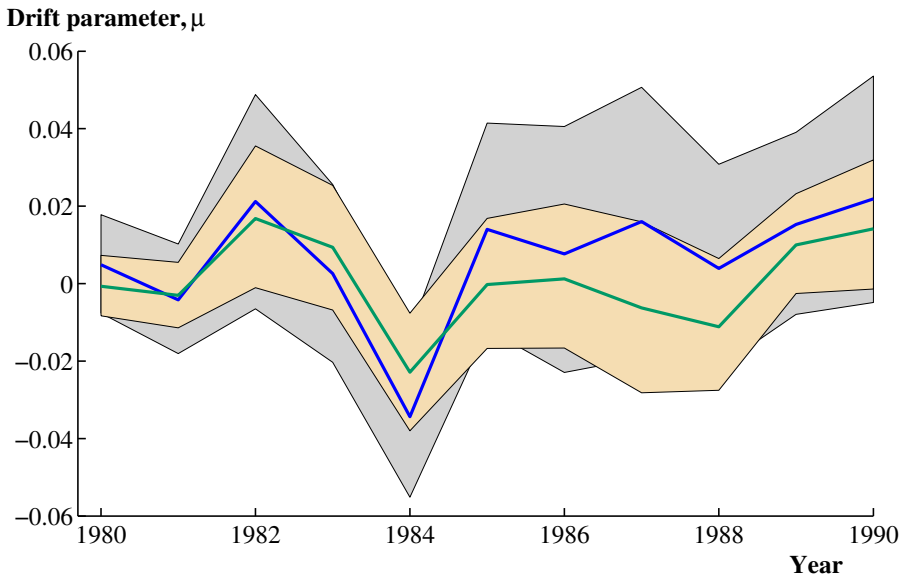
- The drift parameter $\mu \equiv \tilde{\mu} + \frac{1}{2}\sigma^2$ shows clear evidence of increasing during the 1980s, especially in the first half of the decade. For example, using the Top 5 percent data, the value of μ rises from 0.005 in 1980 to 0.022 in 1990.

- The confidence intervals can be wide. Partly this reflects the changing sample

[13] As shown below, this leads to estimates of σ of around 0.15 in 1990, corresponding to a variance of the random walk innovations of 0.0225. This is in the right ballpark based on the empirical literature.

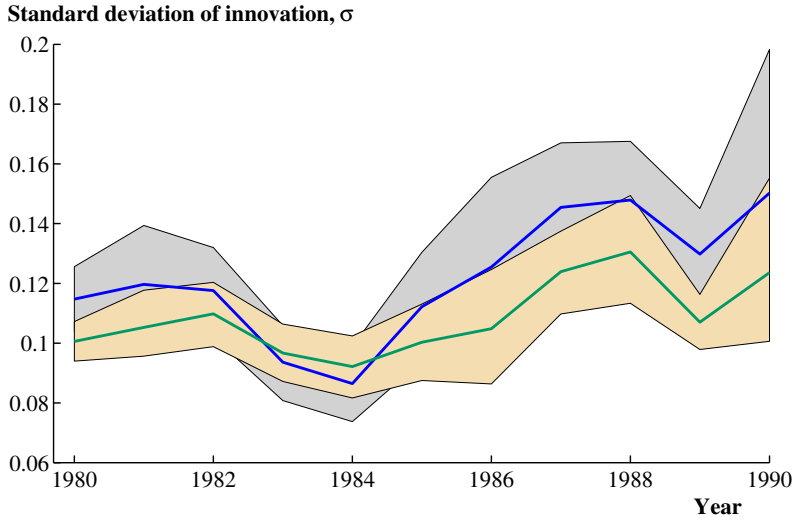
[14] We experimented with more structural models, including specifying the distribution of ν_t as normal and estimating a 2-state mixture model of normal random variables for each cross-section at a point in time. This yields broadly similar results but is viewed as inferior because it yields a very high variance for ν_t, potentially allowing incomes to *increase* when a destruction shock occurs.

Figure 10: Estimates of μ from U.S. IRS Public Use Panel



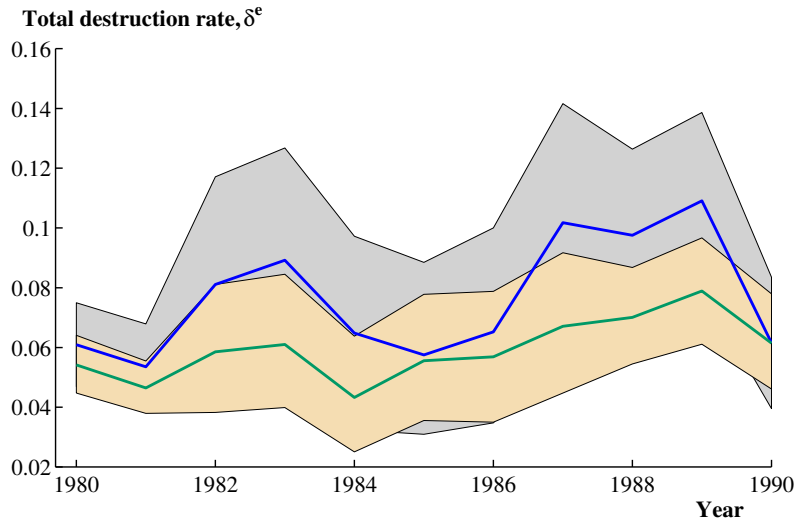
Note: This figure shows the estimates of μ when using the Top 5 (blue/grey) and Top 10 (green/tan) percentiles of wage/salary income from the IRS public use panel data. 95 percent confidence intervals obtained with 5000 bootstrap iterations are shown.

Figure 11: Estimates of σ from U.S. IRS Public Use Panel



Note: This figure shows the estimates of the standard deviation parameter for the innovation to the Brownian motion, σ, when using the Top 5 (blue/grey) and Top 10 (green/tan) percentiles of wage/salary income from the IRS public use panel data. 95 percent confidence intervals obtained with 5000 bootstrap iterations are shown.

Figure 12: Estimates of δ^e from U.S. IRS Public Use Panel



Note: This figure shows the estimates of the overall destruction rate $\delta^e = \delta + \bar{\delta}$ when using the Top 5 (blue/grey) and Top 10 (green/tan) percentiles of wage/salary income from the IRS public use panel data. This parameter is estimated by the fraction of growth rates that reduce top incomes by more than 40 percent. 95 percent confidence intervals obtained with 5000 bootstrap iterations are shown.

sizes. For example, for the Top 5 percent estimation, the sample size for estimating μ and σ in 1980 is 1067 observations, falling to a low of 213 observations in 1983, and then ending with 455 observations in 1990.

- As shown in Figure 11, the standard deviation of the innovation, σ, shows some evidence of increasing, but the confidence intervals are wide. Using the Top 5 percent data, the value is 0.115 in 1980 and rises to 0.150 in 1990.

- The estimates of δ_t^e are shown in Figure 12. The mean value is 0.077, and while there is some indication of a slight upward trend (which would reduce top inequality), confidence intervals are wide.

A useful way to summarize the movements in these parameters is to use our theoretical model to compute their implications for long-run top income inequality. More specifically, for each year, we use the parameter estimates from that year to compute the value of η implied by equation (32). There are two caveats to this calculation. First, the value of η is actually the value that would apply in the long run, not immediately, as

this is a steady-state result. However, bearing this caveat in mind, the calculation as a summary still seems useful. Second, given values for μ and σ, the calculation requires a value of $\bar{\delta}$. We obtain this by noting that $\bar{\delta} = \delta^e - \delta$. That is, we simply need to subtract the endogenous rate of creative destruction from the point estimate. For illustration, we assume a constant value of $\delta = .02$.

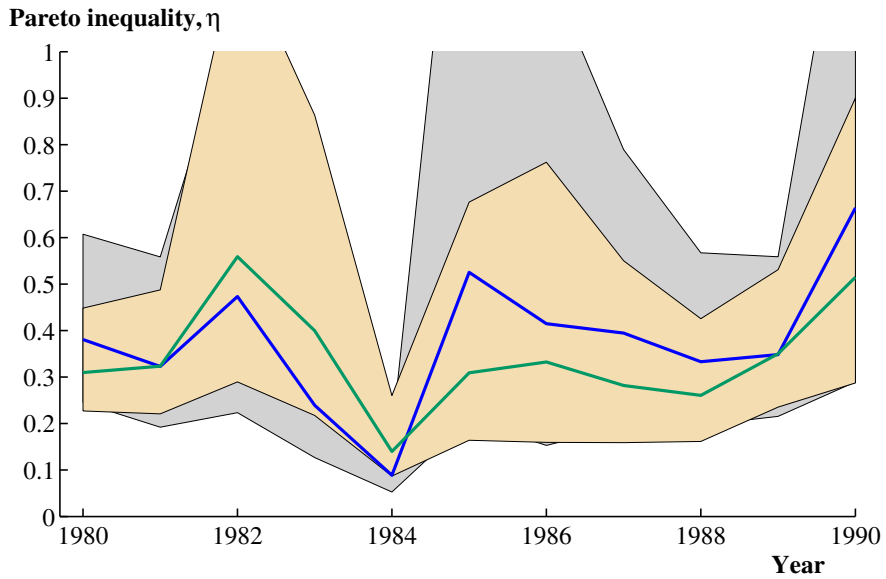
Figure 13 then shows the estimates of η implied by this calcuation, together with bootstrapped confidence intervals. Two key results are suggested by the figure. First, the estimates of Pareto inequality are on average close in magnitude to the numbers we observe in the data. Second, there is some suggestion of a rise in top inequality, particularly using the Top 5 percent sample: η rises from an average of 0.27 in 1982–84 to 0.45 in 1988–90, albeit with extremely wide confidence intervals. One way to get a simple feel for these numbers is to return to the simple model at the start of the paper. There, recall, $\eta = \mu/\delta$. So a value of μ that rises from 0.025 to 0.045 with a value of δ around 0.10 is roughly consistent with these facts. Alternatively, the rise in η for the U.S. data is from around 0.33 to 0.55, which would correspond to μ rising from 0.033 to 0.055, for example.

The evidence from the IRS public use panel of tax returns is obviously limited in many ways. Nevertheless, the evidence from estimation using these data is revealing. It suggests that the basic model used in this paper is capable of explaining a substantial rise in U.S. top income inequality. Access to the restricted administrative tax data from the IRS and from tax authorities in other countries could be used to make additional progress and is a clear priority for future research.

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 from the IRS panel data from the previous section. For example, we assume that σ for the United States rises from 0.10 to 0.15, broadly consistent with the evidence in Figure 11. We assume $\bar{\delta} = 0.06$ and $\log \gamma = 1$, so that $\delta = 0.02$ when the economy's growth rate is 2 percent, and therefore $\bar{\delta} + \delta = 0.08$, broadly consistent

Figure 13: Implied Long-Run Pareto Inequality from the IRS Public Use Panel



Note: This figure shows the estimates of the implied steady-state Pareto inequality, η, based on the parameter values estimated for the stochastic income process. The calculation assumes the endogenous creative destruction rate δ equals 0.02. Estimates correspond to using the Top 5 (blue/grey) and Top 10 (green/tan) percentiles of wage/salary income from the IRS public use panel data. 95 percent confidence intervals obtained with 5000 bootstrap iterations are shown.

with the evidence in Figure 12.

These remain examples, however, for three main reasons. First, the U.S. panel evidence we have on μ, σ, and δ^e applies only to the years 1980 to 1990, whereas we attempt in this section to shed light on inequality movements all the way until 2007. Second, 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. Finally, for reasons discussed earlier, our numerical exercises do not consider transition dynamics and instead report a sequence of steady 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.

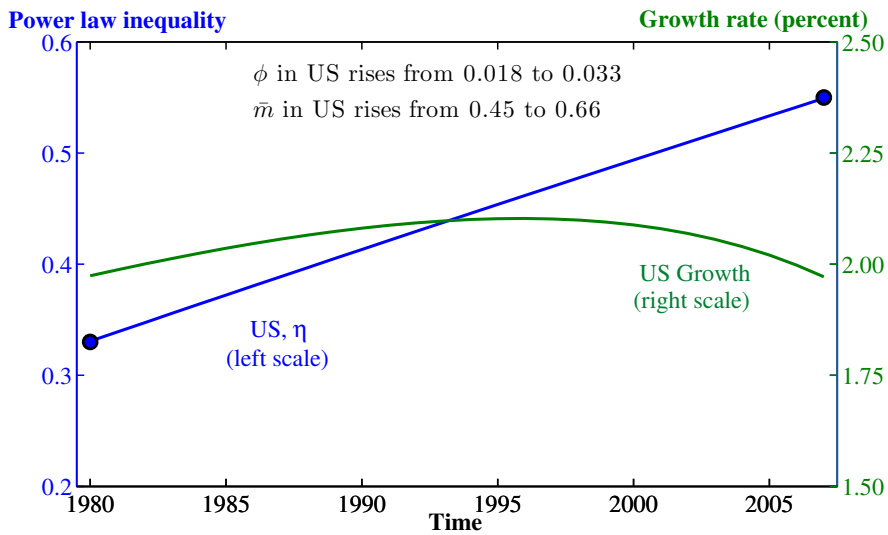
7.1. Matching U.S. Inequality

Figure 14 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 — $\rho = 0.03, \bar{L} = 6, \tau = 0, \theta = 2/3, \beta = 0.08, \lambda = 0.033, \bar{z} = 0, \sigma = 0.10, \bar{\delta} = 0.06, \phi = .0185$, and $\bar{m} = 0.45$ — which match U.S. Pareto inequality in 1980. Next, we allow σ to change roughly as it does in the data; specifically, we allow it to rise linearly over time to a value of 0.15 in 2007. Finally we assume that ϕ, the technology parameter converting entrepreneurial effort into growth in a variety's productivity, rises over time. We calibrate the change in ϕ to match the rise in the U.S. top inequality between 1980 and 2007. The values of ϕ we recover produce values of μ ranging from 0.010 to 0.024, broadly consistent with the evidence in Figure 10.

Although the increases in ϕ and σ match the changes in inequality, they have a negative impact on growth. An increase in ϕ leads to a rise in entrepreneurial effort and to a rise in inequality, while σ raises inequality directly. Because prospective researchers are risk averse, more inequality among entrepreneurs reduces the number of researchers, thereby limiting overall growth. To offset this, we vary another parameter in the U.S. in Figure 14. In particular, we assume that the subsidy to research, $\bar{m}$, rises from 0.45 to 0.66, which stabilizes U.S. growth on average.[15]

[15]There are other ways to achieve this same end. For example, one could let $\bar{L}$ increase instead, possibly reflecting globalization.

Figure 14: Numerical Example: Matching U.S. Inequality



Note: Baseline parameter values are given in the text. Over time in this simulation, ϕ rises linearly from 0.0185 to 0.0325, σ rises linearly from 0.10 to 0.15, and $\bar{m}$ rises linearly from 0.45 to 0.66. The rise in $\bar{m}$ is required to keep growth rates from falling sharply with the rise in ϕ and σ.

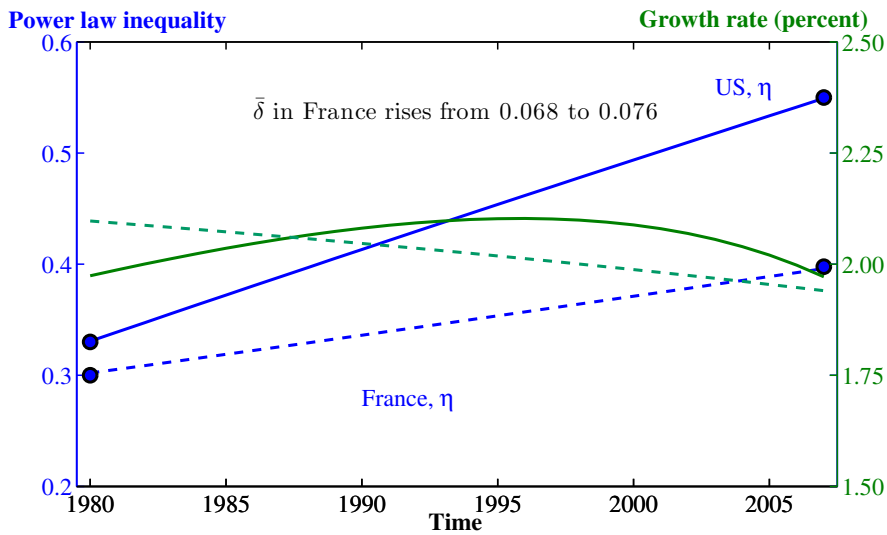
7.2. Matching Inequality in France

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 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 a rise in $\bar{\delta}$. In particular, a linear increase in $\bar{\delta}$ from 0.068 to 0.076 (versus a constant value of 0.06 for the United States) allows us to match the evidence on French inequality, as shown in Figure 15.

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

Figure 15: Numerical Example: U.S. and France



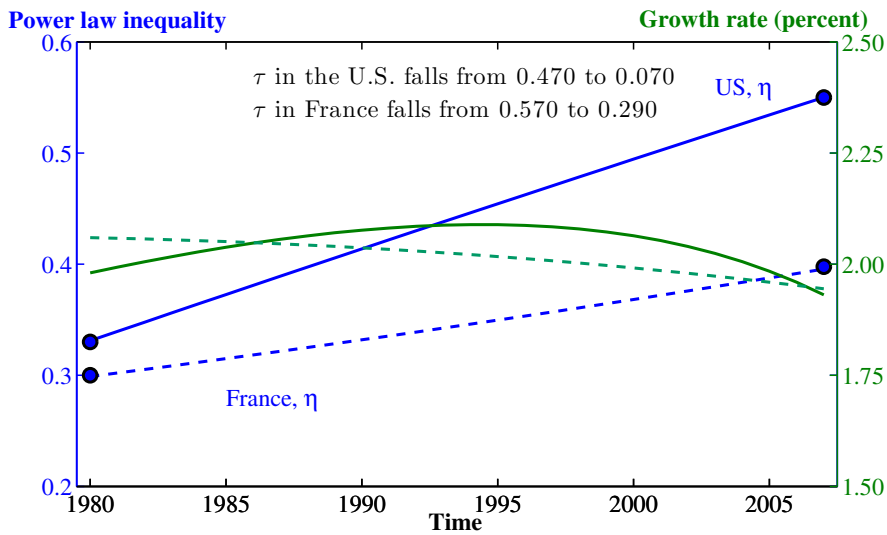
Note: The baseline parameter values described in the text are used; for example, the U.S. parameter values are the same as in Figure 14. In France, we assume ϕ rises in exactly the same way. To offset the large increase in inequality that would be implied, we assume $\bar{\delta}$ rises from 0.068 to 0.076. For France, we assume constant values of $\sigma = .10$ and $\bar{m} = .45$.

τ. A literal interpretation of this parameter is that it 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 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 16. The baseline parameter values used in this example are identical to those used in Figure 15, with two exceptions. First, we raise the (constant) level of ϕ to 0.035 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

[16]Kim (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 some a modest portion of the changes in inequality over time and across countries.

Figure 16: 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. As before, σ rises linearly from 0.10 to 0.15 and $\bar{m}$, rises linearly in the U.S. case from 0.48 to 0.66 to offset what would otherwise be a decline in growth. For France, we assume constant values of $\sigma = .10$ and $\bar{m} = .49$.

States fall from 47 percent to 7 percent generates the U.S. inequality numbers. And a more modest decline from 57 percent to 29 percent can generate the French inequality facts. The U.S. tax change by itself, however, leads to a large *decline* in overall income growth, suggesting that something else most also be going on. As before, we offset this growth effect with a rise in $\bar{m}$, in this case from 0.48 to 0.66.

8. Conclusion

A model in which entrepreneur's expend effort to enhance the productivity of their existing ideas while researchers seek new ideas to replace incumbents in a process of creative destruction generates a Pareto distribution for labor incomes at the top of the income distribution. Moreover, it suggests economic forces that can give rise to changes in top income inequality. Forces that increase the effort of existing entrepreneurs in improving their products — or that increase the efficiency of their effort — can increase top inequality. Forces that enhance creative destruction 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 Appendix: Estimating the U.S. Stochastic Income Process

A1. Data

The data we use are the U.S. Internal Revenue Service public use tax model panel files created by the Statistics of Income Division from 1979 to 1990, hosted by the NBER. See http://www.nber.org/taxsim-notes.html. We restrict our sample to tax units that involved married taxpayers filing jointly in two consecutive years and convert nominal values to 2012 constant dollars using the consumer price index. 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 5 percent of our income measure. As a robustness check, we also consider a Top 10 percent cutoff.

A2. Estimation

For each set of consecutive years, we use our cross-section of income levels and growth rates to estimate the stochastic process in equation (37). Hence, our three parameters $\tilde{\mu}_t$, σ_t, and δ_t^e are indexed by time. We assume that any growth rates that reduce in-

come by more than Δ percent are due to the destruction shocks. For our benchmark estimates, we assume $\Delta = 40$; we also considered values of 50 and 60 and got similar results. Our estimate of δ_t^e is therefore the fraction of growth rates that reduce incomes by more than Δ percent. We then use the remaining growth rates to estimate $\tilde{\mu}_t$ and σ_t using the sample mean and sample standard deviation.[17] As discussed in the text, we estimate σ^2 as 1/3 the sample variance to capture the innovation to the random walk as opposed to temporary income shocks. In our setting, temporary income shocks with a "thinner" tail (e.g. log normal) than the income distribution will not affect our measure of Pareto inequality. We conduct this exercise for every cross section of growth rates between 1980 and 1990 to recover our three parameters in each year. We compute 95 percent confidence intervals using 5000-draw bootstraps in each year. Complete point estimates and sample sizes are shown in Tables A1 and A2.

[17]We also estimated $\tilde{\mu}_t$ and σ_t by fitting a truncated normal distribution, which generated very similar results.

Table A1: Parameter Estimates (Top 5 percentile cutoff)

	— Basic parameters —			Implied	
Year	μ	σ	δ^e	η	N
1980	0.0049	0.1147	0.0609	0.3807	1067
1981	-0.0042	0.1197	0.0535	0.3227	1046
1982	0.0212	0.1176	0.0811	0.4732	222
1983	0.0026	0.0936	0.0892	0.2391	213
1984	-0.0343	0.0865	0.0648	0.0889	216
1985	0.0140	0.1122	0.0575	0.5252	226
1986	0.0077	0.1255	0.0652	0.4148	230
1987	0.0160	0.1454	0.1018	0.3945	226
1988	0.0039	0.1479	0.0976	0.3329	451
1989	0.0153	0.1298	0.1091	0.3483	440
1990	0.0219	0.1502	0.0615	0.6639	455

Note: See text for details. For the implied parameter values, $\mu \equiv \tilde{\mu} + \frac{1}{2}\sigma^2$ and η is computed from the other parameter values using equation (32) with $\bar{\delta} = \delta^e - \delta$ assuming $\delta = .02$.

Table A2: Parameter Estimates (Top 10 percentile cutoff)

	— Basic parameters —			Implied	
Year	μ	σ	δ^e	η	N
1980	-0.0007	0.1006	0.0542	0.3096	2123
1981	-0.0030	0.1053	0.0465	0.3233	2109
1982	0.0168	0.1098	0.0586	0.5591	444
1983	0.0094	0.0967	0.0610	0.3998	426
1984	-0.0229	0.0922	0.0433	0.1398	439
1985	-0.0002	0.1003	0.0556	0.3092	450
1986	0.0012	0.1048	0.0569	0.3325	457
1987	-0.0063	0.1239	0.0671	0.2820	447
1988	-0.0111	0.1305	0.0701	0.2605	899
1989	0.0100	0.1070	0.0789	0.3500	900
1990	0.0141	0.1236	0.0615	0.5147	911

Note: See notes to Table A1.

B 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 (15) yields

$$\frac{\beta}{\Omega - e^*} = \phi V_x(x_t, t) x_t, \tag{A1}$$

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 (A1) as

$$\frac{\beta}{\Omega - e^*} = \phi \zeta_2. \tag{A2}$$

Now substituting (A2) into (15), we have

$$\begin{aligned}(\rho + \delta + \bar{\delta})(\zeta_0 + \zeta_1 t + \zeta_2 \log x_t) \\ = \beta \log\left(\frac{\beta}{\phi\zeta_2}\right) + \zeta_2 \phi e^* - \frac{1}{2}\zeta_2 \sigma^2 + \zeta_1 + (\delta + \bar{\delta}) V^w(t) + \log \psi_t + \log x_t\end{aligned} \tag{A3}$$

Equating the coefficients on $\log x_t$ yields

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

We then substitute ζ_2 into (A2) 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 (A3) 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, (27) and (29) imply that both $\log w_t$ and $\log \psi_t$ are linear functions of n_t. Since $\dot{n}_t$ is constant in the stationary equilibrium, $\log w_t$ and $\log \psi_t$ are linear in t. Therefore, the right-hand side of (A3) will have the same form as our conjecture, and we obtain ζ_0 and

ζ_1 by equating the coefficients and constants in (A3). QED.

Proof of Proposition 2. The Pareto Income Distribution

Substituting our guess $f(x) = Cx^{-\xi-1}$ to (18), 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 (20). QED.

Proof of Proposition 3. 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

$$\left(\frac{Y}{Y_i} \right)^{1-\theta} = p_i. \tag{A4}$$

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

$$\max_{Y_i} p_i(Y_i)Y_i - wL_i = Y^{1-\theta} Y_i^\theta - \frac{w}{\gamma^n x_i^\alpha} Y_i. \tag{A5}$$

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

$$Y_i = \left(\frac{1}{\theta}\frac{w}{\gamma^n x_i^\alpha}\right)^{\frac{1}{\theta-1}} Y. \tag{A6}$$

By plugging (A6) in the final goods production function, we obtain the equilibrium wage equation

$$w = \theta\gamma^n \left(\int_0^1 x_i^{\alpha\frac{\theta}{1-\theta}}\right)^{\frac{1-\theta}{\theta}} \equiv \theta\gamma^n X^\alpha, \tag{A7}$$

where we assume $\alpha\frac{\theta}{1-\theta} = 1$ and $X \equiv \int_0^1 x_i di$. Using this equation we can rewrite (A6) as

$$Y_i = \left(\frac{x_i}{X}\right)^{\frac{1}{\theta}} Y. \tag{A8}$$

Next, combining (A8) and (22) 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:

$$Y = \gamma^n X^\alpha L. \tag{A9}$$

Lastly, the profit π_i is calculated from plugging the optimal solution (A4), (A6), and (A8) into the monopoly problem (A5). QED.

Proof of Proposition 4. Growth and inequality in the $\bar{s}$ case

The proof is provided in the main text. QED.

Proof of Proposition 5. Allocation of Labor

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 (33) can be rewritten as

$$V^w(t) = \int_t^\infty \exp^{-\rho(\tau-t)} \log w_\tau d\tau = \frac{1}{\rho}\log w_t + \frac{g}{\rho^2}, \tag{A10}$$

where the last equality comes from the fact that w_t given in Proposition 3 grows at the constant rate of growth $g \equiv \dot{n}_t \log\gamma$ 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 (34) 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 (15) and (A3). We rewrite (A3) as

$$(\rho+\delta+\bar{\delta})V(x_t,t) = \log\psi_t + \log x_t + C + (\delta+\bar{\delta})V^w(t) + \frac{dV(x,t)}{dt}, \tag{A11}$$

where $C = \beta\log(\Omega - e^*) + \dfrac{\phi e^* - \frac{1}{2}\sigma^2}{\rho+\delta+\bar{\delta}}$ contains constant terms. Differentiating (A11) with respect to time, we obtain

$$\frac{dV(x,t)}{dt} = \frac{1}{\rho+\delta+\bar{\delta}}\left(\frac{\dot{\psi}_t}{\psi_t} + (\delta+\bar{\delta})\frac{dV^w}{dt}\right) = \frac{g}{\rho}. \tag{A12}$$

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 (A12) and (A10) into (A11) yields

$$(\rho+\delta+\bar{\delta})V(x,t) = \log\frac{1-\theta}{\theta} + \log L_t - \log x_0 + \log(1-\eta) + \log x_t + C + (\rho+\delta+\bar{\delta})V^w(t). \tag{A13}$$

Now taking expectations on (A13) and rearranging, we get

$$\mathbb{E}[V(x_t,t)] = \frac{1}{\rho+\delta+\bar{\delta}}\left(\log\frac{1-\theta}{\theta} + \log L_t + \log(1-\eta) + \eta + C\right) + V^w(t), \tag{A14}$$

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 (A13) that

$$V(x_0,t) = \frac{1}{\rho+\delta+\bar{\delta}}\left(\log\frac{1-\theta}{\theta} + \log L_t + \log(1-\eta) + C\right) + V^w(t). \tag{A15}$$

We then rewrite (A14) as

$$\mathbb{E}[V(x_t,t)] = V(x_0,t) + \frac{\eta}{\rho+\delta+\bar{\delta}}. \tag{A16}$$

Next substituting (A16) into (34) and rearranging, we obtain

$$(\rho+\lambda(1-\bar{z})+\bar{\delta}_R)V^R(t) = \log\bar{m} + \underbrace{\log w_t + \frac{g}{\rho}}_{=\rho V^w(t)} + (\lambda(1-\bar{z})+\bar{\delta}_R)V(x_0,t) + \frac{\lambda(1-\bar{z})\eta}{\rho+\delta+\bar{\delta}}. \tag{A17}$$

Lastly we substitute (A15) into (A17) and apply the indifference equation $V^w(t) = V^R(t)$ to (A17) 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 obtain the allocation of labor in (36). 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.

Achdou, Yves, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "Heterogeneous Agent Models in Continuous Time," 2014. Princeton University manuscript.

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, March 1992, *60* (2), 323–351.

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

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.

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.

—, Benjamin Moll, Jean-Michel Lasry, and Pierre-Louis Lions, "The Dynamics of Inequality," July 2014. Princeton University manuscript.

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.

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.

—, Serdar Ozkan, and Jae Song, "The Nature of Countercyclical Income Risk," Technical Report, Board of Governors of the Federal Reserve System (U.S.) September 2013. working paper.

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.

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.

Krusell, Per and Anthony A. Smith, "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, October 1998, *106* (5), 867–896.

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.

— , "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.

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.