

A Schumpeterian Model of Top Income Inequality

Charles I. Jones

Stanford GSB and NBER

Jihee Kim *

KAIST

December 28, 2013 – Version 0.44

Preliminary and Incomplete

Abstract

Top income inequality rose sharply in the United States over the last 30 years but increased only slightly in economies like France and Japan. Why? This paper explores a model in which the interplay between entrepreneurial efforts to grow existing market share and the creative destruction associated with outside innovation lead top incomes to obey a Pareto distribution. The extent to which entrepreneurs, interpreted very broadly, can grow the markets for their existing ideas is a key determinant of top income inequality, while the creative destruction by which one entrepreneur replaces another restrains inequality. Any changes in the economy that make a given amount of entrepreneurial effort more effective at building market share will increase top inequality. Examples might include the world wide web and possibly globalization. On the other hand, policies that restrain the growth of market share by entrepreneurs can limit inequality. Differences in these considerations across countries may help explain the divergent patterns of top income inequality that we see in the data.

*We are grateful to Xavier Gabaix, Mike Harrison, Chris Tonetti, and seminar participants at the Federal Reserve Bank of San Francisco, Harvard, Princeton, Stanford, and USC 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, however. For example, top inequality rose only slightly in France and Japan. Why? What economic forces explain the rise in top income inequality in the United States? And why has top inequality risen by much less in countries like France and Japan?

We consider these questions from a theoretical standpoint. 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 changes in the power law exponent.

We are far from being able to point to a single explanation, say, for explaining the rise in top income inequality in the United States. Instead, we present a model that highlights different economic forces that could potentially contribute to such an explanation.

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 incomes are Pareto and why Pareto inequality is changing over time, rather than why a power law inequality measure is so close to unity.¹ The basic insight in this literature is that exponential growth, tweaked slightly, can deliver a Pareto distribution for outcomes. The tweak is needed for the following reason. Suppose that city populations (or incomes or employment by firms) grow exponentially at 2% per year plus some random normally-distributed shock. In this case, the log of population would follow a normal distribution with a variance that grows over time. To keep the distribution from spreading out forever, we need some kind of “tweak.” For example, a constant probability of death will suffice to render the distribution stationary, as would some minimum size for the city.

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 entrepreneurs increase their productivity and build market share for their new products. The growth rate of this process is tied to entrepreneurial effort, and anything that raises this effort, resulting in faster growth of idiosyncratic productivity, will raise top income inequality. The “death rate” in our setup is naturally tied to creative destruction: researchers create new products that make the previous best-selling song or smartphone application obsolete. A higher rate of creative destruction restrains the random growth process and results in lower top income inequality. In this way, the interplay between existing entrepreneurs growing their markets 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. The next two sections then develop

¹These papers in turn build on a large literature on such mechanisms outside economics. For example, see Reed (2001), Mitzenmacher (2004), and Malevergne, Saichev and Sornette (2010). Gabaix (2009) and Luttmer (2010) have excellent surveys of these mechanisms, written for economists.

the model, first with an exogenous allocation of labor to research (which makes the main mechanism transparent), and then more fully with an endogenous allocation of labor. The final main section studies numerical examples to illustrate the quantitative possibilities of the framework. A conclusion summarizes what remains to be done and how the model could be taken to the data more carefully.

1.1. The existing literature

Before moving on, it is worth pausing to consider a couple of natural candidate explanations in order to understand why they fall short, at least in their simple form. First, consider a basic income tax explanation. It is well-known that top marginal tax rates have fallen sharply in the United States. For example, in the early 1960s, the top marginal tax rate exceeded 90%, but by 2008, the top marginal tax rate was just 35%. Why doesn't this decline explain the rise in top income inequality in the U.S.? The key problem with this explanation is that, at least in its simple form, it cannot work at the very top. Because inequality is rising even at the very top, one needs to explain why the Top 0.01% share is rising more than the Top 0.1% share. But because earnings are so high for both of these groups, they are both facing the same top marginal tax rate. A decline in that rate from 90% to 35% would not tilt the income toward the higher earning group, which is what we require. This point is emphasized by Piketty, Saez and Stantcheva (2011). Kim (2013) shows that the change in marginal tax rates can play a more important role in the presence of a random-growth process but concludes that this is only a part of the story for the U.S.

Next, consider explanations related to fairness. A simple version would be that cultural norms in France, say, seek to keep top CEOs from making more than \$5 million per year, or more than 100 times the average wage. The problem with explanations like this, which involve implicit caps of some kind, is that they would lead to a clumping of incomes around the cap. Top incomes would not look like a Pareto distribution in this case.

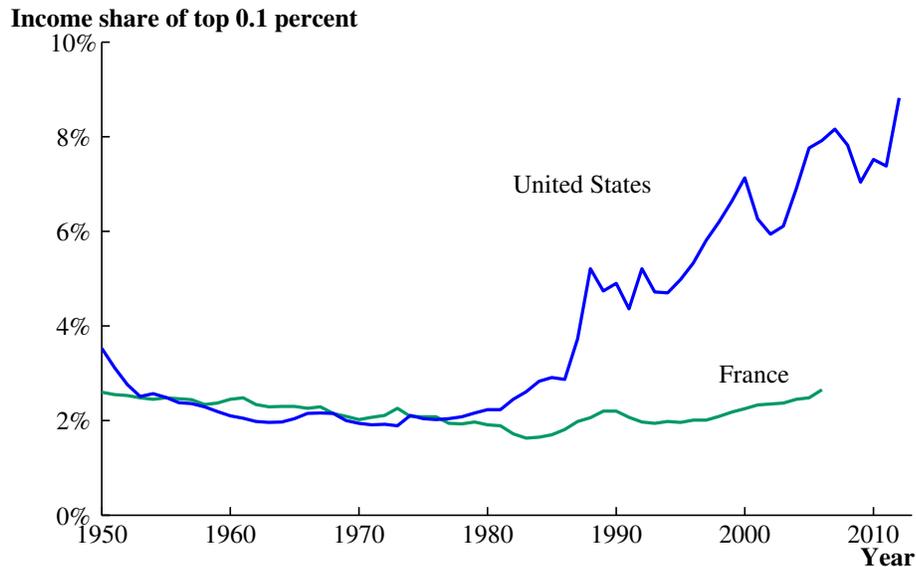
A number of other recent papers contribute to our understanding of the dynam-

ics of top income inequality. Piketty, Saez and Stantcheva (2011) 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. Haskel, Lawrence, Leamer and Slaughter (2012) suggest that globalization may have raised the returns to superstars via a Rosen (1981) mechanism. 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. 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 (2011) 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.

The most closely-related papers to this one are Benhabib, Bisin and Zhu (2011), Nirei (2009), Aoki and Nirei (2013), and Moll (2012). 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.

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



Source: World Top Incomes Database.

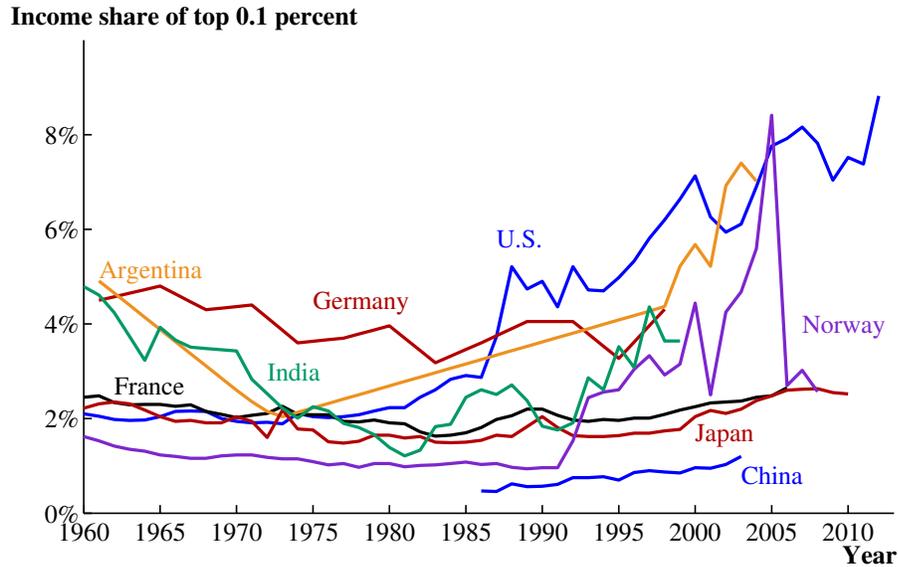
As mentioned earlier, this paper is not able to conclusively distinguish between various explanations that might contribute to the different patterns of top income inequality that we see around the world. Instead, the goal is less ambitious. Drawing on Pareto-generating mechanisms that have been used to study Zipf's Law for cities and the size distribution of firms, we hope to provide a baseline model that explains why top income inequality obeys a power law and suggest some of the forces that may contribute to the heterogeneity we see across countries and over time.

2. Some Basic Facts

Figures 1 through 3 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, compared to the relative stability of inequality in France.

Figure 2 shows these same statistics for more countries. The facts that stand

Figure 2: Top Income Inequality around the World



Source: World Top Incomes Database.

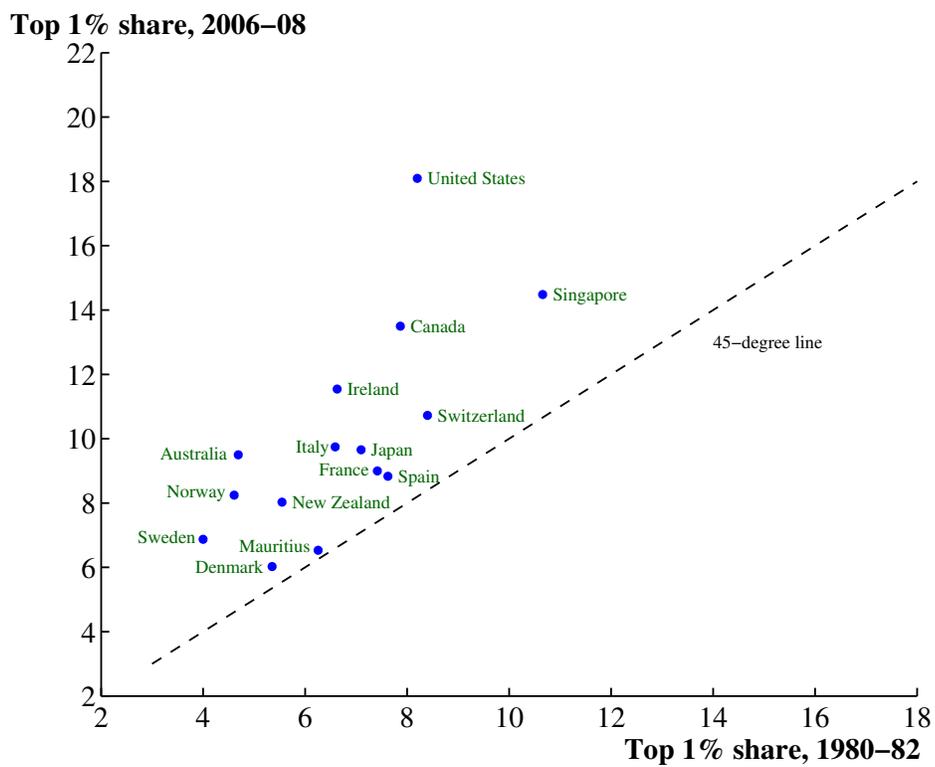
out are (i) heterogeneity, (ii) the basic stability of the distribution between 1960 and 1980, and (iii) the rise in top inequality in a number of countries since 1980.

Finally, Figure 3 shows the rise since 1980 more systematically. The horizontal axis shows the top 1% share averaged between 1980 and 1982, while the vertical axis shows the same share for 2006–2008. Nearly all the economies for which we have data lie above the 45-degree line: that is, top income inequality has risen almost everywhere. The rise is largest in the United States, South Africa, and Norway, but substantial increases are also seen elsewhere, such as in Ireland, Portugal, Singapore, Italy, and Sweden. Japan and France exhibit smaller but still noticeable increases. For example, the top 1% share in France rises from 7.4% to 9.0%.

2.1. Fractal Inequality and the Pareto Distribution

It is well known, dating back to Pareto (1896), that the top portion of the income distribution is accurately characterized by a power law. That is, at high levels, the

Figure 3: 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.

income distribution is approximately Pareto. Saez (2001) documents this fact carefully using U.S. income tax records. If Y is a random variable denoting incomes above some high level (i.e. for $Y \geq y_0$), then

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

where ξ is called the “power law exponent.”

An important property of the Pareto distribution can then be seen easily. In particular, let $\tilde{S}(p)$ denote the share of income going to the top p percentiles. This share is given by $(p/100)^{1-1/\xi}$. A larger power-law exponent, ξ , is associated with lower top income inequality. It is therefore convenient to define the “power-law inequality” exponent as

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

so that

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

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

To see the fractal structure of top inequality under a Pareto distribution, let $\tilde{S}(a, b)$ denote the fraction of income going to the top b percent of people that actually goes to the top a percent. For example, $\tilde{S}(1, 10)$ answers the question, “Of the income going to the top 10 percent of earners, what fraction actually goes to the top 1 percent?” If income follows a Pareto distribution, then

$$\begin{aligned} \tilde{S}(a, b) &= \frac{\tilde{S}(a)}{\tilde{S}(b)} \\ &= \left(\frac{b}{a} \right)^{\eta-1} \end{aligned}$$

It is often useful — certainly given the data in the World Top Incomes Database — to look at shares where b is 10 times larger than a . In this case, let’s define $S(a) \equiv \tilde{S}(a, 10 \cdot a)$. For example, $S(1)$ is the fraction of income going to the top 10% that

actually accrues to the top 1%, and $S(.1)$ is the fraction of income going to the top 1% that actually goes to the top 1 in 1000 earners. Under a Pareto distribution,

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

and

$$\log_{10} S(a) = \eta - 1. \quad (5)$$

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.² Figure 4 shows the $S(a)$ shares directly. At the very top, the fractal prediction holds remarkably well, and $S(.01) \approx S(.1) \approx S(1)$. The share of all income going to the Top 10 percent, $S(10)$ is slightly higher than the others prior to 1980. But after 1980, even this share fits the fractal pattern quite well. 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.

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

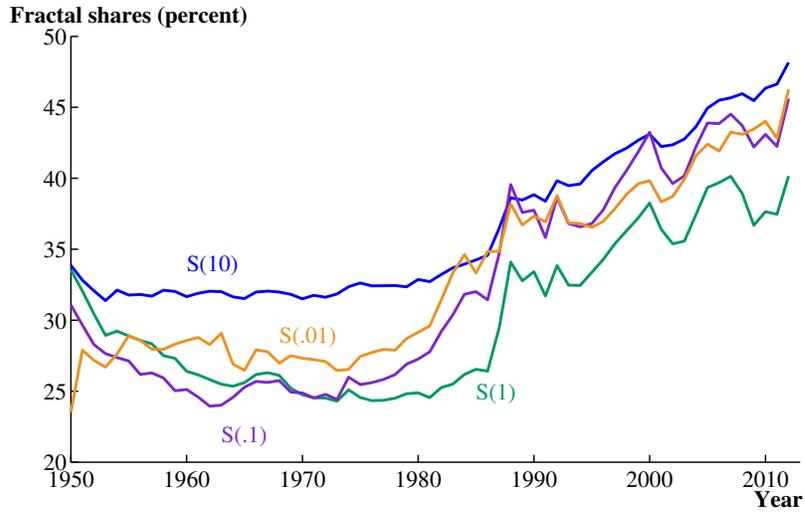
$$\Delta \log_{10} S(a) = \Delta \eta. \quad (6)$$

The corresponding Pareto inequality measures are shown in Figure 5.

Figure 6 breaks down the overall top 0.1% of earners by occupation using data

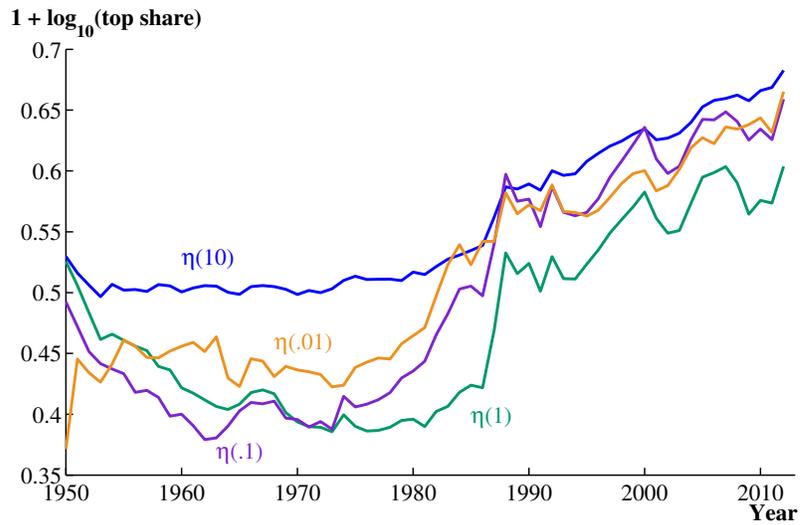
²Others have noticed this before. For example, see Aluation.wordpress.com (2011).

Figure 4: Fractal Inequality Shares in the United States



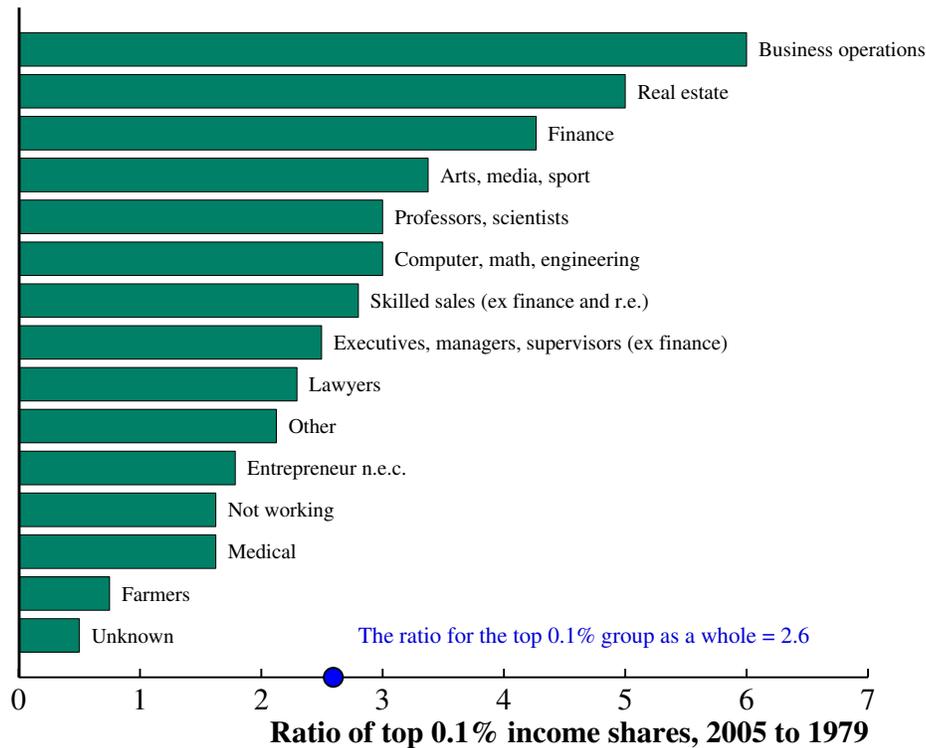
Note: $S(a)$ denotes the share of income going to the top .01 percent of earners as a share of that going to the top $10a$ percent. For example, $S(10)$ is the share of all income earned by the Top 10%, while $S(1)$ is the share of the Top 1% in the Top 10% income, etc. Source: World Top Incomes Database.

Figure 5: The Power-Law Inequality Exponent η , United States



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

Figure 6: The Rise in Top Income Inequality by Occupation



Note: The graph shows the ratio of the top 0.1% income share in 2005 to its share in 1979, by occupation. Source: Data are from Table 7 in Bakija, Cole and Heim (2010).

from Bakija, Cole and Heim (2010) and shows that the rise in top income inequality is not driven by any particular occupation, such as finance. Specifically, the figure shows the factor by which an occupation's share of national income rose between 1979 and 2005. For example, taxpayers working the "Arts, media, and sports" within the top 0.1% of earners saw their share of national income increase from 0.08 percent to 0.27 percent, a factor of 3.4. Executives, managers, and supervisors (non-finance) — the largest category among the top 0.1% — saw their share rise from 1.37 percent to 3.42 percent, a factor of 2.5. This is approximately equal to the 2.6-fold increase in the share received by the top 0.1% of earners overall.

2.2. Summary

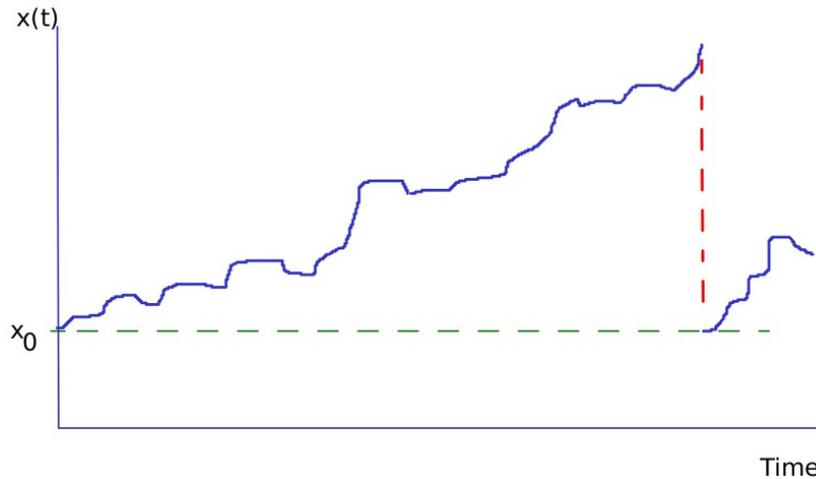
Here then are the basic facts related to top income inequality that we'd like to be able to explain:

1. Between 1960 and 1980, top income inequality was relatively low and stable in both the United States and France.
2. Since around 1980, top income inequality has increased sharply in countries like the United States, Norway, and Portugal, while it has increased only slightly in some other countries like France, Japan, and Spain.
3. Top income inequality in the United States follows a fractal pattern, where the share of the top X percent of income going to the top $X/10$ percent of earners is similar for different values of X . This share is a simple function of the power-law inequality exponent.
4. Changing top income inequality corresponds to a change in the power-law inequality exponent.
5. According to Piketty and Saez (2003), the rise in inequality since 1980 is primarily associated with labor income, not capital income.

3. A Model of Top Income Inequality

It is now well-known that exponential growth and Pareto distributions are tightly linked, and this link is at the heart of the basic mechanism in this paper. To see why, consider Figure 7. Suppose x_{it} corresponds to the idiosyncratic productivity of an entrepreneur's production function, and suppose this productivity grows exponentially subject to some random shocks. In addition, there is a constant probability of creative destruction — that the entrepreneur's run as a leader comes to an end and s/he is replaced by a new entrepreneur at the starting value x_0 . With a continuum of entrepreneurs obeying a similar process, this setup will converge to a stationary Pareto distribution for x_i . The Pareto inequality parameter will be an increasing function of the growth rate of the process and a decreasing function of the rate of

Figure 7: Basic Mechanism: Random growth and creative destruction



Note: A random growth process, subject to some lower bound x_0 and occasional “creative destruction,” can lead to a stationary Pareto distribution of x across a continuum of such entrepreneurs.

creative destruction. This 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. The remainder of this section develops this setup more formally.

3.1. 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 market share/quality/productivity, denoted by x . The entrepreneur then expends effort to improve x . We explain later

how x affects firm productivity and profitability. For the moment, however, it is sufficient to just assume that the entrepreneur's income is proportional to x , as it will be in general equilibrium.

The entrepreneur's problem is

$$\max_{\{c_t, e_t\}} \mathbb{E} \int_0^\infty e^{-(\rho+\delta+\bar{\delta})t} [\log c_t + \beta \log \ell_t] dt \quad (7)$$

subject to

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

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

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

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

The entrepreneur gets utility from consumption and leisure. 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 market share variable 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.³

Equation (10) describes how effort improves idiosyncratic 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.

The entrepreneur discounts future utility for three reasons. First, there is the standard rate of time preference ρ . Second, 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

³Because consumption enters in log form, 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. See Kim (2013).

δ . Because our model will turn out to be stationary, we omit the additional notation required to permit δ to be time varying. Finally, there is an exogenous piece to destruction as well, which occurs at a constant rate $\bar{\delta}$.

3.2. Bellman approach

The Bellman equation for the entrepreneur is

$$\rho V(x) = \max_e \log \psi + \log x + \beta \log(\Omega - e) + \frac{\mathbb{E}[dV(x)]}{dt} + (\delta + \bar{\delta})(V^w - V(x)) \quad (12)$$

subject to (10), where $\Omega \equiv 1 - \tau$ and $\frac{\mathbb{E}[dV(x)]}{dt}$ is short-hand for the Ito calculus terms, i.e. $\frac{\mathbb{E}[dV(x)]}{dt} = \mu(e)xV'(x) + \frac{1}{2}\sigma^2x^2V''(x)$. $V(x)$ 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 .

zzz Bellman equation missing a g/ρ term associated with growth of ψ_t

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 (12) and is given by*

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

This proposition implies that entrepreneurial effort is an increasing function of the technology parameter ϕ but decreases whenever τ , β , ρ , δ , or $\bar{\delta}$ are higher.

3.3. The Stationary Distribution of Entrepreneurial Income

We assume there are a continuum of entrepreneurs of unit measure at any point in time, corresponding to the usual unit measure of product varieties. The ini-

tial distribution of idiosyncratic productivity x is given by $f_0(x)$, and the distribution evolves according to the geometric Brownian motion process given above. Entrepreneurs can be displaced in one of two ways. Endogenous creative destruction (the Poisson process at rate 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 entrepreneur and gives 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.⁴ If a variable follows a Brownian motion, like x above, the density of the distribution $f(x, t)$ satisfies a Kolmogorov forward equation. In particular, the density satisfies

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

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

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

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

⁴For more detailed discussion, see Reed (2001), Mitzenmacher (2004), Gabaix (2009), and Luttmer (2010). Malevergne, Saichev and Sornette (2010) is closest to the present setup.

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

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

where

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

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

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

Comparative statics: Taking δ^* as exogenous for the moment, the comparative static results are as follows: power-law inequality, η^* , increases if effort is more effective at growing market share (a higher ϕ), decreases if the time endowment is reduced by government policy (a higher τ and hence lower Ω), 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 $\bar{\delta}$).⁵

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

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.

⁵The effect of σ^2 on power-law inequality is more subtle. If $\eta^* < \mu^*/\delta^*$, then a rise in σ^2 increase η^* , while if $\eta^* > \mu^*/\delta^*$, then the effect goes in the other direction. zzz clean up

3.4. 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:

$$Y = \left(\int_0^1 Y_i^\theta di \right)^{1/\theta}. \quad (18)$$

Each variety is produced by an entrepreneur using a production function that exhibits constant returns to the only nonrivalrous input, labor L_i :

$$Y_i = \gamma^{n_t} x_i^\alpha L_i. \quad (19)$$

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

The second term is the key place where an entrepreneur's effort decisions pay off: 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 key resource constraint in this environment involves labor:

$$L_t + R_t + 1 = \bar{N}, \quad L_t \equiv \int_0^1 L_{it} di \quad (20)$$

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 increases the productivity of a randomly chosen variety by a proportion $\gamma > 1$. The researcher is awarded a patent to be the exclusive producer of that variety. In addition, as explained above, the new idea generates spillovers that raise productivity in all other varieties as well.

The flow rate of innovation is therefore

$$\dot{n}_t = \lambda R_t \quad (21)$$

and this also gives the rate of creative destruction:

$$\delta_t = \dot{n}_t. \quad (22)$$

3.5. The Allocation of Resources

There are 12 key quantities in this economic environment: $Y, Y_i, x_i, L_i, L, R, n, \delta, e_i, c_i, \ell_i, \psi$. The entrepreneur's choice problem laid out earlier pins down c, ℓ , and e for each entrepreneur. Production functions determine Y, Y_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, the following conditions hold:

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 $\alpha = (1 - \theta)/\theta$.⁶ The equilibrium with monopolistic competition leads to*

$$Y_t = \gamma^{n_t} X^\alpha L \quad (23)$$

$$w_t = \theta \gamma^{n_t} X^\alpha \quad (24)$$

$$\pi_{it} = (1 - \theta) \gamma^{n_t} X^\alpha \left(\frac{x_i}{X} \right) \quad (25)$$

where $X \equiv \int_0^1 x_i di$ 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 .

We can now determine the value of ψ_t , the parameter that relates entrepreneurial income to x . Entrepreneurs earn the profits from their variety, minus whatever costs have to be paid to the researcher who discovered the idea in the first place (i.e. a payment from the entrepreneur to the researcher for the patent). We specify this payment as a rental contract where the entrepreneur pays the fraction p of profits in every period.⁷ The entrepreneur's income is then given by $\pi_i(1 - p)$. In the entrepreneur's problem, we previously stated that the entrepreneur's income is $\psi_t x_i$, so these two equations define ψ_t as

$$\psi_t = (1 - p)(1 - \theta) \gamma^{n_t} X^{\alpha-1}. \quad (26)$$

⁶This is merely a simplifying assumption that makes profits a linear function of x_i . It can be relaxed with a bit more algebra.

⁷The assumption that researchers are an exogenous and constant share of the labor force implicitly means that p is forced to adjust whenever we consider a comparative static exercise. When we endogenize the research allocation below, p will be the key price that is chosen so that researchers and raw labor earn the same wage.

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 (23) implies that growth in output per person is $\dot{n}_t \log \gamma = \lambda \bar{s} \bar{L} \log \gamma$ 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.

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*

$$g_y^* = \lambda \bar{s} \bar{L} \log \gamma \quad (27)$$

$$\delta^* = \lambda \bar{s} \bar{L}. \quad (28)$$

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

As a reminder, from Proposition 2, Power-law inequality is given by

$$\eta^* \equiv 1/\xi^*, \text{ where } \xi^* = -\frac{\tilde{\mu}^*}{\sigma^2} + \sqrt{\left(\frac{\tilde{\mu}^*}{\sigma^2}\right)^2 + \frac{2\bar{\delta}}{\sigma^2}} \quad (29)$$

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

3.6. Growth and Inequality: Comparative Statics

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

First, a technological change that increases ϕ will increase top income inequality in the long run. This corresponds to anything that increases the effectiveness of entrepreneur's 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 internet allows successful entrepreneurs to growth their markets 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 3.

Interestingly, this technological change has no effect on the long-run growth rate of the economy, at least as long as \bar{s} is held fixed. The reason is instructive about how the model works. In the long run, there is a stationary distribution of 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 effects. Long-run growth only comes about through the arrival of new ideas, not through the productivity growth associated with developing an existing idea. (And this is true despite the presence of a linear differential equation for x !)

The parameters $\Omega \equiv 1 - \tau$ 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 differed with respect to β . An economy where preferences are such that entrepreneurs put more weight on leisure will spend less time building businesses and feature lower top income inequality in the long run.

The two key parameters in the model that affect both growth and top income inequality are λ and \bar{s} , and they work the same way. If researchers are more productive at finding new ideas (a higher λ) or if simply a larger fraction of the labor works in research, 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.

?? $\bar{\delta}$ – discuss as a govt policy distortion?

?? Firm size distribution?

Figure 8 shows numerical results when \bar{s} exogenously pins down the allocation of labor to research.

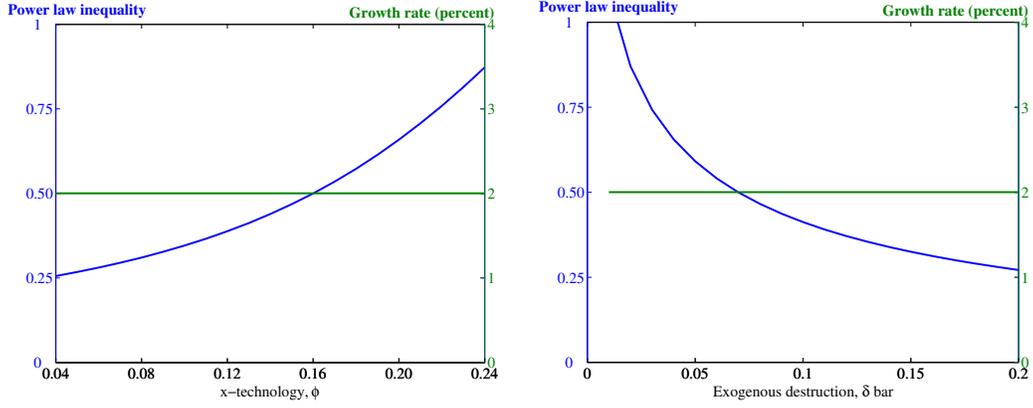
4. Endogenizing R&D

We endogenize the allocation of labor to research (s) in a standard fashion...

Endogenizing $s = R/\bar{L}$

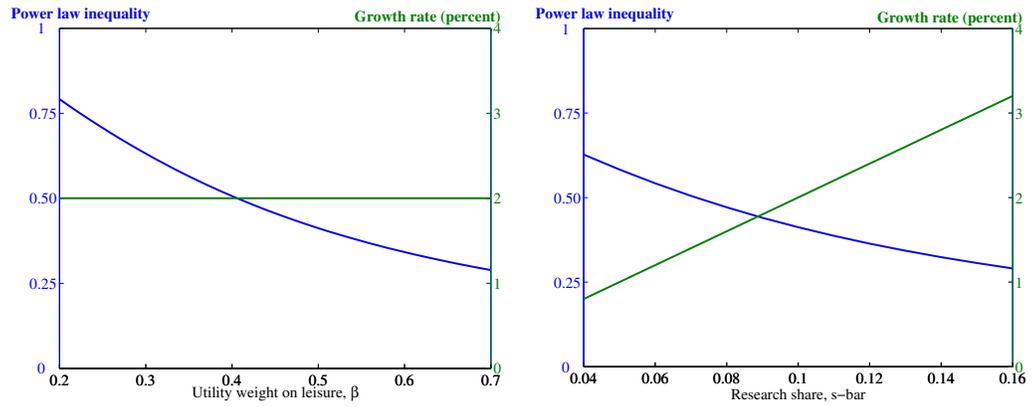
- Researcher can sell a new idea at price P_t
 - Assume x not observed until after idea used
 - So P does not depend on x .
 - Entrepreneurs pay a constant fraction of their profits, p , to rent an idea
 - People are indifferent ex ante to being researcher, worker, or entrepreneur
 - Worker and researcher: $w_t = \lambda P_t(1 - \tau_R)$
 - Worker and entrepreneur: $\mathbb{E}[V(x, p)] = V^w$
-

Figure 8: Numerical Examples: Exogenous \bar{s}



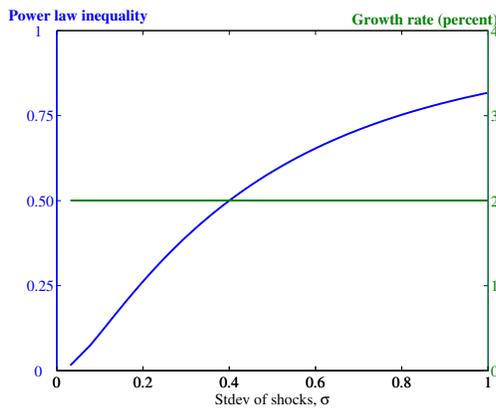
(a) Varying ϕ

(b) Varying $\bar{\delta}$



(c) Varying β

(d) Varying \bar{s}



(e) Varying σ

- Price of an idea ($P_t = \text{pdv of rental payments } \psi_t x_t p$)

$$P_t = \frac{\psi_t p}{r + \delta + \bar{\delta} - g - \mu} \cdot \frac{x_0}{1 - \eta}$$

- Value function for worker (no leisure)

$$V^w = \frac{1}{\rho} \left(\log w_t + \frac{g}{\rho} \right)$$

- Value function for entrepreneur

$$\begin{aligned} (\rho + \delta + \bar{\delta}) \mathbb{E}V(x, p) = & \log x_0 + \eta + \beta \log(\Omega - e^*) + \log \psi_t(p) \\ & + \frac{\tilde{\mu}}{\rho + \delta + \bar{\delta}} + (\delta + \bar{\delta}) V^w \end{aligned}$$

Equilibrium solution

Drift of log x $\tilde{\mu}^* = \phi \Omega - \beta(\rho + \delta^* + \bar{\delta}) - \frac{1}{2} \sigma^2$

Pareto inequality $\eta^* = 1/\xi^*, \quad \xi^* = -\frac{\tilde{\mu}^*}{\sigma^2} + \sqrt{\left(\frac{\tilde{\mu}^*}{\sigma^2}\right)^2 + \frac{2\bar{\delta}}{\sigma^2}}$

Research allocation $\frac{s^*}{1-s^*} = \frac{1-\theta}{\theta} \cdot \frac{\delta^* p^* (1-\tau_R)}{\rho + \delta^* + \bar{\delta} - \mu}$

Creative destruction $\delta^* = \lambda s^* \bar{L}$

Growth $g^* = \delta^* \log \gamma$

Rental price of idea $p^* = 1 - \frac{\theta}{1-\theta} \cdot \frac{\exp(-\eta)}{1-\eta} \cdot \frac{\exp\left\{\frac{g}{\rho} - \frac{\tilde{\mu}}{\rho + \delta^* + \bar{\delta}}\right\}}{(1-s^*) \bar{L} (\Omega - e^*)^\beta}$

Understanding Real World Inequality?

	1980	2007
US	.44	.65
France	.40	.47

- Pareto inequality is rising throughout the world
 - Technological change: $\uparrow \phi$ – World wide web, etc. make it easier for an entrepreneur to grow her market
 - Does rising top inequality always reflect positive changes?
 - No! $\uparrow \bar{s}$ raises growth but reduces inequality (more creative destruction).
-
-

Why does η for France and Japan rise less?

- Technological change needs to be offset by something
 - Rise in \bar{s} and creative destruction?
 - Rise in misallocation: $\uparrow \bar{\delta}$
 - Globalization: rise in δ^* because of opening up to competition from U.S., etc.
 - flat inequality could result from a “good” policy
 - Policies that encourage entrepreneurs to exert less effort
 - Rise in utility weight on leisure in France
 - Lost decade in Japan?
-

5. Numerical Examples

Figure 9 shows numerical results when $s \equiv R/\bar{L}$ is endogenously determined. Finally, Figure 10 shows numerical examples that illustrate different ways in which the model can match the time series behavior of top income inequality in the United States and France.

6. Notes

Firm size distribution issues...

Guvenen AER 2007 Are Labor Income Shocks Really Very Persistent? Also RED 2009 (suggests that a unit root process is not out of the question).

Future work: estimate the stochastic process for top incomes. Does it look like the random growth process with creative destruction? Have the parameters changed over time in the U.S.? Can we do this as well for a country like France or Japan for comparison? This would be a great way to verify the basic mechanism of the paper...

Evidence that hours worked by top earners has been rising over time in U.S., in contrast to median/others?

KFE intuition? Always converge? Pareto is unique solution??

“And so we come again to the gains of the top earners, clearly the big story told by the data. Its worth noting that over this same period of time, inequality of work hours increased too. The top earners worked a lot more and most other Americans worked somewhat less.” –T Cowen.

Meghir pistorferri. RED volume. BPEA Kaplan just recently (comment)

Within firms? What is happening to wage inequality within firms??

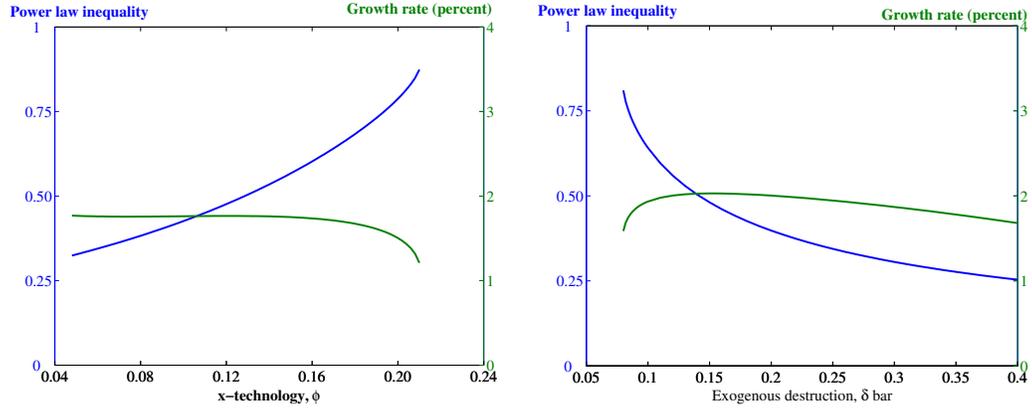
Experience matters. Human k. Age earnings. Flatter in France.

Mike Harrison

$V(x,t)$ – be careful with time dependence in the Prop 1

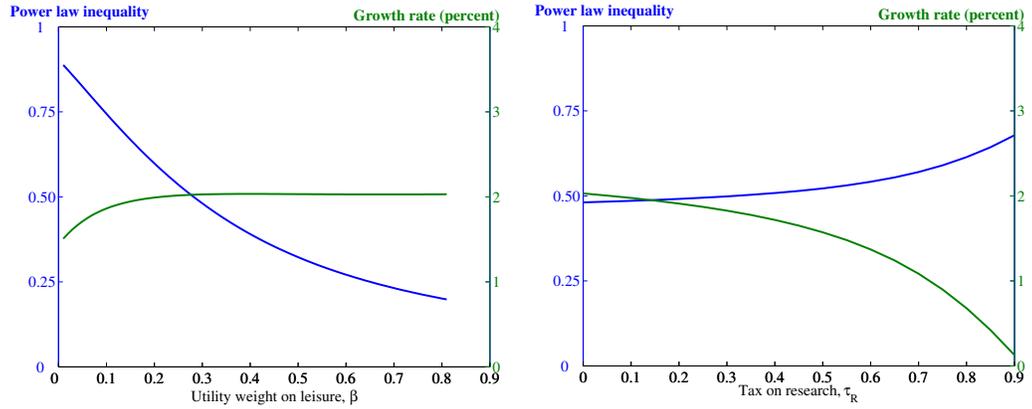
Ensemble arguments: how to go from stationary distn for a given entrepreneur to the distn across varieties for all entrepreneurs at a point in time?

Figure 9: Numerical Examples: Endogenous s



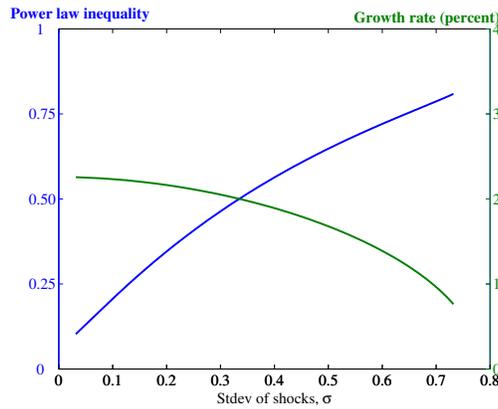
(a) Varying ϕ

(b) Varying $\bar{\delta}$



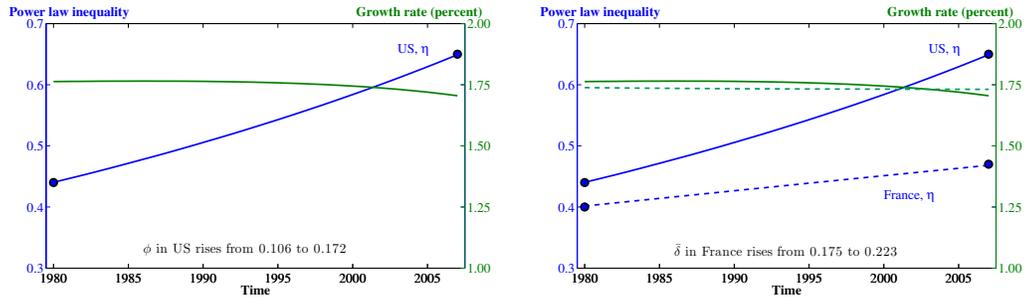
(c) Varying β

(d) Varying τ_R



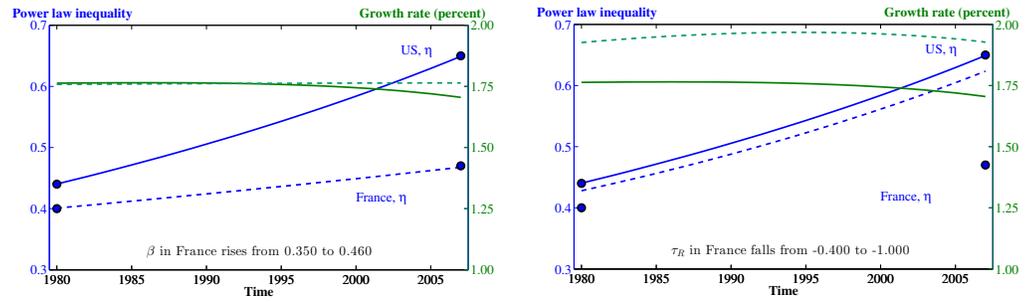
(e) Varying σ

Figure 10: Numerical Examples: Matching Inequality in U.S. and France



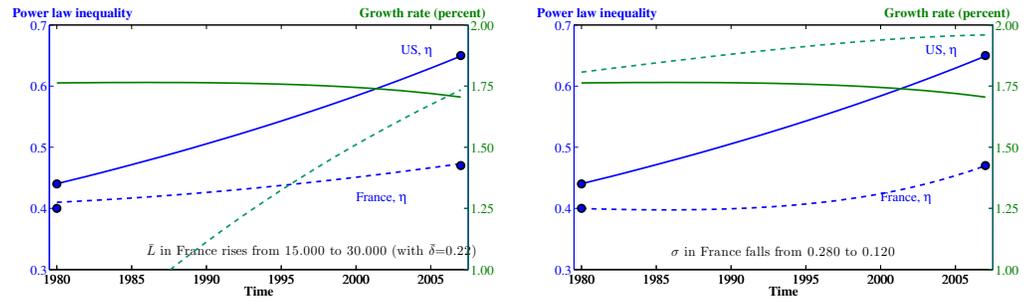
(a) Varying ϕ to match U.S. η

(b) Varying $\bar{\delta}$



(c) Varying β

(d) Varying τ_R



(e) Varying \bar{L}

(f) Varying σ

Proving convergence and uniqueness: Harrison in his book does this for a standard diffusion w/ drift, which is what I have for $\log x$. He prefers a probabilistic argument to the KFE because it connects directly with the primitives of the problem (maybe use Harrison for formal proof in appdx and keep KFE in text for analytic directness?).

Look in his book at section on reflected Brownian motion where he “divide future into regenerative blocks.” These blocks in my setting will be the exponential lengths of time that an entrepreneur is active (exponential because killing times are poisson). Renewal theory?

Harrison also mentioned liking the Preferential Attachment generator. Maybe this could be connected to Klette and Kortum? But of course I do not want firm size to be changing over time...

Using avgincome99 etc to make progress?

Scale effects

Is x human capital? Possibly. Not obvious that human capital depreciates rapidly when new entrepreneur replaces you? That seems more about ideas...

Within firm income inequality? e.g. Frydman?

A Appendix: Proofs of the Propositions

Proof of Proposition 1. Entrepreneurial Effort

See pp. 1-3 of notes. Needs to be typed in...

References

- Acemoglu, Daron and David Autor, “Skills, Tasks and Technologies: Implications for Employment and Earnings,” in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, June 2011, chapter 12, pp. 1043–1171.
- Aghion, Philippe and Peter Howitt, “A Model of Growth through Creative Destruction,” *Econometrica*, March 1992, 60 (2), 323–351.
- 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.
- 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 Multi-tasking,” Working Paper 18936, National Bureau of Economic Research April 2013.
- Benhabib, Jess, 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.
- 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.

- 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.
- 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.
- 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.
- Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, 65 (6), 1389–1419.
- Lucas, Robert E. and Benjamin Moll, "Knowledge Growth and the Allocation of Time," NBER Working Papers 17495, National Bureau of Economic Research, Inc October 2011.
- 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.
- Malevergne, Y., A. Saichev, and D. Sornette, "Zipf's law and maximum sustainable growth," Quantitative Finance Papers 1012.0199, arXiv.org December 2010.
- 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.
- 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*, forthcoming 2014, pp. xxx–xxx.

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 Stefanie Stantcheva, "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," Technical Report November 2011. NBER Working Paper 17616.

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.