Exploring the Dynamics of Top Income Inequality

Charles I. Jones
Stanford GSB and NBER

Jihee Kim *
Stanford University

November 16, 2012 – Version 0.2
Preliminary and Incomplete

Abstract

Top income inequality has risen sharply in the United States over the last 30 years but remained low and stable in economies like France and Japan. Why? This paper explores three theoretical mechanisms that endogenize the amount of top income inequality. The first model suggests that the rate of growth of top incomes is a key determinant. A rise in the returns to experience or an increase in effort devoted to accumulating human capital may increase top income inequality. The second model studies the direction of technological change and considers why it might be "talent biased," at least along a transition path. Finally, the third model considers the allocation of talent. If the matching process between firms and talent is distorted, top income inequality can decline. The rise in top income inequality in the United States, then, may reflect an improvement in the allocation of talent.

*We are grateful to seminar participants at the Federal Reserve Bank of San Francisco and Stanford 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, in some extreme cases like Japan and France, top income inequality has been relatively constant. Why? That is, what economic forces explain the rise in top income inequality in the United States? And why has top income 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. 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 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 exlaining the rise in top income inequality in the United States. Instead, we present three different economic mechanisms that could potentially contribute to such an explanation.

Before summarizing those mechanisms, however, 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).

Similarly, 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.

This paper explores three mechanisms that endogenize top income inequality. The first we describe as Katz and Murphy (1992) meet Gabaix (1999). In the context of cities and Zipf's Law, Gabaix shows how an exponential growth process hit by random shocks gives rise to a Pareto distribution. We apply this framework to income inequality to show that the Pareto measure of top income inequality is proportional to the growth rate of incomes. We then consider two examples. In the first, the growth rate of incomes is captured by an exogenous return to experience, and the Pareto inequality measure depends on the ratio of the returns to experience for college graduates to the growth rate of college graduates. These forces are reminiscent of the Katz and Murphy (1992) supply-and-demand explanation for inequality between college graduates and high school graduates, only here it applies to top income inequality.

The second example endogenizes the growth rate of individual income using a Lucas (1988) model of human capital accumulation. The growth rate is proportional to the amount of effort that people put into accumulating human capital. Kim (2012) suggests that this effort could in turn be a function of taxes, resurrecting the intriguing link between top income inequality and top marginal tax rates.

The second mechanism for understanding top income inequality that we consider focuses on the direction of technological change and is reminiscent of the Rosen (1981) superstar model. There are two kinds of ideas in the model: “neutral” ideas that raise the productivity of all workers and “talent-biased” ideas that disproportionately benefit people of higher talent. Crucially, this last statement is true for any two talent levels: the higher talent benefits more from talent-biased technical change than the lower talent. This is necessary to explain why inequality between the 99.5th percentile and the 99.9th percentile increases.

Following Acemoglu (1998), researchers can choose which kind of idea to invent. Along the transition path, both types of ideas are invented, and the presence of talent-biased technical change leads to rising top income inequality. Eventually, however, all technological change is neutral and the level of top income inequality settles down to a constant which governs the rate at which talent-biased technical change becomes more difficult over time.

Finally, the third mechanism we study relates top income inequality to the misallocation of talent. We consider a model in which firms and talented workers must match in order to produce. The misallocation of talent in this matching process reduces the efficiency of production and reduces top income inequality.

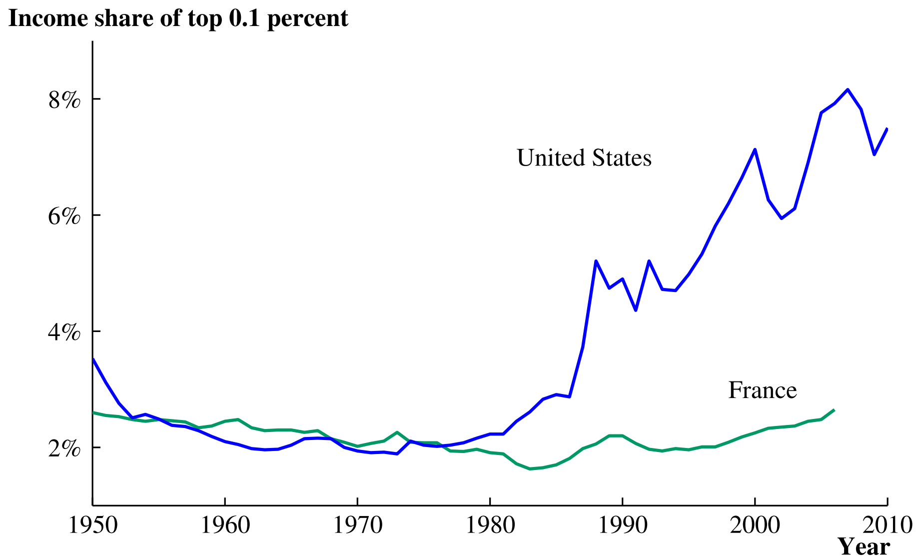
A problem with the simplest version of this model, in which talent itself obeys a Pareto distribution, is that it predicts that the inequality of top incomes should mirror the inequality in the size distribution of firms. After all, if the best talent is not matched with the most productive firm, that firm will hire fewer workers. While this is an interesting prediction, it is soundly rejected by the data, for two reasons. First, there is much more inequality in the firm size distribution than at the top of the income distribution: just think about the size of Walmart versus the size of the convenience store on the corner. Second, a very robust fact is that the firm size distribution in the United States has a Pareto exponent equal to one, and that this distribution is stable over time. Even though income inequality has risen sharply in the last 30 years, the size distribution of firms has remained stable.

Fortunately, studying misallocation and income inequality in the Gabaix and Landier (2008) framework addresses these problems. The previous mechanisms we’ve considered in this paper either generate (in the first mechanism) or assume (in the second) that talent obeys a Pareto distribution. In contrast, Gabaix and Landier argue that the distribution of talent is far from Pareto, and is in fact bounded. Studying misallocation and top inequality in their framework resolves the puzzle we just outlined. Because talent is bounded, the size distribution of firms is determined solely by the exogenous distribution of underlying productivity and is unaffected by the misallocation of talent. However, the misallocation of talent still affects top income inequality. This model therefore suggests a final possibility. Perhaps the rise in top inequality in the United States is associated with an improved allocation of talent over the last 30 years. Examples of changing allocations that are consistent with this explanation include the enormous rise of women in high skill occupations and the fact that top talent from around the world seems to be increasingly attracted to the U.S. labor market.

A number of other recent papers contribute to our understanding of the dynamics 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) style 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 (2008) 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. 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.

This paper proceeds as follows. Section 2 presents some basic facts of top income inequality, emphasizing that the rise in top inequality in the United States is accurately characterized as a change in the power law parameter. The remaining sections then develop our different economic mechanisms for understanding the

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



Source: World Top Incomes Database.

dynamics of top income inequality.

2. Some Basic Facts

2.1. Fractal Inequality and the Pareto Distribution

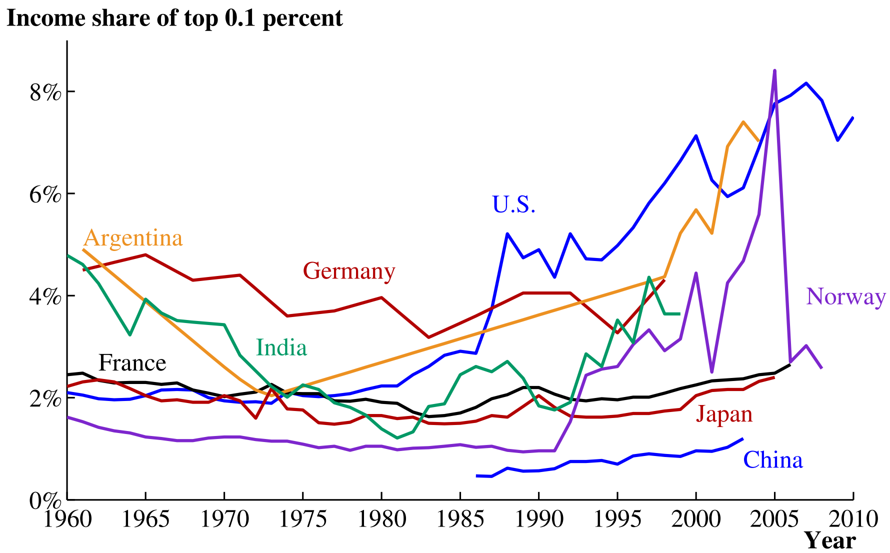
It is well known, dating back to Pareto (1896), that the top portion of the income distribution is accurately characterized by a power law. That is, for high levels of income, income follows a Pareto distribution; for a recent reference, see Saez (2001). If Y is a random variable denoting income, then for $Y \geq y_0$,

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

where ξ is called the "power law exponent."

An important property of the Pareto distribution can then be seen easily. Let

Figure 2: Top Income Inequality around the World



Source: World Top Incomes Database.

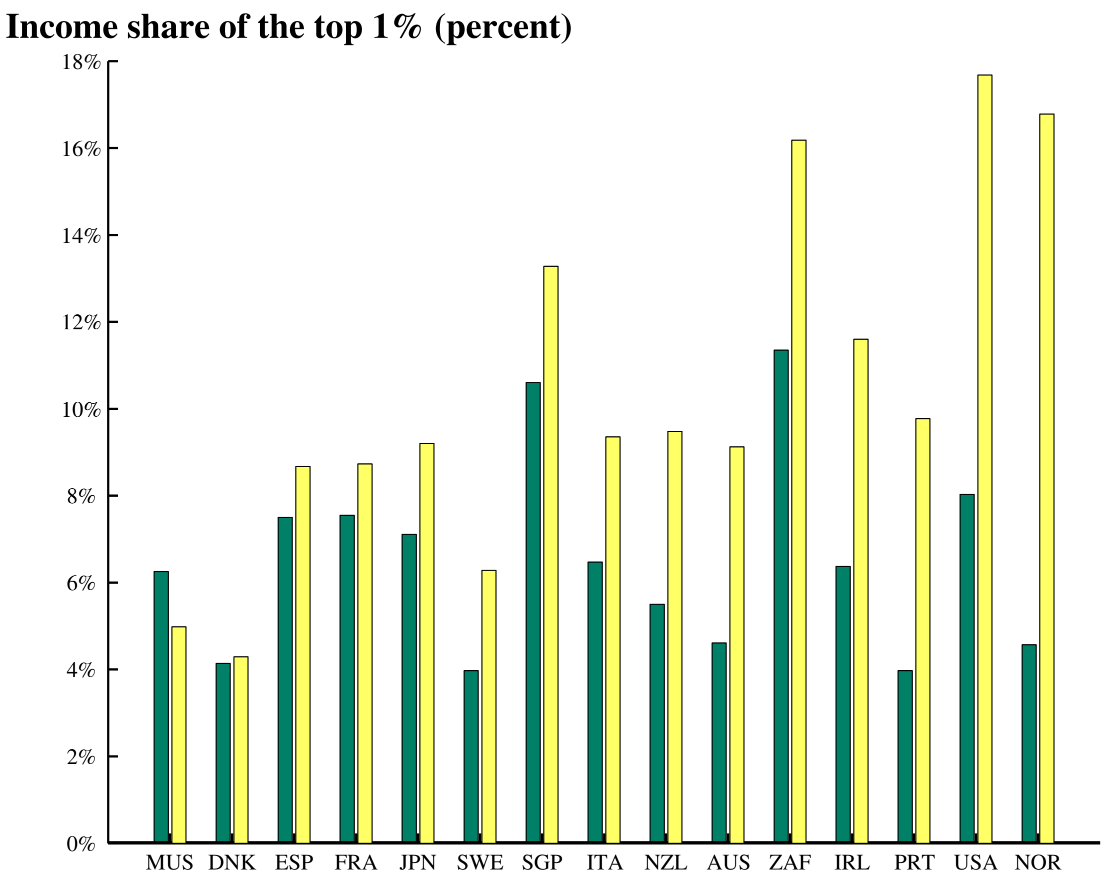
$\tilde{S}(a,b)$ denote the fraction of income going to the top b percent of people that actually goes to the top a percent. For example, $\tilde{S}(1,10)$ answers the question, "Of the income going to the top 10 percent of earners, what fraction actually goes to the top 1 percent?" If income follows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

$$\tilde{S}(a,b) = \left(\frac{a}{b}\right)^{\frac{\xi-1}{\xi}}. \tag{2}$$

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,

$$S(a) = 10^{\frac{1}{\xi}-1}. \tag{3}$$

Figure 3: Top Income Inequality around the World, 1981 and 2005



Note: The bars show the income share of the top 1% of earners in 1981 and 2005, sorted according to the change in inequality. Source: World Top Incomes Database.

From this equation, it is clear that 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

$$\eta \equiv \frac{1}{\xi} \tag{4}$$

so that

$$S(a) = 10^{\eta-1} \tag{5}$$

and

$$\boxed{\log_{10} S(a) = \eta - 1.} \tag{6}$$

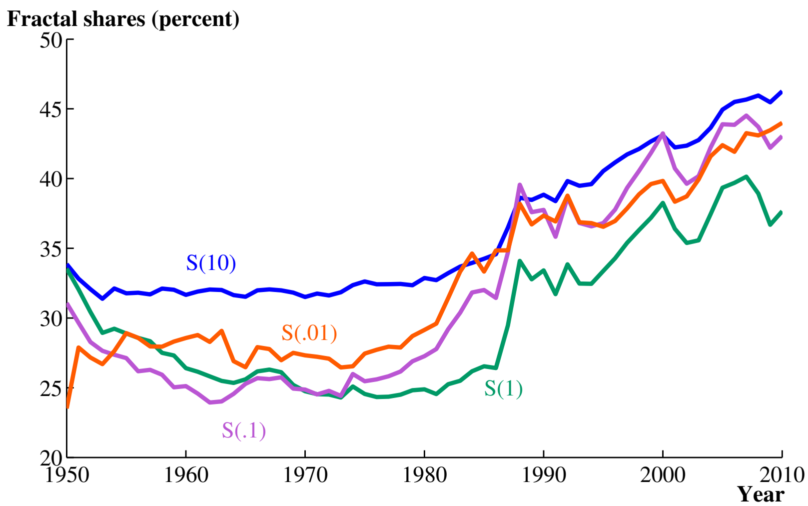
Notice that this last result holds for all values of a, or at least for all values for which income follows a Pareto distribution. This means that top income inequality obeys a *fractal* pattern: the fraction of the Top 10 percent's income going to the Top 1 percent is the same as the fraction of the 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.[1] 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 (6). Or, put another way, the *change* in the fractal shares is precisely equal to the change in the PL inequality

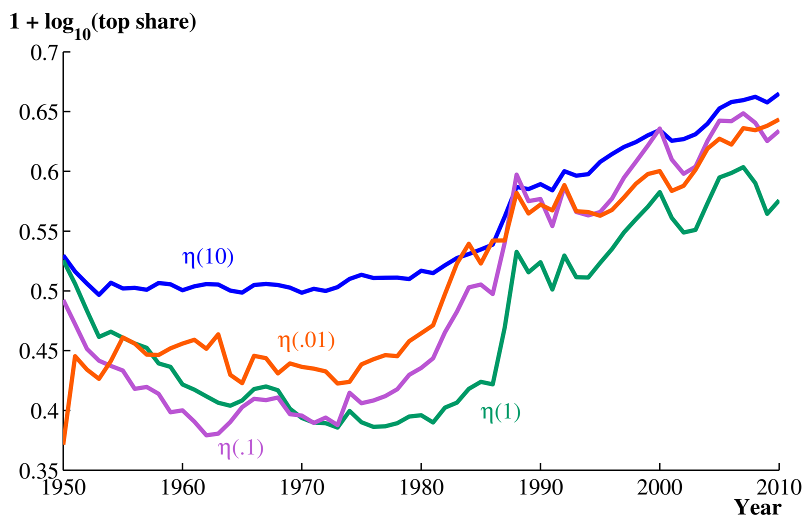
[1]Others have noticed this before. For example, see Aluation.wordpress.com (2011).

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 (6) in the text.

exponent:

$$\Delta \log_{10} S(a) = \Delta\eta. \tag{7}$$

The corresponding Pareto inequality measures are shown in Figure 5.

2.2. The Simple Math of Power-Law Inequality

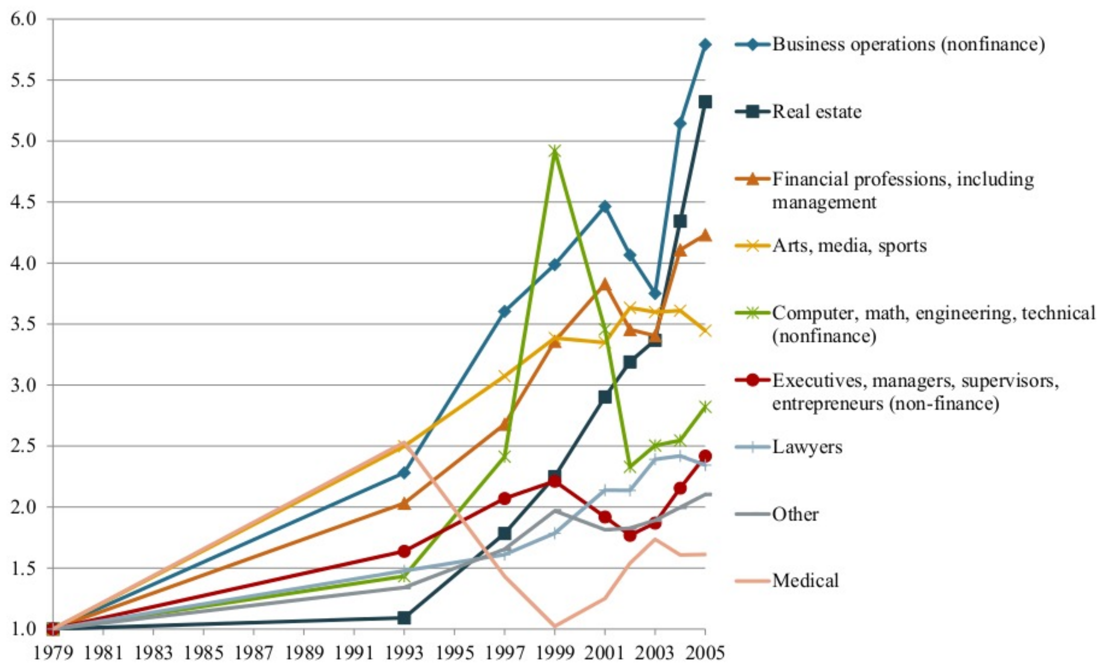
Before proceeding to the models, there is some simple math of power-law inequality that is worth pointing out, as it makes it much easier to understand the theoretical results. These rules are simply the inequality-exponent versions of the rules for power law exponents given in Gabaix (2009, p. 259).

Suppose $x_1, \ldots, x_N$ are N independent random variables with power-law inequality exponents given by η_i, and let α be some positive constant. The following statements hold:

1. If $y = x^\alpha$, then $\eta_y = \alpha\eta_x$.
2. If $y = \sum_i x_i$, then $\eta_y = \max_i\{\eta_i\}$.

Figure 6: Heim (2012) Graph by Occupation

Figure 10 -- Percentage of national income (excluding capital gains) going to top 0.1 percent by occupation, relative to 1979



Note: Top income inequality has increased in virtually all occupations. It is not driven solely by finance, for example.

3. If $y = \Pi_i x_i$, then $\eta_y = \max_i\{\eta_i\}$.
4. If $y = \alpha x$, then $\eta_y = \eta_x$.
5. If $y = \max_i\{x_i\}$, then $\eta_y = \max_i\{\eta_i\}$.

In particular, if Y and X are two power-law random variables with $\eta_Y \geq \eta_X$, then $X + Y$, $X \cdot Y$, and $\max\{X, Y\}$ preserve the power-law characteristic, inheriting the highest level of inequality, η_Y.

2.3. 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 stayed relatively low in 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 First Model: Katz and Murphy Meet Gabaix

A classic model of wage inequality is Katz and Murphy (1992), which provides a supply-and-demand explanation of the college wage premium. A classic model of power laws is Gabaix (1999), which provides an explanation for Zipf's Law for city sizes driven by random growth. Our first model of top income inequality builds on Gabaix's insights to yield an explanation for the dynamics of top income inequality that is reminiscent of the logic in Katz and Murphy.

To begin, we consider the simplest possible model that leads to a power law in the income distribution.[2] While unrealistic in many ways, this model illustrates very clearly how exponential growth can lead to a Pareto distribution for incomes. We will then enrich the simple framework in several ways to make it more informative.

Suppose the flow of new college graduates in the economy is deterministic and growing at rate n. Time is continuous, and new college graduates earn income equal to y_0. As college graduates age, their income rises with experience at rate g, so the income of a college graduate with experience (or more accurately, age since graduation) a is $y(a) = y_0 e^{ga}$. One could also allow general technological progress to raise the income of all college graduates over time at some different rate — that is, y_0 may be rising over time. But since this raises the income of all college graduates in the same way, it will not affect top income inequality among college graduates, which is the focus here.

What fraction of people in this economy have income greater than some level y? This is simply equal to the fraction of people with at least $a(y)$ years of experience, where

$$a(y) = \frac{1}{g} \log\left(\frac{y}{y_0}\right). \tag{8}$$

Given that new college graduates grow at rate n, experience in this economy follows an exponential distribution. That is,

$$\begin{aligned} \text{Pr}\left[\text{Income} > y\right] &= \text{Pr}\left[\text{Age} > a(y)\right] \\ &= e^{-na(y)} \\ &= \left(\frac{y}{y_0}\right)^{-\frac{n}{g}}. \end{aligned} \tag{9}$$

The power law exponent for income in this model is then $\xi_y = n/g$, and power

[2]Gabaix (2009) presents a version of the model developed below in the context of Zipf's Law for cities and attributes the original development of this model for that application to Steindl (1965).

law inequality is given by

$$\eta_y = \frac{g}{n}. \tag{10}$$

In this model, top income inequality can change for two reasons. First, an increase in the return to experience for college graduates will raise top income inequality. The higher is the growth rate of incomes with age, the higher is the ratio of top incomes to the incomes of a recent college graduate. Second, a decline in the arrival rate of college graduates n will raise top income inequality, as the distribution is more heavily weighted toward older and therefore richer individuals.

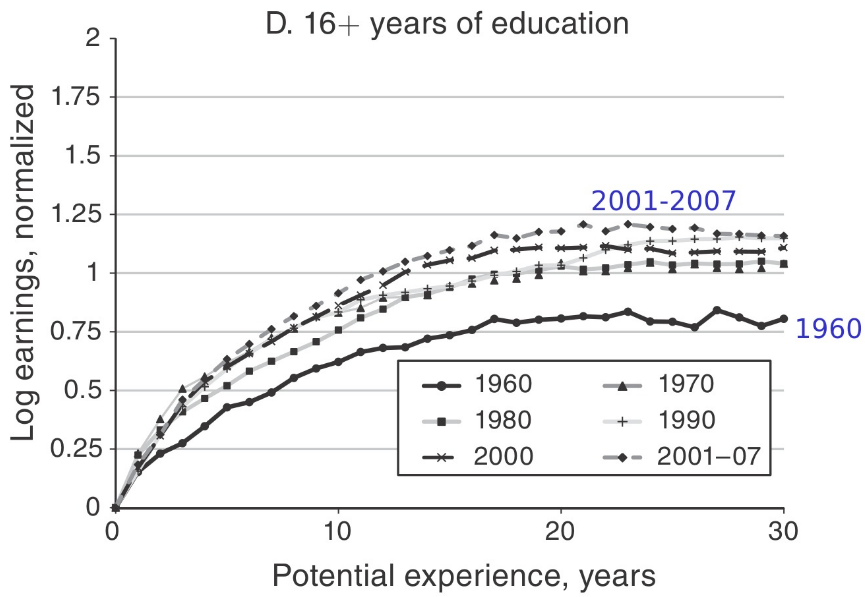
These two forces are somewhat reminiscent of the demand and supply forces in Katz and Murphy (1992). The rise in the return to experience functions like an increase in the demand for skilled labor, while a rise in the arrival rate of college graduates functions like an increase in the supply. On the other hand, the economic forces behind these effects in the Gabaix-like model are distinctly different.

How can this model help us understand the dynamics of top income inequality? One possibility is the following. Perhaps the returns to experience, g, are rising everywhere. Elsby and Shapiro (2012) provide evidence that the returns to experience have risen in the United States since 1960; see Figure 7. Differences in the dynamics of inequality could then be driven by differences in the growth rate of new college graduates. In the United States, college enrollment rates have stabilized during the last 30 years. In contrast, enrollment rates have been increasing in France during this same time period; see Figure 8. This rise in supply could offset the rising returns to experience and keep top inequality flat in France and Japan.

3.1. Extension: Lucas (1988)

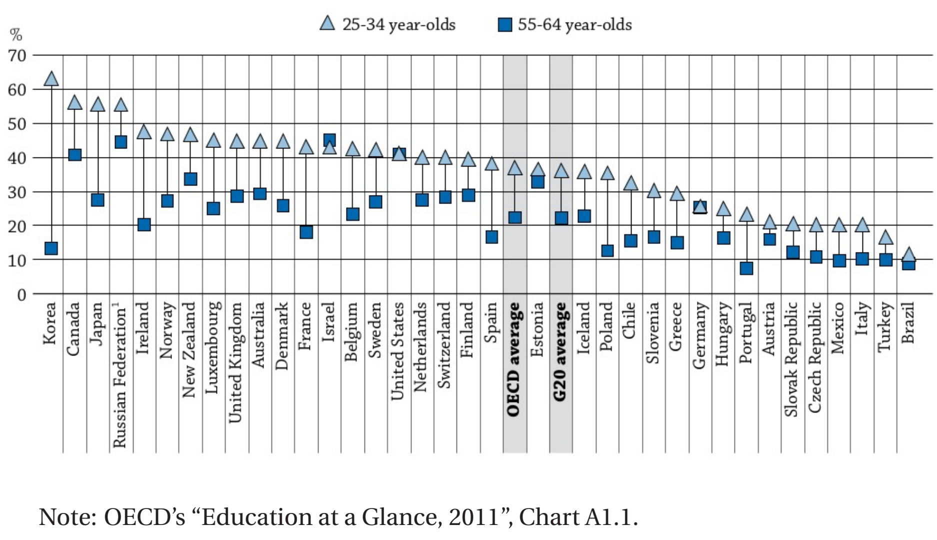
The simple model just provided is useful for several reasons. First, it explains how the Pareto income distribution itself might arise. Second, it identifies some key economic variables, such as the growth rate of an individual's income over time, that determine top income inequality. Of course, the growth rate of an individual's income is arguably an endogenous variable, so one might want to think about how this growth itself is determined.

Figure 7: Returns to Experience?



Source: Elsby and Shapiro (2012 AER).

Figure 8: Percent of Population with a College Education, 2009



Note: OECD's "Education at a Glance, 2011", Chart A1.1.

A very natural model to consider in this context is the Lucas (1988) model of human capital.[3] Suppose income is proportional to human capital and suppose an individual accumulates human capital over time according to

$$\dot{h}_t = \lambda e h_t \tag{11}$$

where e denotes effort the individual expends to accumulate human capital and λ is a productivity parameter. In this case,

$$g = g_h = \lambda e \tag{12}$$

and therefore top income inequality is

$$\eta_y = \frac{\lambda e}{n}. \tag{13}$$

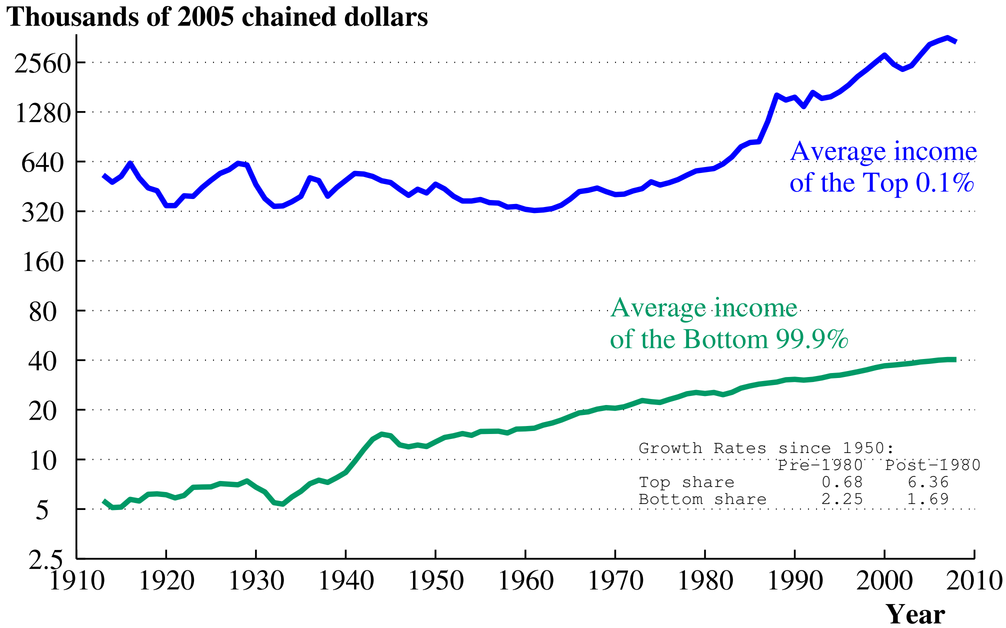
Changes in the effort that people expend accumuating human capital can then change top income inequality.

This gives rise to an interesting hypothesis, explored in detail by Kim (2012). In particular, consider top earners and the effort they put forward accumulating human capital on the job. (Here we think of this human capital arising simply through experience on the job — learning by working; for high skilled jobs, this seems very reasonable.) The top marginal tax rate in the United States has fallen sharply since the 1960s, from more than 90% to around 35% in the 2000s. This strong decline may have increased human capital accumulation and raised top inequality through this channel. Moreover, the top marginal tax rate in France, for example, declined by relatively little over the last 30 years, potentially explaining why top inequality was more stable there.

One potential problem with this story is that it seems to suggest that U.S. growth rates should have risen over this same period. The growth rate of GDP per worker has been relatively stable. However, perhaps this is looking too broadly. Figure 9

[3]Lucas (2009) and Robert E. Lucas and Moll (2011) provide richer models along these lines and note that they give rise to Pareto inequality.

Figure 9: U.S. Income Growth for the Top 0.1% and Bottom 99.9%



Note: Constructed using the Piketty and Saez inequality numbers applied to the Maddison and BEA GDP numbers. A similar point can be made with the actual average income numbers in Piketty and Saez, but the average growth rates are lower, e.g. because of the exclusion of benefits from their income numbers.

shows average income going to the Top 0.1% versus the Bottom 99.9% since 1913. While overall GDP per worker grows at a stable rate over this period, the growth rates for these different income groups are very different. In particular, there was a sharp increase in the growth rate of incomes for the Top 0.1% after 1980. Between 1950 and 1980, this group's average income grew at a paltry 0.68 percent per year. After 1980, though, the growth rate shot up to a China-like 6.36 percent per year. Therefore, an explanation along these lines might be successful if it also explains why effort devoting to accumulating human capital has not risen (and may have even declined) outside the top of the income distribution.

3.2. Extending the model: random growth with births and deaths

The very simple model we have just described has an exact connection between age and income, which is obviously unrealistic. Allowing for randomness in the growth process severs this precise link and introduces another parameter that could change Pareto inequality. The model we develop below is a version of the random growth model developed by Malevergne, Saichev and Sornette (2010); they applied the model to firm size and we apply it to top inequality.

Suppose the flow of new college graduates in the economy at time t is a Poisson process with the arrival rate $N_0 e^{nt}$. Therefore, on average, the number of new college graduates is growing at rate n. New college graduates who enter the labor market at time t start with an income equal to $y_t(0)$, where $y_t(0)$ is an iid random draw from a common random variable $y(0)$ such that $\mathbf{E}[y(0)^{\frac{1}{\eta_y}}] < \infty$ with a lower bound $\underline{y}(0)$.

As college graduates age, their incomes evolve as $\dfrac{dy_t(\tau)}{y_t(\tau)} = gd\tau + \sigma dz_\tau$ where z_τ is a standard Brownian motion. If the income goes below $y_{\text{min}}(< \underline{y}(0))$, people drop out of the labor market. Finally, there is also an exogenous constant rate at which people drop out of the labor market, given by $\delta > 0$.

Note that the total working population N_t satisfies

$$
\begin{aligned}
& \frac{dN_t}{dt} = N_0 e^{nt} - \delta N_t \\
\Rightarrow \quad & N_t = \frac{N_0}{n+\delta}[e^{nt} - e^{-\delta t}] \\
\Rightarrow \quad & \text{(inflow rate)} = \frac{N_0 e^{nt}}{N_t} = \frac{n+\delta}{1-e^{-(n+\delta)t}} \to n+\delta \text{ as } t \to \infty
\end{aligned}
$$

Applying a result from Malevergne, Saichev and Sornette (2010), we obtain[4]

$$
\eta_y = \frac{2}{1 - \frac{2g}{\sigma^2} + \sqrt{\left(1 - \frac{2g}{\sigma^2}\right)^2 + \frac{8(n+\delta)}{\sigma^2}}}
$$

Notice that top inequality is increasing in g and decreasing in n as before. It

[4] Our setup here corresponds to the case where $c_0 = c_1 = 0, d > 0$ in their paper.

also now decreases in δ. Finally, top inequality also depends on the variance of the shocks to the income process. In particular, if $gn + \delta$, then η_y is decreases in σ^2.

4. A Second Model: Talent-Biased Technical Change

This model combines the insights of Katz and Murphy (1992) on skill-biased technical change with the Rosen (1981) superstar models to gain insight into top income inequality. Katz and Murphy (1992) noted that a bias in technological change toward skilled labor as opposed to unskilled labor could increase inequality. Here, the issue is more delicate. There are potentially a large number of different talent levels, and we need the technological change to disproportionately benefit the more talented relative to the less talented at every level.

The key insight can be seen by considering the following production function, reminiscent of the Lucas (1978) span of control model:

$$Y(q) = Aq^n L(q)^\alpha \tag{14}$$

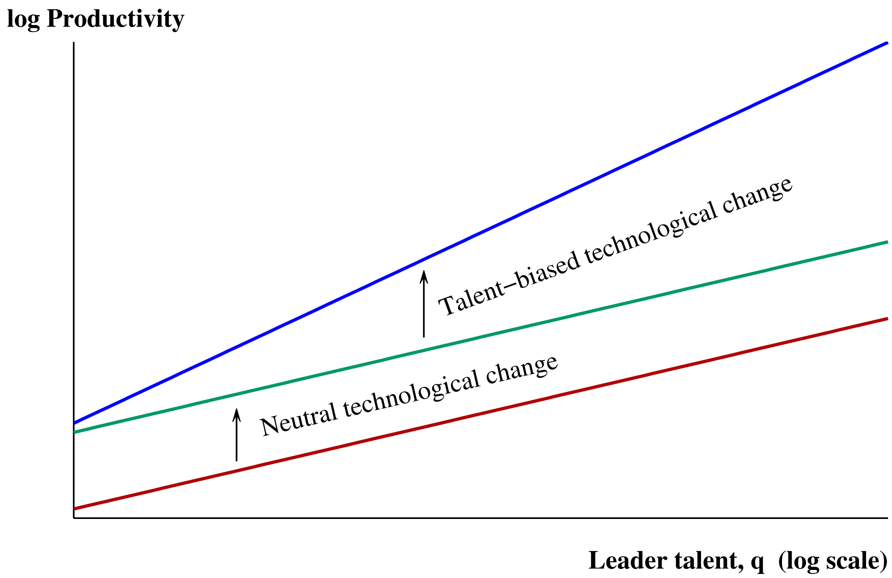
where $q > 1$ is the talent of the "leader" in the firm, A is *neutral* technological change which is the same for all firms, and $L(q)$ is the number of unskilled workers at the firm. The novelty here is the exponent on q: the endogenous variable n denotes the level of *talent-biased technological change.*

Figure 10 illustrates these two kinds of technology by plotting the log of productivity, Aq^n, against the talent of the leaders. Neutral technological change raises productivity for every firm, while talent-biased technical change (TBTC) rotates productivity, disproportionately benefitting leaders who are more talented.

Assume there are a continuum of leaders in the economy with talent distribution $F(q)$. There are $\bar{L}$ unskilled workers in the economy. To keep things simple, suppose the output from different firms aggregates up via a CES technology $Y = (\int Y(q)^\rho dF(q))^{1/\rho}$. In this case, the optimal allocation of unskilled workers will satisfy

$$\max_{L(q)} \left(\int Y(q)^\rho dF(q) \right)^{1/\rho} \text{ s.t. } \int L(q) dF(q) = \bar{L} \tag{15}$$

Figure 10: Talent-Biased Technological Change



which leads to $L(q) \propto Y(q)$. Substituting this into the production function for firm q gives

$$Y(q) = \zeta_1 (Aq^n)^{\frac{1}{1-\alpha\rho}} \equiv \zeta_1 (Aq^n)^{\sigma}. \tag{16}$$

That is, more talented leaders will manage larger firms. Also, defining $\sigma \equiv 1/(1-\alpha\rho)$ simplifies notation later.[5]

For this model, we will assume $F(q) = 1 - \left(\frac{q}{\gamma}\right)^{-1/\eta_q}$ for $q > \gamma > 1$. That is, talent follows a Pareto distribution with PL inequality η_q.

With this distribution of talent and with unskilled labor allocated optimally, evaluating the integrals gives

$$Y = A\gamma^n \bar{L}^{\alpha} \left(\frac{1}{1-\rho\sigma\eta_q n}\right)^{\frac{1}{\rho\sigma}}. \tag{17}$$

Finite output requires n sufficiently small. The reason is that as n gets large, the mean of q^n raised to some power may not exist. If TBTC raises the effective talent of the top leaders too much, the tail of the distribution is so thick that output becomes

[5] ζ_1 depends on aggregate output Y and $\bar{L}$.

infinite. This issue will arise in a slightly different context below and will prove important.

At this point, we embed this production setup into an endogenous growth model to study the direction of technological change. Research can be directed toward neutral technological change or talent-biased technological change. The former will leave top income inequality unaffected. But talent-biased technical change — which raises n — will change an exponent and therefore raise top income inequality.

In modeling the direction of technical change, this paper draws on Acemoglu (1998). Further inspiration comes from the quality ladder models of growth associated with Aghion and Howitt (1992) and Grossman and Helpman (1991). In particular, growth in those models involves an exponent that evolves endogenously, which is what we need to study top income inequality.

4.1. Directed Technical Change

We can now describe the remainder of the economic environment, focused on directed technical change. First, it is convenient to introduce new notation related to neutral technological change so that the two directions are more directly parallel. Let $A \equiv \theta^z$. Here, $\theta > 1$ corresponds to the step size in a standard quality ladder model, and z denotes the number of rungs of the ladder that have been ascended. Researchers can spend their time increasing z, the neutral technology, or n, the talent-biased technology.

Suppose S_z scientists work on the neutral technology and S_n scientists work on TBTC. The probability that a given researcher succeeds — and therefore raises z or n by one unit — is

$$\mu_z \equiv \Pr\left[z \text{ innovation}\right] = \zeta_z \cdot \frac{S_z^{\lambda-1}}{A^{1-\phi}} \tag{18}$$

and

$$\mu_n \equiv \Pr\left[n \text{ innovation}\right] = \zeta_n \cdot \frac{S_n^{\lambda-1}}{B^{1-\phi}} \tag{19}$$

In these expressions, $A \equiv \theta^n$ and B (to be defined shortly) capture the extent to which the cumulative number of discoveries make it harder to innovate in the future, assuming $\phi < 1$.[6]

Because TBTC innovations interact with heterogeneous talent q, it is not completely obvious how to define B. We do what seems most natural and continue to parallel the treatment of neutral technical change:

$$\begin{aligned} B &= \left(\int q^{n\omega} dF(q)\right)^{\frac{1}{\omega}} \\ &= \gamma^n \left(\frac{1}{1 - \omega\eta_q n}\right)^{\frac{1}{\omega}}. \end{aligned} \tag{20}$$

Just like A is θ^n, B is the average of the q^n. With the heterogeneity, we can take the average in many ways. The generalized (CES) mean allows the average to be the arithmetic mean ($\omega = 1$) or to put more weight on either the lower part ($\omega < 1$) or the upper part ($\omega > 1$) of the distribution. More discoveries — either n or z — make it harder to make the next proportional improvement in productivity. An important difference between A and B, however, is that there is a finite value $\bar{n} \equiv 1/\eta_q\omega$ at which $B \to \infty$ so that TBTC becomes impossible. The reason for this is interesting. Because talent follows a Pareto distribution, it has a thick tail. Similarly, $q^{n\omega}$ has PL inequality equal to $omega\eta_q$. As this distribution's upper tail gets thicker and thicker, inequality rises and fewer and fewer moments of the distribution exist. In particular, when $n \to 1/\omega\eta_q$, the mean itself ceases to exist and B goes to infinity. The most talented workers are so productive that it becomes impossible for the research technology to improve productivity with another step up the quality ladder. This feature of the setup turns out to be important.

We next assume that the realized number of innovations in each direction precisely equals the expected number from the Poisson process. Given the well-known use of such processes in growth models, nothing is gained by considering the residual stochasticity or adding another continuum to handle this formally. With this motivation, the actual idea production functions are specified as

$$\dot{z} = \mu_z S_z = \frac{\zeta_z S_z^{\lambda}}{\theta^{(1-\phi)z}} \tag{21}$$

[6]This approach is similar to Segerstrom (1998) and is also related to Kortum (1997).

$$\begin{aligned}\dot{n} = \mu_n S_n &= \frac{\zeta_n S_n^\lambda}{B(n)^{1-\phi}} \\ &= \frac{\zeta_n S_n^\lambda}{\gamma^{(1-\phi)n}} \cdot (1 - \omega\eta_q n)^{\frac{1-\phi}{\omega}}.\end{aligned} \tag{22}$$

4.2. The Optimal Allocation

Assume the measure of scientists, leaders, and unskilled workers grows exogenously at a constant rate of population growth β. Consider the optimal allocation of resources that maximizes expected utility behind the veil of ignorance, before people know their talent and whether or not they will be scientists, leaders, or unskilled workers. In this case, consumption will be equalized across all occupations, so we can consider the welfare of a representative agent.

We've already considered the optimal allocation of unskilled labor. The only remaining allocative decision relates to the direction of technical change. Let $v_t \equiv S_{nt}/S_t$ be the fraction of scientists working on TBTC. The optimal intertemporal allocation then solves

$$\max_{\{v_t\}} \int_0^\infty u(c_t) e^{-\kappa t} dt \tag{23}$$

subject to

$$c_t = \zeta_c \, \theta^{z_t} \gamma^{n_t} \left(\frac{1}{1 - \rho\sigma\eta_q n_t} \right)^{\frac{1}{\rho\sigma}}. \tag{24}$$

$$\dot{z}_t = \frac{\zeta_z (1 - v_t)^\lambda S_t^\lambda}{\theta^{(1-\phi) z_t}} \tag{25}$$

$$\dot{n} = \frac{\zeta_n v_t^\lambda S_t^\lambda}{\gamma^{(1-\phi) n_t}} \cdot (1 - \omega\eta_q n_t)^{\frac{1-\phi}{\omega}} \tag{26}$$

where $S_t \equiv \bar{S} e^{\beta t}$ is the total number of scientists available at date t.

Proposition 1 (BGP under the Optimal Allocation): *Assume $\omega > \rho\sigma$. Then the opti-*

mal allocation features a balanced growth path such that

$$g_c = \bar{g} \equiv \frac{\lambda\beta}{1-\phi} \tag{27}$$

$$\dot{z} = \frac{\bar{g}}{\log\theta} \tag{28}$$

$$\dot{n} = 0, \quad v^* = 0 \tag{29}$$

$$n^* = \frac{1}{\omega\eta_q} \tag{30}$$

What is going on and why? The first thing to recall is that if n gets too large, then aggregate output goes to infinity. In particular, this occurs if n reaches $\bar{n} \equiv \frac{1}{\rho\sigma\eta_q}$.

What prevents n from getting to large? Looking back at the TBTC production function in equations (19) and (26), $B^{1-\phi}$ captures the extent to which it becomes harder and harder to produce proportional improvements in TBTC. Recall that B is a generalized mean of the effective talents, q^n, where ω governs the extent to which this mean leans toward the top of the distribution versus the bottom. In particular, the key condition in the proposition that $\omega > \rho\sigma$ requires the generalized mean to focus on the top of the distribution in gauging the diminishing returns to TBTC production. With ω sufficiently large, $n_t \to n^* < \bar{n}$. That is TBTC eventually comes to a halt, at a level that is consistent with finite consumption.

Then, the growth implications of the model are straightforward. Looking back at the production function in (24), the growth rate of consumption will equal $\dot{z}\log\theta$. The fact that $v_t \to 0$ means that all scientists eventually work on neutral technologies. And from (25), a constant $\dot{z}$ requires S_t^λ and $\theta^{(1-\phi)z_t}$ to grow at the same rate, which in turn gives the condition in (28), which ties the growth rate of the economy to the growth rate of the number of scientists.

Notice that this result is robust to the precise shape of the idea production functions. In particular, it is the λ and ϕ in the neutral production functions that pin down growth. The λ and ϕ in the TBTC production function could differ without changing the result.

4.3. Top Income Inequality

The power law exponent for top income inequality in the TBTC model is $\eta_q n_t$. The economy begins with some initial endowment n_0. Overtime, TBTC means that top income inequality is growing, but eventually at a decreasing rate. Ultimately, top income inequality settles down to a constant $\eta_q n^*$. The reason is that the top talented people are so phenomenally productive in this economy that it becomes exceedingly difficult to raise their productivity by a proportional amount, and TBTC comes to a halt.

5. A Third Model: Misallocation

A third framework for studying the dynamics of top income inequality builds on the assignment model literature of Sattinger (1975, 1993), Heckman and Sedlacek (1985), Moscarini (2005), Gabaix and Landier (2008), and Tervio (2008). Suppose production involves matching between firms and talented workers. The workers have talent x drawn from a Pareto distribution with power-law exponent ξ_x. The collection of fundamental productivity and other inputs (capital, materials, and raw labor, for example) for the firms is denoted A and is drawn from another Pareto distribution with power-law exponent ξ_A. The talented workers could be CEOs, star lawyers, famous authors, top surgeons, and so on.

A firm with productivity A matching with a worker of talent x produces output $y = Ax$. We could of course put exponents on A and x and have these exponents sum to one; that would not change anything. Assume that the talented worker gets paid an amount that is proportional to y.

What does inequality among the talented workers look like in this setting? There are two special cases that lead immediately to the key result of this section. To see this, suppose for simplicity that the distribution of talent and productivity is exactly the same, even upon realization; in particular, $\eta_A = \eta_x = \bar{\eta}$. Because productivity and talent are complements in production, the optimal assignment features perfect assortative matching. That is, we assign the most talented worker to the most productive firm, and so on down the line. And since $A = x$, we have $y = x^2$ and therefore $\eta_y = 2\bar{\eta}$, using the rules for power-law inequality from Section 2.2.

Now consider what happens if instead we mis-allocate talent. First, suppose we allocate talent in the worst possible way, assigning the most talented worker to the *least* productive firm. The distribution of output and therefore wages to the talented workers will obviously be much flatter in this case. Since the Pareto distribution is bounded from below, the least productive firm has some minimum level of productivity (call it γ), so that at the top, $y = \gamma x$, and therefore $\eta_y = \bar{\eta}$. That is, inequality is determined solely by the top worker's talent (or by the top firm's productivity, which has the same distribution, by assumption). The power-law inequality exponent is reduced by a factor of two by this misallocation relative to the perfect matching case.

Alternatively, suppose we randomly assign workers to firms. In this case, $y = Ax$ is the product of two iid Pareto random variables, so the power-law exponent for y is the same as that for x, and $\eta_y = \bar{\eta}$. The inequality exponent is once again half what it is under the optimal assignment.

What this example shows is that the misallocation of talent can substantially reduce top income inequality. Conversely, a rise in top income inequality could potentially be explained in a setting like this by an improvement in the allocation of talent.

5.1. Shortcomings of this Simple Model

While this example conveys the key intuition that misallocation can affect inequality, the example suffers from several shortcomings. First notice that it implies that the rise in inequality among talented workers should reflect a rise in inequality between firms as well. In particular, in this simple example (and in richer models built upon it), the rise in top income inequality should parallel a rise in Pareto inequality for firm sales. However, it is well-documented that the size distribution of firms is remarkably stable and features a much larger degree of inequality than top income inequality. For example, the power-law exponent for the size distribution of firms is a number like 1.06, meaning the inequality exponent is also very close to one. Moreover, there is no rise in inequality among firms in the U.S. since 1980 — the rise in top income inequality occurs amidst a stable firm-size distribution. In contrast, we saw in Section 2. above, the inequality exponent for top incomes in the United States is a number close to 1/2 and it has been rising sharply over time.

A second — and related — limitation of the simple model is that it assumes that the talent distribution itself is Pareto. This could be a reasonable approximation, but one might also question this assumption.[7] For example, Gabaix and Landier (2008) use CEO compensation data to back out the distribution of top talent. Their analysis suggests that, far from being Pareto, the distribution of top talent is in fact bounded from above. Specifically, the suggest that the distribution of top talent is

$$G(t) = 1 - \left(\frac{T_{max} - t}{T_{max} - T_{min}}\right)^{-1/\eta_t} \tag{31}$$

where $\eta_t < 0$ and $T_{min} < t < T_{max}$. (The reason for using notation that requires a negative parameter value will become clear shortly.) This distribution — and a Pareto distribution for comparison — is shown in Figure 11.

Fortunately, both of these problems can be addressed. In the remainder of this section, we apply the Gabaix and Landier (2008) framework to study misallocation and top income inequality. The inequality results from the simple model carry over to this richer setting, and the limitations we've just discussed disappear.

5.2. Misallocation and Inequality in Gabaix-Landier (2008)

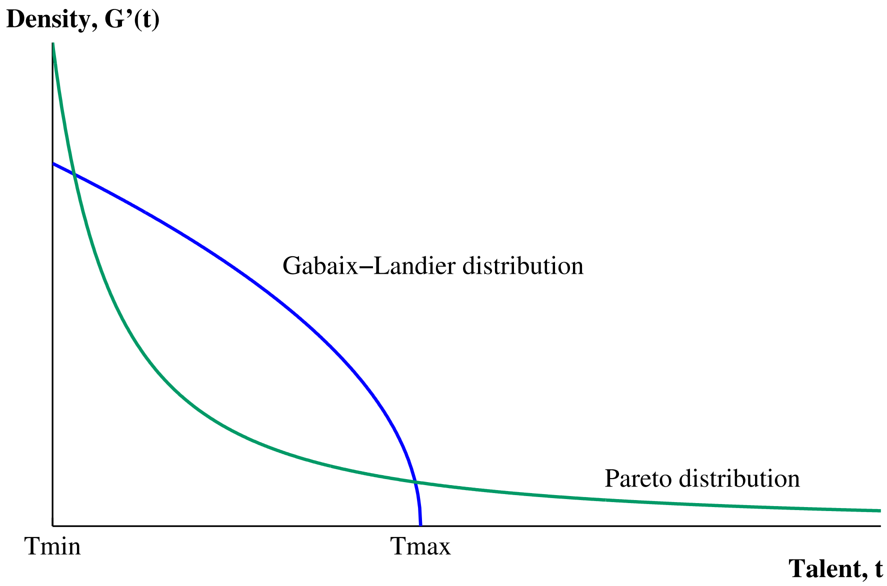
Firms indexed by n produce a homogeneous output good with a production function given by

$$Y_n = A_n L_n^{\alpha} T_n^{1-\alpha} \tag{32}$$

where Y_n is output, A_n is exogenous productivity, L_n is the number of workers hired for unskilled positions, in which talent is for simplicity assumed not to matter. In

[7]The distribution of IQ scores? Lognormal at the top??? thin tail??? What evidence and what citation? Check Garett Jones work.

Figure 11: Talent Distributions



contrast, T_n is the talent of the worker hired for the single position at the firm where talent is crucial. This could be the CEO of a firm but could also be the top surgeon in a medical group, the key rainmaker in a law firm, the star on a sports team, the leader of a band, or even the line manager on a particular production line within a firm. To be concrete, we'll refer to this person as the "leader." In a richer setting (which would be interesting to pursue), the talent of all the workers in the firm would matter, of course. But we keep it simple for now to emphasize top wage inequality. The production setup here is reminiscent of the Lucas (1978) span of control model.

The productivities A_n obey a Pareto distribution, with a power-law inequality exponent $\eta_a > 0$. As in Gabaix and Landier (2008), assume firms are ordered so that the firm with the highest A is at position $n = 0$ and the firm with the lowest A is at position $n = 1$. We think of n as denoting the quantiles, from the top down, of the distribution of A. Therefore,

$$A_n = \phi_a n^{-\eta_a} \tag{33}$$

for some constant ϕ_a.[8]

There are $\bar{L}$ people in this economy. Each person has talent, t, drawn from some distribution $G(t)$. In the efficient allocation of workers, the most talented people in the economy will work as the leaders, while the less talented workers fill the unskilled positions.

Actual talent, however, is not observed. Instead, there is a "perceived" talent distribution: talent has been "scrambled" in a way that will be described in more detail below.

In maximizing expected profits, it is the expectation of "talent raised to the power $1-\alpha$" that enters, so we define

$$T(m)^{1-\alpha} := \mathbb{E}\left[\, t^{1-\alpha} \,|\, m\right] \tag{34}$$

where m is the (upper) quantile of the perceived distribution of talent.

Then, firms hire workers and choose their leader in order to maximize expected profits:

$$\max_{L_n, m} A_n L_n^{\alpha} T_n(m)^{1-\alpha} - w_L L_n - w(m) \tag{35}$$

where w_L is the wage of workers in the unskilled position, and $w(m)$ is the wage earned by the leader at the mth position.

The first order conditions from this problem are both crucial:

$$\alpha \frac{Y_n}{L_n} = w_L \tag{36}$$

$$(1-\alpha)\frac{Y_n}{T_n(m)} T_n'(m) = w'(m). \tag{37}$$

In equilibrium, the most productive firms will be matched to the leaders with the best perceived talent. That is, there is positive assortative matching: $n = m$.

Using this result and substituting (36) into (32) gives the output of the firm as a function of productivity and the leader's talent, showing the multiplier effect that

---

[8]To see more details, let Pr$\,[A > a] \equiv \bar{G}(a) = \left(\frac{a}{\underline{a}}\right)^{-1/\eta_a}$ for $a > \underline{a}$. Then $n(a) = \bar{G}(a)$ is the percentile of some productivity level a. Inverting, $a(n) = \underline{a} n^{-\eta_a}$.

each of these delivers for firm size:

$$Y_m = \left(\frac{\alpha}{w_L}\right)^{\frac{\alpha}{1-\alpha}} A_m^{\frac{1}{1-\alpha}} T(m). \tag{38}$$

Notice that this implies that Y_m exhibits a power law. The nature of this power law depends on the distribution of talent, as we can see in the following result.

Proposition 2 (The Size Distribution of Firms): *Let $\eta_a \equiv \eta_A/(1-\alpha)$. The size distribution of firms obeys a power law with the following inequality exponent:*

1. *If the (perceived) distribution of talent is itself a power law with parameter $\eta_T > 0$, then*

$$\eta_Y = \eta_a + \eta_T$$

2. *If the (perceived) distribution of talent is the (bounded) Gabaix-Landier distribution with $\eta_T < 0$, then*

$$\eta_Y = \eta_a.$$

Next, substituting (38) into (37) gives

$$w'(m) = \phi_w A_m^{\frac{1}{1-\alpha}} T'(m) \tag{39}$$

where ϕ_w is a constant.[9] This equation is a canonical wage equation from assignment models, like Sattinger (1975) and Tervio (2008).

It is useful to contrast this canonical wage equation with the equation governing the size distribution of firms, equation (38). The wage equation says that the wages of top talent depend on the spacings of talent, $T'(m)$, whereas the size distribution of firms depends directly on the talent distribution itself, $T(m)$.

To go further, we use the key insight of Gabaix and Landier (2008): the spacing of extreme values at the top of any "regular" distribution is well-approximated by

$$T'(m) = -\phi_t m^{-\eta_T - 1} \tag{40}$$

[9] $\phi_w \equiv (1-\alpha)\left(\frac{\alpha}{w_L}\right)^{\frac{1}{1-\alpha}}$.

where ϕ_t is some constant. The parameter η_T governs top inequality in the distribution of perceived talent, precisely like the power law exponents such as η_a used in the rest of this paper (which is why we've chosen a similar notation).

In this case, however, η_T can be negative. More generally, results in Gabaix and Landier (2008) imply that

- η_T is the power law inequality exponent if the distribution of perceived talent $F(T)$, exhibits a power law
- $\eta_T = 0$ if $F(T)$ is lognormal, normal, or exponential
- $\eta_T = -1$ if $F(T)$ is uniform
- $\eta_T < 0$ if $F(T)$ is Weibull or the GL talent distribution itself

Given this discussion of the talent distribution and $T'(m)$, we are now ready to return to the wage distribution for leaders. In particular, we now substitute the power law equation for A_n from (33) and the talent spacing equation for $T'(m)$ from (40) into the wage assignment equation (39):

$$w'(m) = -\phi m^{-(\eta_a+\eta_T)-1} \tag{41}$$

where $\phi = \phi_a \phi_t \phi_w$ is just a constant.

Assuming $\eta_a + \eta_T > 0$ (which is not innocuous since $\eta_T < 0$ is allowed, but which we argue below holds for plausible parameter values), one can integrate this equation to get the wage distribution for top talents:

$$w(m) = \text{Constant} + \frac{\phi}{\eta_a + \eta_T} m^{-(\eta_a+\eta_T)}. \tag{42}$$

In this wage equation for top talents, $w(m)$ decays as a power function of m at rate $\eta_a + \eta_T$. It is easy to show that this implies that the wage distribution exhibits a power law with

$$\boxed{\eta_w = \eta_a + \eta_T.} \tag{43}$$

In other words, top wage inequality is the sum of two other inequality parameters: the degree of inequality among firm productivity, η_a, and the degree of inequality at the top of the distribution of perceived talent, η_T.

5.3. Misallocation and Talent

We now explain how misallocation can alter the spacing of talent and therefore affect top wage inequality. It is optimal in this setting for the best talent to match with the most productive firm, the firm with the highest A, and so on. To model misallocation, we suppose that talent is instead matched according to the order of δA, where δ is the "noise" in the matching process, drawn from a Pareto distribution with power law inequality $\eta_\delta > 0$. This noise is independent of the Pareto distribution of productivity.

Lemma 1 *Let $x \equiv \delta A$ be the product of these two independent Pareto random variables. Then, the distribution of x, $F(x)$, is given by*

$$1 - F(x) = x^{-\xi_\delta} + x^{-\xi_A} \tag{44}$$

where $\xi_A \equiv 1/\eta_A$ and $\xi_\delta \equiv 1/\eta_\delta$ and we are ignoring two irrelevant coefficients on the power functions.

There is then positive assortative matching between true talent, t, and the noisy index of productivity, δA, and we get the following result:

Proposition 3 (The Misallocation of Talent): *Let there be positive assortative matching between t and $x \equiv \delta A$. That is,*

$$1 - G(t) = m = 1 - F(\delta A).$$

Then, depending on the talent distribution (and ignoring the coefficients on the power functions), we get two results:

1. *If the distribution of actual talent, t, is Pareto with parameter $\eta_t > 0$, then*

$$t(x) = \gamma_t \left(x^{-\xi_\delta} + x^{-\xi_A}\right)^{-\eta_t}$$

2. *If the distribution of actual talent, t is the (bounded) Gabaix-Landier distribution with $\eta_t < 0$, then*

$$t(x) = \bar{T} - \phi \left(x^{-\xi_\delta} + x^{-\xi_A} \right)^{-\eta_t}$$

Given this matching result, we can now determine the spacing of talent in the upper tail. First, we state precisely what we mean by $T(m)$:

$$T(m)^{1-\alpha} := \mathbb{E}_\delta \left[t(\delta A)^{1-\alpha} \,|\, A = A(m) \right] \tag{45}$$

It is the expected talent that gets matched to the firm at the m^{th} quantile.

And we can now state the key results for wage inequality and firm size.

Proposition 4 (Misallocation and Inequality): *For either the Pareto talent distribution or the Gabaix-Landier talent distribution,*

$$\eta_T = \begin{cases} \eta_t & \text{if } \eta_\delta \leq \eta_A \quad \text{“small noise”} \\ \frac{\eta_A}{\eta_\delta} \cdot \eta_t & \text{if } \eta_\delta > \eta_A \quad \text{“big noise”} \end{cases}$$

and given this value, we have the following key results:

1. *If the distribution of t is Pareto with parameter $\eta_t > 0$, then $\eta_T > 0$ and*

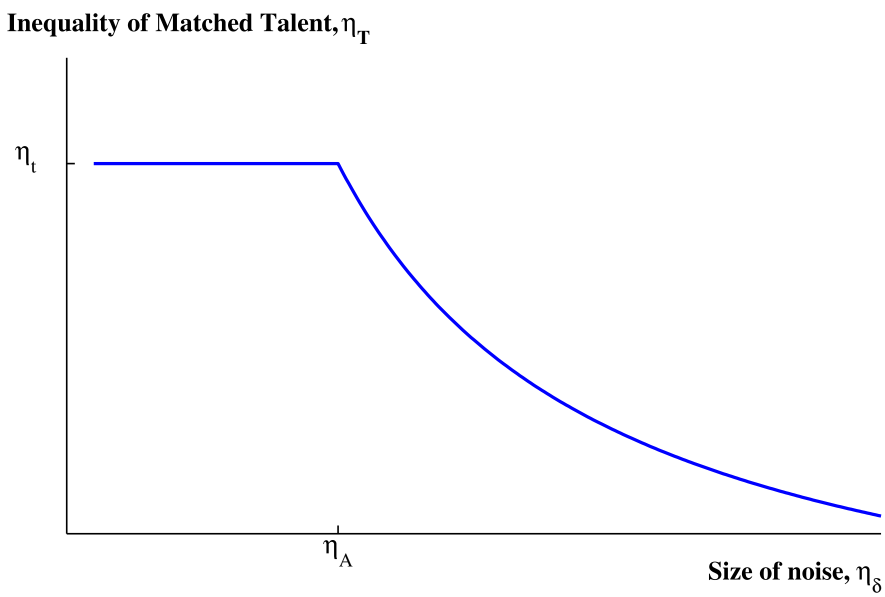
$$\eta_w = \eta_a + \eta_T = \eta_Y$$

2. *If the distribution of t is the (bounded) Gabaix-Landier distribution with $\eta_t < 0$, then $\eta_T < 0$ and*

$$\eta_Y = \eta_a \text{ and } \eta_w = \eta_Y + \eta_T < \eta_Y$$

The way in which inequality varies with the noise is shown in Figure 12. If the noise has an inequality exponent that is smaller than productivity, then the noise is irrelevant: the top talent shines through despite the noise and the best people are matched with the best firms. This is easy to understand if you imagine the density of the upper tail for the noise and for productivity. If the tail for the noise is thinner,

Figure 12: Inequality of Scrambled Talent, η_T



then at the very top it is easy to see who the most productive firms are. However, if the noise has at least as much inequality as productivity, then the noise will distort the allocation, and the inequality of matched talent will be reduced relative to true talent inequality.

5.4. Remarks

This motivates several remarks. First, power law inequality in the firm size distribution is large, and $\eta_y \approx 1$ is the conventional wisdom, whether firm size is measured as sales revenue or employment. For example, Axtell (2001) is a classic citation arguing that $\xi_y = 1.06$ in the United States, implying $\eta_y = 1/1.06 = .94$. Moreover, this finding is remarkably stable over time.

This fact poses a problem for the misallocation model if talent has a Pareto distribution. As shown in the proposition above, in this case, top wage inequality and firm size inequality should be the same! This prediction is soundly at odds with the U.S. facts.

But this is why the Gabaix and Landier (2008) setup is so useful. In a setting with no misallocation of talent, so that $\eta_T = \eta_t$, Gabaix and Landier estimate this parameter for CEOs of large, publicly traded companies in the United States and find $\eta_T = -2/3$. Equation (43) then suggests $\eta_w \approx 1/3$. This value is in the ballpark of the estimates for top income inequality shown in Figure 5 for the United States between 1950 and 1980; it seems low by as much as a factor of two for recent years.

But with $\eta_t < 0$, it is the second part of Proposition 4 that is relevant. In this case, since talent is bounded, firm size inequality is invariant to misallocation: $\eta_Y = \eta_a$. Top wage inequality is strictly less than firm size inequality, and can be affected by misallocation. In particular, if misallocation is sufficiently large (that is, if $\eta_\delta > \eta_A$), then the spacing of talents exhibits less inequality than actual talent: $\eta_T = \frac{\eta_A}{\eta_\delta} \cdot \eta_t < \eta_t$. And the larger is the noise in the matching process, the lower is top wage inequality.

How can this help us understand the facts about top income inequality? First, regarding the levels of inequality, the model suggests that countries that allocate top talent most efficiently will have the highest top income inequality. There is a tradeoff between efficiency and equity.

Second, improvements in the allocation of talent will increase inequality, at least up to a point. (Once noise is sufficiently small, in particular once $\eta_\delta \leq \eta_A$, inequality will remain the same as $\eta_T = \eta_t$.) Rising top income inequality in the United States since 1980 could reflect an improvement in the allocation of talent. Similarly, low levels of inequality and the lack of an increase in top inequality in France and Japan may reflect the inflexibility of the labor markets in those countries.

What evidence is there that the allocation of talent has improved in the United States? One source of evidence is Hsieh, Hurst, Jones and Klenow (2012). That paper observes that the occupational distribution for men and women has changed sharply over the last 50 years. As one example, in 1960, 94 percent of doctors and lawyers were white men. By 2008, this fraction had fallen to 62 percent. At the very least, this suggests the possibility that the allocation of top talent may have improved.

Informally, the top of the U.S. labor market also seems increasingly open to immigrants. As just one example, think about how the nature of the economics profession has changed since 1980. Related observations have been made with respect to the founders of Fortune 500 companies and superstar athletes. Increasingly, it seems that top talent from around the world is attracted to the United States.

6. Conclusion

This paper suggests several mechanisms for understanding the dynamics of top income inequality. Changes in the return to experience (the growth rate of individual human capital), changes in the entry rate of new college graduates, talent-biased technical change, and misallocation among the most talented workers in firms are all possible forces that affect top income inequality. Empirical work will be needed to quantify the importance of these and other possible explanations. What we hope this paper has achieved, however, is an exposition of precisely how these forces can affect the Pareto exponent of the income distribution, which seems to be a useful summary statistic of top income inequality.

References

Acemoglu, Daron, "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *Quarterly Journal of Economics*, 1998, *113*, 1055–1089.

— 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.

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.

Axtell, Robert L., "Zipf Distribution of U.S. Firm Sizes," *Science*, September 2001, *293*, 1818–1820.

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," Department of Economics Working Papers 2010-24, Department of Economics, Williams College March 2008.

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.

Elsby, Michael W. L. and Matthew D. Shapiro, "Why Does Trend Growth Affect Equilibrium Employment? A New Explanation of an Old Puzzle," *American Economic Review*, June 2012, *102* (4), 1378–1413.

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, August 1999, *114* (3), 739–767.

— , "Power Laws in Economics and Finance," *Annual Review of Economics*, 2009, *1* (1), 255–294.

— and Augustin Landier, "Why Has CEO Pay Increased So Much?," *Quarterly Journal of Economics*, 2008, *123* (1), 49–100.

Gordon, Robert J. and Ian Dew-Becker, "Controversies about the Rise of American Inequality: A Survey," NBER Working Papers 13982, National Bureau of Economic Research, Inc May 2008.

Grossman, Gene M. and Elhanan Helpman, *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press, 1991.

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.

Heckman, James J. and Guilherme Sedlacek, "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market," *Journal of Political Economy*, 1985, *93* (6), 1077–1125.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," 2012. unpublished paper.

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 and Kevin Murphy, "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *Quarterly Journal of Economics*, February 1992, *107* (1), 35–78.

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, "Job Market Paper," 2012. unpublished paper.

Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, 1997, *65* (6), 1389–1419.

Lucas, Jr. Robert E. and Benjamin Moll, "Knowledge Growth and the Allocation of Time," NBER Working Papers 17495, National Bureau of Economic Research, Inc October 2011.

Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 1988, *22* (1), 3–42.

— , "Ideas and Growth," *Economica*, February 2009, *76* (301), 1–19.

Lucas, Robert E., Jr., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 1978, *9*, 508–523.

Malevergne, Y., A. Saichev, and D. Sornette, "Zipf's law and maximum sustainable growth," Quantitative Finance Papers 1012.0199, arXiv.org December 2010.

Moscarini, G., "Job Matching and the Wage Distribution," *Econometrica*, 2005, *73* (2), 481–516.

Pareto, Vilfredo, *Cours d'Economie Politique*, Geneva: Droz, 1896.

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.

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.

Sattinger, Michael, "Comparative advantage and the distributions of earnings and abilities," *Econometrica*, 1975, pp. 455–468.

— , "Assignment models of the distribution of earnings," *Journal of Economic Literature*, 1993, *31* (2), 831–880.

Segerstrom, Paul, "Endogenous Growth Without Scale Effects," *American Economic Review*, December 1998, *88* (5), 1290–1310.

Steindl, Joseph, *Random Processes and the Growth of Firms*, New York: Hafner, 1965.

Tervio, Marko, "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review*, 2008, *98* (3), 642–668.