A Schumpeterian Model of Top Income Inequality

Charles I. Jones

Stanford University and National Bureau of Economic Research

Jihee Kim

Korea Advanced Institute of Science and Technology

Top income inequality rose sharply in the United States over the last 40 years but increased only slightly in France and Japan. Why? We explore a model in which heterogeneous entrepreneurs, broadly interpreted, exert effort to generate exponential growth in their incomes, which tends to raise inequality. Creative destruction by outside innovators restrains this expansion and induces top incomes to obey a Pareto distribution. Economic forces that affect these two mechanisms—including information technology, taxes, and policies related to innovation blocking—may explain the varied patterns of top income inequality that we see in the data.

I. Introduction

As documented extensively by Piketty and Saez (2003) and Atkinson, Piketty, and Saez (2011), top income inequality—such as the share of in-

For many helpful comments and suggestions, we are grateful to Daron Acemoglu, Philippe Aghion, Ufuk Akcigit, Jess Benhabib, Sebastian Di Tella, Xavier Gabaix, Mike Harrison, Pete Klenow, Ben Moll, Chris Tonetti, Alwyn Young, and Gabriel Zucman; seminar participants at the American Economic Association annual meetings, Brown, Chicago Booth, Centre de Recerca en Economia Internacional, the Federal Reserve Bank of San Francisco, Groningen, Hong Kong University of Science and Technology, Institute for International Economic Studies Stockholm, Korea University, the NBER Economic Fluctuations and Growth group, the NBER Income Distribution group, Princeton, the Society for Economic Dynamics 2015 meeting, Stanford, University of California Santa Cruz, University of Southern California, University of Washington, Yale, Yonsei University, and Zurich; and especially to the editor and four referees. Data and programs are provided as supplementary material online.

Electronically published September 4, 2018
[*Journal of Political Economy*, 2018, vol. 126, no. 5]

come going to the top 1 percent or top 0.1 percent of earners—has risen sharply in the United States since around 1980. The pattern in other countries is different and heterogeneous. For example, top inequality rose only slightly in France and Japan. Why? What economic forces explain the varied patterns in top income inequality that we see around the world?

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

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

The basic insight in this literature is that exponential growth, tweaked appropriately, can deliver a Pareto distribution for outcomes. The tweak is needed for the following reason. Suppose that city populations (or incomes or employment by firms) grow exponentially at 2 percent per year plus some random normally distributed shock. In this case, the log of population would follow a normal distribution with a variance that grows over time. To keep the distribution from spreading out forever, we need an additional force. For example, a constant probability of death will suffice to render the distribution stationary.

In the model we develop below, researchers create new ideas: new computer chips or manufacturing techniques, but also best-selling books, smart-

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

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

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

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

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

The most closely related papers to this one are Levy (2003), Nirei (2009), Benhabib, Bisin, and Zhu (2011), Moll (2012), Piketty and Saez (2013), Toda (2014), Piketty and Zucman (2015), Benhabib and Bisin (2016), Hubmer, Krusell, and Smith (2016), and Nirei and Aoki (2016). 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. Geerolf (2016) connects both top income inequality and firm size inequality in a Garicano (2000) style model of hierarchies, building on the assignment model of Gabaix and Landier (2008).[3]

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

Finally, Gabaix et al. (2016) show that the basic random growth model has trouble matching the transition dynamics of top income inequality. Build-

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

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

[4] Classic papers on generating Pareto distributions for income include Champernowne (1953), Simon (1955), and Mandelbrot (1960).

ing on Luttmer (2011), they suggest that a model with heterogeneous mean growth rates for top earners will be more successful, and we incorporate their valuable insights, as discussed further below.

II. Some Basic Facts

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

Figure 2 shows the dynamics of top income inequality for a range of countries, illustrating that the United States and France are large countries close to the two extremes. The horizontal axis shows the share of aggregate income earned by the top 1 percent, averaged between 1980 and 1982, while the vertical axis shows the same share for 2006–8. All the economies for which we have data lie above the 45-degree line; that is, top income inequality has risen everywhere. The rise is the largest in the United States, South Africa, and the United Kingdom, but substantial increases are also seen elsewhere, such as in Ireland, Norway, Singapore, Italy, and Sweden. Japan and France exhibit smaller but still noticeable increases. For example, the top 1 percent share in France rises from 7.4 percent to 9.0 percent.

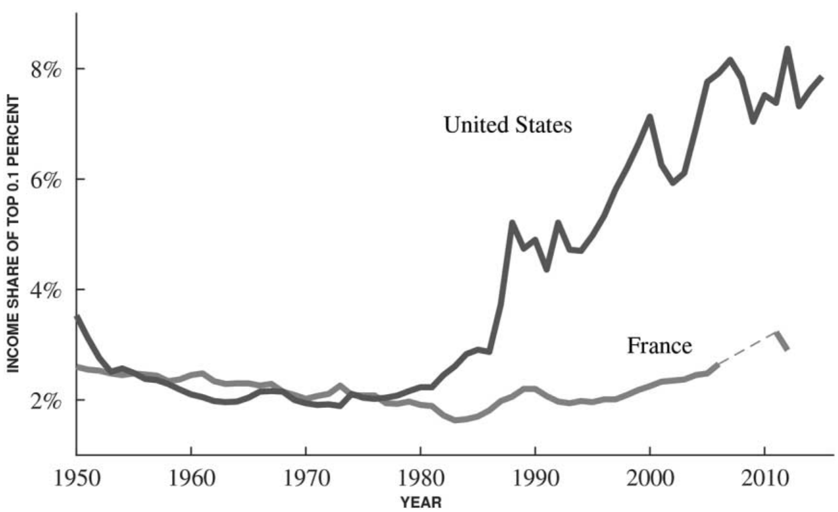

FIG. 1.—Top income inequality in the United States and France. Source: World Wealth and Income Database (http://www.wid.world/). Includes interest and dividends but not capital gains. Color version available as an online enhancement.

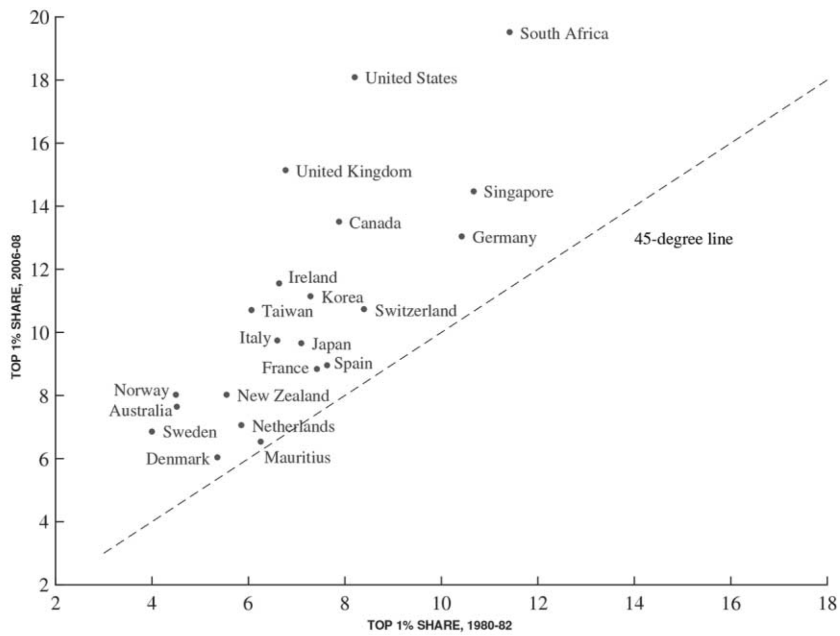

Fig. 2.—Top income inequality around the world, 1980–82 and 2006–8. Top income inequality has increased since 1980 in every country for which we have data. The size of the increase varies substantially, however. Source: World Wealth and Income Database (http://www.wid.world/). Color version available as an online enhancement.

A. *The Role of Labor Income*

As discussed by Atkinson et al. (2011) and Piketty, Saez, and Zucman (2016), a substantial part of the rise in US top income inequality represents a rise in labor income inequality, particularly if one includes "business income" (i.e., profits from sole proprietorships, partnerships, and S corporations) in the labor income category. Given our focus on entrepreneurs, our ideal income measure would always include entrepreneurial income. From now on, when we speak of "labor income," we will include entrepreneurial income as well. Figure 3 shows an updated version of a graph from Piketty and Saez (2003) for the period since 1950, supporting the observation that much of the rise in top income inequality is associated with this broad concept of labor income.

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

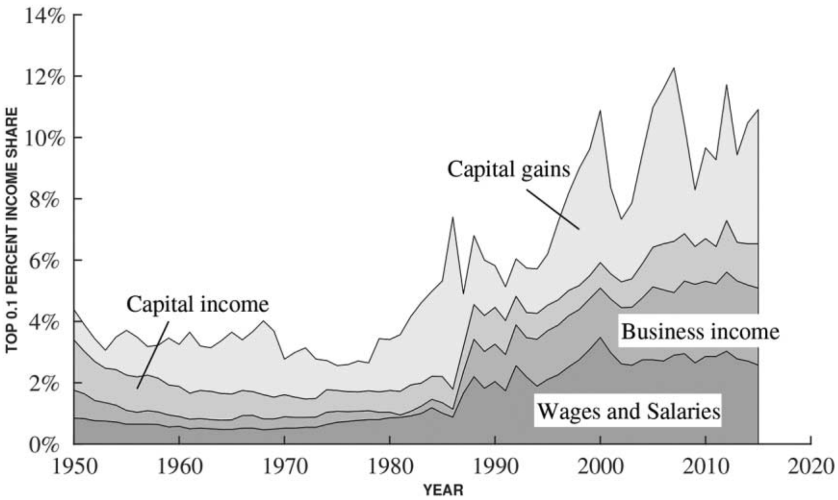

FIG. 3.—The composition of the top 0.1 percent income share. Source: These data are taken from the "data-Fig4B" tab of the June 2016 update of the spreadsheet appendix to Piketty and Saez (2003). Color version available as an online enhancement.

mately Pareto. In particular, if Y is a random variable denoting incomes, then, at least above some high level (i.e., for $Y \geq y_0$),

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

where ξ is called the "power law exponent."

Saez (2001) shows that wage and salary income from US income tax records in the early 1990s is well described by a Pareto distribution. Figure 4 replicates his analysis for 1980 and 2005 for a broader income concept that includes both wage and salary income as well as entrepreneurial income from businesses. In particular, the figures plot mean income above some threshold as a ratio to the threshold itself. If income obeys a Pareto distribution like that in (1), then this ratio should equal the constant $\xi/(\xi - 1)$, regardless of the threshold. That is, as we move to higher and higher income thresholds, the ratio of average income above the threshold to the threshold itself should remain constant.[5] Figure 4 shows that this property holds reasonably well in 1980 and 2005 and also illustrates that the ratio has risen substantially over this period, reflecting the rise in top income inequality.

[5] This follows easily from the fact that the mean of a Pareto distribution is $\xi y_0/(\xi - 1)$ and that the conditional mean just scales up with the threshold.

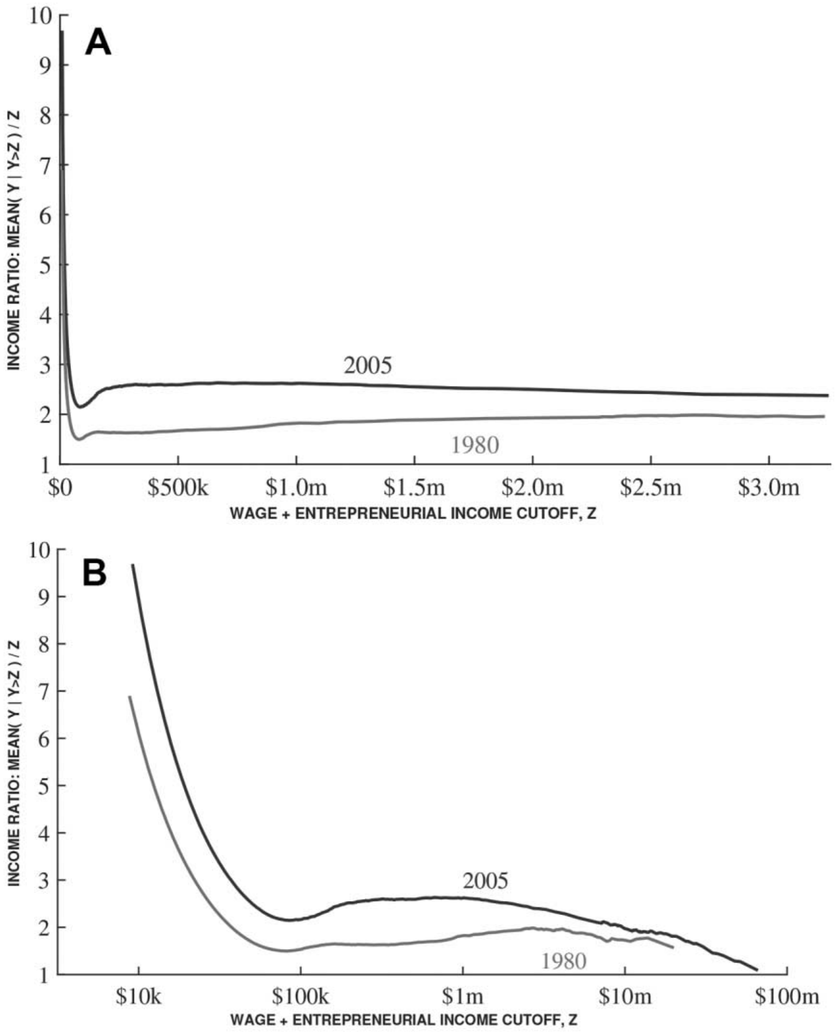

FIG. 4.—The Pareto nature of labor income (broadly defined). *A*, Linear scale, up to $3 million. *B*, Log scale. The figures plot the ratio of average wage plus entrepreneurial income above some threshold z to the threshold itself. For a Pareto distribution with Pareto inequality parameter η, this ratio equals $1/(1-\eta)$. Saez (2001) produced similar graphs for 1992 and 1993 for wage and salary income using the IRS public use tax files available from the NBER at www.nber.org/taxsim-notes.html. The figures here replicate these results using the same data source and a broader income concept for 1980 and 2005. Color version available as an online enhancement.

B. Fractal Inequality and the Pareto Distribution

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

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

so that

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

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

An important property of Pareto distributions is that they exhibit a fractal pattern of top inequality. To see this, let $S(a) = \tilde{S}(a)/\tilde{S}(10a)$ denote the fraction of income earned by the top $10 \times a$ percent of people that actually goes to the top a percent. For example, $S(1)$ is the fraction of income going to the top 10 percent that actually accrues to the top 1 percent, and $S(0.1)$ is the fraction of income going to the top 1 percent that actually goes to the top one in 1,000 earners. Under a Pareto distribution,

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

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

Not surprisingly, top income inequality is well characterized by this fractal pattern, as shown in figure 5.[6] At the very top, the fractal prediction holds remarkably well, and $S(0.01) \approx S(0.1) \approx S(1)$. Prior to 1980, the fractal shares are around 25 percent: one-quarter of the top X percent's income goes to the top $X/10$ percent. By the end of the sample in 2015, this fractal share is closer to 40 percent.

The rise in fractal inequality shown in figure 5 can be related directly to the power law inequality exponent using equation (4) and taking logs. The corresponding Pareto inequality measures are shown in figure 6. This figure gives us the quantitative guidance that we need for theory. The goal is to build a model that explains why top incomes are Pareto and that

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

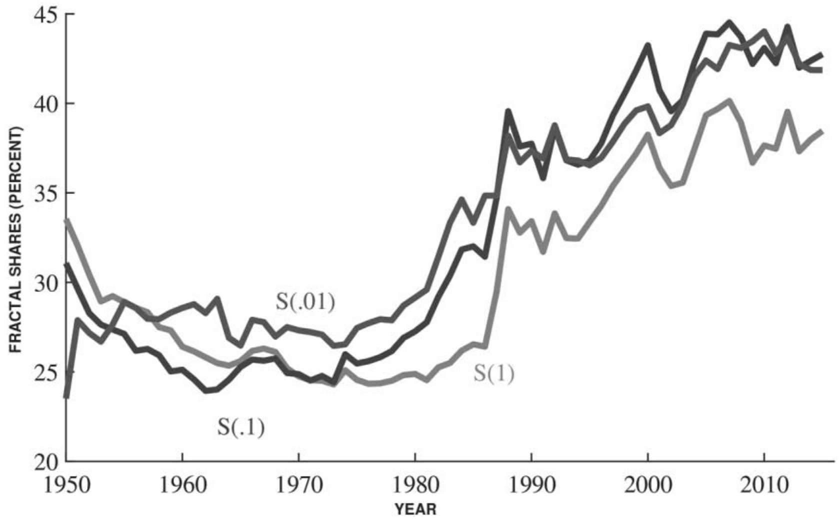

FIG. 5.—Fractal inequality of US income. The term $S(a)$ denotes the fraction of income going to the top $10a$ percent of earners that actually goes to the top a percent. For example, $S(1)$ is the share of the top 10 percent's income that accrues to the top 1 percent. Source: World Wealth and Income Database (http://www.wid.world/). Includes interest and dividends but not capital gains. Color version available as an online enhancement.

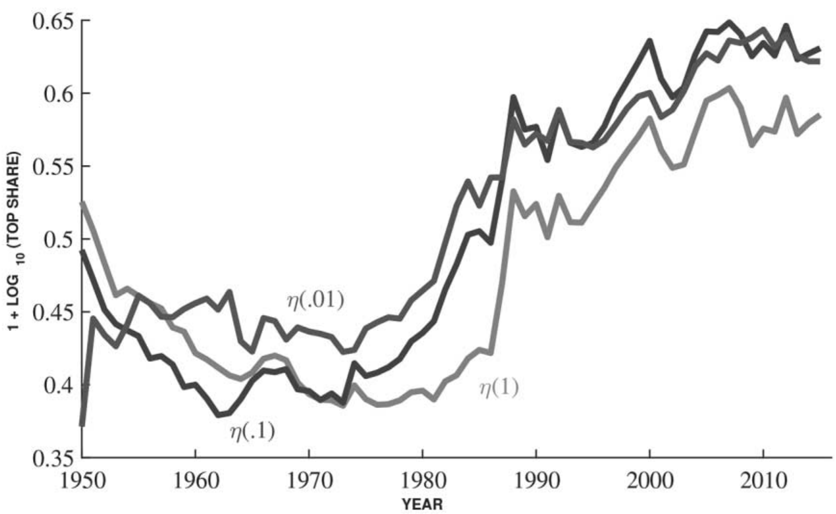

FIG. 6.—The power law inequality exponent η, United States. The term $\eta(a)$ is the power law inequality exponent obtained from the fractal inequality income shares in figure 5 assuming a Pareto distribution. See equation (4) in the text. Color version available as an online enhancement.

generates a Pareto exponent that rises from around 0.4 to around 0.6 for the United States but by much less in France and other countries.

C. *Skill-Biased Technical Change?*

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

Suppose the economy consists of a large number of homogeneous low-skilled workers with fixed income $\bar{y}$. High-skilled people, in contrast, are heterogeneous: income for highly skilled person i is $y_i = \bar{w}x_i^{\alpha}$, where x_i is person i's skill and $\bar{w}$ is the wage per unit of skill (ignore α for now). If the distribution of skill across people is Pareto with inequality parameter η_x, then the income distribution at the top will be Pareto with inequality parameter $\eta_y = \alpha\eta_x$. That is, if $\Pr[x_i > x] = x^{-1/\eta_x}$, then $\Pr[y_i > y] = (y/\bar{w})^{-1/\eta_y}$. An increase in $\bar{w}$—a skill-biased technical change that increases the return to skill—shifts the Pareto distribution right, increasing the gap between high-skilled and low-skilled workers. But it does not change Pareto inequality η_y: a simple story of skill-biased technical change is not enough.

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

D. *Summary*

Here then are the basic facts related to top income inequality that we would like to be able to explain. Between 1960 and 1980, top income inequality was relatively low and stable in both the United States and France. Since around 1980, however, top inequality has increased sharply in countries such as the United States, Norway, and the United Kingdom, while it has increased only slightly in others, including France and Japan. Finally, labor income is well described by a Pareto distribution, and rising top income inequality is to a great extent associated with rising labor income inequality. Changing top income inequality corresponds to a change in the

[7] Gabaix et al. (2016) explore an alternative approach they call “scale dependence” in an extension of Gabaix and Landier (2008), viewing a rise in α as a convexification in the returns to skill.

power law inequality exponent, and the US data suggest a rise from about 0.4 in the 1970s to about 0.6 by 2015. The remainder of this paper develops and analyzes a model to help us understand these facts.

III. A Simple Model of Top Income Inequality

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

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

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

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

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

With a Poisson replacement process, it is well known that the distribution of experience for a given individual follows an exponential distribution, that is, $\Pr[\text{Experience} > x] = e^{-\delta x}$. Let us take for granted that the stationary distribution of experience across a population of entrepreneurs is this same exponential distribution; this is shown more formally in appendix A (apps. A–D are available online). Then the remainder of the argument is straightforward:

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

which is a Pareto distribution!

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

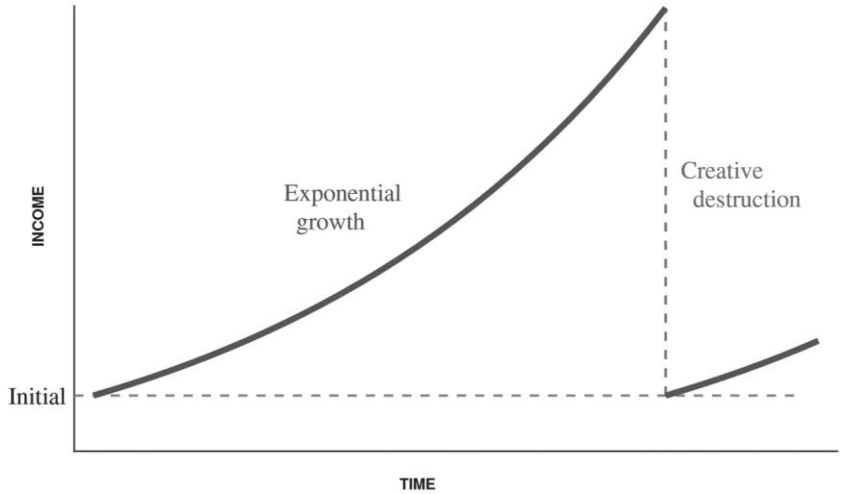

FIG. 7.—Basic mechanism: exponential growth with death ⇒ Pareto. Color version available as an online enhancement.

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

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

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

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

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

IV. A Schumpeterian Model of Top Income Inequality

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

A. *Entrepreneurs*

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

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

Given an x, the entrepreneur maximizes the expected present discounted value of flow utility, $u(c, \ell) = \log c_t + \beta \log \ell_t$, subject to the following constraints:

$$c_t = \psi_t x_t, \tag{8}$$

$$e_t + \ell_t + \tau = 1, \tag{9}$$

$$dx_t = \mu(e_t) x_t dt + \sigma x_t dB_t, \tag{10}$$

$$\mu(e) = \phi e. \tag{11}$$

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

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

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

The Bellman equation for the entrepreneur is

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

subject to (10), where $\Omega \equiv 1 - \tau$ and $\mathbb{E}[dV(x_t, t)]/dt$ is shorthand for the Ito calculus terms, that is,

$$\frac{\mathbb{E}[dV(x_t, t)]}{dt} \equiv \mu(e_t) x_t V_x(x_t, t) + \frac{1}{2}\sigma^2 x_t^2 V_{xx}(x_t, t) + V_t(x_t, t).$$

The term $V(x, t)$ is the expected utility of an entrepreneur with quality x and rate of time preference ρ. The flow of the value function depends on the "dividend" of utility from consumption and leisure, the "capital gain" associated with the expected change in the value function, and the possible loss associated with creative destruction, in which case the entrepreneur becomes a worker with expected utility V^w.

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

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

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

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

B. *The Stationary Distribution of Entrepreneurial Income*

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

We have set up the stochastic process for x so that we can apply a well-known result in the literature for generating Pareto distributions.[9] If a variable follows a Brownian motion, like x above, the density of the distribution $f(x, t)$ satisfies a Kolmogorov forward equation:[10]

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

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

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

[9] For more detailed discussion, see Reed (2001), Mitzenmacher (2004), Gabaix (2009), and Luttmer (2010). Malevergne et al. (2013) is closest to the present setup.

[10] This is the stochastic generalization of an equation like (A2) in app. A, related to the simple model at the start of the paper.

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

PROPOSITION 2 (The Pareto income distribution). The stationary distribution of (normalized) entrepreneurial income is given by

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

where

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

and

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

Power law inequality is therefore given by $\eta^* \equiv 1/\xi^*$.

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

Comparative statics.—Taking δ^* as exogenous for the moment, the comparative static results are as follows: power law inequality, η^*, increases if effort is more effective at growing entrepreneurial income (a higher ϕ), decreases if the time endowment is reduced by government policy (a higher τ), decreases if entrepreneurs place more weight on leisure (a higher β), and decreases if either the endogenous or exogenous rates of creative destruction rise (a higher δ^* or $\bar{\delta}$).[11]

The analysis so far shows how one can endogenously obtain a Pareto-shaped income distribution. We have 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.

C. Heterogeneous Mean Growth Rates

As pointed out by Luttmer (2011) and Gabaix et al. (2016), the basic random growth framework that forms the heart of the model so far has trouble explaining features of the data associated with transition dynamics. For example, in the firm dynamics studied by Luttmer (2011), Google and Microsoft become billion-dollar companies seemingly overnight, much

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

faster (and more frequently) than occurs in plausibly calibrated basic random growth models. Gabaix et al. (2016) also note that the speed of convergence to the stationary distribution is very slow in such models, making it hard for those models to match the rapid rise in top income inequality observed in the data.

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

We follow the implementation by Gabaix et al. (2016) and augment our basic setup to include two growth states for entrepreneurs.[12] When researchers discover a new idea, a fraction of them inherit the high-growth ϕ_H parameter. They then face a Poisson process with arrival rate $\bar{p}$ for transitioning permanently down to the more normal ϕ_L low-growth parameter. In addition, we allow the variance of the shocks to also depend on the state, distinguishing σ_H and σ_L. This change is easily introduced and has a straightforward effect on the analysis we have done so far, as shown in the next proposition.

Proposition 3 (Pareto inequality with heterogeneous mean growth rates). Extending the model to include high- and low-growth rates as in Luttmer (2011) and Gabaix et al. (2016), for ϕ_H sufficiently large, the stationary distribution of (normalized) entrepreneurial income has an upper tail with a Pareto inequality exponent $\eta^* \equiv 1/\xi_H$, where

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

and

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

That is, Pareto inequality is determined just as before, only with the key parameters replaced by those in the high-growth case. The addition of ϕ_H allows some entrepreneurs to grow very rapidly, addressing the Google/Microsoft problem. And the speed of convergence to steady state is governed by the Poisson "death rate." Here, the relevant death rate includes $\bar{p}$, the rate at which entrepreneurs "die" out of the high-growth state. We later

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

estimate this rate to be very rapid, thereby substantially speeding up the transition to the stationary distribution.

D. *Production and General Equilibrium*

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

The remainder of the setup is a relatively conventional model of endogenous growth with quality ladders and creative destruction, in the tradition of Grossman and Helpman (1991) and Aghion and Howitt (1992). A fixed population of people choose to be basic laborers, outside researchers (searching for a new idea), or entrepreneurs (who have found an idea and are in the process of improving it).

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

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

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

$$Y_i = \gamma^{n_t} x_i^\alpha L_i. \tag{20}$$

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

The main resource constraint in this environment involves labor:

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

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

Outside researchers discover new ideas through a Poisson process with arrival rate λ per researcher. Research is undirected, and a successful discovery, if implemented, increases the productivity of a randomly chosen variety by a proportion $\gamma > 1$. Once the research is successful, the researcher becomes the entrepreneur of that variety, replacing the old entrepreneur by endogenous creative destruction. In addition, as explained above, the new idea generates spillovers that raise productivity in all other varieties as well. Existing entrepreneurs, however, may use the political process to block new ideas. We model this in a reduced-form way: a fraction $\bar{z}$ of new ideas are successfully blocked from implementation, preserving the monopoly (and productivity) of the existing entrepreneur.

The flow rate of innovation is therefore

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

and this also gives the rate of creative destruction:

$$\delta_t = \dot{n}_t. \tag{23}$$

E. *The Allocation of Resources*

There are 12 key endogenous variables in this economic environment: Y, Y_i, x_i, L_i, L, R, n, δ, e_i, c_i, ℓ_i, and ψ. (Appendix table A1 summarizes the notation used in the paper.) The entrepreneur's choice problem laid out earlier pins down c, ℓ, and e for each entrepreneur. Production functions and resource constraints determine Y, Y_i, L, x_i, n, and δ. This leaves us needing to determine R, L_i, and ψ.

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

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

PROPOSITION 4 (Output, wages, and profits). Let w denote the wage per unit of raw labor, and let π_i denote the profit earned by the entrepreneur selling variety i. Assume now and for the rest of the paper that $\alpha = (1 - \theta)/\theta$.[13] The equilibrium with monopolistic competition leads to

$$Y_t = \gamma^{n_t} X_t^{\alpha} L_t, \tag{24}$$

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

$$w_t = \theta\gamma^{n_t} X_t^{\alpha}, \tag{25}$$

and

$$\pi_{it} = (1 - \theta)\gamma^{n_t} X_t^{\alpha}\left(\frac{x_{it}}{X_t}\right) L_t, \tag{26}$$

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

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

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

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

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

PROPOSITION 5 (Growth and inequality in the $\bar{s}$ case). If the allocation of research is given exogenously by $R/\bar{L} = \bar{s}$ with $0 < \bar{s} < 1$, then along a balanced growth path, the growth of final output per person, g_y, and the rate of creative destruction are given by

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

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

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

F. *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

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

divided into those that affect top income inequality only and those that also affect economic growth.

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

Interestingly, this technological change has no effect on the long-run growth rate of the economy, at least as long as $\bar{s}$ is held fixed. The reason is instructive about how the model works. In the long run, there is a stationary distribution of entrepreneurial productivity x. Some varieties are extraordinarily successful, while most are not. Even though an increase in ϕ_H increases the rate of growth of x, this serves only to widen the stationary distribution, as we showed back in Section IV.B. There is a level effect on overall GDP (working through X), but no growth effect. Long-run growth comes about only through the arrival of new external ideas from outside research, not through the productivity growth associated with improving an existing idea. In light of the endogenous growth literature, it is interesting that the log-linear differential equation inherent in the geometric Brownian motion leads only to level effects in the model rather than to growth effects. The ultimate reason underlying this fact is the "death rate" δ and the decay rate $\bar{p}$ that cause entrepreneurs to exit, generating a stationary distribution of x. This was the logic shown in the toy model back in Section III.

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

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

V. Endogenizing R&D

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

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

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

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

The value function for a researcher at time t is

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

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

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

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

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

Steady-state comparative statics.—Figure 8 shows the effect of various parameters on steady-state growth and inequality when $s \equiv R/\bar{L}$ is endogenously determined. The effects on Pareto inequality are similar to those from the exogenous $\bar{s}$ case. Now, however, we can also study the effects on economic growth. For example, consider the effect of an increase in the technology parameter ϕ_H, shown in figure 8*A*: an increase in ϕ_H raises Pareto inequality, as discussed earlier, but—perhaps surprisingly—causes a decline in the long-run growth rate of GDP per person. Similar results occur throughout figure 8: parameter changes that increase Pareto inequality tend to reduce economic growth.

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

One can, of course, imagine writing down the model in a different way. For example, if research is undertaken by risk-neutral firms, then this effect would not be present. Ultimately, this question must be decided by empirical work. Our model, however, makes it clear that this additional force

TABLE 1
KEY EQUATIONS CHARACTERIZING THE STATIONARY GENERAL EQUILIBRIUM

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

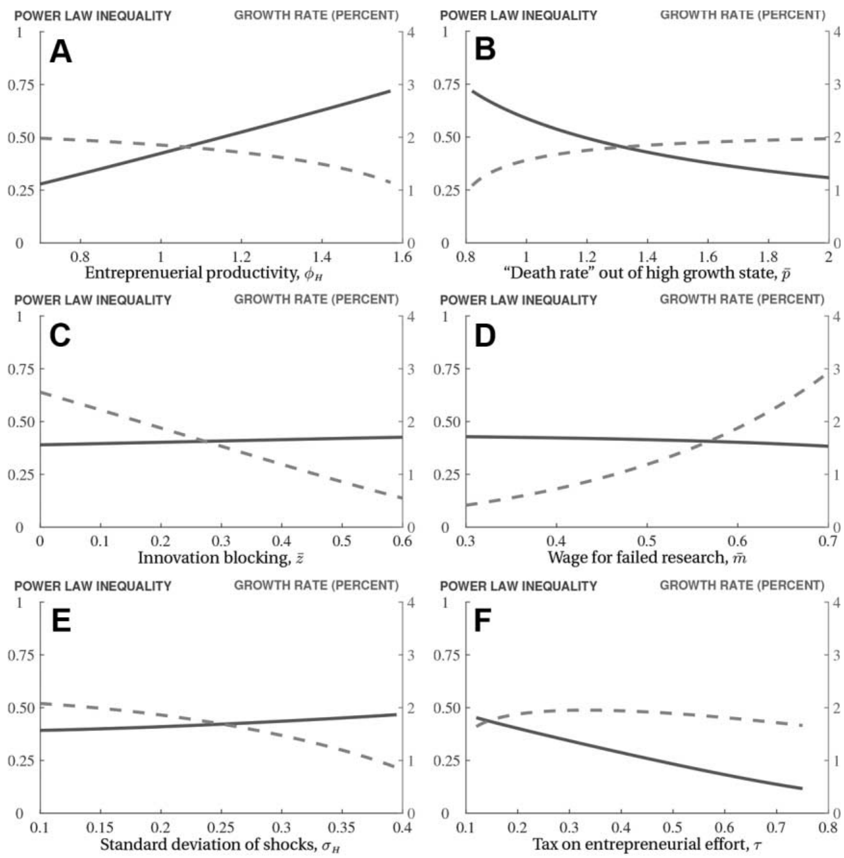

FIG. 8.—Steady-state comparative statics. The figures show the steady-state values of Pareto inequality (solid line) and long-run growth (dashed) when a single parameter changes away from its baseline value. The baseline values are $\rho = .01$, $\theta = 2/3$, $\gamma = 1.4$, $\lambda = .02$, $\phi_H = 0.955$, $\beta = 1$, $\sigma_H = 0.15$, $\bar{\delta} = .08$, $\bar{m} = .6$, $\bar{z} = 0.2$, $\tau = 0.2$, $\bar{L} = 30$, $\bar{p} = 1.5$, and $\bar{q} = .9875$. These values will be discussed in more detail in Section VII. Color version available as an online enhancement.

is present, so that increases in Pareto inequality that result from positive technological changes need not increase the rate of growth.

The model generally features a negative relationship between long-run growth and top income inequality for two reasons. First is the reason just given: higher inequality tends to reduce growth by making research riskier. The second completes the cycle of feedback: faster growth leads to more creative destruction, which lowers inequality. Along a transition path, however, this negative effect on long-run growth is temporarily offset by a positive level effect (e.g., associated with the improved technology). Section VII below shows that this effect can be large, even for periods as long as 20 or 30 years.

VI. Micro Evidence

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

With unlimited access to microdata, our model makes some clear predictions that could be tested. In particular, one could estimate the stochastic process for incomes around the top of the income distribution. In addition to the geometric random walk with heterogeneous drifts, one could estimate the creative destruction parameters: to what extent do high income earners see their incomes drop by a large amount in a short time? Guvenen, Ozkan, and Song (2014) provide evidence for precisely this effect, stating that "individuals in higher earnings percentiles face persistent shocks that are more negatively skewed than those faced by individuals that are ranked lower, consistent with the idea that the higher an individual's earnings are, the more room he has to fall" (20).

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

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

A. *The Distribution of Top Income Growth Rates*

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

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

B. *Empirical Results Based on SSA Data*

Our data and estimation are discussed in more detail in appendix B. In brief, the parameters are estimated from the distribution of growth rates for top earners, as suggested in a stylized way in figure 9. The upper tail of the growth rate distribution is used to estimate $\tilde{\mu}_H$, σ_H, and $\bar{p}$. We estimate $\tilde{\mu}_H$ as the median of growth rates above the 95th percentile of the growth rate distribution, that is, as the growth rate at the 97.5th percentile. We es-

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

[16] Their evidence suggests very thick Pareto-like tails for the growth rate distribution, a fact that our simple model cannot match. Our model predicts that growth rates would exhibit two overlapping normal distributions, together with a thick left tail associated with $\delta + \bar{\delta}$ exits. We suspect that a model with more states for the heterogeneous growth rates could do a better job of matching these data.

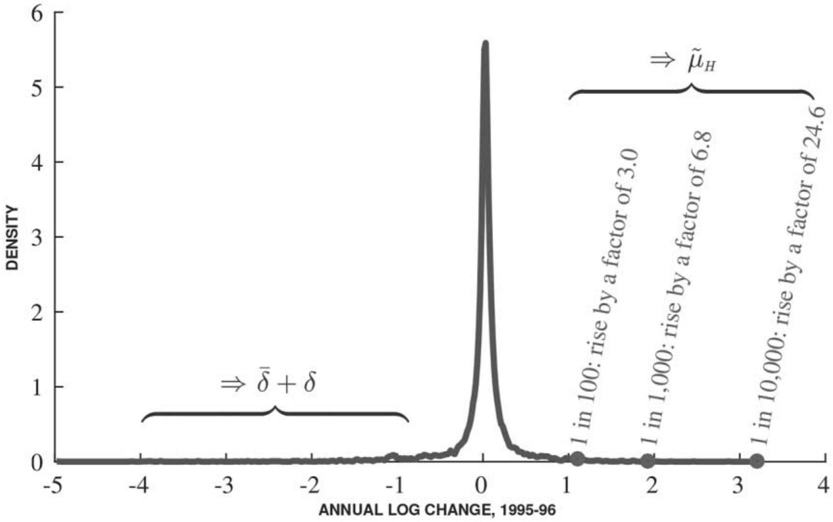

FIG. 9.—The distribution of top income growth rates, 1995–96 (SSA data). This figure shows the density of annual log income changes in 1995–96 for the 90th percentile of top earners aged 45–50, obtained from the spreadsheet data appendix of Guvenen et al. (2016) for wage and salary income. Color version available as an online enhancement.

timate σ_H as the standard deviation of growth rates above the 95th percentile of the growth rate distribution.[17] If top incomes fall by more than 40 percent, we consider this a destruction event. The fraction of growth rates below this cutoff is an estimate of δ^e, which is estimated to be around 13 percent in the SSA data.

To estimate $\bar{p}$, we again use moments provided by Guvenen et al. (2016). For each year and each percentile of the income distribution and for each percentile of the change in log earnings in year t, they report the fraction of people in a cell who have a growth rate above the 95th percentile of the growth rate distribution in year $t+1$. Call this "ProbStayHigh." In our model, this probability equals $e^{-(\bar{p}+\delta^e)}$, which is used to recover an estimate of $\bar{p}$ in each year. Empirically, ProbStayHigh is relatively small: only around 5 percent of top earners with growth rates above the 95th percentile in year t continue in that high-growth state in year $t+1$. This yields estimates of $\bar{p}$ that average about 2.8. From this value, one can see how

[17] Motivated by the empirical income dynamics literature, we make an adjustment for the presence of temporary income shocks, which are absent from our theory. Calculations from Blundell, Pistaferri, and Preston (2008) and Heathcote, Perri, and Violante (2010) suggest that the variance of the random walk innovation accounts for only about one-sixth to one-third of the variance of income growth rates. It is unclear how this applies to top incomes. Hence we make the following correction: we calculate σ^2, the variance of the random walk innovation, to be one-fourth of the variance calculated from the highest decile of the growth rate distribution. Because our estimate of η is relatively insensitive to σ, this adjustment does not play a significant role.

transition dynamics are much faster when based on $\bar{p}$ than when involving only $\bar{\delta}$, a key point made by Gabaix et al. (2016). In fact, the implied half-life for the mean of high-growth incomes is less than half a year with this value.[18]

The heavy solid black line in figure 10 shows the implied steady-state measure of Pareto inequality, based on (smoothed) values of the parameter estimates in each year using the formula in equation (18).[19] The thinner colored lines in the graph show how the steady-state η would evolve if only one parameter changed, with the others held constant at their 1981 values. Of course this ignores transition dynamics, which we turn to in Section VII.

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

Next, the colored lines in figure 10 provide a decomposition of the overall movements in η. For example, the initial rise in η is driven by increases in $\tilde{\mu}_H$ in the 1980s and changes in $\bar{p}$ after that: the largest increases in η come from an increase in ProbStayHigh: the high-growth state appears to have gotten more persistent in these data (i.e., a decline in the decay rate $\bar{p}$). Interestingly, there is little evidence in the SSA wage and salary data for a sustained increase in $\tilde{\mu}_H$. Also, changes in the death rate δ or in the idiosyncratic variance σ_H play a much smaller role, according to this decomposition. For example, reducing σ_H all the way to zero at the 2000 parameter values lowers η^* only from 0.536 to 0.500. Luck matters in our calibration, but it is luck in the form of becoming and remaining a high-growth entrepreneur that is most crucial.

C. *Empirical Results Based on IRS Data*

Similar calculations are possible using the IRS public use microdata, as explained in more detail in appendix C. Unfortunately, the sample sizes are smaller (we can follow between 200 and 1,100 earners in the top 5 percent across a given 2-year period; hence we focus on the top 5 percent in the IRS data rather than the top 1 percent, which was possible in the SSA

[18] Using eq. (29) from Gabaix et al. (2016), the speed of convergence for incomes in the high-growth state is $\lambda_H(-1) = -\tilde{\mu}_H - ½ \cdot \sigma_H^2 + \bar{p} + \bar{\delta}$, and the half-life is ln(2) divided by this value.

[19] We smooth the parameter estimates using a Hodrick-Prescott filter with the smoothing parameter equal to 100, using the data from 1981 until 2006 to avoid the financial crisis having an undue influence on the smoothed values.

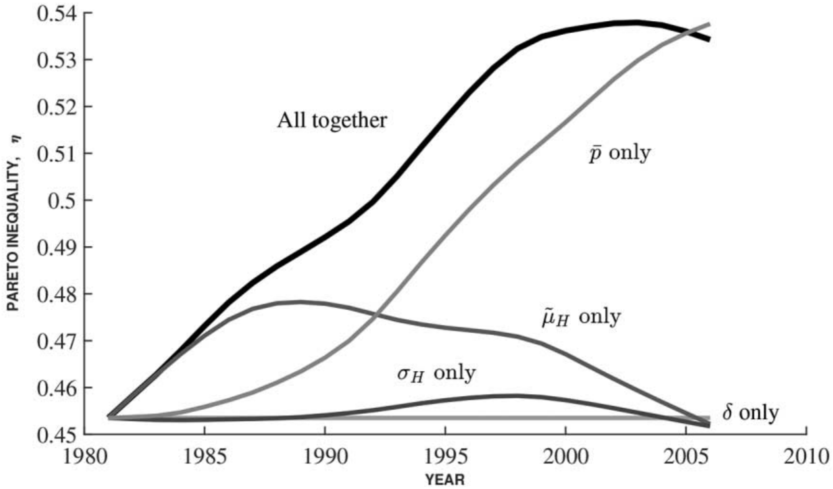

FIG. 10.—Decomposing Pareto inequality: SSA data. Estimates are based on the wage and salary data obtained from Guvenen et al. (2016). Color version available as an online enhancement.

data) and the time frame is shorter (1979–90). But these data have the advantage that we can see beyond wage and salary income to include business and entrepreneurial income. We follow Piketty and Saez (2003) and define entrepreneurial income to be the sum of income from Schedule C, partnerships, S corporations, and farm income.[20]

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

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

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

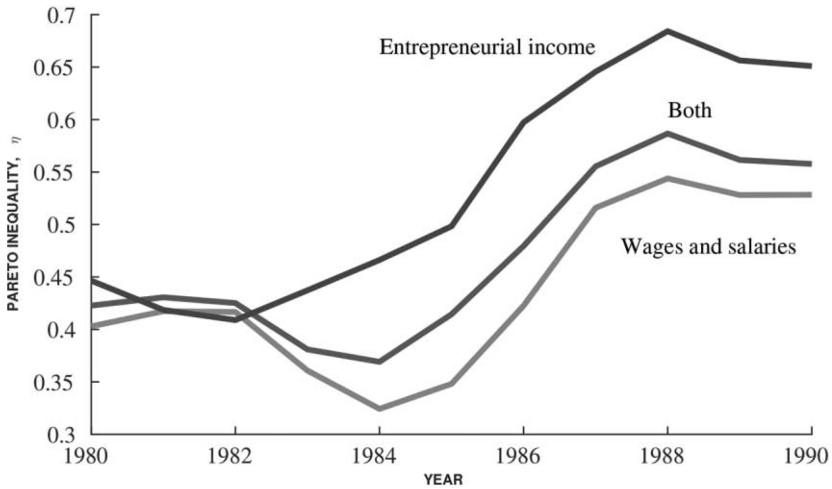

FIG. 11.—Pareto inequality: IRS data. Steady-state Pareto inequality is calculated from equation (18) using the IRS public use microdata, as explained in the text and in appendix C. Color version available as an online enhancement.

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

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

VII. Transition Dynamics

We now explore the role of transition dynamics in this Schumpeterian model. This is important for two reasons. First, Gabaix et al. (2016) suggest that transition dynamics in models like this one can be very slow, mak-

[22] The timing is surprisingly different between figs. 10 and 11. The underlying moments of the data look somewhat different, partly because the larger sample size in the SSA data lets us look higher up in the distribution and potentially in part because of the sampling error in the IRS data.

ing the steady-state calculations of the preceding section potentially misleading. Second, "level effects" in GDP per person may be important, as opposed to just the long-run growth rate. The first set of examples below explores one-time shocks to ϕ_H, $\bar{p}$, and τ, while the second set feeds in the shocks recovered from the IRS and SSA data in the previous section. We trace through the transition dynamics associated with the Kolmogorov forward equation to study how top inequality and GDP per person evolve in response to these shocks.

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

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

Finally, we solve for only an approximation to the true transition dynamics of our model. It is notoriously difficult to solve for transition dynamics in heterogeneous agent models, in part because of the large state space they imply.[23] The simplification we use to make our problem computationally tractable is to assume that control variables (i.e., the effort choice by entrepreneurs and the research allocation) jump immediately to their steady-state values, while state variables evolve according to their laws of motion. The transition dynamics then come only from the Kolmogorov equation as the distribution of x evolves slowly over time. Importantly, this allows us to study transition dynamics in GDP per person as well, since the mean of the x distribution will also evolve slowly. These transition dynamics turn out to be quite important.

A. *One-Time Shocks to ϕ_H, $\bar{p}$, and τ*

Our first set of examples consider one-time shocks to ϕ_H, $\bar{p}$, and τ. We start with a set of baseline parameters that match US Pareto inequality in 1975.

[23] To be more precise, the Hamilton-Jacobi-Bellman and Kolmogorov forward equations (KFE) are coupled, with one running forward in time and one running backward. A full solution requires solving for a fixed point in the time path of two general equilibrium objects (the "wage" of the entrepreneurs ψ_t and the rate of creative destruction), which is at the frontier of existing methods. We are grateful to Ben Moll for advice on the solution technique we undertake.

Where possible, these are chosen to be consistent with the empirical estimates of the previous section. For example, we assume that σ_H for the United States is constant and equal to 0.15, broadly consistent with the evidence in Section VI. We assume $\bar{\delta} = 0.08$ and $\gamma = 1.4$, so that $\delta \approx 0.06$ when the economy's growth rate is 2 percent, and therefore $\bar{\delta} + \delta \approx 0.14$, similar to what we estimated using the SSA data in the previous section.[24]

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

Several findings stand out. First, consider Pareto inequality. One-time shocks to all three parameter values can generate rising Pareto inequality that roughly matches the US experience. Some of the subtleties are interesting. For example, while the shock occurs in 1975, Pareto inequality does not start rising immediately and in fact declines at first. The reason is that top incomes take time to accumulate, and "slightly less than top" incomes accumulate first. In addition, the transition dynamics are remarkably slow, even given the very high decay rates out of the high-growth state. As explained by Gabaix et al. (2016), the reason is that convergence rates at the top are much slower than convergence rates for the mean.

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

This exercise suggests that something like these shocks—or perhaps some combination of them or even a sequence of such shocks—can poten-

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

[25] To solve the KFE, we discretize the state space into 2,000 grid points and consider a time interval of $dt = 1/25$ periods (about 2 weeks). We then apply the finite difference method described in Moll (2016) and used by Gabaix et al. (2016), e.g., in their "fig4b.m" and "fig5b.m" programs. In particular, the discretized KFE becomes a (sparse) Markov transition problem once we discretize the state space, and this equation can be solved in Matlab with standard techniques. For the second case in fig. 13, where we consider the sequence of shocks recovered from the IRS and SSA data, the transition matrix changes over time, but otherwise the same solution method applies.

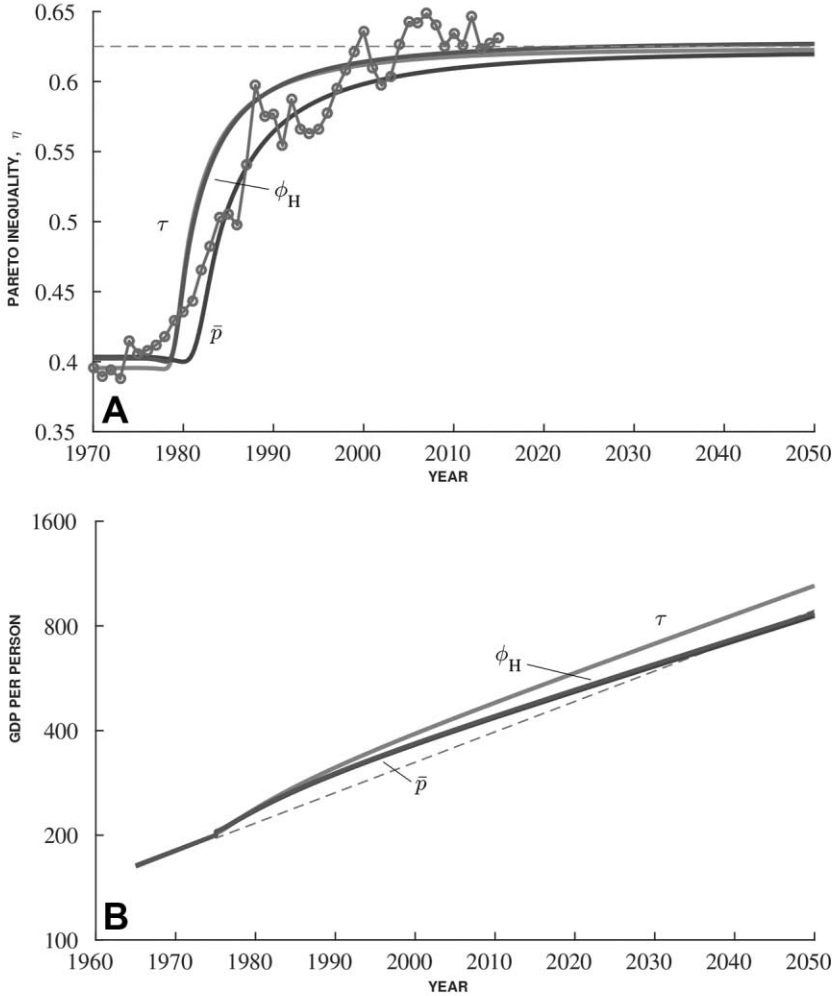

FIG. 12.—One-time shocks to ϕ_H, $\bar{p}$, and τ. *A*, Pareto inequality, η. *B*, GDP per person, log scale. The figures show the US data on top inequality (the circles) as well as the dynamic response of Pareto inequality and GDP per person to a one-time shock to ϕ_H, $\bar{p}$, and τ that occurs in 1975, using the approximation method described in the text. The shocks are sufficient to move Pareto inequality from an initial steady-state value of 0.39 to a new steady-state value of 0.63. All other parameters are held constant. Color version available as an online enhancement.

tially explain the patterns that we see in top income inequality across countries.[26]

[26] The exercise also confirms that the "Luttmer rocket" with high decay rates can successfully address the problems in simple random growth models of slow transition dynamics and "very old" people at the top of the distribution. For example, in the steady state of these simulations, the average length of time that someone in the top 1 percent has been

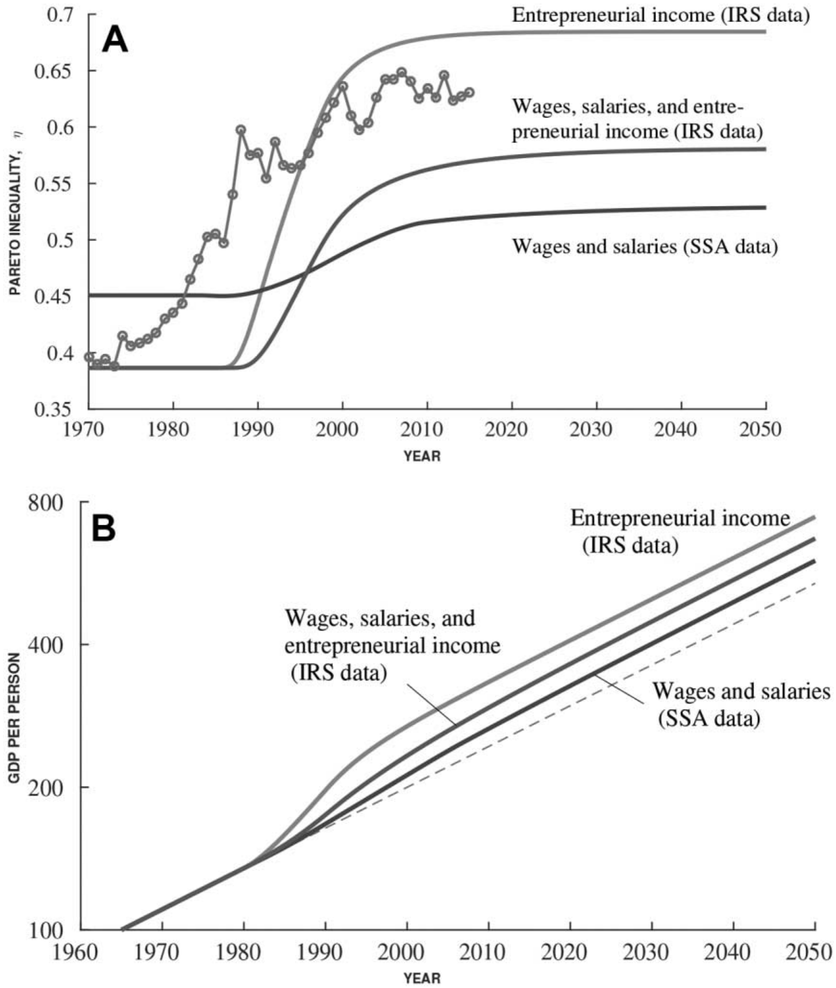

FIG. 13.—The dynamic response to IRS/SSA-inspired shocks. *A*, Pareto inequality, η. *B*, GDP per person, log scale. The figures show the dynamic response of Pareto inequality and GDP per person to the Hodrick-Prescott-filtered shocks to $\tilde{\mu}_H$, $\bar{p}$, $\bar{\delta}$, and σ_H estimated in Section VI, using the approximation method described in the text. The shocks start in and continue until 2006 for the SSA data and until 1990 for the other two cases, which use the IRS data. For the IRS cases, we hold $\bar{p}$ constant at the value that delivers an initial value of η of 0.39. All other parameters are held constant. Color version available as an online enhancement.

B. Transition Dynamics and the IRS/SSA-Inspired Shocks

An alternative set of examples is shown in figure 13. Here, we take the (smoothed) time paths of $\tilde{\mu}_H$, $\bar{p}$, $\bar{\delta}$, and σ_H that we observe in the IRS and

an entrepreneur is around 20 years, and the average for someone in the top 1 percent who is also in the high-growth state is 3.7 years.

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

The results for GDP per person confirm what we saw earlier: there are substantial "level effects" that emerge over 25 years or more that can easily mask any long-run growth effect during that time.

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

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

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

VIII. Conclusion

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

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

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

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

Appendix

TABLE A1
Guide to Notation for the Full Model

η	Pareto inequality measure (inverse of Pareto exponent)
x	Idiosyncratic productivity of an entrepreneur's variety
$f(x, t)$	Distribution of idiosyncratic productivity across entrepreneurs
e	Entrepreneur's effort
$\tilde{\mu}$	$\tilde{\mu} \equiv \phi e - \frac{1}{2}\sigma^2$, drift of log x
ℓ	Entrepreneur's leisure
τ	"Tax" on the entrepreneur's time endowment
ψ	Entrepreneur's "wage" per unit of x
ϕ	Technology parameter: how effort translates into growth of x
σ	Variance of the idiosyncratic shocks to x
Ω	$\Omega \equiv 1 - \tau$
β	Weight on (log) leisure in utility
ρ	Rate of time preference
δ	Endogenous rate of creative destruction
$\bar{\delta}$	Exogenous destruction of entrepreneurs
$\bar{p}$	Rate at which high-growth entrepreneurs decay to low-growth
$\bar{q}$	Fraction of new entrepreneurs at x_0 who begin in the high-growth state
α	Exponent on x_i in production of variety i
θ	Constant elastiticity of substitution curvature parameter in final goods production
n_t	"Height" up the quality ladder; productivity is γ^{n_t}
γ	Step size for the quality ladder, $\gamma > 1$
$\bar{N}$	Aggregate labor endowment
L_t	Aggregate allocation of labor to goods production
R_t	Aggregate allocation of labor to idea production
$\bar{L}$	$\bar{L} \equiv \bar{N} - 1$; labor endowment net of entrepreneurs
λ	Research productivity
$\bar{z}$	Fraction of innovations that are exogenously blocked
V, V^R, V^w	Expected lifetime utility for entrepreneurs, researchers, and workers
$\bar{m}$	Fraction of equilibrium wage paid to failed researchers
X	Mean of the distribution of idiosyncratic productivity, x
w_t	Wage of labor in producing goods
π_{it}	Flow profit in variety i
s_t	$s_t \equiv R/\bar{L}$; fraction of labor engaged in research
g_y	Growth rate of GDP per person (y)

References

Abowd, John M., and David Card. 1989. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica* 57 (2): 411–45.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, vol. 4, edited by O. Ashenfelter and D. Card, chap. 12. Amsterdam: Elsevier.

Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hemous. 2015. "Innovation and Top Income Inequality." Working Paper no. 21247, NBER, Cambridge, MA.

Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen. 2016. "Living the American Dream in Finland: The Social Mobility of Innovators." Technical report, Harvard Univ.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60 (2): 323–51.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas. 2016. "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age." Manuscript, Univ. Chicago.

Aluation.wordpress.com. 2011. "Fractals, Fractiles and Inequality." Technical report. http://aluation.wordpress.com/2011/08/16/fractals-fractiles-and-inequality/.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *J. Econ. Literature* 49 (1): 3–71.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the U.S. Labor Market." *A.E.R.* 96 (2): 189–94.

Baker, Michael. 1997. "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings." *J. Labor Econ.* 15 (2): 338–75.

Baker, Michael, and Gary Solon. 2003. "Earnings Dynamics and Inequality among Canadian Men, 1976–1992: Evidence from Longitudinal Income Tax Records." *J. Labor Econ.* 21 (2): 267–88.

Bakija, Jon, Adam Cole, and Bradley Heim. 2010. "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data." Working paper, Indiana Univ.

Banerjee, Abhijit V., and Andrew F. Newman. 1993. "Occupational Choice and the Process of Development." *J.P.E.* 101 (2): 274–98.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2016. "The Lifecycle of Inventors." Manuscript, Stanford Univ.

Bell, Brian, and John Van Reenen. 2014. "Bankers and Their Bonuses." *Econ. J.* 124 (574): F1–F21.

Benabou, Roland, and Jean Tirole. 2016. "Bonus Culture: Competitive Pay, Screening, and Multitasking." *J.P.E.* 124 (2): 305–70.

Benhabib, Jess. 2014. "Wealth Distribution Overview." Teaching slides, New York Univ. http://www.econ.nyu.edu/user/benhabib/wealth%20distribution%20theories%20overview3.pdf.

Benhabib, Jess, and Alberto Bisin. 2016. "Skewed Wealth Distributions: Theory and Empirics." Manuscript, New York Univ.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu. 2011. "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents." *Econometrica* 79 (1): 123–57.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." *A.E.R.* 98 (5): 1887–1921.

Cantelli, F. P. 1921. "Sulle applicazioni del calcolo delle probabilita alla fisica molecolare." *Metron* 1 (3): 83–91.

Champernowne, David G. 1953. "A Model of Income Distribution." *Econ. J.* 63 (250): 318–51.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Q.J.E.* 114 (3): 739–67.

———. 2009. "Power Laws in Economics and Finance." *Ann. Rev. Econ.* 1:255–94.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Q.J.E.* 123 (1): 49–100.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality." *Econometrica* 84 (6): 2071–2111.

Galor, Oded, and Joseph Zeira. 1993. "Income Distribution and Macroeconomics." *Rev. Econ. Studies* 60 (1): 35–52.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *J.P.E.* 108 (5): 874–904.

Geerolf, Francois. 2016. "A Theory of Pareto Distributions." Manuscript, Univ. California, Los Angeles.

Gordon, Robert J., and Ian Dew-Becker. 2008. "Controversies about the Rise of American Inequality: A Survey." Working Paper no. 13982, NBER, Cambridge, MA.

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

Guvenen, Fatih. 2007. "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?" *A.E.R.* 97 (3): 687–712.

———. 2009. "An Empirical Investigation of Labor Income Processes." *Rev. Econ. Dynamics* 12 (1): 58–79.

Guvenen, Fatih, Greg Kaplan, and Jae Song. 2014. "The Glass Ceiling and the Paper Floor: Gender Differences among Top Earners, 1981–2012." Working Paper no. 20560, NBER, Cambridge, MA.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2016. "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?" Manuscript, Univ. Minnesota.

Guvenen, Fatih, Serdar Ozkan, and Jae Song. 2014. "The Nature of Countercyclical Income Risk." *J.P.E.* 122 (3): 621–60.

Haskel, Jonathan, Robert Z. Lawrence, Edward E. Leamer, and Matthew J. Slaughter. 2012. "Globalization and U.S. Wages: Modifying Classic Theory to Explain Recent Facts." *J. Econ. Perspectives* 26 (2): 119–40.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967–2006." *Rev. Econ. Dynamics* 13 (1): 15–51.

Hubmer, Joachim, Per Krusell, and Anthony A. Smith. 2016. "The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation." Working Paper no. 23011, NBER, Cambridge, MA.

Kaplan, Steven N., and Joshua Rauh. 2010. "Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?" *Rev. Financial Studies* 23 (3): 1004–50.

Katz, Lawrence F., and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics*, vol. 3, edited by O. Ashenfelter and D. Card, chap. 26. Amsterdam: Elsevier.

Koenig, Michael, Jan Lorenz, and Fabrizio Zilibotti. 2016. "Innovation vs. Imitation and the Evolution of Productivity Distributions." *Theoretical Econ.* 11 (3): 1053–1102.

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

Levy, Moshe. 2003. "Are Rich People Smarter?" *J. Econ. Theory* 110 (1): 42–64.

Lillard, Lee A., and Robert J. Willis. 1978. "Dynamic Aspects of Earning Mobility." *Econometrica* 46 (5): 985–1012.

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

Lucas, Robert E., and Benjamin Moll. 2014. "Knowledge Growth and the Allocation of Time." *J.P.E.* 122 (1): 1–51.

Luttmer, Erzo G. J. 2007. "Selection, Growth, and the Size Distribution of Firms." *Q.J.E.* 122 (3): 1103–44.

———. 2010. "Models of Growth and Firm Heterogeneity." *Ann. Rev. Econ.* 2:547–76.

———. 2011. "On the Mechanics of Firm Growth." *Rev. Econ. Studies* 78 (3): 1042–68.

———. 2014. “An Assignment Model of Knowledge Diffusion and Income Inequality.” Working Paper no. 715, Fed. Reserve Bank Minneapolis.
———. 2016. “Further Notes on Micro Heterogeneity and Macro Slow Convergence.” Manuscript, Univ. Minnesota.
MaCurdy, Thomas E. 1982. “The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis.” *J. Econometrics* 18 (1): 83–114.
Malevergne, Y., A. Saichev, and D. Sornette. 2013. “Zipf’s Law and Maximum Sustainable Growth.” *J. Econ. Dynamics and Control* 37 (6): 1195–1212.
Mandelbrot, Benoit. 1960. “The Pareto-Levy Law and the Distribution of Income.” *Internat. Econ. Rev.* 1 (2): 79–106.
Meghir, Costas, and Luigi Pistaferri. 2004. “Income Variance Dynamics and Heterogeneity.” *Econometrica* 72 (1): 1–32.
———. 2011. “Earnings, Consumption and Life Cycle Choices.” In *Handbook of Labor Economics*, vol. 4, edited by O. Ashenfelter and D. Card, chap. 9. Amsterdam: Elsevier.
Mitzenmacher, Michael. 2004. “A Brief History of Generative Models for Power Law and Lognormal Distributions.” *Internet Math.* 1 (2).
Moll, Benjamin. 2012. “Inequality and Financial Development: A Power-Law Kuznets Curve.” Working paper, Princeton Univ.
———. 2016. “Lectures 6 and 7: Theories of Top Inequality.” Teaching slides, Princeton Univ. http://www.princeton.edu/~moll/ECO521 2016/Lectures6_7_ECO521.pdf.
Nirei, Makoto. 2009. “Pareto Distributions in Economic Growth Models.” Working Paper no. 09-05, Inst. Innovation Res., Hitotsubashi Univ.
Nirei, Makoto, and Shuhei Aoki. 2016. “Pareto Distribution of Income in Neoclassical Growth Models.” *Rev. Econ. Dynamics* 20:25–42.
Pareto, Vilfredo. 1896. *Cours d’economie politique*. Geneva: Droz.
Perla, Jesse, and Christopher Tonetti. 2014. “Equilibrium Imitation and Growth.” *J.P.E.* 122 (1): 52–76.
Philippon, Thomas, and Ariell Reshef. 2012. “Wages and Human Capital in the U.S. Finance Industry: 1909–2006.” *Q.J.E.* 127 (4): 1551–1609.
Piketty, Thomas, and Emmanuel Saez. 2003. “Income Inequality in the United States, 1913–1998.” *Q.J.E.* 118 (1): 1–39.
———. 2013. “A Theory of Optimal Inheritance Taxation.” *Econometrica* 81 (5): 1851–86.
Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva. 2014. “Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities.” *American Econ. J.: Econ. Policy* 6 (1): 230–71.
Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2016. “Distributional National Accounts: Methods and Estimates for the United States.” Working Paper no. 22945, NBER, Cambridge, MA.
Piketty, Thomas, and Gabriel Zucman. 2015. “Wealth and Inheritance in the Long Run.” In *Handbook of Income Distribution*, vol. 2B, edited by Anthony B. Atkinson and François Bourguignon, chap. 15. Amsterdam: North-Holland.
Reed, William J. 2001. “The Pareto, Zipf and Other Power Laws.” *Econ. Letters* 74 (1): 15–19.
Rosen, Sherwin. 1981. “The Economics of Superstars.” *A.E.R.* 71 (5): 845–58.
Rothschild, Casey, and Florian Scheuer. 2016. “Optimal Taxation with Rent-Seeking.” *Rev. Econ. Studies* 83 (3): 1225–62.
Saez, Emmanuel. 2001. “Using Elasticities to Derive Optimal Tax Rates.” *Rev. Econ. Studies* 68:205–29.

Simon, Herbert A. 1955. "On a Class of Skew Distribution Functions." *Biometrika* 42 (3/4): 425–40.

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

Toda, Alexis Akira. 2014. "Incomplete Market Dynamics and Cross-Sectional Distributions." *J. Econ. Theory* 154:310–48.

Topel, Robert H., and Michael P. Ward. 1992. "Job Mobility and the Careers of Young Men." *Q.J.E.* 107 (2): 439–79.