

# The Dynamics of Inequality\*

Xavier Gabaix, Jean-Michel Lasry, Pierre-Louis Lions, Benjamin Moll

June 25, 2015

## Abstract

The past forty years have seen a rapid rise in top income inequality in the United States. While there is a large number of existing theories of the Pareto tails of the income and wealth distributions at a given point in time, almost none of these address the fast rise in top inequality observed in the data. We show that standard theories, which build on a random growth mechanism, generate transition dynamics that are an order of magnitude too slow relative to those observed in the data. We then suggest parsimonious deviations from the basic model that can explain such changes, namely heterogeneity in mean growth rates or deviations from Gibrat's law. These deviations are consistent with theories in which the increase in top income inequality is driven by the rise of "superstar" entrepreneurs or managers.

---

\*NYU Stern, Dauphine, Collège de France, Princeton. We thank Fernando Alvarez, Roland Bénabou, Fatih Guvenen, Chad Jones, Greg Kaplan, Erzo Luttmer, Makoto Nirei, Ezra Oberfield, Jonathan Parker, John Shea, Joe Sullivan and Gabriel Zucman for their insights, and seminar participants at UQAM, Queen's University, Cornell, the Chicago Fed, Princeton and the University of Maryland for useful comments. We also thank Cristian Alonso, Joshua Bernstein, Nik Engbom and Jason Ravit for excellent research assistance.

# 1 Introduction

The past forty years have seen a rapid rise in top income inequality in the United States (Piketty and Saez, 2003; Atkinson, Piketty and Saez, 2011). Since Pareto (1896), it is well-known that the upper tail of the income distribution follows a power law, or equivalently, that top inequality is “fractal”, and the rise in top income inequality has coincided with a “fattening” of the right tail of the income distribution. That is, the “super rich” have pulled ahead *relative* to the rich. This rise in top inequality requires an understanding of the forces that have led to a fatter Pareto tail. There is also an ongoing debate about the dynamics of top wealth inequality.<sup>1</sup> To the extent that wealth inequality has also increased, we similarly need to understand the dynamics of its Pareto tail.

What explains the observed rapid rise in top inequality is an open question. While there is a large number of existing theories of the Pareto tails of the income and wealth distributions at a point in time, almost none of these address the fast rise in top inequality observed in the data, or any fast change for that matter.

The main contributions of this paper are: first, to show that the most common framework (a simple Gibrat’s law for income or wealth dynamics) cannot explain rapid changes in tail inequality, and second, to suggest parsimonious deviations from the basic model that can explain such changes. Our analytical results bear on a large class of economic theories of top inequality, so that our results shed light on the ultimate drivers of the rise in top inequality observed in the data.

The first result of our paper is negative: standard random growth models, like those considered in much of the existing literature, feature extremely slow transition dynamics and cannot explain the rapid changes that arise empirically. To address this issue, we consider the following thought experiment: initially at time zero, the economy is in a steady state with a stationary distribution that has a Pareto tail. At time zero, there is a change in the underlying economic environment that leads to higher top inequality in the long-run, say because capital income taxes fall. The question is: what can we say about the speed of this transition? Will this increase in inequality come about quickly or take a long time? We present two answers to this question. First, we derive an analytic formula for a measure of the “average” speed of convergence throughout the distribution. We argue that, when calibrated to be consistent with microeconomic evidence, the implied half life is an order of magnitude too high to explain the observed rapid rise in top income inequality. It is also too

---

<sup>1</sup>See e.g. Piketty (2014), Saez and Zucman (2014), Bricker et al. (2015) and Kopczuk (2015).

high to explain even the relatively gradual rise in top wealth inequality suggested by some empirical analyses. Second, we derive a measure of the speed of convergence for the part of the distribution we are most interested in, namely its upper tail. We argue that, in standard theories, transitions are even slower in the tail and, additionally, that our low measure of the average speed of convergence overestimates the speed of convergence in the upper tail. We also show that allowing for jumps in the income or wealth process, while useful for descriptively matching micro-level data, does not help with generating fast transitions.

Given this negative result, we are confronted with a puzzle: what, then, explains the observed rise in top income and (potentially) wealth inequality? We develop a “generalized random growth model” that features two parsimonious departures from the canonical model that do generate fast transitions. The first departure is cross-sectional heterogeneity in mean growth rates and, in particular, a “high growth regime”.<sup>2</sup> For instance, some highly skilled entrepreneurs or managers may experience much higher average earnings growth rates than other individuals over short- to medium horizons. We argue analytically and quantitatively that this first departure can explain the observed fast rise in income inequality. The second departure consists of deviations from Gibrat’s law (the assumption that the distribution of income growth rates is independent of their levels), which we also refer to as “superstar shocks”. These are shocks that disproportionately affect high incomes, e.g. shocks to the shape production function of talent (e.g. shocks to the “span of control” as in Garicano and Rossi-Hansberg (2006)).<sup>3</sup> This second departure generates infinitely fast transitions in inequality.

To obtain our analytic formulas for the speed of convergence, we employ tools from ergodic theory and the theory of partial differential equations. Our measure of the average speed of convergence is the first non-trivial eigenvalue or “spectral gap” of the differential operator governing the stochastic process for income or wealth. One of the main contributions of this paper is to derive an analytic formula for this first non-trivial eigenvalue

---

<sup>2</sup>Güvener (2007) argues that heterogeneity in mean growth rates is an important feature of the data on income dynamics. Luttmer (2011) studies a similar framework applied to firm dynamics and argues that persistent heterogeneity in mean firm growth rates is needed to account for the relatively young age of very large firms at a given point in time (a statement about the stationary distribution rather than transition dynamics as in our paper).

<sup>3</sup>Technically, the shocks to log income affect it multiplicatively, rather than additively, as in the usual random growth model. We use the term “deviations from Gibrat’s law” only to refer to these multiplicative shocks and not to the model with heterogeneous growth rates. Technically, also the latter constitutes such a deviation, although a local version of Gibrat’s law still holds.

(i.e. the second eigenvalue) for a large variety of random growth processes. See Hansen and Scheinkman (2009) for a related application of operator methods in economics. We obtain our measure of the speed of convergence in the tail of the distribution by making use of the fact that the solution to the Kolmogorov Forward equation for random growth processes can be characterized tightly by calculating the Laplace transform of this equation. The use of Laplace transforms to characterize the transition dynamics of a distribution is another methodological contribution of our paper.

A large theoretical literature builds on random growth processes to theorize about the upper tails of income and wealth distributions. Early theories of the income distribution include Champernowne (1953) and Simon (1955), with more recent contributions by Nirei (2009), Toda (2012), Kim (2013), Jones and Kim (2014) and Luttmer (2015). Similarly, random growth theories of the wealth distribution include Wold and Whittle (1957) and more recently Benhabib, Bisin and Zhu (2011, 2013, 2014), Piketty and Zucman (2014*b*), Jones (2015), and Acemoglu and Robinson (2015). All of these papers focus on the income or wealth distribution *at a given point in time* by studying stationary distributions, and none of them analyze transition dynamics. Aoki and Nirei (2015) are a notable exception, who examine the dynamics of the income distribution and ask whether tax changes can account for the rise in top income inequality observed in the United States. Our paper differs from theirs in that we obtain a number of analytic results providing a tight characterization of transition dynamics in random growth models whereas their analysis of transition dynamics is purely numerical.<sup>4</sup>

Our finding that heterogeneity in mean growth rates delivers fast dynamics of top inequality is also related to Guvenen (2007), who has argued that an income process with heterogeneous income profiles provides a better fit to the micro data than a model in which all individuals face the same income profile. In our model variant with multiple growth regimes, we also allow for heterogeneity in the standard deviation of income innovations in different regimes which is akin to the mixture specification advocated by Guvenen et al. (2015). One key difference between our model with multiple growth regimes and the standard random growth model is that, in the standard model, the key determinant for an individual's place in the income distribution is her age. In contrast, in a model with multiple growth regimes

---

<sup>4</sup>Also related is Luttmer (2012) who studies the speed of convergence of aggregates like GDP in response to shocks in an economy with a power law firm size distribution. His analysis differs from ours in that it studies the convergence of aggregates which, in his framework, can be characterized independently of the cross-sectional distribution.

another important determinant is the individual’s “growth type” which may represent her occupation or her talent as an entrepreneur. This is consistent with salient patterns of the tail of the income distribution in the United States (Güvener, Kaplan and Song, 2014).<sup>5</sup>

One of the most ubiquitous regularities in economics and finance is that the empirical distribution of many variables is well approximated by a power law. For this reason, theories of random growth are an integral part of many different strands of the literature beside those studying the distributions of income and wealth. For example, they have been used to study the distribution of city sizes (Gabaix, 1999) and firm sizes (Luttmer, 2007), the shape of the production function (Jones, 2005), and in many other contexts (see the review by Gabaix, 2009). The tools and results presented in this paper should therefore also prove useful in other applications.

The paper is organized as follows. Section 2 states the main motivating facts for our analysis, and Section 3 reviews random growth theories of the income and wealth distributions at a point in time. In Section 4, we present our main negative results on the slow transitions generated by such models and we explore their empirical implications for the dynamics of income inequality. Section 5 presents two theoretical mechanisms for generating fast transitions, and shows that these have the potential to account for the fast transitions observed in the data. Section 6 extends our results to the case of wealth inequality, and Section 7 concludes.

## 2 Motivating Facts

In this section, we briefly review some facts regarding the evolution of top income and wealth inequality in the United States. We return to these in Sections 4 and 5 when comparing various random growth models and their ability to generate the trends observed in the data.

Panel (a) of Figure 1 displays the evolution of measures of the top 1% income share. Panel (a) shows the large and rapid increase in the top 1% income share that has been extensively documented by Piketty and Saez (2003), Atkinson, Piketty and Saez (2011) and others.<sup>6</sup> As already noted, the upper tails of the income and wealth distributions follow power laws, or equivalently top inequality is fractal in nature. For an exact power law, the top 0.1% are

---

<sup>5</sup>Luttmer (2011) makes a similar observation about the relationship between a firm’s age and its place in the firm size distribution. As Luttmer puts it succinctly: “Gibrat implies 750-year-old firms.”

<sup>6</sup>The series is from the “World Top Incomes Database.” Here, we plot total income (salaries plus business income plus capital income) excluding capital gains. The series display a similar trend when we include capital gains or focus on salaries only (though the levels are different).

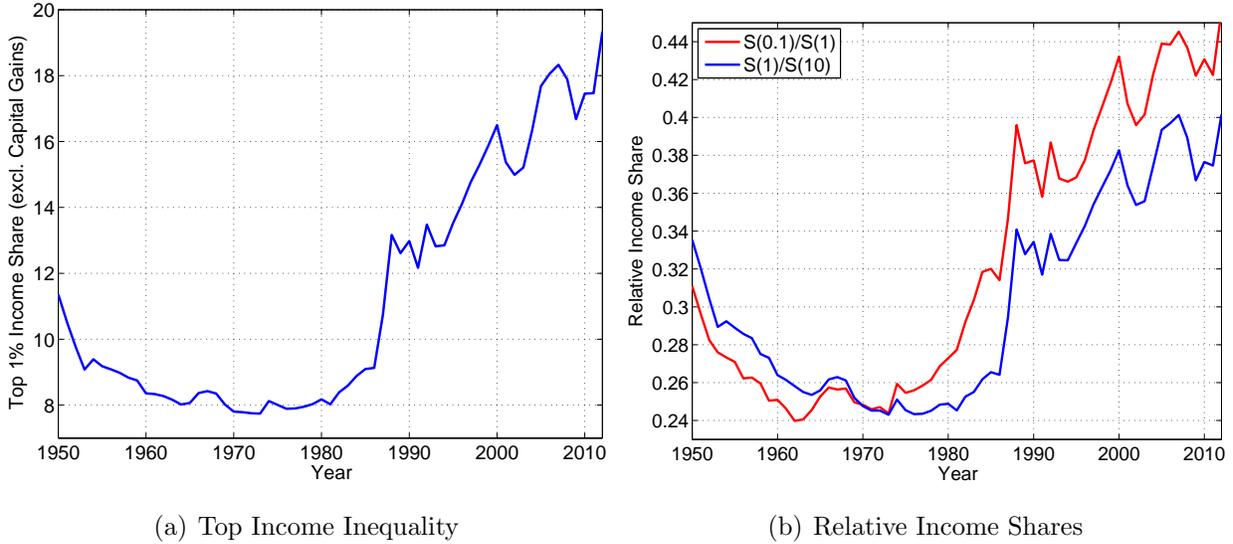


Figure 1: Evolution of Top 1% Income Share and “Fractal Inequality” in U.S.

$X$  times richer on average than the top 1% who are, in turn,  $X$  times richer than the top 10%, where  $X$  is a fixed number. Equivalently, the top 0.1% income share is a fraction  $Y$  of the top 1% income share, which, in turn, is a fraction  $Y$  of the top 10% income share, and so on. We now explore this fractal pattern in the data using a strategy borrowed from Jones and Kim (2014). Panel (b) of Figure 1 plots the income share of the top 0.1% relative to that of the top 1% and the income share of the top 1% relative to that of the top 10%. As expected, the two lines track each other relatively closely. More importantly, there is an upward trend in both lines. That is, there has been a *relative* increase in top income shares. As we explain in more detail below, this increase in “fractal inequality” implies equivalently a “fattening” of the Pareto tail of the income distribution.

In Appendix B, we present an analogous exercise for top wealth inequality. As discussed there in more detail, the facts about the evolution of top wealth inequality are more ambiguous than those for top income inequality. Different data series suggest different conclusions, and top wealth shares appear to have increased, although it is unclear by how much.

There are two main takeaways from this section. First, top income shares have increased dramatically since the late 1970s. Second, the Pareto tail of the income distribution has gotten fatter over time.

### 3 Random Growth Theories of Income and Wealth Inequality

Our starting point is the existing theories that can explain top inequality at a point in time, meaning that they can generate stationary distributions that have Pareto tails. Many of these share the same basic mechanism for generating power laws, namely proportional random growth. In this section, we give a brief overview of such theories. We start with two concrete examples corresponding to our two main applications: a simple model of income dynamics and a simple model of wealth accumulation. We then present a unifying framework that nests these two examples as special cases and that will be the focus of our analysis of transition dynamics in the next section.

#### 3.1 Income Dynamics

Time is continuous, and there is a continuum of workers. A worker's wage is given by  $w_{it} = \omega h_{it}$  where  $\omega$  is an exogenous skill price and  $h_{it}$  is her human capital or her skills. Workers die (retire) at rate  $\delta$ , in which case they are replaced by a young worker with human capital  $h_{i0}$ . A worker's human capital evolves as

$$dh_{it} = g(I_{it}, h_{it})dz_{it}$$

where  $dz_{it}$  denotes an ability shock and  $I_{it}$  investment into human capital. We make the following four assumptions. First, ability shocks are i.i.d. over time and given by  $dz_{it} = \bar{z}dt + \tilde{\sigma}dZ_{it}$  where  $Z_{it}$  is a standard Brownian motion.<sup>7</sup> Second, the function  $g$  has constant returns. Third, an individual's investment is proportional to her human capital  $I_{it} = \theta h_{it}$ .<sup>8</sup> And fourth, initial human capital is the same for everyone  $h_{i0} = \bar{h}_0$ . Given these assumptions, it is easy to show that the resulting wage dynamics are given by

$$dw_{it} = \bar{z}g(\theta, 1)w_{it}dt + \tilde{\sigma}g(\theta, 1)w_{it}dZ_{it}. \tag{1}$$

A large number of models of the upper tail of the income distribution end up with a similar reduced form. In some of these models, investment into human capital or skills  $\theta$  is derived

---

<sup>7</sup>That is, income innovations  $dZ_{it}$  are normally distributed. Approximately,  $dZ_{it} \simeq \varepsilon_{it}\sqrt{\Delta t}$ ,  $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ , for a small  $\Delta t$ .

<sup>8</sup>Such an investment rule can also be derived from optimizing behavior. Results available upon request.

from an individual optimization problem, in which case it may depend on labor income (or other) taxes faced by individuals, that is  $\theta = \theta(\tau)$  where  $\tau$  is a tax rate.<sup>9</sup>

A large literature estimates reduced-form labor income processes similar to (1) using panel data.<sup>10</sup> In particular, (1) implies that the logarithm of income follows a Brownian motion which is simply the continuous-time version of a random walk. Hence (1) is the special case of the widespread “permanent-transitory model” of income dynamics, but with only a permanent component. As a result, good estimates are available for its parameter values. The process could easily be extended to feature a transitory component, e.g. by introducing jumps that are distributed i.i.d. over time and across individuals.<sup>11</sup>

### 3.2 Wealth Accumulation

The following simple model captures the main features of a large number of models of the upper tail of the wealth distribution.<sup>12</sup> Time is continuous and there is a continuum of individuals that are heterogeneous in their wealth  $\tilde{w}_{it}$ . At the individual level, wealth evolves as

$$d\tilde{w}_{it} = (1 - \tau)\tilde{w}_{it}dR_{it} + (y_t - c_{it}) dt$$

where  $\tau$  is the capital income tax rate,  $dR_{it}$  is the rate of return on wealth which is stochastic and  $y_t$  is labor income, and  $c_{it}$  is consumption. To keep things simple, we make the following assumptions. First, capital income is i.i.d. over time and, in particular,  $dR_{it} = \tilde{r}dt + \tilde{\nu}dZ_{it}$ , where  $\tilde{r}$  and  $\tilde{\nu}$  are parameters, and  $Z_{it}$  is a standard Brownian motion, which reflects idiosyncratic returns to human capital or to financial capital (this idiosyncratic shock captures the undiversified ownership of an entrepreneur, for instance).<sup>13</sup> Second, we assume that individuals consume an exogenous fraction  $\tilde{\theta}$  of their wealth at every point in time,  $c_{it} = \tilde{\theta}\tilde{w}_{it}$ .<sup>14</sup> Third, we assume that all individuals earn the same labor income  $y_t$ , which grows determin-

---

<sup>9</sup>See e.g. Champernowne (1953), Simon (1955), Nirei (2009), Toda (2012), Aoki and Nirei (2015) and Luttmer (2015) for models with similar reduced forms and Kim (2013), Jones and Kim (2014) for such models that additionally feature taxes.

<sup>10</sup>See e.g. MaCurdy (1982), Heathcote, Perri and Violante (2010) and Meghir and Pistaferri (2011).

<sup>11</sup>Results are available upon request.

<sup>12</sup>See e.g. Wold and Whittle (1957), Benhabib, Bisin and Zhu (2011, 2013, 2014), Piketty and Zucman (2014b), Jones (2015) and Acemoglu and Robinson (2015).

<sup>13</sup>Benhabib, Bisin and Luo (2015) argue that, in the data, such uninsured capital income risk is the main determinant of the wealth distribution’s right tail. In Section 5.3, we additionally consider common shocks.

<sup>14</sup>A consumption rule with such a constant marginal propensity to consume can also be derived from optimizing behavior, at least for large wealth levels  $w_{it}$ . Results available upon request.

istically at a rate  $g$ ,  $y_t = ye^{gt}$ . Given these assumptions, it is easy to show that detrended wealth  $w_{it} = \tilde{w}_{it}e^{-gt}$  follows the stochastic process

$$dw_{it} = [y + (r - g - \tilde{\theta})w_{it}]dt + \sigma w_{it}dZ_{it} \quad (2)$$

where  $r = (1 - \tau)\tilde{r}$  is the after-tax average rate of return on wealth and  $\sigma = (1 - \tau)\tilde{\nu}$  is the after-tax wealth volatility. Many other shocks (e.g. demographic shocks or shocks to saving rates) result in a similar reduced form.

### 3.3 Unifying Reduced Form Model

Both of these models share a common reduced form. In particular, income or wealth  $w_{it}$  follows a geometric Brownian motion

$$dw_{it} = \bar{\gamma}w_{it}dt + \sigma w_{it}dZ_{it}. \quad (3)$$

where  $\bar{\gamma}$  and  $\sigma$  are parameters (for the moment, we ignore the additive term  $ydt$  in (2)). All theories of top inequality add a “friction” to the random growth process (3) to ensure the existence of a stationary distribution (Gabaix, 1999). In the absence of such a “friction”, the cross-sectional variance of  $w_{it}$  grows without bound. We consider four such “frictions”:

**Friction 1** *Death at rate  $\delta$  together with reinjection at some  $w_0 > 0$ , normalized to  $w_0 = 1$ .*

**Friction 2** *A reflecting barrier at  $\underline{w} > 0$ , normalized to  $\underline{w} = 1$ .*

**Friction 3** *The combination of Frictions 1 and 2.*

**Friction 4** *The addition of an additive term  $ydt$  to (3) with  $y > 0$ .*

The simple model of income dynamics (1) is the special case of (3) with  $\bar{\gamma} = \bar{z}g(\theta, 1)$  and  $\sigma = \tilde{\sigma}g(\theta, 1)$  together with the first friction. Similarly, the simple model of wealth accumulation (2) is the special case of (3) with  $\bar{\gamma} = r - g - \theta$  together with the fourth friction. For the remainder of the paper’s theoretical analysis, we will focus on the income application for simplicity, and we will, therefore, refer to  $w_{it}$  as “income.” The reader should, however, keep in mind that all our results apply equally to the case where  $w_{it}$  is “wealth”. In Section 6, we explicitly return to the case of wealth dynamics.

We will later find it useful to conduct much of the analysis in terms of the logarithm of income  $w_{it}$ , which we denote by  $x_{it}$ . Applying Ito’s formula to (3),  $x_{it} = \log w_{it}$  follows

$$dx_{it} = \mu dt + \sigma dW_{it}, \quad \text{where} \quad \mu = \bar{\gamma} - \frac{\sigma^2}{2}. \quad (4)$$

The properties of the stationary distribution of the income process (3) are well understood. In particular, under certain parameter restrictions, this stationary distribution has a Pareto tail<sup>15</sup>

$$\mathbb{P}(w_{it} > w) \sim Cw^{-\zeta}$$

where  $C$  is a constant and  $\zeta > 0$  is a simple function of the parameters  $\mu$  (equivalently  $\bar{\gamma}$ ),  $\sigma$  and the particular friction (see e.g. Gabaix (2009)). With Friction 1 (the relevant friction for the case of income dynamics)

$$\zeta = \frac{-\mu + \sqrt{\mu^2 + 2\sigma^2\delta}}{\sigma^2}. \quad (5)$$

The constant  $\zeta$  is called the “power law exponent”, with a smaller  $\zeta$  corresponding to a fatter tail. We also find it useful to refer to the inverse of the power law exponent  $\eta = 1/\zeta$  as “top inequality”. Intuitively, tail inequality is increasing in  $\bar{\gamma} = \bar{z}g(\theta, 1)$  and  $\sigma = \bar{\sigma}g(\theta, 1)$  and decreasing in the death rate  $\delta$ . In Appendix C, we provide a complete characterization of the stationary distributions for all four frictions. It will be useful later to note that, equivalently, the logarithm of  $w$  has an exponential tail,  $\mathbb{P}(x_{it} > x) \sim Ce^{-\zeta x}$ .

To make the connection to the empirical evidence in the introduction, note that if the distribution of  $w$  has a Pareto tail above the  $p$ th percentile, then the share of the top  $p/10$ th percentile relative to that of the  $p$ th percentile is given by  $\frac{S(p/10)}{S(p)} = 10^{\eta-1}$ . There is, therefore, a one-to-one mapping between the relative income shares in panel (b) of Figure 1 and the top inequality parameter  $\eta = 1/\zeta$ .<sup>16</sup> Most existing contributions focus on the stationary distribution of the process (3) and completely ignore the corresponding transition dynamics. It is unclear whether these theories can explain the observed dynamics of the tail parameter  $\eta$ . This is what we turn to in the next section.

### 3.4 Other Theories of Top Income Inequality

This paper studies the dynamics of inequality in theories that can generate power laws, and we explicitly confine ourselves to studying such theories only. This is because we consider the fractal feature of top inequality discussed in Section 2 an important empirical regularity that deserves special attention. Beside random growth theories, there is one other class of

---

<sup>15</sup>Here and elsewhere “ $f(x) \sim g(x)$ ” for two functions  $f$  and  $g$  means there are positive constants  $k, k'$  such that  $kg(x) \leq f(x) \leq k'g(x)$  for  $x$  large enough.

<sup>16</sup>See Jones and Kim (2014) and Jones (2015) for two papers that use this fact extensively.

theories that can generate power laws, namely those building on “superstar” mechanisms.<sup>17</sup> We return to these theories below.

## 4 The Baseline Random Growth Model Generates Slow Transitions

Changes in the parameters of the income process (4) lead to changes in the fatness of the right tail of its stationary distribution. For example, an increase in the standard deviation of income innovations  $\sigma$  leads to an increase in stationary tail inequality  $\eta$  in (5). But this leaves the question as to whether this increase in inequality will come about quickly or will take a long time to manifest itself unanswered. The main message of this section is that the simple random growth model (4) gives rise to very slow transition dynamics.

Throughout this section, we conduct the following thought experiment. Initially at time  $t = 0$ , the economy is in a Pareto steady state corresponding to some initial parameters  $\mu_0, \sigma_0$  and so on. At time  $t = 0$ , a parameter changes; for example, the innovation variance  $\sigma^2$  may increase. Asymptotically as  $t \rightarrow \infty$ , the distribution converges to its new stationary distribution. The question is: what can we say about the speed of this transition? We present two sets of results corresponding to different notions of the speed of convergence. The first notion measures an “average” speed of convergence throughout the distribution. The second notion captures differential speeds of convergence across the distribution, allowing us in particular to put the spotlight on its upper tail.

Throughout the remainder of the paper, we denote the cross-sectional distribution of the logarithm of income  $x$  at time  $t$  by  $p(x, t)$ , the initial distribution by  $p_0(x)$  and the stationary distribution by  $p_\infty(x)$ . In order to talk about convergence, we also need a measure of distance between the distribution at time  $t$  and the stationary distribution. Throughout the paper we use the  $L^1$ -norm or total variation norm  $\|\cdot\|$  defined as

$$\|p(x, t) - p_\infty(x)\| := \int_{-\infty}^{\infty} |p(x, t) - p_\infty(x)| dx. \quad (6)$$

For frictions 1 to 3, the cross-sectional distribution  $p(x, t)$  satisfies the Kolmogorov Forward equation

$$p_t = -\mu p_x + \frac{\sigma^2}{2} p_{xx} - \delta p + \delta \delta_0. \quad (7)$$

---

<sup>17</sup>See Rosen (1981), Gabaix and Landier (2008), Tervio (2008) and Geerolf (2014) among others.

with initial condition  $p(x, 0) = p_0(x)$ , where we use the compact notation  $p_t := \frac{\partial p(x,t)}{\partial t}$  (similarly for  $x$ ) and where  $\delta_0$  denotes the Dirac delta function, i.e. a point mass at  $x = 0$ .<sup>18</sup> The first two terms on the right hand side capture the evolution of  $x$  due to diffusion with drift  $\mu$  and variance  $\sigma^2$ . The third term captures death and, hence, an outflow of individuals at rate  $\delta$ , and the fourth term captures birth, namely that every “dying” individual is replaced with a newborn at  $x = 0$ .<sup>19</sup> When there is a reflecting barrier (frictions 2 and 3),  $p$  must additionally satisfy the boundary condition

$$0 = -\mu p + \frac{\sigma^2}{2} p_x, \quad \text{at } x = 0. \quad (8)$$

It is often convenient to write this partial differential equation in terms of a differential operator

$$p_t = \mathcal{A}^* p + \delta \delta_0, \quad \mathcal{A}^* p = -\delta p - \mu p_x + \frac{\sigma^2}{2} p_{xx}. \quad (9)$$

This formulation is quite flexible and can be extended in a number of ways. For instance, in Section 4.4 we extend our framework to include Poisson jumps.

## 4.1 Average Speed of Convergence

We now state Proposition 1, one of the two main theoretical results of our paper. Before doing so we make the following assumption.

**Assumption 1** *The income process (4) has a unique invariant distribution  $p_\infty(x)$  and the initial and invariant distributions,  $p_0(x)$  and  $p_\infty(x)$ , satisfy*

$$\int_{-\infty}^{\infty} \frac{(p_0(x))^2}{p_\infty(x)} dx < \infty. \quad (10)$$

Note that (10) is a relatively weak restriction. For instance, assume that both  $p_0$  and  $p_\infty$  have Pareto tails  $p_0(x) \sim c_0 e^{-\alpha x}$  and  $p_\infty(x) \sim c e^{-\zeta x}$  for large  $x$ . Then (10) is equivalent to  $\alpha > \zeta/2$ . In particular, it is satisfied in all cases where top inequality in the new steady state is larger than that in the initial steady state,  $\zeta < \alpha$ , the main application we are interested in.<sup>20</sup>

<sup>18</sup>To be clear, the death rate  $\delta$  is a number, while  $\delta_0$  is a (generalized) function.

<sup>19</sup>More generally, we can allow for the income of newborns to be drawn from an arbitrary (thin-tailed) distribution  $\psi(x)$ , in which case the fourth term reads  $+\delta\psi(x)$ . We consider the special case in which this distribution is a point mass,  $\psi(x) = \delta_0(x)$ , i.e. all newborns have the same income for simplicity.

<sup>20</sup>Proposition 1 can also be extended to the case where  $p_0$  does not decay fast enough, i.e. if  $\alpha < \zeta/2$ . In particular, one can bound the speed of convergence, which becomes lower. Results are available upon request.

**Proposition 1** Consider the income process (4). Under Assumption 1, the cross-sectional distribution  $p(x, t)$  converges to its stationary distribution exponentially in the total variation norm, that is  $\|p(x, t) - p_\infty(x)\| \sim ke^{-\lambda t}$  for constants  $k$  and  $\lambda$ . The rate of convergence

$$\lambda = - \lim_{t \rightarrow \infty} \frac{1}{t} \log \|p(x, t) - p_\infty(x)\|$$

depends on whether there is a reflecting barrier at  $x = 0$ . Without a reflecting barrier (Friction 1)

$$\lambda = \delta. \tag{11}$$

With a reflecting barrier (Frictions 2 and 3)

$$\lambda = \frac{1}{2} \frac{\mu^2}{\sigma^2} \mathbf{1}_{\{\mu < 0\}} + \delta \tag{12}$$

where  $\mathbf{1}_{\{\cdot\}}$  is the indicator function.

We shall see that Proposition 1 implies that the traditional canonical model delivers convergence that is far too slow:  $\lambda$  is too low compared to the empirical estimate.

The proof of the Proposition for the case without a reflecting barrier analyzes directly the  $L^1$ -norm (6) by means of a differential equation for  $|p(x, t) - p_\infty(x)|$ . The rough idea of the proof of the Proposition for the case with a reflecting barrier is as follows. The entire dynamics of the process for  $x_{it}$  are summarized by the operator  $\mathcal{A}^*$  defined in (9). This operator is the appropriate generalization of a transition matrix for a finite-state process to processes with a continuum of states such as (4), and it can be analyzed in an exactly analogous way. In particular, the critical property of  $\mathcal{A}^*$  governing the speed of convergence of  $p$  is its largest non-trivial eigenvalue. The key contribution of the Proposition and the main step of the proof is then to obtain an explicit formula for the second eigenvalue of  $\mathcal{A}^*$  in the form of (12).<sup>21</sup>

The intuition for formulas (11) and (12) is as follows. In the case without a reflecting barrier, the speed is simply given by the death intensity  $\delta$ . This is intuitive: the higher  $\delta$  is, the more churning there is in the cross-sectional distribution and the faster the distribution settles down to its invariant distribution. In the extreme case where  $\delta \rightarrow \infty$ , the distribution jumps to its steady state immediately. Next, consider the case with a reflecting barrier,  $\mu < 0$ ,

---

<sup>21</sup>Linetsky (2005) derives a related result for the special case of a reflected Brownian motion with  $\mu < 0$  and  $\delta = 0$ . For the same case one can also derive the formula for the speed of convergence by “brute force” from the standard formulas for reflected Brownian motion (see e.g. Harrison, 1985). Our results are considerably more general.

and no Poisson death,  $\delta = 0$  (Friction 2). From (5), stationary tail inequality for this case is  $\eta = 1/\zeta = -\frac{\sigma^2}{2\mu}$  and therefore the speed of convergence can also be written as

$$\lambda = \frac{1}{8} \frac{\sigma^2}{\eta^2}. \quad (13)$$

This expression is again intuitive. It states that the transition is faster the higher is the standard deviation of growth rates  $\sigma$  and the lower is the tail inequality  $\eta$ ; that is, high inequality goes hand in hand with slow transitions. The interpretation of the formula for the case with a reflecting barrier and  $\delta > 0$  is similar.

In section 4.3 we show that when the parameters  $\mu, \sigma$  and  $\delta$  are calibrated to be consistent with the micro data and the observed inequality at a point in time, the implied speed of convergence is an order of magnitude too low to explain the observed increase in inequality in the data.

## 4.2 Speed of Convergence in the Tail

In the preceding section we characterized a measure of the average speed of convergence across the entire distribution. The purpose of this section is to examine the possibility that different parts of the distribution may converge at different speeds. In particular, we show that convergence is particularly slow in the upper tail of the distribution. That is, the formula in Proposition 1 *overestimates* the speed of convergence of parts of the distribution. In this section we focus on the case with Poisson death, but without a reflecting barrier (Friction 1) because it is possible to obtain clean analytic formulas for this case. While clean formulas are not available if the income process features a reflecting barrier, we show by means of numerical computations that the general lessons from our analysis carry over to the other two cases of frictions as well.

### 4.2.1 An Instructive Special Case: the Steindl Model

To explain the main result of this section in the most accessible way, we first examine the restrictive but instructive special case where  $\sigma = 0$  and  $\mu, \delta > 0$ . In this model, originally due to Steindl (1965), the logarithm of income  $x_{it}$  grows at rate  $\mu$  and gets reset to  $x_{i0} = 0$  at rate  $\delta$ . The Steindl model has recently also been examined by Jones (2015). The distribution  $p(x, t)$  then satisfies the Kolmogorov Forward equation (7) with  $\sigma = 0$  for  $x > 0$ . The corresponding stationary distribution is a Pareto distribution

$$p_\infty(x) = \zeta e^{-\zeta x}, \quad \text{with} \quad \zeta = \frac{\delta}{\mu}.$$

For concreteness, consider an economy starting in a steady state with some growth rate  $\mu_0$  (and death rate  $\delta_0$ ). At  $t = 0$  the growth rate changes permanently to  $\mu > \mu_0$  (and death rate  $\delta$ ). Then, the new steady state distribution is more fat-tailed,  $\zeta < \alpha$ .

**Lemma 1** (Closed form solution for the transition in the Steindl model) *The time path of  $p(x, t)$  is the solution to (7) with  $\sigma = 0$  and initial condition  $p_0(x) = \alpha e^{-\alpha x}$ ,  $\alpha = \delta_0/\mu_0$  which is given by*

$$p(x, t) = \zeta e^{-\zeta x} \mathbf{1}_{\{x \leq \mu t\}} + \alpha e^{-\alpha x + (\alpha - \zeta)t} \mathbf{1}_{\{x > \mu t\}} \quad (14)$$

where  $\mathbf{1}_{\{\cdot\}}$  is the indicator function.

The solution is depicted in Figure 2. Consider, in particular, the local power law exponent

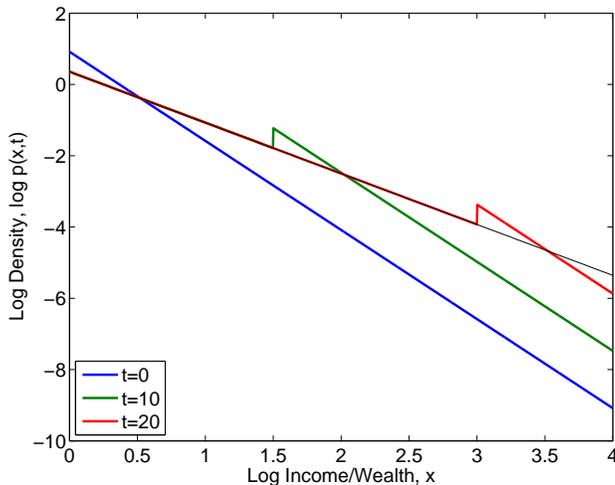


Figure 2: Transition of Distribution in Steindl Model,  $\sigma = 0, \mu, \delta > 0$

$\zeta(x, t) = -\partial \log p(x, t) / \partial x$ . Since the Figure plots the log density,  $\log p(x, t)$ , against log income  $x$ , this local power law exponent is simply the slope of the line in the Figure. The time path of the distribution features a “traveling discontinuity”. Importantly, the local power law exponent (the slope of the line) first changes only for low values of  $x$ . In contrast, for high values of  $x$ , the distribution shifts out in parallel and the slope of the line does not move at all. More precisely, for a given point  $x$ , the distribution fully converges at time  $\tau(x) = x/\mu$ , but does not move at all when  $t < \tau(x)$ . In the Steindl model, the convergence of the distribution is slower the further out in the tail we look. In particular, note from the Figure that the *asymptotic* (for large  $x$ ) power law exponent  $\zeta(t) = -\lim_{x \rightarrow \infty} \partial \log p(x, t) / \partial x$  takes an *infinite* time to converge to its stationary distribution. In the special case of the

Steindl model, this slow convergence in the tail is particularly stark in that some parts of the distribution do not move at all. We show below that, while less stark, the general insight that convergence is slower in the tail also carries over to the model with  $\sigma > 0$ .

Consider the behavior of top income shares in response to the permanent increase in  $\mu$  considered above. Lemma 1 implies that the relative income of the 0.1% vs 1% income quantiles is *constant* for a while; it budes only when the “traveling discontinuity” hits the top 1% quantile. In contrast, the level of the top 1% income quantile and the 0.1% income quantile increase quickly after the shock (to be more precise, after any time  $t > 0$ , they have moved, in parallel). Hence, the ratio of the 0.1% to 1% share moves slowly (indeed, not at all for a while), though the top 1% share moves fairly fast. This phenomenon will be confirmed in the next subsection.

For completeness, we note that it is also possible to characterize the time path of the distribution for  $\sigma > 0$  and for more general initial conditions  $p_0(x)$ :

**Lemma 2** (Closed form solution for general model without reflecting barrier) *In the case with no reflecting barrier, we have*

$$p(x, t) = p_\infty(x) + e^{-\delta t} \mathbb{E} [p_0(x - g_t) - p_\infty(x - g_t)] \quad (15)$$

where  $g_t := \mu t + \sigma Z_t$ , and the expectation is taken over the stochastic realizations of  $g_t$ .

Note that when  $\sigma = 0$ ,  $p(x, t) = p_\infty(x) + e^{-\delta t} [p_0(x - \mu t) - p_\infty(x - \mu t)]$ , which coincides with (14) when the initial distribution is  $p_0(x) = \alpha e^{-\alpha x} \mathbf{1}_{\{x > 0\}}$ .

#### 4.2.2 Speed of Convergence in the Tail for the General Model

As noted earlier, the distribution  $p(x, t)$  satisfies the Kolmogorov Forward equation (7). As in the first part of this section, we continue to focus on the case with Poisson death, but without a reflecting barrier. One can show (see e.g. Gabaix, 2009) that in this case the stationary distribution is a double Pareto distribution

$$p_\infty(x) = c \min\{e^{-\zeta_- x}, e^{-\zeta_+ x}\} \quad (16)$$

where  $c = -\zeta_- \zeta_+ / (\zeta_+ - \zeta_-)$  and where  $\zeta_- < 0 < \zeta_+$  are the two roots of

$$0 = \frac{\sigma^2}{2} \zeta^2 + \zeta \mu - \delta. \quad (17)$$

Apart from the stationary distribution, the solution to the Kolmogorov Forward equation is cumbersome.

The key insight of this section is that, for the case without a reflecting barrier, the entire time path of the solution to the Kolmogorov Forward equation can be characterized conveniently in terms of the so-called ‘‘Laplace transform’’ of  $p$

$$\widehat{p}(\xi, t) := \int_{-\infty}^{\infty} e^{-\xi x} p(x, t) dx = \mathbb{E} [e^{-\xi X_t}], \quad (18)$$

where  $\xi$  is a real number and  $X_t$  represents the random variable (log income) with distribution  $p(x, t)$ .<sup>22</sup> For  $\xi \leq 0$ , the Laplace transform has the natural interpretation of the  $-\xi$ th moment of the distribution of income, that is  $\mathbb{E}[w^{-\xi}]$ , where  $w = e^X$  is income. We show momentarily that we can obtain a clean analytic formula for the entire time path of this object for all  $t$ . This is useful because a complete characterization of a function’s Laplace transform is equivalent to a complete characterization of the function itself. This is because by varying the variable  $\xi$ , we can trace out the behavior of different parts of the distribution. In particular, the more negative  $\xi$  is, the more we know about the distribution’s tail behavior. In a similar vein, our analysis using Laplace transforms will allow us to characterize tightly the behavior of a weighted version of the  $L^1$ -norm in (6) :

$$\|p(x, t) - p_{\infty}(x)\|_{\xi} := \int_{-\infty}^{\infty} |p(x, t) - p_{\infty}(x)| e^{-\xi x} dx. \quad (19)$$

In the special case  $\xi = 0$ , this distance measure coincides with the  $L^1$ -norm defined in (6). But by taking  $\xi < 0$ , (19) puts more weight on the behavior of the distribution’s tail, the main focus of the current section. Note that the Laplace transform (18) ceases to exist if  $\xi$  is too negative or too positive. To ensure that the Laplace transform exists we impose the restriction that  $-\zeta_+ < \xi < -\zeta_-$  where  $\zeta_-, \zeta_+$  are the tail parameters of the stationary distribution (16).<sup>23</sup>

We apply the Laplace transform to the Kolmogorov Forward equation (7). We use the usual rules:  $\widehat{p}_x = \xi \widehat{p}$ ,  $\widehat{p_{xx}} = \xi^2 \widehat{p}$ , and obtain:

$$\frac{\partial \widehat{p}(\xi, t)}{\partial t} = -\lambda(\xi) \widehat{p}(\xi, t) + \delta \quad \text{where} \quad \lambda(\xi) := \mu \xi - \frac{\sigma^2}{2} \xi^2 + \delta \quad (20)$$

with initial condition  $\widehat{p}(\xi, 0) = \widehat{p}_0(\xi)$ , the Laplace transform of  $p_0(x)$ . Here, we used that for any  $x_0$ , the Laplace transform of the Dirac delta function  $\delta_{x_0}$  is  $e^{-\xi x_0}$  and hence the Laplace

---

<sup>22</sup>Note that we here work with the ‘‘bilateral’’ or ‘‘two-sided’’ Laplace transform which integrates over the entire real line. This is in contrast to the one-sided Laplace transform defined as  $\int_0^{\infty} e^{-\xi x} p(x, t) dx$ .

<sup>23</sup>To see this, calculate the Laplace transform of (16)  $\widehat{p}_{\infty}(\xi) = \frac{c}{\zeta_+ + \xi} - \frac{c}{\zeta_- + \xi} = \frac{\zeta_- - \zeta_+}{(\zeta_- + \xi)(\zeta_+ + \xi)}$ . We therefore need  $-\zeta_+ < \xi < -\zeta_-$ .

transform of  $\delta_0$  equals 1. Importantly, note that for fixed  $\xi$ , (20) is a simple ordinary differential equation for  $\widehat{p}$  that can be solved analytically.

**Proposition 2** (Speed of convergence in the tail) *Consider the Laplace transform of the income distribution  $\widehat{p}(\xi, t)$  defined in (18) for  $-\zeta_+ < \xi < -\zeta_-$  where  $\zeta_-, \zeta_+$  are the two roots of (17). Its time path is given by*

$$\widehat{p}(\xi, t) = \widehat{p}_\infty(\xi) + (\widehat{p}_0(\xi) - \widehat{p}_\infty(\xi)) e^{-\lambda(\xi)t}, \quad (21)$$

$$\lambda(\xi) := \mu\xi - \frac{\sigma^2}{2}\xi^2 + \delta, \quad (22)$$

$$\widehat{p}_\infty(\xi) := \frac{\delta}{\mu\xi - \frac{\sigma^2}{2}\xi^2 + \delta}. \quad (23)$$

Furthermore, the weighted  $L^1$ -norm defined in (19) converges at rate  $\lambda(\xi)$ :

$$\|p(x, t) - p_\infty(x)\|_\xi \sim ke^{-\lambda(\xi)t}. \quad (24)$$

Consider first the speed of convergence of the weighted distance measure in (24). For the special case  $\xi = 0$ , we have  $\lambda(\xi) = \delta$ ; when the weighted  $L^1$ -norm places no additional weight on the behavior of the distribution's tail, we recover our original result from Proposition 1, as expected. As we take  $\xi$  more and more negative, the weighted norm places more and more weight on the behavior of the distribution's tail, and the corresponding speed of convergence is given by  $\lambda(\xi)$ . Note that for  $\mu > 0$ , the speed of convergence  $\lambda(\xi)$  is always lower the lower  $\xi$  is, for all  $\xi \leq 0$ . If  $\mu < 0$ , the same is true for all  $\xi$  less than some critical value. The formula for  $\lambda(\xi)$  therefore indicates that convergence is slower the more weight we put on observations in the distribution's tail.

Next consider (21) which provides a closed form solution for the evolution of the Laplace transform or equivalently for the evolution of all moments of the cross-sectional income distribution. These moments converge at the same rate  $\lambda(\xi)$  as the weighted norm in (24). Hence, the closed form solution for the Laplace transform in (21) shows that high moments converge more slowly than low moments.

Figure 3 provides a graphical illustration of these theoretical results. As in the Steindl case of Figure 2, the power law exponent  $\zeta$  (equivalently top inequality  $\eta$ ) does not change at first and the distribution instead shifts out in parallel.<sup>24</sup>

---

<sup>24</sup>This is more than a numerical result. Defining the local power law exponent  $\zeta(x, t) := -P_x(x, t)/P(x, t)$  where  $P$  is the CDF corresponding to  $p$ , one can show using (7) that this local power law exponent does not move on impact following a shock,  $\zeta_t(x, t)|_{t=0} = 0$  for all  $x$ .

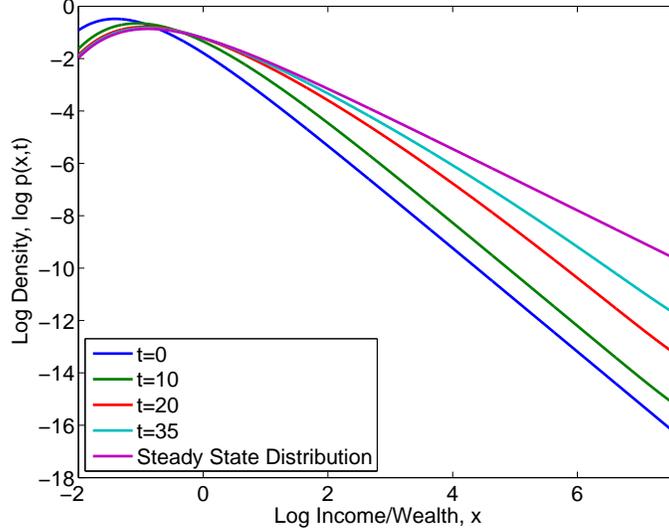


Figure 3: Slow Convergence in the Tail in Standard Random Growth Model

Finally, consider the expression for the stationary Laplace transform (23). Reassuringly, one can check that it coincides with the result obtained from directly Laplace-transforming the stationary distribution  $\hat{p}_\infty(\xi)$ .<sup>25</sup> Also note that if  $\xi = -\zeta_+$  or  $\xi = -\zeta_-$  so that  $\mu\xi - \frac{\sigma^2}{2}\xi^2 + \delta = 0$ , then  $\hat{p}_\infty(\xi)$  ceases to exist. This, in fact, suggests a general strategy for identifying a distribution's Pareto tail from knowledge of its Laplace transform only that will be useful later: the tail parameter is simply the critical value  $\zeta > 0$  such that  $\hat{p}_\infty(\xi)$  ceases to exist for all  $\xi \leq -\zeta$ .<sup>26</sup>

### 4.3 The Baseline Model Cannot explain the Fast Rise in Income Inequality

We now revisit Figure 1 from Section 2 and ask: can standard random growth models generate the observed increase in income inequality? We find that they cannot. In particular, the transition dynamics generated by the model are too slow relative to the dynamics observed

<sup>25</sup>We have

$$\hat{p}_\infty(\xi) = \frac{c}{\zeta_+ + \xi} - \frac{c}{\zeta_- + \xi} = \frac{\zeta_- - \zeta_+}{(\zeta_- + \xi)(\zeta_+ + \xi)}. \quad (25)$$

From the quadratic for  $\zeta$  (17), we have that  $\zeta_- \zeta_+ = -\delta/(\sigma^2/2)$ ,  $\zeta_- + \zeta_+ = -\mu/(\sigma^2/2)$  and hence  $(\zeta_- + \xi)(\zeta_+ + \xi) = \zeta_- \zeta_+ + \xi(\zeta_- + \zeta_+) + \xi^2 = -\delta/(\sigma^2/2) - \mu\xi/(\sigma^2/2) + \xi^2$ . Substituting into (25) we obtain (23).

<sup>26</sup>In particular, one can show that for any distribution  $p$  with a Pareto tail, that is  $p(x) \sim ce^{-\zeta x}$   $x \rightarrow \infty$  for constants  $c$  and  $\zeta$ , the Laplace transform  $\hat{p}(\xi) \sim \frac{c}{\zeta + \xi}$  as  $\xi \downarrow -\zeta$ .

in the data. This confirms, with calibrated data, the theoretical results in the preceding two sections.

More precisely, we ask whether an increase in the variance of the permanent component of wages  $\sigma^2$  can explain the increase in income inequality observed in the data. That an increase in the variance of permanent earnings has contributed to the rise of inequality observed in the data has been argued by Moffitt and Gottschalk (1995), Kopczuk, Saez and Song (2010) and DeBacker et al. (2013) (however, Guvenen, Ozkan and Song (2014) examine administrative data and dispute that there has been such a trend – either way our argument is that an increase in  $\sigma$  cannot explain the rise in top inequality). The particular experiment we consider is an increase in the variance of permanent earnings  $\sigma^2$  from 0.01 in 1973 to 0.025 today (implying that the standard deviation  $\sigma$  increases from 0.1 to 0.158 broadly consistent with evidence in Heathcote, Perri and Violante (2010)).

We calibrate the remaining parameters  $\mu$  and  $\delta$  as follows. First, note that our model is stationary whereas the U.S. economy features long-run growth. Since we normalized the starting wage of labor force entrants to  $w_0 = 1$ ,  $\mu$  is the growth rate of individual incomes over the lifecycle relative to this long-run trend. We set  $\delta = 1/30$  corresponding to an expected work life of thirty years and calibrate  $\mu$  to match the observed tail inequality in 1973,  $\eta_{1973} = 0.39$  or equivalently  $\zeta_{1973} = 1/\eta_{1973} = 2.56$  which yields  $\mu = 0.002$ , i.e. individual income growth 0.2% above the economy’s long-run growth rate.<sup>27</sup> Note that this calibration is conservative. In particular, given that convergence rates are increasing in  $\delta$ , a longer expected work life would result in even slower transitions.

Since the income process is not subject to a reflecting barrier, Proposition 1 implies that the average speed of convergence is simply  $\lambda = \delta$  and the corresponding half-life is  $t_{1/2} = \log(2)/\delta = 20.8$  years. As shown in Section 4.2, the speed of convergence in the tail can be much slower. In particular, consider the formula for the speed of convergence  $\lambda(\xi) = \mu\xi - \frac{\sigma^2}{2}\xi^2 + \delta$  from Proposition 2 where the reader should recall that by varying  $\xi$ , we can trace out the speed of convergence of all moments of the distribution and  $\lambda(\xi)$  is the speed of convergence of the  $-\xi$ th moment. Equivalently,  $-\xi$  is the weight on the tail in the weighted  $L^1$ -norm (19), with the maximum admissible weight equal to the tail exponent of the new stationary distribution which here equals  $\zeta = 1.6$ .

Figure 4 plots the corresponding half-life  $t_{1/2}(\xi) = \log(2)/\lambda(\xi)$  for the parameter values

---

<sup>27</sup>We compute  $\eta$  from the relative income shares in panel (b) of Figure 1 as  $\eta(p) = \log S(p/10)/S(p)$ . We here use  $\eta(1) = S(0.1)/S(1)$ . To understand how  $\mu$  is calibrated, note that  $\zeta$  satisfies (17). Therefore, given  $\sigma^2 = 0.01$  and  $\delta = 1/30$ , we need  $\mu = \delta/\zeta - \zeta\sigma^2/2 = 0.002$ .

used in our experiment as a function of the moment under consideration  $-\xi$ . Consider first

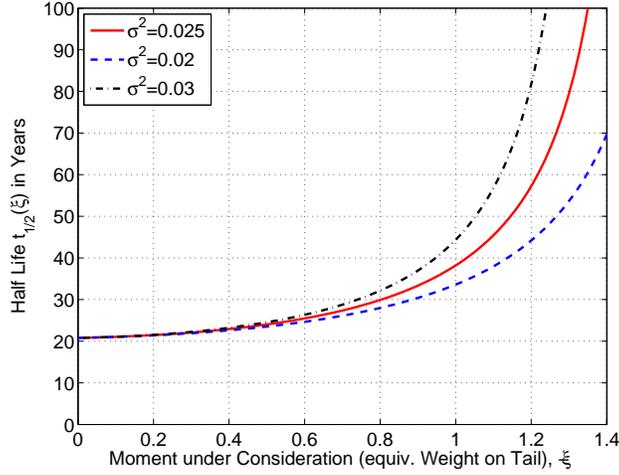


Figure 4: Theoretical Speed of Convergence of Different Moments of Income Distribution

the red solid line which plots the half life  $t_{1/2}(\xi)$  for  $\sigma^2 = 0.025$ , the variance of the permanent component of wages used in our experiment. For  $\xi = 0$ , the speed of convergence is given by  $t_{1/2} = \log(2)/\delta = 20.8$  years as expected. There are two main takeaways from the figure. First, even for relatively low moments the speed of convergence is considerably lower. For example, the half-life of convergence of the first moment ( $\xi = -1$ ) is around 40 years, i.e. twice as much as the average speed. Second, the speed of convergence becomes slower and slower the higher the moment under consideration, with half lives of 100 years close to the highest admissible moment  $\zeta = 1.6$ . The figure also shows that the speed of convergence is quite sensitive to the value of the variance  $\sigma^2$ , particularly the speed in the tail.

Figure 5 plots the time path for the top 1% income share (panel (a)) and the empirical power law exponent (panel (b)) generated by the baseline random growth model and compares it to that in the data.<sup>28</sup> Not surprisingly given our analytical results, the model fails spectacularly.<sup>29</sup> An increase in the variance in the permanent component of income  $\sigma^2$  is therefore not a promising candidate for explaining the observed increase in top income inequality. Similarly, consider a simple tax story like that mentioned in Section 3.1 in which the intensity of investment is affected by taxation  $\theta = \theta(\tau)$  where  $\tau$  is a tax rate, and therefore so are the reduced form parameters  $\mu$  and  $\sigma$ . Because this simple story collapses in a

<sup>28</sup>We solve the Kolmogorov Forward equation (7) numerically using a finite difference method.

<sup>29</sup>Note that the power law exponent in panel (b) is completely flat on impact, consistent with Figures 2 and 3, and footnote 24.

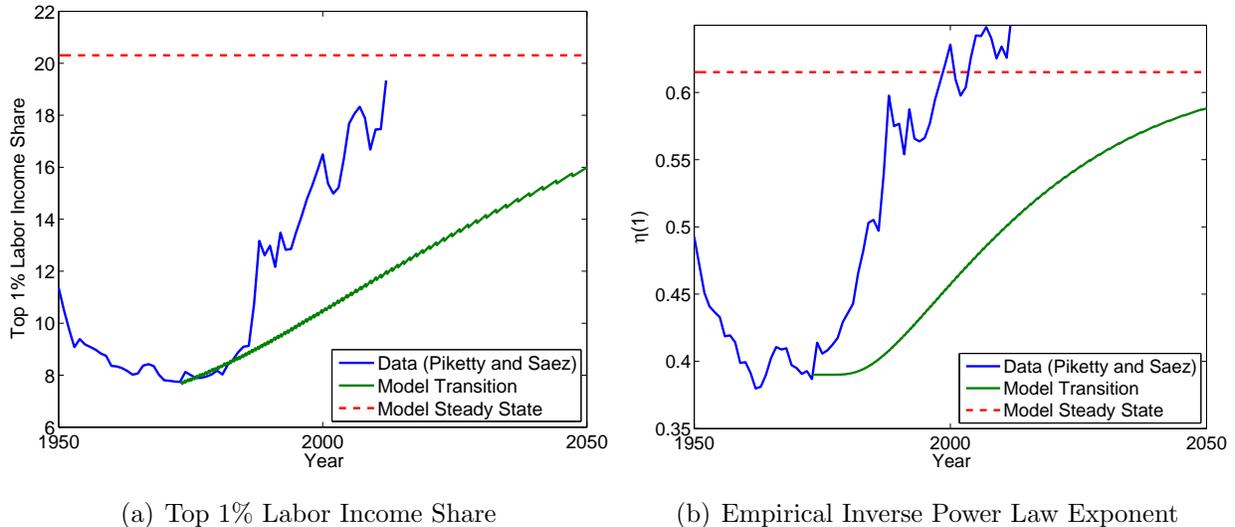


Figure 5: Dynamics of Income Inequality in the Baseline Model

reduced form to a standard random growth model, it cannot generate fast transitions either (results are available upon request).

#### 4.4 Jumps Cannot Explain the Fast Rise in Inequality Either

The standard random growth model we have studied so far assumes that income innovations are log-normally distributed as implied by the assumption that the income process is a geometric Brownian motion. Recent research suggests that this is a quite imperfect description of the data. For instance, Guvenen et al. (2015) using administrative data, document that earnings innovations are very fat-tailed and much more so than a normally distributed random variable. In particular, they show that the distribution of income *growth rates* itself has Pareto tails (and not just the income distribution). In our continuous time setup, the most natural way of generating such kurtosis is by introducing jumps.<sup>30</sup>

In this section, we ask whether departing from the standard log-normal framework by introducing jumps can help resolve the puzzle raised in Section 4.3 that random growth processes cannot explain the fast rise income inequality observed in the data. We find that

<sup>30</sup>It is not surprising that income innovations will be leptokurtic if the distribution from which jumps are drawn features kurtosis itself. Interestingly, this is not necessary for income innovations to be leptokurtic: even normally distributed jumps that arrive with a Poisson arrival rate can generate kurtosis in data observed at discrete time intervals. The same logic is used in the theory of so-called “subordinated stochastic processes.”

they cannot: while jumps are useful descriptively for capturing certain features of the data, they do not increase the speed of convergence of the cross-sectional income distribution.

To introduce jumps, we extend the income process (4) as follows:

$$dx_{it} = \mu dt + \sigma dZ_{it} + g_{it} dN_{it} \quad (26)$$

where  $dN_{it}$  is a jump process with intensity  $\phi$ . That is, there is a jump in  $[t, t + dt)$  (i.e.,  $dN_{it} = 1$ ) with probability  $\phi dt$  and no jump (i.e.,  $dN_{it} = 0$ ) with probability  $1 - \phi dt$ . The innovations  $g_{it}$  are drawn from an exogenous distribution  $f$ . The distribution  $f$  can have arbitrary support and it may be either thin-tailed (e.g. a normal distribution) or thick-tailed (e.g. a double-Pareto distribution consistent with the evidence in Guvenen et al. (2015)). With jumps, the Kolmogorov Forward equation (7) becomes

$$p_t = -\delta p - \mu p_x + \frac{\sigma^2}{2} p_{xx} + \phi \mathbb{E} [p(x - g) - p(x)] + \delta \delta_0.$$

Relative to (7), the new term is the expectation  $\mathbb{E} [p(x - g) - p(x)]$ , which is taken over the random jump  $g$  and is multiplied by  $\phi$ , the arrival rate of jumps. Note that this term can also be written as

$$\mathbb{E} [p(x - g) - p(x)] = \int_{-\infty}^{\infty} [p(x - g) - p(x)] f(g) dg = (p * f)(x) - p(x)$$

where  $*$  is the convolution operator. Our analysis of transition dynamics using Laplace transforms in Section 4.2 can easily be extended to the case of jumps because integral transforms like the Laplace transform are the ideal tool for handling convolutions. In particular, the Laplace transform of a convolution of two functions is the product of the Laplace transforms of the two functions:  $\widehat{(p * f)}(\xi) = \widehat{p}(\xi) \widehat{f}(\xi)$ . Using this fact, Proposition 2 extends immediately to the case of jumps.

**Proposition 3** (Speed of convergence when there are jumps in the growth rate) *Consider the Laplace transform of the cross-sectional income distribution  $\widehat{p}(\xi, t)$  defined in (18), where jumps are allowed. Its time path is still given by (21), but now with rate of convergence*

$$\lambda(\xi, \phi) := \xi \mu - \xi^2 \frac{\sigma^2}{2} + \delta - \phi(\widehat{f}(\xi) - 1), \quad (27)$$

$$\widehat{f}(\xi) := \int_{-\infty}^{\infty} e^{-\xi g} f(g) dg, \quad (28)$$

*the Laplace transform of the distribution of jumps  $f(g)$  which satisfies  $\widehat{f}(0) = 1$  and  $\widehat{f}(\xi) > 1$  for all  $\xi < 0$ . The weighted  $L^1$ -norm converges at rate  $\lambda(\xi, \phi)$ :*

$$\|p(x, t) - p_{\infty}(x)\|_{\xi} \sim k e^{-\lambda(\xi, \phi)t}.$$

In particular, the average speed of convergence as measured by the unweighted  $L^1$ -norm equals  $\lambda(0, \phi) = \delta$  for all  $\phi$ , i.e. it is entirely unaffected by the presence of jumps. With  $\xi < 0$ , jumps make the speed of convergence lower than in the absence of jumps:  $\lambda(\xi, \phi) < \lambda(\xi, 0)$ ,  $\phi > 0$ .

It is worth reemphasizing the last part of the Proposition: if we confine attention to the average speed of convergence  $\|p(x, t) - p_\infty(x)\|$  jumps have no effect whatsoever. If instead, we put more weight on observations in the distribution's tail,  $\xi < 0$ , then the rate of convergence becomes worse, not better. This follows from (27) and the fact that  $\hat{f}(\xi) > 1$  for  $\xi < 0$ . Furthermore for  $\xi < 0$ ,  $\lambda(\xi, \phi)$  is *decreasing* in  $\phi$ , that is the higher is the jump intensity, the lower is the rate of convergence. We conclude that jump processes, though very useful for the purpose of capturing salient features of the data, are not helpful in terms of providing a theory of fast transitions.

## 5 Models that Generate Fast Transitions

Given the negative results of the preceding section, it is natural to ask: what then explains the observed fast rise in top income inequality? We argue that fast transitions require very specific departures from the standard random growth model. We extend the standard random growth model along two dimensions. First, we allow for heterogeneity in mean growth rates, in particular a “high growth regime.” Second, we consider deviations from Gibrat’s law, a feature which we argue arises naturally in “superstar theories”. We discuss the role of these two additions in turn in Sections 5.2 and 5.3. In Section 5.4, we then revisit the rise in income inequality and argue that our generalized random growth model can generate transitions that are as fast as those observed in the data.

### 5.1 The Augmented Random Growth Model

In its most general form, we consider a random growth model with distinct “growth regimes” indexed by  $j = 1, \dots, J$ , deviations from Gibrat’s law captured by a process  $S_t$ , and jumps denoted by  $dN_{jit}$ . In particular, the dynamics of income  $x_{it}$  of individual  $i$  in regime  $j$  are given by

$$\begin{aligned} x_{it} &= e^{b_j S_t} y_{it}, \\ dy_{it} &= \mu_j dt + \sigma_j dZ_{it} + g_{jit} dN_{jit} + \text{Injection} - \text{Death}, \end{aligned} \tag{29}$$

where  $dN_{jit}$  is a Poisson process with intensity  $\phi_j$  and  $g_{jit}$  is a random variable with distribution  $f_j$ . As before, we assume that workers retire at rate  $\delta$  and get replaced by labor

entrants with  $x = 0$ . We assume that a fraction  $\theta_j$  of labor force entrants are born in regime  $j$  and workers switch from regime  $j$  to regime  $k$  at rate  $\phi_{j,k}$ . Different “growth regimes” differ in the mean growth rate  $\mu_j$  and the standard deviation of income changes  $\sigma_j$ . Guvenen (2007) has argued that an income process with heterogeneous income profiles provides a better fit of the micro data than a model in which all individuals face the same income profile, and he finds large heterogeneity in the slope of income profiles. The model above also allows for heterogeneity in the standard deviation of income innovations  $\sigma_j$ . This has a similar flavor to the mixture specification advocated by Guvenen et al. (2015). We build on Luttmer (2011), who studies a related framework applied to firm dynamics and argues that persistent heterogeneity in mean firm growth rates is needed to account for the relatively young age of very large firms at a given point in time (a statement about the stationary distribution rather than transition dynamics as in our paper). Aoki and Nirei (2015) present a related and more complex economic model with entrepreneurs and workers that are subject to different income growth rates, and Jones and Kim (2014) examine a model with different types of entrepreneurs.

Deviations from Gibrat’s law are captured by  $S_t$ , which is an arbitrary stochastic process satisfying  $\lim_{t \rightarrow \infty} \mathbb{E}[S_t] < \infty$ . To see this, note that (29) can be written as

$$dx_{it} = \tilde{\mu}_{jt}dt + \tilde{\sigma}_{jt}dZ_{it} + b_jx_{it}dS_t + g_{jit}d\tilde{N}_{jit} + \text{Injection} - \text{Death} \quad (30)$$

where  $\tilde{\mu}_{jt} = \mu_j e^{b_j S_t}$ ,  $\tilde{\sigma}_{jt} = \sigma_j e^{b_j S_t}$  and  $d\tilde{N}_{jit} = dN_{jit} e^{b_j S_t}$ . If  $b_j dS_t > 0$ , the growth rate of income  $x_{it}$  is increasing in income, constituting a deviation from Gibrat’s law.<sup>31</sup> Our baseline model analyzed in section 4 is the special case in which  $J = 1$  and  $dS_t = dN_{jit} = 0$ .

## 5.2 The Role of Heterogeneity in Mean Growth Rates

First, consider the special case of (29) with multiple distinct “growth regimes”, but without deviations from Gibrat’s law or jumps  $dS_t = dN_{it} = 0$ . Here we focus on a simple case with two regimes, a high-growth regime and a low-growth regime, but our results can be extended to three or more regimes.

Denote the density of individuals who are currently in the high and low growth states by  $p^H(x, t)$  and  $p^L(x, t)$  and the cross-sectional wage distribution by  $p(x, t) = p^H(x, t) + p^L(x, t)$ . We assume that a fraction  $\theta$  of individuals start their career in the high-growth regime and

---

<sup>31</sup>Also note that  $Z_{it}$  is an idiosyncratic stochastic process whereas  $S_t$  is an aggregate or common shock that hits all individuals simultaneously.

the remainder in the low-growth regime, that individuals switch from high to low growth with intensity  $\psi$  and that low growth is an absorbing state that is only left upon retirement. Then, the densities satisfy the following system of Kolmogorov Forward equations

$$\begin{aligned} p_t^H &= -\mu_H p_x^H + \frac{\sigma_H^2}{2} p_{xx}^H - \psi p^H - \delta p^H + \beta_H \delta_0, \\ p_t^L &= -\mu_L p_x^L + \frac{\sigma_L^2}{2} p_{xx}^L + \psi p^H - \delta p^L + \beta_L \delta_0, \end{aligned} \quad (31)$$

with initial conditions  $p^H(x, 0) = p_0^H(x)$ ,  $p^L(x, 0) = p_0^L(x)$  and where  $\beta_H = \theta\delta$  and  $\beta_L = (1 - \theta)\delta$  are the birth rates in the two regimes.<sup>32</sup>

While we are not aware of an analytic solution method for the system of partial differential equations (31), this system can be conveniently analyzed by means of Laplace transforms as in Section 4.2. In particular,  $\widehat{p}^H(\xi, t)$  and  $\widehat{p}^L(\xi, t)$  satisfy

$$\widehat{p}_t^H(\xi, t) = -\lambda_H(\xi)\widehat{p}^H(\xi, t) + \beta_H, \quad \lambda_H(\xi) := \xi\mu_H - \xi^2\frac{\sigma_H^2}{2} + \psi + \delta, \quad (32)$$

$$\widehat{p}_t^L(\xi, t) = -\lambda_L(\xi)\widehat{p}^L(\xi, t) + \psi\widehat{p}^H(\xi, t) + \beta_L, \quad \lambda_L(\xi) := \xi\mu_L - \xi^2\frac{\sigma_L^2}{2} + \delta, \quad (33)$$

with initial conditions  $\widehat{p}^H(\xi, 0) = \widehat{p}_0^H(\xi)$ ,  $\widehat{p}^L(\xi, 0) = \widehat{p}_0^L(\xi)$ . Importantly, for fixed  $\xi$ , this is again simply a system of ordinary (rather than partial) differential equations and it can be solved analytically. In particular, note that the system is triangular so that one can first solve the equation for  $\widehat{p}^H(\xi, t)$  and then the one for  $\widehat{p}^L(\xi, t)$ .<sup>33</sup>

**Proposition 4** (Speed of convergence with heterogeneity in mean growth rates) *Consider the cross-sectional distribution  $p(x, t) := p^H(x, t) + p^L(x, t)$ . The stationary distribution  $p_\infty(x) = p_\infty^H(x) + p_\infty^L(x)$  has a Pareto tail for large  $x$  with tail exponent  $\zeta = \min\{\zeta_L, \zeta_H\}$  where  $\zeta_H$  is the positive root of  $0 = \zeta^2\frac{\sigma_H^2}{2} + \zeta\mu_H - \psi - \delta$  and  $\zeta_L$  is the positive root of  $0 = \zeta^2\frac{\sigma_L^2}{2} + \zeta\mu_L - \delta$ . Next consider the Laplace transforms of  $p^H(x, t)$  and  $p(x, t) := p^H(x, t) + p^L(x, t)$ . Their time*

<sup>32</sup>Assuming that the fraction of individuals in the high-growth regime is stationary, it equals  $\theta\delta/(\psi + \delta)$ .

<sup>33</sup>Proposition 4 can easily be extended to the case where the system is not triangular, i.e. if the low state is not an absorbing state and low types can switch to being high types. This is achieved by writing the analogue of (32) in matrix form. The speed of convergence is then governed by the eigenvalues of that matrix. In the triangular case, these eigenvalues are simply  $-\lambda_L(\xi)$  and  $-\lambda_H(\xi)$ . The triangularity simplifies the analysis but it is by no means a crucial assumption. Results for the general case are available on request.

paths are given by

$$\widehat{p}^H(\xi, t) - \widehat{p}_\infty^H(\xi) = e^{-\lambda_H(\xi)t}(\widehat{p}_0^H(\xi) - \widehat{p}_\infty^H(\xi)), \quad (34)$$

$$\widehat{p}(\xi, t) - \widehat{p}_\infty(\xi) = c_H(\xi)e^{-\lambda_H(\xi)t} + c_L(\xi)e^{-\lambda_L(\xi)t}, \quad (35)$$

$$\lambda_H(\xi) := \xi\mu_H - \xi^2\frac{\sigma_H^2}{2} + \psi + \delta, \quad (36)$$

$$\lambda_L(\xi) := \xi\mu_L - \xi^2\frac{\sigma_L^2}{2} + \delta, \quad (37)$$

where  $\widehat{p}_\infty^H(\xi)$  and  $\widehat{p}_\infty(\xi)$  are the Laplace transforms of the stationary distributions and  $c_H(\xi)$  and  $c_L(\xi)$  are constants of integration.<sup>34</sup> Finally, the weighted  $L^1$ -norm of the distribution of individuals in the high-growth regime satisfies  $\|p^H(x, t) - p_\infty^H(x)\|_\xi \sim ke^{-\lambda_H(\xi)t}$ .

The transition dynamics of the income distribution therefore take place on two different time scales: part of the transition happens at rate  $\lambda_H(\xi)$  and another part at rate  $\lambda_L(\xi)$ . The model therefore has the theoretical potential to explain fast short-run dynamics – something we explore quantitatively in Section 5.4.

A natural assumption is that the rate of switching from the high- to the low-growth regime  $\psi$  is large enough to swamp any differences between the  $\mu$ 's and  $\sigma$ 's in the two states and so  $\lambda_H(\xi) > \lambda_L(\xi)$  in (36) and (37). Consider now a change in parameter values that triggers transition dynamics. In contrast to the baseline random growth model of section 4, these now take place on two different time scales: part of the transition happens quickly at rate  $\lambda_H(\xi)$ , but the other part of the transition happens at a much slower pace  $\lambda_L(\xi)$ . In the short-run, the dynamics governed by  $\lambda_H(\xi)$  dominate whereas in the long-run the slower dynamics due to  $\lambda_L(\xi)$  determine the dynamics of the income distribution. We argue in the next section that such a model has the potential to explain the observed rise in income inequality.

---

<sup>34</sup>These are given by

$$\begin{aligned} \widehat{p}_\infty^H(\xi) &= \frac{\beta_H}{\lambda_H(\xi)}, & \widehat{p}_\infty(\xi) &= \frac{\beta_H}{\lambda_H(\xi)} + \frac{\beta_L}{\lambda_L(\xi)} + \psi \frac{\beta_H}{\lambda_L(\xi)\lambda_H(\xi)} \\ c_H(\xi) &:= \frac{\lambda_H(\xi) - \lambda_L(\xi) - \psi}{\lambda_H(\xi) - \lambda_L(\xi)} (\widehat{p}_0^H(\xi) - \widehat{p}_\infty^H(\xi)) \\ c_L(\xi) &:= (\widehat{p}_0^L(\xi) - \widehat{p}_\infty^L(\xi)) + \frac{\psi}{\lambda_H(\xi) - \lambda_L(\xi)} (\widehat{p}_0^H(\xi) - \widehat{p}_\infty^H(\xi)). \end{aligned}$$

### 5.3 Superstar Shocks: Deviations from Gibrat’s Law

Next consider the special case of (29) with deviations from Gibrat’s law  $S_t \neq 0$ , but without jumps  $dN_{it} = 0$ . We call the shocks  $S_t$  “aggregate shocks to high incomes” or “superstar shocks”: the reason is that they affect high incomes more than low incomes and more than proportionately so. We explain why a model with such “superstar shocks” has the potential to generate fast transition dynamics like those observed in the data. We then provide an economic interpretation of such “superstar shocks”.

**The model with superstars generates fast transitions** Indeed, suppose that the process for the underlying process  $y_{it}$  does not change, but  $S_t$  varies. Given that  $x_{it} = e^{S_t} y_{it}$ , we have:

$$\zeta_t^x = e^{-S_t} \zeta_t^y$$

instantaneously. Suppose that the distribution of  $y_{it}$  is constant over time (so  $\zeta_t^y$  is constant). The power law of  $x$  changes immediately when there is a shock to  $S_t$ . Hence, the process is extremely fast – it features instantaneous transitions in the power law exponent, and if  $S_t$  has a secular trend the power law exponent inherits this trend.

**Proposition 5** (Infinitely fast adjustment in models with aggregate shocks to high incomes, aka “superstar shocks”) *Consider the process (29), reproduced here as:  $x_{it} = e^{S_t} y_{it}$ , where  $y_{it}$  is assumed to have a constant law of motion (and distribution) and  $S_t$  is an aggregate shock to high incomes. This process has an infinitely fast speed of adjustment:  $\lambda = \infty$ . Indeed, we have  $\zeta_t^x = e^{-S_t} \zeta_t^y$ , where  $\zeta_t^x, \zeta_t^y$  are the power law exponent of incomes  $x_{it}$  and  $y_{it}$ .*

**Proof.** The mechanism is so basic that the proof is very simple: if  $\mathbb{P}(y_{it} > z) = K e^{-\zeta^y z}$ ,

$$\mathbb{P}(x_{it} > x) = \mathbb{P}(e^{S_t} y_{it} > x) = \mathbb{P}(y_{it} > e^{-S_t} x) = K e^{-\zeta^y e^{-S_t} x} \Rightarrow \zeta_t^x = e^{-S_t} \zeta_t^y. \square$$

Hence, those processes are promising, as they do generate a fast transition.

**Those fast transitions are empirically relevant at high frequencies** There is supportive evidence for the aggregate shocks to high incomes (29) in Parker and Vissing-Jorgensen (2010). They find that in good (respectively bad) times, the incomes of top earners increase (respectively decrease), in a manner consistent with (29) : the sensitivity to the shock at time  $t$  is proportional to  $x_{it}$ , as in

$$dx_{it} = x_{it} dS_t + \mu dt + \sigma dZ_{it}.$$

Note that the shock  $x_{it}dS_t$  to log income is multiplicative in log income, as opposed to additive as in traditional random growth model. This finding appears to have been amplified and broadly confirmed by Guvenen (2015, p.40).

We conclude that those aggregate shocks to high incomes are an empirically grounded source of fast transitions. Next, we present an illustrative model than gives them a concrete theoretical interpretation.

**A microfoundation for “superstar shocks”** We here provide a microfoundation for the  $e^{S_t}$  term in equation (29). Here we adopt the model of Gabaix and Landier (GL, 2008), which is a tractable, calibratable version of the “superstars economics” ideas of Rosen (1981). There is a continuum of firms of different size and a continuum of managers with different talent. A CEO of talent  $T$ , matched with firm  $S$ , produces an improvement in firm value  $CTS^\gamma$ , where  $C$  is a constant. We use the term “CEO” but it could correspond to top lawyers, financiers, etc. Since talented CEOs are more valuable in larger firms, the  $n$ th most talented manager is matched with the  $n$ th largest firm in competitive equilibrium, and earns the following competitive equilibrium pay (really  $n$  is a quantile, so that a low  $n$  means a high talent or firm size). We assume a Pareto firm size distribution with exponent  $1/\alpha$  so that a firm of rank  $n$  has size:  $S(n) = An^{-\alpha}$ . A manager of talent rank  $n$  has talent  $T(n) = T_{\max} - \frac{B}{\beta}n^\beta$ . Hence, the value added of the manager is  $\left(T_{\max} - \frac{BC}{\beta}n^\beta\right)A^\gamma n^{-\alpha\gamma}$ .

GL (summarized in the appendix) derive that for a CEO of upper quantile  $n$ , the market equilibrium pay is:

$$w(n) = e^{a_t}n^{-\chi_t},$$

where  $n$  is the rank (quantile) of that CEO’s talent, and we define:

$$\chi_t = \alpha_t\gamma_t - \beta_t. \tag{38}$$

Denoting the log wage by  $x_{it} = \log w_{it}$  and the log quantile by  $y_{it} = -\log n_{it}$ , we have

$$x_{it} = \chi_t y_{it} + a_t. \tag{39}$$

Note that  $\chi_t$  and  $a_t$  depend on economy-wide forces, while  $x_{it}, y_{it}$  are specific to each CEO. CEO  $i$ ’s talent  $y_{it}$  evolves stochastically as  $dy_{it} = \mu_t dt + \sigma_t dZ_{it}$  with death rate  $\delta_t$ .<sup>35</sup> Hence, this generates a process exactly like (29) with  $e^{S_t} := \chi_t$ .

---

<sup>35</sup>By construction, the quantile  $n_{it} = e^{-y_{it}}$  has a uniform distribution. Therefore  $y_{it}$  must have an exponential distribution with exponent 1 which imposes the restriction  $\mu_t + \frac{1}{2}\sigma_t^2 - \delta_t = 0$ .

A natural variant is one with two growth regimes corresponding to two groups of individuals: high types who have the chance of becoming CEOs, and for whom  $b_H = 1$ , and low types who never become CEOs and who are, therefore, also not affected by “superstar shocks,”  $b_L = 0$ .

A shock to  $\chi_t$  can be a shock to  $\alpha_t, \beta_t, \gamma_t$ . We find it simplest to think about a shock to  $\beta_t$  or  $\gamma_t$ , i.e. to the (perceived) importance of talent in the production function.<sup>36</sup> In the baseline GL calibration,  $\beta_t \simeq 2/3$  and  $\gamma_t \simeq 1$ . When  $\gamma_t$  is higher or  $\beta_t$  is lower, the marginal impact of talent,  $CT'(n)S(n) = A^\gamma BCn^{1-\chi_t}$  is higher. Increases in  $\chi_t$  correspond to increases in the “span of control” or “scales managed by talent”, somewhat as in Garicano and Rossi-Hansberg (2006).

## 5.4 Quantitative Exploration: Fast Transitions in the Augmented Model

We now use the framework of this section to revisit the rise in income inequality. We argued in section 4.3 that the standard random growth model fails spectacularly in terms of explaining the rise in top income inequality in the United States. We now argue that, in contrast, the model with heterogeneity in mean growth rates presented in the preceding sections has the potential to explain the observed rise in top income inequality.

We conduct an analogous exercise to that in Section 4.3. The shock we consider in the present exercise is an increase in the mean growth rate of individuals in the high growth regime  $\mu_H$  (while  $\mu_L$  is unchanged). This is motivated in part by casual evidence of very rapid income growth rates since the 1980s, for instance for Bill Gates, Mark Zuckerberg, hedge fund managers and the like – their growth is very high for a while, then tails off. This impression is confirmed by Jones and Kim (2014), who find that there has been a substantial increase in the average growth rate in the upper tail of the growth rate distribution since the late 1970s.<sup>37</sup> We follow a similar calibration strategy as in Section 4.3. First, note from

---

<sup>36</sup>Indeed,  $1/\alpha_t$  is the tail exponent of the firm size distribution, so it is closely pinned to  $\alpha_t = 1$  (Zipf’s law).

<sup>37</sup>Jones and Kim (2014) proxy  $\mu_H$  with the median of the upper decile, i.e. the 95th percentile, of the distribution of income growth rates. Combining evidence from the IRS public use panel of tax returns and from Guvenen, Ozkan and Song (2014), they show that this measure of  $\mu_H$  has increased substantially from 1979-81 to 1988-90, and then again from 1988-90 to 1995-96. Jones and Kim note that this evidence should be viewed as suggestive due to limited sample sizes in the IRS data and comparability of the IRS and the Social Security Administration data used by Guvenen, Ozkan and Song (2014). Below we discuss ongoing work and directions for future work that could improve on these estimates. In the meantime, Jones and Kim

Proposition 2 that, if  $\mu_H$  is sufficiently bigger than  $\mu_L$ , the Pareto tail of the stationary income distribution is determined only by the dynamics in the high regime and given by

$$\zeta = \min\{\zeta_L, \zeta_H\} = \frac{-\mu_H + \sqrt{\mu_H^2 + 2\sigma_H^2(\delta + \psi)}}{\sigma_H^2}, \quad (40)$$

and the parameters  $\sigma_L$  and  $\mu_L$  do not matter for top inequality in either the initial or terminal steady state. As before, we set  $\delta = 1/30$  and impose that the economy is initially in a Pareto steady state with  $\zeta_{1973} = 2.56$ . We set  $\sigma_H = 0.15$ , which is a conservative estimate.<sup>38</sup> We do not have precise estimates for  $\psi$ , the rate of switching from the high- to the low-growth regime. For our baseline results, we set  $\psi = 1/6$ , corresponding to an expected duration of the high-growth regime of 6 years, and we report results under alternative parameter values. Given values for  $\sigma_H, \delta$  and  $\psi$ , we calibrate the initial  $\mu_H$  so that (40) yields  $\zeta_{1973} = 2.56$ . In the initial steady state, the difference in mean growth rates between high- and low growth types is  $\mu_H - \mu_L = 0.06$ .

Our baseline exercise considers a once-and-for-all increase in  $\mu_H$  by 8 percentage points. This increase in  $\mu_H$  is smaller in total size than the increase documented by Jones and Kim (2014), but also much more abrupt. The resulting gap of  $\mu_H - \mu_L = 0.14$  is broadly consistent with empirical evidence in Guvenen, Kaplan and Song (2014).<sup>39</sup> Figure 6 plots the corresponding results. The difference to the earlier experiment in Figure 5 is striking. The model with heterogeneity in mean growth rates can generate transition dynamics that can replicate the rapid rise in income inequality observed in the United States.

The key parameters that govern the speed of transition are  $\mu_H$  and  $\psi$ , the growth rate in the high-growth regime and the probability of leaving it. Since we do not have precise estimates of these two parameters, we have explored a number of alternative parameterizations. In particular, we have computed results for both higher and lower switching rates  $\psi = 1/3$  and  $\psi = 1/12$ . In each case, we use (40) to recalibrate the initial  $\mu_H$  so as to match the tail inequality observed in the data. As expected given our theoretical results, transitions are

---

provide the best available evidence documenting potential drivers of the increase in top income inequality.

<sup>38</sup>Larger values of  $\sigma_H$  lead to even faster transition dynamics. We set  $\sigma_L = 0.1$  based on the evidence discussed in Section 4.3. We view  $\sigma_H = 0.15$  as conservative because the growth rates of parts of the population may be much more volatile (think of startups).

<sup>39</sup>Guvenen, Kaplan and Song (2014) document differences in average growth rates of different population groups as large as 0.23 log points per year. See in particular their Figure 7 which presents the age-earnings profiles of three groups: the top 0.1%, the next 0.9% and the bottom 99% of *lifetime* earnings. Over a ten year period (ages 25 to 35), the earnings of the top 0.1% grow by 2.3 log points (0.23 per year), those of the next 0.9% by 1.4 log points (0.14 per year) and those of the bottom 99% by 0.5 log points (0.05 per year).

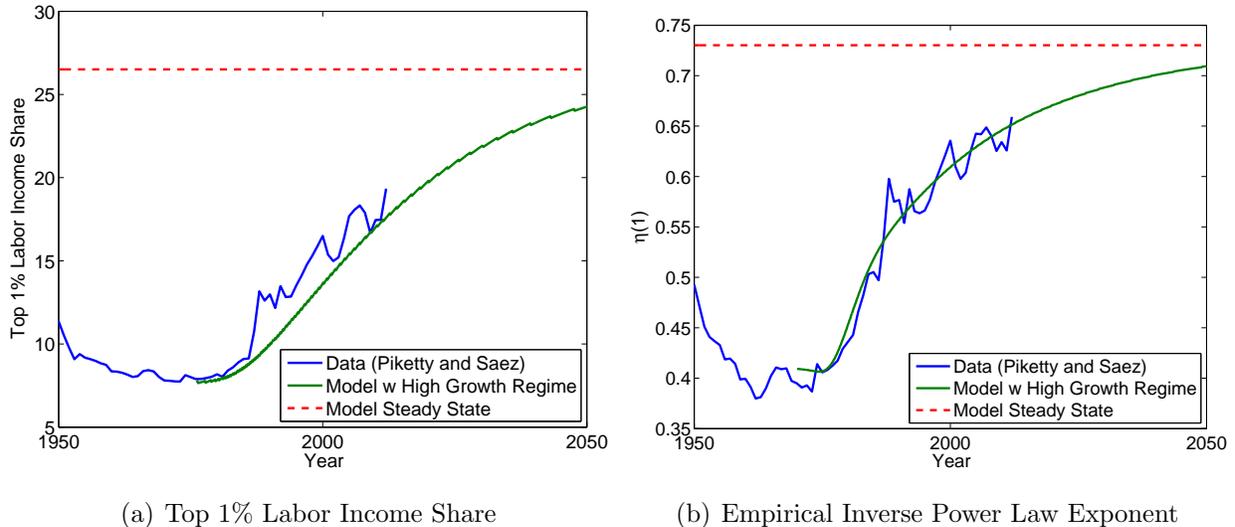


Figure 6: Transition Dynamics in Model with Heterogeneity in Mean Growth Rates

fastest when  $\psi$  and  $\mu_H$  are high, i.e. when individuals can experience very short-lived, very high-growth spurts, what one may call “live-fast-die-young dynamics”.<sup>40</sup> In their ongoing work using a very similar model, Jones and Kim (2014) propose such a “live-fast-die-young” calibration with very low  $\psi$  and high  $\mu_H$ .<sup>41</sup> Second, we have also conducted experiments in which we feed in a gradual increase in  $\mu_H$  rather than a once-and-for-all increase. We find that in this case a higher switching rate  $\psi$  is needed than the one used in our baseline experiment.<sup>42</sup>

In summary, the model with heterogeneity in mean growth rates is capable of generating fast transition dynamics of top inequality for a number of alternative parameterizations that are broadly consistent with the micro data. The common feature of these parameterizations is a combination of relatively high growth rates for part of the population (high enough  $\mu_H$ ) over relatively short time horizons (high enough  $\psi$ ). In the absence of better micro estimates of these critical parameter values, the quantitative explorations in this section

<sup>40</sup>On the other hand, the speed of convergence becomes close to that in the benchmark model as  $\psi$  and  $\mu_H$  become small. Indeed, as  $\psi \rightarrow 0$  the model collapses to the one-regime model of Section 4.3. This is because in this case we need  $\mu_H \rightarrow \mu_L$  so as to still match data on the tail exponent  $\zeta$  (see (40)).

<sup>41</sup>We have computed transition dynamics using their proposed calibration and confirm that it can generate fast transitions, as also expected given our theoretical results. Results available upon request.

<sup>42</sup>A parameter combination that generates time paths quite similar to those in Figure 6 is a gradual increase over a time period of 20 years (1973 to 1993) of  $\mu_H$  by 11%, together with  $\psi = 1/4$ . This is still more conservative than the calibration of Jones and Kim who feature a larger increase of  $\mu_H$  and higher  $\psi$ .

should be viewed as suggestive. An advantage of our tractable theoretical approach is that it pinpoints clearly the important parameters. Better empirical evidence is a clear priority for future research. For instance, it would be extremely valuable to extend the work of Guvenen, Kaplan and Song (2014) and Guvenen et al. (2015), who use administrative data from the Social Security Administration to document the drivers of the increase in top inequality. Better statistical methods could also help. For example, our model with multiple growth regimes suggests estimating a version of Hamilton’s (1989) regime-switching model using labor income panel data.

## 6 The Dynamics of Wealth Inequality

We now show how our theoretical results can be extended to the simple model of top wealth inequality in Section 3.2. We then ask whether an increase in  $r - g$ , the gap between the after-tax average rate of return and the growth rate, can explain the increase in top wealth inequality observed in some datasets as suggested by Piketty (2014).

### 6.1 Stationary Wealth Inequality

The properties of the stationary wealth distribution are again well understood. Applying the standard results in Appendix C, one can show that the stationary wealth distribution has a Pareto tail with tail inequality

$$\eta = \frac{1}{\zeta} = \frac{\sigma^2/2}{\sigma^2/2 - (r - g - \theta)} \quad (41)$$

provided that  $r - g - \theta - \sigma^2/2 < 0$ . Intuitively, tail inequality is increasing in the gap between the after-tax rate of return to wealth and the growth rate  $r - g$ . Similarly, tail inequality is higher the lower the marginal propensity to consume  $\theta$  and the higher the after-tax wealth volatility  $\sigma$  are. Given that  $r = (1 - \tau)\tilde{r}$  and  $\sigma = (1 - \tau)\tilde{\sigma}$ , top wealth inequality is also decreasing in the capital income tax rate  $\tau$ . Intuitively, a higher gap between  $r$  and  $g$  works as an “amplifier mechanism” for wealth inequality: for a given structure of shocks ( $\sigma$ ), the long-run magnitude of wealth inequality will tend to be magnified if the gap  $r - g$  is higher (Piketty and Zucman, 2014*b*). However, this leaves unanswered the question whether increases in top wealth inequality triggered by an increase in  $r - g$  will come about quickly or take many hundreds of years to materialize.

## 6.2 Dynamics of Wealth Inequality: Theoretical Results

We now show how our theoretical results can be extended to the case of wealth dynamics. Individual wealth satisfies (2) or, equivalently, (3) in combination with an additive income term  $ydt$  (friction 4). From Ito's formula, the logarithm of wealth  $x_{it} = \log w_{it}$  satisfies

$$dx_{it} = (ye^{-x_{it}} + \mu)dt + \sigma dZ_{it} \quad (42)$$

where the reader should recall that  $y$  denotes labor income. The addition of this labor income term  $y$  introduces some difficulties for extending Proposition 1. However, note that for large wealth levels this term becomes negligible, which makes it possible to derive a tight *upper bound* on the speed of convergence of the cross-sectional distribution.

**Proposition 6** (Speed of convergence for wealth dynamics (friction 4)) *Consider the wealth process (42). Under Assumption 1, and if  $\mu < 0$ , the cross-sectional distribution  $p(x, t)$  converges to its stationary distribution exponentially in the total variation norm, that is  $\|p(x, t) - p_\infty(x)\| \sim ke^{-\lambda t}$  for constants  $k$  and  $\lambda$ . The rate of convergence satisfies*

$$\lambda \leq \frac{1}{2} \frac{\mu^2}{\sigma^2} + \delta$$

where  $\mathbf{1}_{\{\cdot\}}$  is the indicator function, and with equality for  $|\mu|$  below a threshold  $|\mu^*|$ .

We conjecture that with  $\mu > 0$ ,  $\lambda = \delta$ , as in Proposition 1.

In the case of friction 4, it is not possible to obtain an exact formula for the speed of convergence. However, the speed of convergence is bounded above and, in particular, is equal to or less than the speed with a reflecting barrier from Proposition 1. For the case of the process (42), it is also no longer possible to characterize the corresponding Kolmogorov Forward equation by using Laplace transforms (due to the presence of the term  $ye^{-xt}$ ). Numerical experiments nevertheless confirm our results from section 4.2 that the speed of convergence in the tail can be substantially lower than the average speed of convergence characterized in Proposition 6.

## 6.3 Wealth Inequality and Capital Taxes

In this section, we ask whether an increase in  $r - g$ , the gap between the (average) after-tax rate of return on wealth and the economy's growth rate, can explain the increase in wealth inequality observed in some data sets, as suggested by Piketty (2014). To do so, we first

construct a measure of the time series of  $r - g$ . This requires three data inputs: on the average pre-tax rate of return, on capital income taxes, and on a measure of the economy’s growth rate. We use data on the average before-tax rate of return from Piketty and Zucman (2014a), the series of top marginal capital income tax rates from Auerbach and Hassett (2015) and data on the growth rate of per capita GDP of the United States from the Penn World Tables. Panel (a) of Figure 7 plots our time series for  $r - g$ , displaying a strong upward trend starting in the late 1970s, which coincides with the time when top wealth inequality started to increase (Figure 1).<sup>43</sup> The Figure therefore suggests that, a priori, the theory that variations in  $r - g$  are a potential candidate for explaining increasing wealth inequality.

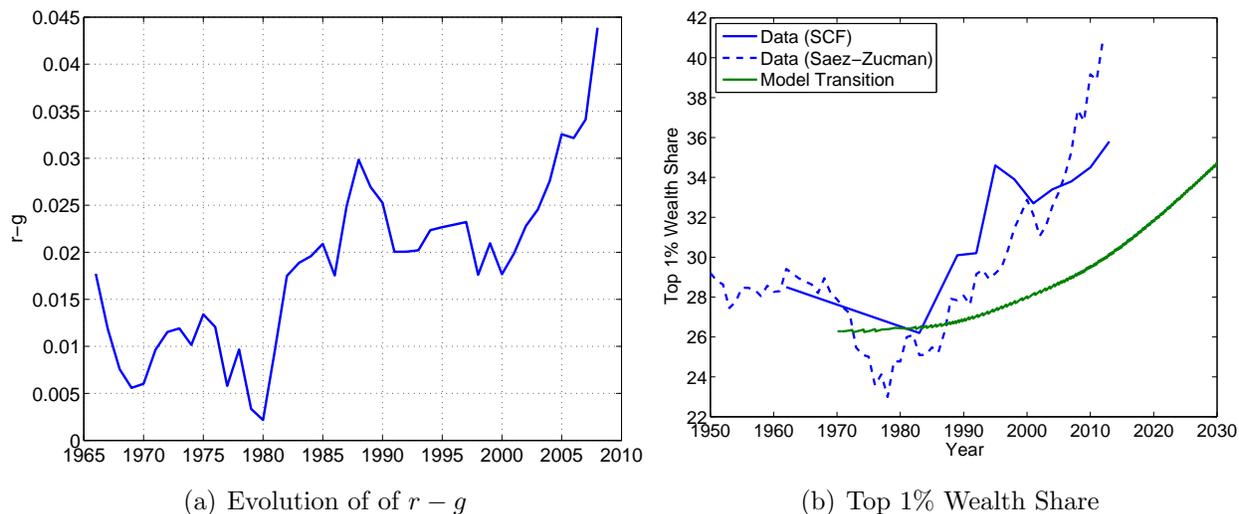


Figure 7: Dynamics of Wealth Inequality in the Baseline Model

We now ask whether the simple model of wealth accumulation from Section 3 has the potential to explain the different data series for wealth inequality in Figure 1. To this end, recall equation (2) and note that the dynamics of this parsimonious model are described by two parameter combinations only,  $r - g - \theta$ , where  $\theta$  is the marginal propensity to consume out of wealth, and the cross-sectional standard deviation of the return to capital  $\sigma$ . Our exercise proceeds in three steps. First, we obtain an estimate for  $\sigma$ . We use  $\sigma = 0.3$ , which is on the upper end of values estimated or used in the existing literature.<sup>44</sup> Second, given  $\sigma$

<sup>43</sup>We have tried a number of alternative exercises with different data series for the return on capital and taxes, e.g. we set the pre-tax  $r$  equal to the yields of 10-year government bonds as in Auerbach and Hassett (2015) and Piketty and Zucman (2014b). Results are very similar and available upon request.

<sup>44</sup>Overall, good estimates of  $\sigma$  are quite hard to come by and relatively dispersed. Campbell (2001) provide the only estimates for an exactly analogous parameter using Swedish wealth tax statistics on asset returns.

and our data for  $r - g$  in 1970, we calibrate the marginal propensity to consume  $\theta$  so as to match the amount of tail inequality observed in the data in 1970,  $\eta = 0.6$ . Third, we feed the time path for  $r - g$  from panel (a) of Figure 7 into the calibrated model.

Before comparing the model’s prediction to the evolution of top wealth inequality in the data, we make use of our analytic formulas from section 4 to calculate measures of the speed of convergence. To this end, revisit the average speed of convergence in Proposition 1, and in particular the formula in terms of inequality (13). To operationalize this formula, we use the tail exponent observed in 2010 in the SCF of  $\eta = 0.65$  together with our other parameter values.<sup>45</sup> With these numbers in hand we obtain a half life of

$$t_{1/2} \geq \frac{\log(2) \times 8 \times \eta^2}{\sigma^2} = \frac{\log(2) \times 8 \times (0.65)^2}{0.3^2} \approx 26 \text{ years.}$$

That is, on average, the distribution takes 26 years to cover half the distance to the new steady state. Panel (b) of Figure 7 displays the results of our experiment using the parameter values just discussed. The main takeaway is that the baseline random growth model cannot even explain the gradual rise in top wealth inequality found in the SCF. It fails even more obviously in explaining the rise in top wealth inequality found by Saez and Zucman (2014).

## 6.4 Fast Dynamics of Wealth Inequality

What, then, explains the dynamics of wealth inequality observed in the data? The lessons of Section 5 still apply. In particular, processes of the form (29) that feature heterogeneity in mean growth rates or deviations from Gibrat’s law have the potential to deliver fast transitions. We view both as potentially relevant for the case of wealth dynamics. First, average growth rates may differ across individuals if some individuals have higher average returns, e.g. because they can afford better investment advice, and the rates of return of “superstar investors” relative to those of others may have increased over time, e.g. because their effective capital income taxes have decreased. Second, there may be deviations from Gibrat’s law, for example, because the saving rates of the super wealthy relative to those of the wealthy may change over time (Saez and Zucman, 2014).<sup>46</sup>

---

They estimate an average  $\sigma$  of 0.18. Moskowitz and Vissing-Jorgensen (2002) argue for  $\sigma$  of 0.3.

<sup>45</sup>Ideally one would use an estimate of tail inequality in the new stationary distribution  $\eta$ . Since  $\lambda$  is decreasing in inequality, we use the tail exponent observed in 2010 in the SCF of  $\eta = 0.65$ , which provides an upper bound on the speed of convergence  $\lambda$ . Since inequality in the new stationary distribution may be even higher, true convergence may be even slower.

<sup>46</sup>Finally, it is natural to ask whether the extension to multiple distinct growth regimes of Section 5.2 can generate fast transition dynamics in response to the increase in  $r - g$  from Section 6.3. Numerical experiments

## 7 Conclusion

This paper makes two contributions. First, it finds that the standard random growth models cannot explain rapid changes in tail inequality, for robust analytical reasons. This required developing new tools to analyze transition dynamics, as most previous literature could analyze only separate steady states, without being able to assess analytically the speed of transition between them and without identifying the above-mentioned important defect of the standard model. Second, it suggests two parsimonious deviations from the basic model that can explain such fast changes: (i) heterogeneity in mean growth rates and (ii) “superstar shocks” that disproportionately affect top earners (in the sense that the shock to log income is multiplicative in log income, as opposed to additive as in traditional model). We view them as promising, because they have some support in the data (as we argued above, see especially Jones and Kim (2014), Parker and Vissing-Jorgensen (2010) and Guvenen (2015)). We hope that future research explores the importance of heterogeneity in mean growth rates in more detail. The forces we have analyzed in this paper may serve to guide future empirical and theoretical work on the determinants of fast changes in inequality.

## Appendix

### A Proof of Proposition 1

Proposition 1 is concerned with two different cases. The first case involves models with death and reinjection where the dynamics without those terms are not ergodic. The second one concerns ergodic models (with or without insertion and killing). The strategy of the proof in both cases is different and we therefore present the two cases separately. In the first case (“non-ergodic case”), the rate of convergence is obtained by directly analyzing the dynamics of the  $L^1$  norm (6) of the cross-sectional distribution. In the second case (“ergodic case”), the convergence is to a real invariant measure and the rate of convergence is obtained by a spectral analysis (in particular it is given by the so-called “spectral gap”).

#### A.1 Proof of Proposition 1: Case without a reflecting barrier

We here study the case without the reflecting barrier, starting with a generally useful lemma.

---

suggest that the answer is no.

**Lemma 3** Suppose that a function  $q(x, t)$  solves  $q_t = \mathcal{A}q$  with  $\mathcal{A}q = a(x, t)q + b(x, t)q_x + c(x, t)q_{xx}$  with  $c(x, t) \geq 0$  for all  $x$ . Then  $|q(x, t)|$  is a “subsolution” of the same equation, that is

$$|q|_t \leq \mathcal{A}|q|. \quad (43)$$

**Proof of Lemma 3:** The key is that  $|q|$  is a convex function of  $q$ . Assume  $\varphi$  is a  $C^2$  convex function and set  $z = \varphi(q)$ . Then  $z_t = \varphi'(q)q_t$ ,  $z_x = \varphi'(q)q_x$ ,  $z_{xx} = \varphi''(q)q_x^2 + \varphi'(q)q_{xx}$ , so:

$$z_t - \mathcal{A}z = \varphi'(q) \underbrace{[q_t - bq_x - cq_{xx}]}_{=aq} - az - c \underbrace{\varphi''(q)}_{\geq 0} \underbrace{q_x^2}_{\geq 0} \leq a(\varphi'(q)q - \varphi(q)).$$

Take  $\varphi(q) = \varphi^{(\varepsilon)}(q) = \sqrt{\varepsilon^2 + q^2}$  for some  $\varepsilon > 0$  and  $z^{(\varepsilon)} = \varphi^{(\varepsilon)}(q)$ . Then  $\varphi'(q)q - \varphi(q) = \frac{q^2}{\sqrt{\varepsilon^2 + q^2}} - \sqrt{\varepsilon^2 + q^2} = \frac{-\varepsilon^2}{\sqrt{\varepsilon^2 + q^2}} \in [-\varepsilon, 0]$ , so  $z_t^{(\varepsilon)} - \mathcal{A}z^{(\varepsilon)} \leq |a(x, t)|\varepsilon$ . As  $\varepsilon \rightarrow 0$ ,  $z^{(\varepsilon)} \rightarrow |q|$ , so this inequality becomes:  $|q|_t - \mathcal{A}|q| \leq 0$ .  $\square$

We next apply Lemma 3 to  $q(x, t) := p(x, t) - p_\infty(x)$  to prove a useful inequality. We note that since  $p_t = \mathcal{A}^*p$  and  $0 = \mathcal{A}^*p_\infty$ , we have  $q_t = \mathcal{A}^*q = -\mu q_x + \frac{\sigma^2}{2}q_{xx} - \delta q$ .

**Lemma 4** The decay rate of the  $L^1$  norm  $m(t) := \|q(\cdot, t)\|$  is at least  $\delta$ :  $\lambda \geq \delta$ .

**Proof of Lemma 4:** We have  $m(t) := \|q(\cdot, t)\| = \int |q(x, t)| dx$  and hence

$$m'(t) = \int |q(x, t)|_t dx \leq \int \left( -\delta |q| - \mu |q|_x + \frac{\sigma^2}{2} |q|_{xx} \right) dx = -\delta \int |q| dx$$

where the inequality follows from Lemma 3 and the last equality from the boundary conditions corresponding to  $p$ . Hence  $m'(t) \leq -\delta \int |q| dx = -\delta m(t)$  and therefore  $m(t) \leq e^{-\delta t} m(0)$  by Grönwall’s lemma.  $\square$

**Lemma 5** The decay rate of the  $L^1$  norm  $m(t) := \|q(\cdot, t)\|$  is at most  $\delta$ :  $\lambda \leq \delta$ .

**Proof of Lemma 5:** Define  $r(x, t)$  by  $q(x, t) = e^{-\delta t} r(x, t)$ . Therefore  $\int_{-\infty}^{+\infty} |q(x, t)| dx = e^{-\delta t} \int_{-\infty}^{+\infty} |r(x, t)| dx$ , i.e.  $r(x, t)$  captures the extra rate of decay additional to  $\delta$  (if any). The remainder of the proof shows that this extra rate is zero. To see this note that  $r$  satisfies

$$r_t = -\mu r_x + \frac{\sigma^2}{2} r_{xx} \quad (44)$$

Note further that

$$\int_{-\infty}^{+\infty} |r(x, t)| dx = \int_{-\infty}^{+\infty} |\tilde{r}(x, t)| dx \quad (45)$$

where  $\tilde{r}(x, t) = r(x + \mu t, t)$ . Note that this works only because the limits of integration are  $\pm\infty$  and hence one can simply “translate everything” by  $\mu t$ , i.e. it does not work in the case with a reflecting barrier. From (44), and using that  $\tilde{r}_t = r_t + \mu r_x$ ,  $\tilde{r}$  solves the standard heat equation

$$\tilde{r}_t = \frac{\sigma^2}{2} \tilde{r}_{xx}. \quad (46)$$

It is well-appreciated in the theory of partial differential equations that the solution to the heat equation does not decay exponentially. For completeness, we provide a proof of this fact (the difficulty being only that  $\tilde{r}(x, t)$  could change sign). Suppose by contradiction that

$$\int_{-\infty}^{\infty} |\tilde{r}(x, t)| dx \leq C e^{-\gamma t}$$

for some constant  $C$ , and some  $\gamma > 0$ . Then, for all  $\beta \in (0, \gamma)$ , we have  $\int_{-\infty}^{\infty} e^{\beta t} |\tilde{r}(x, t)| dx \leq C e^{-(\gamma-\beta)t}$ , so that

$$\int_{t=0}^{\infty} \int_{x=-\infty}^{\infty} e^{\beta t} |\tilde{r}(x, t)| dx dt \leq \int_{t=0}^{\infty} C e^{-(\gamma-\beta)t} dt = \frac{C}{\gamma - \beta} < \infty$$

i.e.  $e^{\beta t} \tilde{r}(x, t) \in L^1(\mathbb{R}_+ \times \mathbb{R})$ . Hence  $R(x) := \int_0^{\infty} e^{\beta t} \tilde{r}(x, t) dt$  is defined and  $R(\cdot) \in L^1$ . Using the heat equation (46) (normalizing without loss of generality  $\sigma^2 = 2$ ), we have

$$\begin{aligned} R''(x) &= \int_0^{\infty} e^{\beta t} \tilde{r}_{xx}(x, t) dt = \int_0^{\infty} e^{\beta t} \tilde{r}_t(x, t) dt = [e^{\beta t} \tilde{r}(x, t)]_0^{\infty} - \int_0^{\infty} \beta e^{\beta t} \tilde{r}(x, t) dt \\ &= -\tilde{r}(0, x) - \beta R(x) = -\tilde{r}_0(x) - \beta R(x) \end{aligned}$$

i.e.  $R'' + \beta R = -\tilde{r}_0$ . Hence, taking the Fourier transform ( $\hat{R}(\xi) := \int e^{-i\xi x} R(x) dx$ ),

$$\hat{R}(\xi) (-|\xi|^2 + \beta) = -\hat{\tilde{r}}_0(\xi). \quad (47)$$

This implies that  $\hat{\tilde{r}}_0(\xi) = 0$  for  $\xi = \beta^{1/2}$  (we just plug the value  $\xi = \beta^{1/2}$  in (47)).

But the argument worked for any  $\beta \in (0, \gamma)$ . Hence,  $\hat{\tilde{r}}_0(\xi) = 0$  for any  $\xi \in (0, \gamma^{1/2})$ . We assume that  $|\tilde{r}_0(x)| \leq A e^{-kx}$  for a constant  $A, k > 0$  (see also Assumption 1), which guarantees that  $\hat{\tilde{r}}_0$  is analytic. Given that  $\hat{\tilde{r}}_0$  is analytic and equal to 0 on a segment, we have  $\hat{\tilde{r}}_0(\xi) = 0$  for all  $\xi$ , and  $\tilde{r}_0 = 0$ . We have reached a contradiction.  $\square$

Gathering the arguments. Putting together Lemmas 4 and 5, we obtain that  $\lambda = \delta$ .

## A.2 Proof of Proposition 1: Case with a reflecting barrier

We next study the “ergodic case”: there is a reflecting barrier and additionally  $\mu < 0$ . Then the process (4) is ergodic even with  $\delta = 0$ . In this case, the cross-sectional distribution

satisfies (9) with boundary condition (8). The key insight is that the speed of convergence of  $p$  is governed by the second eigenvalue of the operator  $\mathcal{A}^*$ , and the key step is to obtain an analytic formula for this second eigenvalue given by  $|\lambda_2| = \frac{1}{2} \frac{\mu^2}{\sigma^2} + \delta$ . Before proceeding with the proof, we review some mathematical concepts that will be useful.

### A.2.1 Mathematical Preliminaries

The following definitions are useful. First, the *inner product* of two continuous functions  $u$  and  $v$  is  $\langle u, v \rangle = \int_{-\infty}^{\infty} u(x)v(x)dx$ . Second, for an operator  $\mathcal{A}$ , the *adjoint* of  $\mathcal{A}$  is the operator  $\mathcal{A}^*$  satisfying  $\langle \mathcal{A}u, p \rangle = \langle u, \mathcal{A}^*p \rangle$ . Third, an operator  $\mathcal{B}$  is *self-adjoint* if  $\mathcal{B}^* = \mathcal{B}$ .<sup>47</sup> The following theorem is standard.

**Theorem A.1:** All eigenvalues of a self-adjoint operator are real.

Finally, the *infinitesimal generator* of a Brownian motion reflected at  $x = 0$  and with death at Poisson rate  $\delta$  is the operator  $\mathcal{A}$  defined by

$$\mathcal{A}u = \mu \frac{\partial}{\partial x} u + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} u - \delta u \quad (48)$$

$$\frac{\partial u(0, t)}{\partial x} = 0. \quad (49)$$

**Lemma 6** *The Kolmogorov Forward operator  $\mathcal{A}^*$  in (9) with boundary condition (8) is the adjoint of the infinitesimal generator  $\mathcal{A}$  in (48) with boundary condition (49).*

**Proof of Lemma 6:** Using Definition 2, we have

$$\begin{aligned} \langle \mathcal{A}u, p \rangle &= \int_0^{\infty} \left( \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2} + \mu \frac{\partial u}{\partial x} - \delta u \right) p dx = \frac{\sigma^2}{2} \int_0^{\infty} \frac{\partial^2 u}{\partial x^2} p dx + \mu \int_0^{\infty} \frac{\partial u}{\partial x} p dx - \delta \int_0^{\infty} u p dx \\ &= \frac{\sigma^2}{2} \frac{\partial u}{\partial x} p \Big|_0^{\infty} - \frac{\sigma^2}{2} u \frac{\partial p}{\partial x} \Big|_0^{\infty} + \mu u p \Big|_0^{\infty} + \int_0^{\infty} u \left( \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} - \mu \frac{\partial p}{\partial x} \right) dx - \delta \int_0^{\infty} u p dx \\ &= \int_0^{\infty} u \left( \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} - \mu \frac{\partial p}{\partial x} - \delta p \right) dx = \langle u, \mathcal{A}^* p \rangle \end{aligned}$$

where the fourth equality follows from the boundary conditions (8) and (49).

### A.2.2 Main Proof

With these preliminaries in hand, we proceed with the proof of the Proposition. We present here only the proof for the case  $\delta = 0$  (friction 2). The extension to the case  $\delta > 0$  is

---

<sup>47</sup>Note that the adjoint is the infinite-dimensional analogue of a matrix transpose.

straightforward. The goal is to analyze the eigenvalues of the infinitesimal generator  $\mathcal{A}$  or equivalently its adjoint  $\mathcal{A}^*$ . The difficulty is that  $\mathcal{A}$  is not self-adjoint,  $\mathcal{A}^* \neq \mathcal{A}$ , and therefore its eigenvalues could, in principle, be anywhere in the complex plane. We therefore construct a self-adjoint transformation  $\mathcal{B}$  of  $\mathcal{A}$  as follows.

**Lemma 7** Consider  $u$  satisfying  $u_t = \mathcal{A}u$  and the corresponding stationary distribution,  $p_\infty = e^{(2\mu/\sigma^2)x}$ . Then  $v = up_\infty^{1/2} = ue^{(\mu/\sigma^2)x}$  satisfies

$$v_t = \mathcal{B}v, \quad \mathcal{B} = \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} - \frac{1}{2} \frac{\mu^2}{\sigma^2} \quad (50)$$

with boundary condition  $\frac{\partial v(0,t)}{\partial x} = \frac{\mu}{\sigma^2} v(0,t)$ . Furthermore,  $\mathcal{B}$  is self-adjoint.

**Proof:** (50) follows from differentiating  $v = ue^{(\mu/\sigma^2)x}$ . To see that  $\mathcal{B}$  is self-adjoint, integrate by parts as in Lemma 6 to conclude that for any  $u, p$ ,  $\langle \mathcal{B}u, p \rangle = \langle u, \mathcal{B}p \rangle$ .

**Lemma 8** The first eigenvalue of  $\mathcal{B}$  is  $\lambda_1 = 0$  and the second eigenvalue is  $\lambda_2 = -\frac{1}{2} \frac{\mu^2}{\sigma^2}$ . All remaining eigenvalues satisfy  $|\lambda| > |\lambda_2|$ . Hence the spectral gap of  $\mathcal{B}$  equals  $|\lambda_2| = \frac{1}{2} \frac{\mu^2}{\sigma^2}$ .

**Proof of Lemma 8** Consider the eigenvalue problem  $\mathcal{B}\varphi = \lambda\varphi$  or equivalently

$$\frac{\sigma^2}{2} \varphi''(x) - \frac{1}{2} \frac{\mu^2}{\sigma^2} \varphi(x) = \lambda \varphi(x), \quad (51)$$

$$\varphi'(0) = \frac{\mu}{\sigma^2} \varphi(0). \quad (52)$$

We are looking for non-positive eigenvalues  $\lambda$  and the question is: for what values of  $\lambda \leq 0$  does (51) have a solution  $\varphi(x)$  that satisfies the boundary condition (52) and is in the domain of  $\mathcal{B}$ . The domain of  $\mathcal{B}$  here is the set of  $C^\infty$  functions  $\phi : [0, \infty) \rightarrow \mathbb{C}$  such that  $\phi$  and all its derivatives have at most polynomial growth: for all integers  $n \geq 0$ , there is a  $K_n$  and a  $k_n > 0$  such that  $|\phi^{(n)}(x)| \leq K_n (1 + x^{k_n})$  for all  $x \geq 0$ .

To answer this question, note that for a given  $\lambda \leq 0$ , the general solution to (51) is  $\varphi(x) = c_1 e^{ax} + c_2 e^{-ax}$  where  $a$  satisfies

$$\frac{\sigma^2}{2} a^2 = \frac{1}{2} \frac{\mu^2}{\sigma^2} + \lambda. \quad (53)$$

Consider four different cases:

1.  $\lambda = 0$ . In this case the solution to (53) is  $a = \frac{\mu}{\sigma^2}$ , i.e.  $\varphi(x) = e^{\frac{\mu}{\sigma^2}x}$  which satisfies (52) and stays bounded as  $x \rightarrow \infty$  (since  $\mu < 0$ ). Hence  $\lambda = 0$  is an eigenvalue of  $\mathcal{B}$ .

2.  $-\frac{1}{2}\frac{\mu^2}{\sigma^2} < \lambda < 0$ . In this case,  $a$  solving (53) is real and positive. We therefore need  $c_1 = 0$  so that  $\varphi$  stays bounded as  $x \rightarrow \infty$ . But then (52) becomes  $-a = \frac{\mu}{\sigma^2}$  which is a contradiction. Hence  $\mathcal{B}$  has no eigenvalues in the interval  $\left(-\frac{1}{2}\frac{\mu^2}{\sigma^2}, 0\right)$ .
3.  $\lambda = -\frac{1}{2}\frac{\mu^2}{\sigma^2}$ . In this case, (51) becomes  $\varphi''(x) = 0$  and (52) implies  $\varphi(x) = x + \frac{\sigma^2}{\mu}$  which has polynomial growth. Hence  $\lambda = -\frac{1}{2}\frac{\mu^2}{\sigma^2}$  is an eigenvalue of  $\mathcal{B}$ .
4.  $\lambda < -\frac{1}{2}\frac{\mu^2}{\sigma^2}$ . In this case,  $a$  solving (53) is a purely imaginary number. We have  $e^{ix} = \cos x + i \sin x$ , so  $\varphi(x) = c_1 e^{ax} + c_2 e^{-ax}$  oscillates but stays bounded as  $x \rightarrow \infty$ . We can therefore choose  $c_1, c_2 \neq 0$  to satisfy (52). Hence any  $\lambda < -\frac{1}{2}\frac{\mu^2}{\sigma^2}$  is also an eigenvalue of  $\mathcal{B}$ .

Summarizing,  $\mathcal{B}$  has an isolated eigenvalue  $\lambda_1 = 0$ , the second highest eigenvalue (in absolute value) is  $\lambda_2 = -\frac{1}{2}\frac{\mu^2}{\sigma^2}$ , and all remaining eigenvalues satisfy  $|\lambda| > |\lambda_2|$ .  $\square$

## References

- Acemoglu, Daron, and James A. Robinson.** 2015. “The Rise and Decline of General Laws of Capitalism.” *Journal of Economic Perspectives*, 29(1): 3–28.
- Aoki, Shuhei, and Makoto Nirei.** 2015. “Zipf’s Law, Pareto’s Law, and the Evolution of Top Incomes in the U.S.” Hitotsubashi University Working Paper.
- Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez.** 2011. “Top Incomes in the Long Run of History.” *Journal of Economic Literature*, 49(1): 3–71.
- Auerbach, Alan, and Kevin Hassett.** 2015. “Capital Taxation in the 21st Century.” University of California, Berkeley Working Paper.
- Benhabib, Jess, Alberto Bisin, and Mi Luo.** 2015. “Wealth Distribution and Social Mobility: An Empirical Approach.” NYU Working Papers.
- Benhabib, Jess, Alberto Bisin, and Shenghao Zhu.** 2011. “The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents.” *Econometrica*, 79(1).
- Benhabib, Jess, Alberto Bisin, and Shenghao Zhu.** 2013. “The Distribution of Wealth in the Blanchard-Yaari Model.” NYU Working Papers.
- Benhabib, Jess, Alberto Bisin, and Shenghao Zhu.** 2014. “The Wealth Distribution in Bewley Models with Investment Risk.” NYU Working Papers.
- Bricker, Jesse, Alice M. Henriques, Jacob Krimmel, and John Sabelhaus.** 2015. “Measuring Income and Wealth at the Top Using Administrative and Survey Data.” Board of Governors Finance and Economics Discussion Series 2015-30.

- Campbell, John Y.** 2001. “Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk.” *Journal of Finance*, 56(1): 1–43.
- Champernowne, D. G.** 1953. “A Model of Income Distribution.” *The Economic Journal*, 63(250): pp. 318–351.
- DeBacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos.** 2013. “Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns.” *Brookings Papers on Economic Activity*, 46(1): 67–142.
- Gabaix, Xavier.** 1999. “Zipf’s Law For Cities: An Explanation.” *The Quarterly Journal of Economics*, 114(3): 739–767.
- Gabaix, Xavier.** 2009. “Power Laws in Economics and Finance.” *Annual Review of Economics*, 1(1): 255–293.
- Gabaix, Xavier, and Augustin Landier.** 2008. “Why Has CEO Pay Increased So Much?” *The Quarterly Journal of Economics*, 123(1): 49–100.
- Garicano, Luis, and Esteban Rossi-Hansberg.** 2006. “Organization and Inequality in a Knowledge Economy.” 121(4): 1383–1435.
- Geerolf, Francois.** 2014. “A Static and Microfounded Theory of Zipf’s Law for Firms and of the Top Labor Income Distribution.” UCLA Working Paper.
- Güvenen, Fatih.** 2007. “Learning Your Earning: Are Labor Income Shocks Really Very Persistent?” *American Economic Review*, 97(3): 687–712.
- Güvenen, Fatih.** 2015. “Income Inequality and Income Risk: Old Myths vs. New Facts.” Lecture Slides available at [https://fguvenendotcom.files.wordpress.com/2014/04/handout\\_inequality\\_risk\\_webversion\\_may20151.pdf](https://fguvenendotcom.files.wordpress.com/2014/04/handout_inequality_risk_webversion_may20151.pdf).
- Güvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song.** 2015. “What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?” National Bureau of Economic Research, Inc NBER Working Papers 20913.
- Güvenen, Fatih, Greg Kaplan, and Jae Song.** 2014. “The Glass Ceiling and The Paper Floor: Gender Differences among Top Earners, 1981-2012.” National Bureau of Economic Research, Inc NBER Working Papers 20560.
- Güvenen, Fatih, Serdar Ozkan, and Jae Song.** 2014. “The Nature of Countercyclical Income Risk.” *Journal of Political Economy*, 122(3): 621 – 660.
- Hansen, Lars Peter, and José A. Scheinkman.** 2009. “Long-Term Risk: An Operator Approach.” *Econometrica*, 77(1): 177–234.
- Harrison, Michael.** 1985. *Brownian Motion and Stochastic Flow Systems*. Wiley & Sons.

- Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante.** 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006." *Review of Economic Dynamics*, 13(1): 15–51.
- Jones, Charles I.** 2005. "The Shape of Production Functions and the Direction of Technical Change." *The Quarterly Journal of Economics*, 120(2): 517–549.
- Jones, Charles I.** 2015. "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality." *Journal of Economic Perspectives*, 29(1): 29–46.
- Jones, Charles I., and Jihee Kim.** 2014. "A Schumpeterian Model of Top Income Inequality." Stanford Working Paper.
- Kato, Tosio.** 1995. *Perturbation Theory for Linear Operators*. Springer.
- Kim, Jihee.** 2013. "The Effect of the Top Marginal Tax Rate on Top Income Inequality." KAIST Working Paper.
- Kopczuk, Wojciech.** 2015. "What Do We Know about the Evolution of Top Wealth Shares in the United States?" *Journal of Economic Perspectives*, 29(1): 47–66.
- Kopczuk, Wojciech, Emmanuel Saez, and Jae Song.** 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937." *The Quarterly Journal of Economics*, 125(1): 91–128.
- Linetsky, Vadim.** 2005. "On the Transition Densities for Reflected Diffusions." *Advances in Applied Probability*, 37(2): 435–460.
- Luttmer, Erzo G. J.** 2007. "Selection, Growth, and the Size Distribution of Firms." *The Quarterly Journal of Economics*, 122(3): 1103–1144.
- Luttmer, Erzo G. J.** 2011. "On the Mechanics of Firm Growth." *Review of Economic Studies*, 78(3): 1042–1068.
- Luttmer, Erzo G.J.** 2012. "Slow convergence in economies with firm heterogeneity."
- Luttmer, Erzo G. J.** 2015. "An Assignment Model of Knowledge Diffusion and Income Inequality." Federal Reserve Bank of Minneapolis Staff Report 509.
- MaCurdy, Thomas E.** 1982. "The use of time series processes to model the error structure of earnings in a longitudinal data analysis." *Journal of Econometrics*, 18(1): 83–114.
- Meghir, Costas, and Luigi Pistaferri.** 2011. "Earnings, Consumption and Life Cycle Choices." Vol. 4 of *Handbook of Labor Economics*, Chapter 9, 773–854. Elsevier.
- Moffitt, Robert A., and Peter Gottschalk.** 1995. "Trends in the Transitory Variance of Male Earnings in the U.S., 1969-1987." Working paper.

- Moskowitz, Tobias J., and Annette Vissing-Jorgensen.** 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review*, 92(4): 745–778.
- Nirei, Makoto.** 2009. "Pareto Distributions in Economic Growth Models." Institute of Innovation Research, Hitotsubashi University IIR Working Paper 09-05.
- Pareto, Vilfredo.** 1896. *Cours d'économie politique*.
- Parker, Jonathan A., and Annette Vissing-Jorgensen.** 2010. "The Increase in Income Cyclicity of High-Income Households and Its Relation to the Rise in Top Income Shares." *Brookings Papers on Economic Activity*, 41(2 (Fall)): 1–70.
- Piketty, Thomas.** 2014. *Capital in the Twenty-First Century*. Harvard University Press.
- Piketty, Thomas, and Emmanuel Saez.** 2003. "Income Inequality In The United States, 1913-1998." *The Quarterly Journal of Economics*, 118(1): 1–39.
- Piketty, Thomas, and Gabriel Zucman.** 2014a. "Capital is Back: Wealth-Income Ratios in Rich Countries 1700-2010." 129(3): 1255–1310.
- Piketty, Thomas, and Gabriel Zucman.** 2014b. "Wealth and Inheritance in the Long Run." In *Handbook of Income Distribution. Handbook of Income Distribution*. forthcoming.
- Rosen, Sherwin.** 1981. "The Economics of Superstars." *American Economic Review*, 71(5): 845–58.
- Saez, Emmanuel, and Gabriel Zucman.** 2014. "The Distribution of US Wealth, Capital Income and Returns Since 1913." University of California, Berkeley Working Paper.
- Sattinger, Michael.** 1993. "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature*, 31(2): 831–80.
- Simon, Herbert A.** 1955. "On a Class of Skew Distribution Functions." *Biometrika*, 42(3/4): pp. 425–440.
- Steindl, J.** 1965. "Random Processes and the Growth of Firms: A Study of the Pareto Law."
- Tervio, Marko.** 2008. "The Difference That CEOs Make: An Assignment Model Approach." *American Economic Review*, 98(3): 642–68.
- Toda, Alexis Akira.** 2012. "The double power law in income distribution: Explanations and evidence." *Journal of Economic Behavior & Organization*, 84(1): 364–381.
- Wold, H. O. A., and P. Whittle.** 1957. "A Model Explaining the Pareto Distribution of Wealth." *Econometrica*, 25(4): pp. 591–595.