Abstract

What explains the evolution of top income and wealth inequality in the United States, for example the rapid rise of the top one percent income share over the past forty years? We study the dynamics of inequality in a large class of theories that have been widely used as explanations of top income and wealth inequality at a point in time, namely theories building on a random growth mechanism. We show that standard random growth models generate transition dynamics that are an order of magnitude too slow relative to those observed in the data. We show that fast transitions require specific departures from this benchmark model such as heterogeneity in mean growth rates or deviations from “Gibrat’s law”. Since only very particular economic stories can generate such departures, we use this fact to eliminate the ones that cannot. For instance, we argue that a rise in the variance of the permanent component of earnings is not a promising theory of increasing top income inequality, but theories emphasizing the role of “superstar” entrepreneurs or managers are.
1 Introduction

In the United States the past forty years have seen a rapid rise in top income inequality (Piketty and Saez, 2003; Atkinson, Piketty and Saez, 2011). At least 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 it turns out that the rise in top income inequality has also coincided with a “fattening” of the right tail of the income distribution. To understand the rise in top inequality, we therefore need to understand the forces that have led to a fatter Pareto tail of this distribution. There is also an ongoing debate about the dynamics of top wealth inequality, and to theorize about the dynamics of the wealth distribution we similarly need to understand its Pareto tail.¹

A number of explanations have been advanced for the observed rise in inequality, ranging from superstar effects to falling taxes.² At the same time, there are a large number of existing theories of the Pareto tails of the income and wealth distributions at a point in time. But almost none of these consider whether these theories can also explain the fast rise in top inequality observed in the data, or any fast change for that matter.

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

We start with existing theories that can explain top inequality at a point in time, meaning that they can generate stationary distributions that have Pareto tails. While different theories differ in the underlying economics, many share the same basic mechanism for generating power laws, namely proportional random growth. We therefore study the transition dynamics of the cross-sectional distribution of income or wealth in theories with a random growth mechanism. We then use our results to try to rule out some economic theories of the rise in inequality that are popularly advanced, and to rule in others.

From a theoretical perspective, the main message of our paper is that standard random growth models, like those considered in much of the existing literature, feature extremely slow transition dynamics – so that unadorned they cannot explain the rapid changes that

¹See e.g. Piketty (2014), Saez and Zucman (2014) and Kopczuk (2015).
²See the end of this introduction for a detailed discussion of the related literature.
arise empirically. 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 capital income taxes fall just to take one example. And the question is: what can we say about the speed of this transition, that is will this increase in inequality come about quickly or take a long time? We present two main results. 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 income inequality, and it is also too high to explain even the relatively gradual rise in 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 transitions are even slower in the tail, that is our low measure of the average speed of convergence additionally overestimates the speed of convergence in the upper tail.

Given this negative result, what then explains the observed rise in top income and (potentially) wealth inequality? We argue that fast transitions require very specific departures from the standard random growth model. We develop a “generalized random growth model” that features two such departures that can generate fast changes in inequality. The first departure is cross-sectional heterogeneity in mean growth rates, and in particular a “high growth regime”. For instance, some college graduates may enter the labor force in occupations with much higher average earnings growth rates than others. The second departure consists of deviations from the assumption that the distributions of the growth rate of income and wealth are independent of their levels (“Gibrat’s law”) or “superstar shocks.” We argue that, in contrast to the standard random growth model, this “generalized random growth model” can explain the observed rise in income inequality. In contrast, we 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.

Our result that fast transitions require very particular departures from the benchmark random growth model has important implications for the question what is the economic

\[\text{Guvenen (2009) argues that heterogeneity in mean growth rates is an important feature of the data on income dynamics. Luttmer (2011) studies a similar framework applied to firm dynamics and argues that persistent heterogeneity in mean firm growth rates is needed to account for the relatively young age of very large firms at a given point in time (a statement about the stationary distribution).}\]
mechanism behind the observed increase in top income inequality. This is because only very particular economic stories can generate such departures, and we can therefore use this result to eliminate the ones that cannot. Put differently, starting with a given number of candidate economic theories for the increase in income and wealth inequality, our analysis allows us to eliminate a certain fraction of these. In particular, we argue that the observed rise in top income inequality cannot been due to an increase in the variance of the permanent component of earnings\textsuperscript{4} or simple stories about the disincentive effect of taxes.

In sum, our analytic results show that two types of stories are viable candidate economic theories for the increase in top income inequality: (i) stories in which the very top earners grow faster and in a more volatile way (perhaps due to technology) (ii) “superstar effects” affecting the returns to talent, with something we call “superstars” shocks that affect top income in a way deviating from Gibrat’s law – i.e. affecting extremely high incomes even more than merely high incomes. Similarly, we argue that the “$r-g$ theory” of Piketty (2014) in which increases in top wealth inequality are triggered by an increase in the average rate of return on wealth increases (e.g. due to falling capital taxes) is not a plausible theory because it cannot generate fast transition dynamics. Instead, promising candidate economic theories include increases in the growth rate or standard deviation of the returns of the “super rich” relative to the rest of the population.

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 second eigenvalue of the differential operator governing the stochastic process for income or wealth.\textsuperscript{5} One of the main contributions of this paper is to derive an analytic formula for this second eigenvalue for a large variety of random growth processes.\textsuperscript{6} We obtain our measure of 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 quite precisely 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

\textsuperscript{4}That an increase in the variance of permanent earnings has contributed to the rise of inequality observed in the data has been argued by Kopczuk, Saez and Song (2010) and DeBacker et al. (2013).

\textsuperscript{5}Since the first eigenvalue is zero, the second eigenvalue also equals the “spectral gap.”

\textsuperscript{6}Linetsky (2005) computes this second eigenvalue for a special case of our framework, a Brownian motion with a reflecting barrier, under the parameter restriction that the volatility equals one.
upper tails of income and wealth distributions. Early theories of the income distribution include Champernowne (1953), Simon (1955), and Mandelbrot (1961), and more recent contributions are by Nirei (2009), Toda (2012), Kim (2013), Jones and Kim (2014). Similarly, random growth theories of the wealth distribution include Wold and Whittle (1957), Stiglitz (1967) and more recently Cowell (1998), Benhabib, Bisin and Zhu (2011, 2013, 2014), Piketty and Zucman (2014), Jones (2014), 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. A notable exception is Aoki and Nirei (2014), who study 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.7

That the empirical distribution of many variables is well approximated by a power law is one of the most ubiquitous regularities in economics and finance. For this reason, theories of random growth are an integral part of many different literatures besides those studying the distributions of income and wealth. For example, they have been used to study the distribution of city sizes (Gabaix, 1999), that of firm sizes (Luttmer, 2007), 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.

Our finding that heterogeneity in mean growth rates is needed to deliver fast dynamics of top inequality is also related to Guvenen (2009), who has argued that an income process with heterogeneous income profiles (HIP) 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. (2013).

The paper is organized as follows. Section 2 briefly states the main motivating facts for our analysis, and Section 3 reviews random growth theories of the income and wealth distribution 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

---

7 An early contribution by Blinder (1973) studies the dynamics of the wealth inequality, but in his model the distribution of wealth does not feature a Pareto tail.
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 examines the effects of an increase in $r - g$ on top wealth inequality. 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.

Figure 1 displays the evolution of measures of the top 1% income share (panel (a)) and of the top 1% wealth share (panel (b)). Panel (a) shows the large and rapid increase in

\[ \begin{align*}
\text{(a) Top Income Inequality} && \text{(b) Top Wealth Inequality}
\end{align*} \]

Figure 1: Evolution of Top 1% Income and Wealth Shares in U.S.

the top 1% income share that has been extensively documented by Piketty and Saez (2003), Atkinson, Piketty and Saez (2011) and others.\(^8\) Panel (b) shows the time path of the top 1% wealth share from two different data sources. The first is the Survey of Consumer Finances (SCF) and the second is a series constructed by Saez and Zucman (2014) by capitalizing capital income data.\(^9\) The two series suggest quite different conclusions. In particular data

\(^8\)The series is from the “World Top Incomes Database.” We here plot total income (salaries plus business income plus capital income) but excluding capital gains. The series display a similar trend when we include capital gains or focus on salaries only (though the levels are different).

\(^9\)The SCF data for 1989 to 2013 is from the online Appendix of Saez and Zucman (2014). The SCF
from the SCF suggest a relatively gradual rise in the top 1% wealth share, whereas Saez and Zucman’s estimates suggest a much more dramatic rise, a discrepancy that has generated some controversy (see e.g. Kopczuk, 2015). Finally, comparing panels (a) and (b) one can also see that wealth is much more unequally distributed than income.

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 \( X \) times richer on average than the top 1% who, in turn are \( X \) times richer than the top 10%, and 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). Figure 2, panel (a), 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, top income shares have increased relative to each other. As we explain in more detail below, this increase in “fractal inequality” implies equivalently a “fattening” of the Pareto tail of the income distribution. For wealth,
data for 1962 and 1983 is from Wolff (1987, Table 3). The 1962 dataset is called the “Survey of Financial Characteristics of Consumers” or SFCC.

\(^{10}\)Kopczuk (2015) notes that a third method of measuring top wealth shares, the estate-tax multiplier technique, suggests an even smaller increase in the top one percent wealth share than the SCF. Also see Auerbach and Hassett’s (2015) critique of Piketty (2014).
the finding depends again on the underlying data source, with the SCF showing no clear pattern and the capitalization method suggesting a large thickening of the tail of the wealth distribution.

There are five main takeaways from this section. First, top income shares have increased dramatically since the late 1970s. Second, top wealth shares appear to have increased though it is unclear by how much. Third, the Pareto tail of the income distribution has gotten fatter over time. Fourth, it is ambiguous whether the tail thickness of the tail of the wealth distribution has increased over time. And finally, wealth is more unequally distributed than income and, related, the wealth distribution has a fatter Pareto tail than the income distribution.

3 Random Growth Theories of Income and Wealth Inequality

Our starting point are 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

$$\frac{dh_{it}}{dt} = z_{it} g(i_{it}, h_{it})$$

where $z_{it}$ denotes ability and $i_{it}$ investment into human capital. We make the following four assumptions. First ability is i.i.d. over time and given by $z_{it} dt = \bar{z} dt + \tilde{\sigma} dZ_{it}$ where $Z_{it}$
is a standard Brownian motion. Second, the function $g$ has constant returns. Third, an individual’s investment is proportional to her human capital $i_{it} = \theta h_{it}$. 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

$$dw_{it} = zg(\theta, 1)w_{it}dt + \sigma g(\theta, 1)w_{it}dZ_{it}. \quad (1)$$

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 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.\textsuperscript{11}

### 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.\textsuperscript{12} 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

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

where $y_t$ is labor income, $\tau$ is the capital income tax rate, $R_{it}$ is the rate of return on wealth which is stochastic and $c_{it}$ denotes consumption. To keep things simple, we make the following assumptions. First, capital income is i.i.d. over time, and in particular $R_{it}dt = \bar{r}dt + \bar{\sigma}dZ_{it}$, where $\bar{r}$ and $\bar{\sigma}$ are parameters, and $Z_{it}$ is a standard Brownian motion. Second, we assume that individuals consume an exogenous fraction $\theta$ of their wealth at every point in time, $c_{it} = \theta \tilde{w}_{it}$.\textsuperscript{13} Third, we assume that all individuals earn the same labor income $y_t$, which grows deterministically 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

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

\textsuperscript{11}See e.g. Champernowne (1953), Simon (1955), Mandelbrot (1961), Nirei (2009), Toda (2012) Aoki and Nirei (2014) for models with similar reduced forms and Kim (2013), Jones and Kim (2014) for such models that additionally feature taxes.


\textsuperscript{13}A consumption rule with such a constant marginal propensity to consume can also be derived from optimizing behavior, at least for large wealth levels $w_{it}$. See Appendix XXX.
where $r = (1 - \tau)\tilde{r}$ is the after-tax average rate of return on wealth and $\sigma = (1 - \tau)\tilde{\sigma}$ is the after-tax wealth volatility. As discussed at length by Piketty (2015) 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 $w_{it}$ grows at a stochastic rate $\frac{dw_{it}}{w_{it}} = \gamma_{it}$ where $\gamma_{it}dt = \tilde{\gamma}dt + \sigma dZ_{it}$ and $\tilde{\gamma}$ and $\sigma$ are parameters (for the moment, we ignore the additive term $ydt$ in (2)). Therefore, $w_{it}$ follows a geometric Brownian motion

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

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

**Friction 2** Death at rate $\delta$ together with reinjection at some $w_0 > 0$, normalized to $w_0 = 1$

**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 (3) with $\tilde{\gamma} = \tilde{g}(\theta, 1)$ and $\sigma = \tilde{\sigma}g(\theta, 1)$ together with the second friction. Similarly, the simple model of wealth accumulation (2) is the special case of (3) with $\tilde{\gamma} = r - g - \theta$ 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 3.2 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

$$dx_{it} = \mu dt + \sigma dW_{it}, \quad \text{where} \quad \mu = \tilde{\gamma} - \frac{\sigma^2}{2}. \quad (4)$$
The properties of the stationary distribution of the income process (3) are well understood. In particular Gabaix (1999) has shown that, under certain parameter restrictions, this stationary distribution has a Pareto tail

\[ \Pr(w_{it} > w) \sim Cw^{-\zeta} \]

where \( C \) is a constant and \( \zeta \) is a simple function of the parameters \( \mu \) (equivalently \( \bar{\gamma} \)), \( \sigma \) and the particular friction. With friction 2 (the relevant friction for the case of income dynamics)

\[ \zeta = \frac{-\mu + \sqrt{\mu^2 + 2\sigma^2\delta}}{\sigma^2} \]  

(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 \( \eta = 1/\zeta \) as “top inequality.” Intuitively, tail inequality is increasing in \( \bar{\gamma} = \bar{\gamma}g(\theta, 1) \) and \( \sigma = \bar{\sigma}g(\theta, 1) \) and decreasing in the death rate \( \delta \). It will be useful later to note that, equivalently, the logarithm of \( w \) has an exponential tail, \( \Pr(x_{it} > x) \sim Ce^{-\zeta x} \).

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 wealth and income shares in Figure 2 and the top inequality parameter \( \eta \).\(^{15} \) Most existing contributions focus on the stationary distribution of the process (3) and completely ignore the corresponding transition dynamics. It is therefore 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. Besides random growth theories, there is one other class of theories that can generate power laws, namely those building on “superstar” mechanisms.\(^{16} \)

We return to these theories below.

\(^{14}\)Here and elsewhere “\( f(x) \sim g(x) \)” for two functions \( f \) and \( g \) means \( \lim_{x \to \infty} f(x)/g(x) = 1 \).

\(^{15}\)See Jones and Kim (2014) and Jones (2014) for two papers that use this fact extensively.

4 Slow Transitions in Baseline Random Growth Model

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 unanswered the question whether this increase in inequality will come about quickly or will take a long time to manifest itself. 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 \( \sigma \) may increase. Asymptotically as \( t \to \infty \), the distribution converges to its new stationary distribution. And 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 \( p_\infty(x) \). To talk about convergence and so on, 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

\[
||p(x, t) - p_\infty(x)|| = \int_{-\infty}^{\infty} |p(x, t) - p_\infty(x)| dx.
\]

For all four frictions, the cross-sectional distribution \( p(x, t) \) satisfies the Kolmogorov Forward equation

\[
\frac{\partial p(x, t)}{\partial t} = -\mu \frac{\partial p(x, t)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p(x, t)}{\partial x^2} - \delta p(x, t) + \delta \delta_0(x)
\]

with initial condition \( p(x, 0) = p_0(x) \), where \( \delta = 0 \) in the case of frictions 1 and 4, and where \( \delta_0 \) denotes the Dirac delta function, i.e. a point mass at \( x = 0 \). 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 \( x = 0 \).\(^{17}\) Throughout this paper we will use the more compact expression \( p_t := \frac{\partial p(x, t)}{\partial t} \) and

\(^{17}\)More generally, we can allow for the income of newborns to be drawn from an arbitrary (thin-tailed)
Similarly for the other partial derivatives so that the equation becomes:

\[ p_t = -\mu p_x + \frac{\sigma^2}{2} p_{xx} - \delta p + \delta \delta_0 \quad (6) \]

\[ = \mathcal{A}^* p + \delta \delta_0 \quad (7) \]

\[ \mathcal{A}^* p = -\delta p - \mu p_x + \frac{\sigma^2}{2} p_{xx} \quad (8) \]

Frictions 1 to 3 manifest themselves in different boundary conditions (see the Appendix).

If we extend to a model with jumps arriving with intensity \( \phi \), we have: \( p_t = \mathcal{A}^* p + \delta \delta_0 \) with:

\[ \mathcal{A}^* p = -\delta p - \mu p_x + \frac{\sigma^2}{2} p_{xx} + \phi E [p(x - g) - p(x)] \]

where the expectation is taken over the random jump \( g \), with intensity \( \phi \).

Suppose that we have

\[ q(x, t) := p(x, t) - p_\infty(x) \]

We want to study how fast the distribution converges \( p(x, t) \) to \( p_\infty(x) \), i.e. how fast \( q(x, t) \) converges to 0. We observe that

\[ q_t = \mathcal{A}^* q \quad (9) \]

### 4.1 Average Speed of Convergence

The following Proposition 1 is one of the two main theoretical results of our paper.

**Proposition 1** Consider the income process (4). If a stationary distribution \( p_\infty(x) \) exists, then 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

\[ \lambda = - \lim_{t \to \infty} \frac{1}{t} \ln ||p(x, t) - p_\infty(x)|| \]

depends on the friction ensuring the existence of a stationary distribution. With frictions 1 and 3

\[ \lambda = \frac{1}{2} \mu^2 1_{\{\mu < 0\}} + \delta \quad (10) \]

where \( 1_{\{\}} \) is the indicator function. With friction 2,

\[ \lambda = \delta. \quad (11) \]

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.
We shall see that Proposition 1 implies that the traditional canonical model delivers a much too low speed of convergence: \( \lambda \) is too low compared to the empirical estimate.

We show a proof sketch in the main text, because it uses bounding arguments that are novel or at least quite unusual in economics, so that they can be of general interest.

**Proof of Proposition 1 when there is no reflecting barrier (friction 2)** The key argument is the following.

**Lemma 5** Suppose that a function \( q(x, t) \) satisfies \( q_t = A^*q \). Then, we have

\[
|q|_t \leq A^*|q|
\]

(12)

Note that above the operator \( A^* \) is applied to the function \( |q(x, t)| \). The proof of the Lemma is in the Appendix’s Section A.

We next consider the \( L^1 \) norm of \( q \):\[
m(t) := \|q(\cdot, t)\| = \int |q(x, t)| \, dx
\]

(13)

**Lemma 6** The \( L^1 \) norm \( m(t) := \|q(\cdot, t)\| \) decays at a rate weakly greater than \( \delta \). We have \( m'(t) \leq -\delta m(t) \) and \( m(t) \leq m(0)e^{-\delta t} \). This implies that the decay rate is at least \( \delta \): \( \lambda \geq \delta \).

**Proof of Lemma 6** We have

\[
m'(t) = \int |q(x, t)|_t \, dx \\
\leq \int (A^*|q(x)|)(x, t) \, dx \quad \text{by Lemma 5} \\
= \int \left( -\delta |q| - \mu |q|_x + \frac{\sigma^2}{2} |q|_{xx} + \phi E[|q|(x-g) - |q|(x)] \right) \, dx := H
\]

Now, observe that for any function \( Q \),\(^{18}\)

\[
\int Q_x \, dx = \int Q_{xx} \, dx = \int (EQ(x-g) - Q(x)) \, dx = 0
\]

\(^{18}\)Indeed, for instance \( \int Q_x \, dx = [Q(x,t)]_{x=-\infty}^{x=\infty} = 0 \), and

\[
\int (EQ(x-g) - Q(x)) \, dx = \int \left( \int Q(x-g) \, dx \right) f(g) \, dg - \int Q(x) \, dx \\
= \left( \int Q(y) \, dy \right) \left( \int f(g) \, dg \right) - \int Q(x) \, dx = 0
\]
Hence, \( m'(t) \leq H = -\delta \int |q| \, dx = -\delta m(t) \) hence: \( m'(t) \leq -\delta m(t) \), and \( m(t) \leq e^{-\delta t} m(0) \) by Gronwall’s lemma. □

We next show that \( \lambda \leq \delta \).

**Lemma 7** The \( L^1 \) norm \( m(t) := \|q(\cdot, t)\| \) decays at a rate weakly less than \( \delta \): \( \lambda \leq \delta \).

**Proof sketch for Lemma 7** Call \( a(t) := \int xq(x, t) \, dx \). Calculations detailed in the proof appendix show that \( a(t) = e^{-\delta t} a(0) \). One can show that this implies that the decay is no greater than \( \delta \) : \( \lambda \leq \delta \). □

The two Lemmas 6 and 7 imply that \( \lambda \geq \delta \) and \( \lambda \leq \delta \) respectively. So finally \( \lambda = \delta \). □

**Case with a reflecting barrier: Frictions 1 and 3** Consider next the speed of convergence with a reflecting barrier (friction 1), \( \mu < 0 \), and no Poisson death \( \delta = 0 \). From (5), stationary tail inequality for this case is \( \eta = 1/\zeta = -\sigma^2/(2\mu) \) and therefore the speed of convergence can also be written as

\[
\lambda = \frac{1}{8} \frac{\sigma^2}{\eta^2}.
\]

(14)

This expression is intuitive. It states that the transition is faster the higher is the standard deviation of growth rates \( \sigma \) and the lower is tail inequality \( \eta \); that is, high inequality goes hand in hand with slow transitions. The interpretation of the formulas for the second and third frictions is similar.

The rough idea of the proof of Proposition 1 is as follows, and most easily explained in the case of a reflecting barrier and no Poisson death (friction 1). In this case, the Kolmogorov Forward equation can be written in terms of a differential operator \( \mathcal{A}^* \) as

\[
p_t = \mathcal{A}^* p, \quad \mathcal{A}^* = -\mu \frac{\partial}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2}
\]

The operator \( \mathcal{A}^* \) summarizes the entire dynamics of the process for \( x_{it} \). It 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 second largest (in absolute value) eigenvalue.\(^{19}\) The key contribution in the Proposition and main step of the proof is then to obtain an explicit formula for the second eigenvalue of \( \mathcal{A}^* \) in the form of

\(^{19}\)To get the intuition for this result, consider briefly the case where \( x_{it} \) is a finite-state Poisson process where \( x_{it} \) can take a discrete number \( N \) of possible values. The distribution of \( x \) is then a vector \( p(t) \in \mathbb{R}^N \)
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 at rate $\delta$ (friction 2) because it is possible to obtain clean analytic formulas for this case. While such clean formulas are not available for the other two frictions ensuring stationarity, we show by means of numerical computations that the general lessons from our analysis carry over to these other two cases as well.

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

To explain the main result of this section in the most accessible fashion, we first examine the restrictive but instructive special case $\sigma = 0$ 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 satisfying $p(t) = A^T p(t)$ where $A$ is the $N \times N$ transition matrix of the Poisson process. Denoting the corresponding eigenvalues by $0 = |\lambda_1| < |\lambda_2| < \ldots < |\lambda_N|$, one can show that if a stationary distribution $p_\infty$ exists (and under certain other conditions), $p(t)$ converges to this stationary distribution at rate $|\lambda_2|$. Consider an example with $N = 2$ and a symmetric transition matrix with diagonal entries $-\phi$ and off-diagonal entries $\phi$ (\(\phi\) is the Poisson switching intensity). The eigenvalues of this matrix are $\lambda_1 = 0$ and $|\lambda_2| = 2\phi$. Intuitively, the convergence speed $|\lambda_2|$ increases in the switching intensity $\phi$.

Linetsky (2005) derives a related result for the case of a reflected Brownian motion (friction 1) and for the special case $\sigma = 1$. Similarly, in the case of a reflected Brownian motion it is possible to 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.

Linetsky (2005) derives a related result for the case of a reflected Brownian motion (friction 1) and for the special case $\sigma = 1$. Similarly, in the case of a reflected Brownian motion it is possible to 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.

In contrast, the first part of the result stating that convergence takes place at an exponential rate is relatively standard. It is the continuous-time version of other well-known results. See for example Theorem 11.12 in Stokey, Lucas and Prescott (1989) which covers the case of discrete-time continuous-state processes. Also related is Theorem 1.4 which states an analogous result for finite-state processes.
\( \delta \). The Steindl model has recently also been examined by Jones (2014). The distribution \( p(x, t) \) then satisfies the Kolmogorov Forward equation

\[
p_t = -\mu p_x - \delta p
\]  

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 at \( t = 0 \) the growth rate changes to \( \mu > \mu_0 \). The time path of \( p(x, t) \) is the solution to (15) with initial condition \( p_0(x) = \alpha e^{-\alpha x}, \alpha = \delta/\mu_0 \) which is given by

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

where again \( \mathbf{1}_{\{\cdot\}} \) is the indicator function. The solution is depicted in Figure 3. Consider in particular, the local power law exponent

\[
\zeta(x, t) = -\frac{\partial \log p(x, t)}{\partial x}
\]

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, convergence of the distribution is therefore slower the further out in the tail we look. In particular note from the Figure that the asymptotic (for large \(x\)) power law exponent \(\zeta(t) = -\lim_{x \to \infty} \partial \log p(x, t) / \partial x\) 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 \(\sigma > 0\).

It is also possible to characterize the time path of the distribution for \(\sigma > 0\) and for more general initial conditions \(p_0(x)\), and we state this result here for completeness.

**Proposition 2**  
In the case with no reflecting barrier, we have

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

where \(g_t := \mu t + \sigma Z_t\).

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 (15) when the initial distribution is \(p_0(x) = \alpha e^{-\alpha x}1_{\{x > 0\}}\).

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

As noted earlier, the distribution \(p(x, t)\) satisfies the Kolmogorov Forward equation (6).

One can show (see e.g. Gabaix, 2009) that the stationary distribution is a double Pareto distribution \(p_\infty(x) = c \min\{e^{-\zeta_1 x}, e^{-\zeta_2 x}\}\) where \(c = -\zeta_1 \zeta_2 / (\zeta_2 - \zeta_1)\) and where \(\zeta_1 < 0 < \zeta_2\) are the two roots of

\[
0 = \frac{\sigma^2}{2} \zeta^2 + \zeta \mu - \delta
\]

Apart from the stationary distribution, the solution to the Kolmogorov Forward equation is cumbersome.
The key insight of this section is that the solution to this partial differential equation can be characterized conveniently in terms of the so-called “Laplace transform” of \( p \)

\[
\hat{p}(\xi, t) := \int_{-\infty}^{\infty} e^{-\xi x} p(x, t) \, dx = \mathbb{E} [e^{-\xi x}]
\]

where \( \xi \) is a real number.\(^{22}\) For \( \xi \leq 0 \), the Laplace transform has the natural interpretation of the \(-\xi\)th moment of the distribution of \( w = \exp(x) \), that is \( \mathbb{E}[w^{-\xi}] \). We show momentarily that we can obtain a clean analytic formula for the entire time path of this object for all time \( t \). 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 is \( \xi \), the more we know about the distribution’s tail behavior.

Integrating the Kolmogorov Forward equation (6), we have

\[
\frac{\partial \hat{p}(\xi, t)}{\partial t} = -\xi \mu \hat{p}(\xi, t) + \xi^2 \frac{\sigma^2}{2} \hat{p}(\xi, t) - \delta \hat{p}(\xi, t) + \delta
\]

with initial condition \( \hat{p}(\xi, 0) = \hat{p}_0(\xi) \), the Laplace transform of \( p_0(x) \). We here used that, for any \( x_0 \), the Laplace transform of the Dirac delta function \( \delta_{x_0} \) is \( e^{-\xi x_0} \). Importantly, note that for fixed \( \xi \), (19) is a simple ordinary differential equation for \( \hat{p} \) that can be solved analytically

\[
\hat{p}(\xi, t) = \frac{\delta}{\lambda(\xi)} + \left( \hat{p}_0(\xi) - \frac{\delta}{\lambda(\xi)} \right) e^{-\lambda(\xi)t}
\]

\[
\lambda(\xi) = \mu \xi - \frac{\sigma^2}{2} \xi^2 + \delta
\]

First consider the stationary solution of (19)

\[
\hat{p}_\infty(\xi) = \frac{\delta}{\mu \xi - \frac{\sigma^2}{2} \xi^2 + \delta}
\]

As noted earlier, for \( \xi \leq 0 \), \( \hat{p}_\infty(\xi) \) is the \(-\xi\)th moment of the stationary distribution \( p_\infty(x) \). We can use this fact to identify the Pareto tail of the stationary distribution as follows. For any distribution with a Pareto tail with tail parameter \( \zeta \), moments of order higher than \( \zeta \) are infinite, i.e. do not exist. This suggests that the tail parameter of the stationary distribution is the critical value \( \zeta > 0 \) such that \( \hat{p}_\infty(\xi) \) ceases to exist for all \( \xi \leq -\zeta \).\(^{23}\) Examining (22),

\(^{22}\)COMMENT ON FACT THAT INTEGRAL STARTS AT \( x = -\infty \) RATHER THAN \( x = 0 \).

\(^{23}\)In particular, one can show that for any distribution \( p \) with a Pareto tail, that is \( p(x) \sim cx^{-\zeta} \) \( x \to \infty \) for constants \( c \) and \( \zeta \), the Laplace transform \( \hat{p}(\xi) \sim \frac{c}{\zeta + \xi} \) as \( \xi \downarrow -\zeta \).
we see that this critical value satisfies the quadratic equation (17), and in particular the right tail is governed by the positive of the two roots, exactly as expected. For the remainder of this section, we impose the restriction $\xi > -\zeta$, that is we only consider moments of the distribution that are finite.

Next, consider the speed of convergence which is the main focus of our paper. To fix ideas, consider again the case in the initial distribution has a Pareto tail $\hat{p}_0(\xi) \sim \frac{c}{\xi + \alpha}$ and where parameters change in such a way that stationary tail inequality is higher than initial tail inequality, that is $\zeta < \alpha$. Examining (20), one can see that the speed of convergence of the $-\xi$th moment is given by $\lambda(\xi)$ in (21). Note that for $\mu > 0$, the speed of convergence is always lower the lower is $\xi$. If $\mu < 0$, the same is true for all $\xi$ less than some critical value. Hence the closed form solution for the Laplace transform in (20) shows that high moments converge more slowly than low moments. Figure 4 provides a graphical illustration of this theoretical result. As in the Steindl case of Figure 3, the power law exponent $\zeta$ (equivalently top inequality $\eta$) does not change at first and the distribution instead shifts out in parallel.

\[ \hat{p}_{\infty}(\xi) = \frac{c}{\zeta_2 + \xi} - \frac{c}{\zeta_1 + \xi} = \frac{\zeta_1 \zeta_2}{(\zeta_1 + \xi)(\zeta_2 + \xi)} \] (23)

From the quadratic for $\zeta$, we have that $\zeta_1\zeta_2 = \frac{-\delta}{\sigma^2/2}, \zeta_1 + \zeta_2 = \frac{-\mu}{\sigma^2/2}$ and hence $(\zeta_1 + \xi)(\zeta_2 + \xi) = \zeta_1\zeta_2 + \xi(\zeta_1 + \zeta_2) + \xi^2 = \frac{-\delta}{\sigma^2/2} - \mu\xi/(\sigma^2/2) + \xi^2$. Substituting into (23) we obtain (22).
4.3 The Baseline Model Cannot explain the Rise in Income Inequality

We now revisit Figures 1 and 2 from Section 2 and ask: can standard random growth models generate the observed increase in income inequality observed in the data? Not surprisingly given the theoretical results in the preceding two sections we argue that they cannot. In particular, the transition dynamics generated by the model are too slow relative to the dynamics observed in the data.

More precisely, we ask whether increase in the variance of the permanent component of wages \( \sigma \) 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 Kopczuk, Saez and Song (2010) and DeBacker et al. (2013). 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 set \( \delta = 1/30 \) corresponding to an expected working age of thirty years, and calibrate \( \mu \) to match the observed tail inequality \( \eta \) in 1973. Since \( \mu > 0 \), Proposition 1 implies that the average speed of convergence is simply \( \lambda = \delta \) and the corresponding half-life is \( t_{1/2} = \ln(2)/\delta = 20.8 \) years. The implied speed of convergence in the tail is again slower. For example for \( \xi = -1.4 \) (the 1.4th moment of the distribution)

\[
t_{1/2}(\xi) = \ln(2) \left( \mu \xi - \frac{\sigma^2}{2} \xi^2 + \delta \right)^{-1} \approx 68 \text{ years}
\]

Figure 5 plots the time path for the top 1% income share (panel (a)) and the power law exponent (panel (b)) generated by the baseline random growth model and compares it to that in the data. Not surprisingly given our analytical results, the model fails spectacularly. An increase in the variance in the permanent component of income \( \sigma \) 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 depends on the tax rate \( \theta = \theta(\tau) \), and therefore so do the reduced form parameters \( \mu \) and \( \sigma \).

Because this simple story collapses in a reduced form to a standard random growth model, it cannot generate fast transitions either (results are available upon request).
Given the negative results of the preceding section, it is natural to ask: what then explains the observed rise in top income? We argue that fast transitions require very specific departures from the standard random growth model. We extend the standard random growth model along three 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”. Third, we generalize our model to feature jump processes, where income innovations are drawn from an exogenous distribution. We discuss the role of each of these three additions in turn in Sections 6.2 to 6.4. In Section 5.5, 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 Generalized Random Growth Model

In its most general form, we consider a random growth model with distinct “growth regimes” indexed by $j = 1, ..., J$, 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
\[ \begin{align*}
x_{it} &= e^{b_j S_t} y_{it} \\
\text{dy}_{it} &= \mu_j dt + \sigma_j dZ_{it} + g_{jit} dN_{jit} + \text{Injection} - \text{Death}
\end{align*} \]

where \( dN_{jit} \) is a Poisson process with intensity \( \varphi_j \) (that is, \( dN_{jit} = 1 \) with probability \( \varphi_j dt \) and \( dN_{jit} = 0 \) with probability \( 1 - \varphi_j dt \)) 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 \( \sigma_j \). Deviations from Gibrat’s law are captured by \( S_t \) which is an arbitrary stochastic process satisfying \( \lim_{t \to \infty} \mathbb{E}[S_t] < \infty \).

To see this, note that (24) and (25) can be written as
\[ dx_{it} = \tilde{\mu}_{jt} dt + \tilde{\sigma}_{jt} dZ_{it} + b_j x_{it} dS_t + g_{jit} d\tilde{N}_{jit} + \text{Injection} - \text{Death} \]

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.\(^{25}\) 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 a variant of the standard random growth model in section 4 with multiple distinct “growth regimes”, but without deviations from Gibrat’s law or jumps \( dS_t = dN_{it} = 0 \). We here 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. For concreteness, we describe the model with our income dynamics application in mind. 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).

Denote by \( p^H(x, t) \) and \( p^L(x, t) \) the density of individuals who are currently in the high and low growth states and by \( p(x, t) = p^H(x, t) + p^L(x, t) \) the cross-sectional wage distribution.\(^{25}\)

\(^{25}\) Also note that \( Z_{it} \) is an idiosyncratic stochastic process whereas \( S_t \) is an aggregate or common shock that hits all individuals simultaneously.
We have

\[ p_t^H = -\mu_H p_x^H + \frac{\sigma_H^2}{2} p_{xx}^H - \phi p^H + \beta_H \delta_0 \]

\[ p_t^L = -\mu_L p_x^L + \frac{\sigma_L^2}{2} p_{xx}^L + \phi p^H - \delta p^L + \beta_L \delta_0 \]

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

While we are not aware of an analytic solution method for the system of partial differential equations (27), this system can be conveniently analyzed by means of Laplace transforms as in section 4.2. In particular, \( \hat{p}^H(\xi, t) \) and \( \hat{p}^L(\xi, t) \) satisfy

\[ \hat{p}_t^H = -\xi \mu_H \hat{p}^H + \xi^2 \frac{\sigma_H^2}{2} \hat{p}^H - \phi \hat{p}^H + \beta_H \]

\[ \hat{p}_t^L = -\xi \mu_L \hat{p}^L + \xi^2 \frac{\sigma_L^2}{2} \hat{p}^L + \phi \hat{p}^H - \delta \hat{p}^L + \beta_L \]

with initial conditions \( \hat{p}^H(\xi, 0) = \hat{p}_0^H(\xi), \hat{p}^L(\xi, 0) = \hat{p}_0^L(\xi) \). Importantly, for fixed \( \xi \), this is again simply a system of ordinary (rather than partial) differential equations and it can be solved analytically.

**Proposition 3** Consider the cross-sectional distribution \( p(x, t) = p^H(x, t) + p^L(x, t) \) and the corresponding Laplace transform \( \hat{p}(x, t) = \hat{p}^H(x, t) + \hat{p}^L(x, t) = \mathbb{E}[e^{-\xi x}] \). We have

\[ \hat{p}(\xi, t) - \hat{p}_\infty(\xi) = c_H(\xi)e^{-\lambda_H(\xi)t} + c_L(\xi)e^{-\lambda_L(\xi)t} \]

where \( c_L(\xi) \) and \( c_H(\xi) \) are constants of integration and

\[ \hat{p}_\infty(\xi) = \frac{\beta_H}{\lambda_H(\xi)} + \frac{\beta_L}{\lambda_L(\xi)} + \phi \frac{\beta_H}{\lambda_L(\xi) \lambda_H(\xi)}, \]

\[ \lambda_H(\xi) = \xi \mu_H - \xi^2 \frac{\sigma_H^2}{2} + \phi, \]

\[ \lambda_L(\xi) = \xi \mu_L - \xi^2 \frac{\sigma_L^2}{2} + \delta. \]

The stationary distribution \( p_\infty(x) = p^H_\infty(x) + p^L_\infty(x) \) has a Pareto tail with tail exponent \( \zeta = \min\{\zeta_L, \zeta_H\} \) where \( \zeta_H \) is the positive root of \( 0 = \zeta^2 \frac{\sigma_H^2}{2} + \zeta \mu_H - \phi \) and \( \zeta_L \) is the positive root of \( 0 = \zeta^2 \frac{\sigma_L^2}{2} + \zeta \mu_L - \delta. \)

The derivation of (29) is instructive and we therefore state it in the main text. The rest of the proof is in the Appendix. Write the system of ordinary differential equations (28) in
matrix form as
\[
\dot{\hat{q}}_t = A\hat{q} + \beta, \quad \text{where} \quad A = \begin{bmatrix}
-\lambda_H(\xi) & 0 \\
\phi & -\lambda_L(\xi)
\end{bmatrix},
\]
and where \( \hat{q} = (\hat{p}^H, \hat{p}^L)' \) and \( \beta = (\beta_H, \beta_L)' \). Because the matrix \( A \) is triangular, its eigenvalues are simply \( \lambda_H(\xi) \) and \( \lambda_L(\xi) \). Hence the solution to (33) is
\[
\hat{p}^j(\xi, t) - \hat{p}^j_{\infty}(\xi) = \bar{c}_He^{-\lambda_H(\xi)t}v_H^j(\xi) + \bar{c}_Le^{-\lambda_L(\xi)t}v_L^j(\xi), \quad j = L, H
\]
where \( \bar{c}_L \) and \( \bar{c}_H \) are constants of integration and where we denote by \( v_H = (v_{HH}, v_{HL})' \) and \( v_L = (v_{LH}, v_{LL})' \) the eigenvectors of \( A \). Adding the two equations we obtain (28).

A natural assumption is that the rate of switching from the high- to the low-growth state \( \phi \) is considerably higher than the rate of retirement \( \delta \) and that the difference 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 (31) and (32). 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.

Guvenen (2009) has argued that an income process with heterogeneous income profiles (HIP) 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 in the two regimes \( \sigma_H \) and \( \sigma_L \). This has a similar flavor to the mixture specification advocated by Guvenen et al. (2013).

5.3 The Role of Deviations from Gibrat’s law/superstars

We call those shocks \( S_t \) in (25) “aggregate shocks to high incomes” or “superstars shocks”: the reason is that they affect more high (log) incomes.

5.3.1 A microfoundation for “superstars shocks”

We here provide a microfoundation for the \( e^{S_t} \) term in equation (25). Here we adopt the model of Gabaix and Landier (GL, 2008), which can be viewed as a tractable, calibratable
version of the “superstars economics” ideas of Rosen (1981). There is a continuum of firms of different size and managers with different talent. A CEO of talent \( T \), matched with firm \( S \), produces an improvement in firm value \( C T S^\gamma \), where \( C \) is a constant. 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 \( 1/\alpha \) a firm of rank \( n \) has size: \( S(n) = An^{-\alpha} \).

A manager of talent rank \( n \) has talent \( T(n) = T_{\text{max}} - B\beta n^\beta \). Hence, the value added of the manager is \( (T_{\text{max}} - BCn^{\beta})A^\gamma n^{-\alpha\gamma} \).

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

\[
 w(n) = e^{\alpha t}n^{-\chi_t}
\]

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

\[
 \chi_t = \alpha_t\gamma_t - \beta_t \tag{34}
\]

Noting \( x_{it} = \ln w_{it} \) the log wage, and \( q_{it} = -\ln n_{it} \) the log quantile, and we have

\[
 x_{it} = \chi_t q_{it} + a_t \tag{35}
\]

Note that \( \chi_t \) and \( a_t \) depend on economy-wide forces, while \( x_{it}, q_{it} \) are specific to the agent.

Hence, the process is: \( dx_{it} = \chi_t dq_{it} + q_{it} d\chi_t + da_t \), i.e.

\[
 dx_{it} = \chi_t dq_{it} + \frac{d\chi_t}{\chi_t} (x_{it} - a_t) + da_t
\]

This generates a process like (25), where we will take \( b_H = 1 \) for the high type group, and \( b_L = 0 \), and \( y_{it} := q_{it} \). Only the high group has the talent considered here. We have:

\[
 e^{S_t} := \chi_t
\]

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.\(^{26}\) 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

\(^{26}\)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).
to increases in the “span of control” or “scales managed by talent”, somewhat as in Garicano and Rossi-Hansberg (2006).

Agent $i$’s talent evolves stochastically. The process for $q_{it}$ is:

$$dq_{it} = \mu^q_{it} dt + \sigma^q_{it} dz_{it}$$

with death rate $\delta^q_{it}$. By construction, the quantile $n_{it} = e^{-q_{it}}$ has a uniform distribution, $q_{it}$ should have an exponential distribution with exponent 1, which in turns implies: $\mu^q_{it} + \frac{1}{2} (\sigma^q_{it})^2 - \delta^q_{it} = 0$.

### 5.3.2 Implications for fast transitions

The model with shocks to the returns to talent generates fast transitions

Indeed, suppose that the process for the underlying process $y_{it}$ does change, but $S_t$ varies. Given that $x_{it} = e^{S_t} y_{it}$, we have $\zeta^x_t = e^{-S_t} \zeta^y_t$. Hence, the process is extremely fast – in fact it features instantaneous transitions in the power law exponent.

**Proposition 4** (Infinitely fast adjustment in models with aggregate shocks to high incomes, aka “superstars shocks”) Consider the process (25), 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 infinite fast speed of adjustment: $\lambda = \infty$. Indeed, we have $\zeta^x_t = e^{-S_t} \zeta^y$, where $\zeta^x_t$, $\zeta^y$ is power law exponent of incomes $x_{it}$ and $y_{it}$.

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

**Those fast transitions are empirically relevant at high frequencies** There is some evidence for the aggregate shocks to high incomes (25) in Parker and Vissing-Jorgensen (2010). They find that in good (resp. bad) times, the incomes of top earners increase (resp. decrease), in a manner quite similar to (25): the sensitivity to the shock at time $t$ is proportional to $x_{it}$, as in

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

This finding appears to have been amplified and broadly confirmed by Guvenen (2015, private communication).

In our model, the interpretation is that when there is a positive technology shock to the use of top resources (e.g. top talent, which corresponds to an increase in $\chi_t$), $e^{S_t} = \chi_t$

\[27\text{Indeed, } P(x_{it} > x) = P(e^{S_t} y_{it} > x) = P(y_{it} > e^{-S_t} x) = Ke^{-\zeta y e^{-S_t} x}.\]
increases, so that very high incomes increase very much, and large in percentage terms that simply high incomes. It would be interesting to know if these effects are important at low frequency (e.g., from one decade to the next), not simply at high frequency (from one year to the next).

We conclude that those aggregate shocks to high incomes are a theoretically legitimate and empirically grounded source of fast transitions.

5.4 The Role of Jumps

Proposition 1, with friction \(2\), included jumps, and already allows us to conclude that jumps do not speed up the convergence: the speed remains \(\lambda = \delta\), whatever the jumps. In this section, we still study in more detail the influence of jumps.

We consider the special case of the generalized random growth model with jumps only, that is \(J = 1\) and \(dS_t = 0\). We argue that jumps are useful descriptively for capturing certain features of the date but they do not increase the speed of transition. For instance, Guvenen et al. (2013) using administrative data, document that earnings innovations are very leptokurtic as opposed to Gaussian which is the assumption in the standard random growth model. Jumps are the natural way of introducing such kurtosis. Kaplan, Moll and Violante (2015) argue that even normally distributed jumps that arrive with a Poisson arrival rate can generate much kurtosis.\(^{28}\)

In the case \(J = 1\) and \(dS_t = 0\), (24) and (25) become

\[
\begin{align*}
    dx_{it} &= \mu dt + \sigma dZ_{it} + g_{it} dN_{it} \\
\end{align*}
\]

where \(dN_{it}\) is a jump process with intensity \(\varphi\) and the innovations \(g_{it}\) are drawn from a distribution \(f\). The following Proposition characterizes the resulting dynamics of the income distribution.

**Proposition 5** In the case with jumps (36), consider the cross-sectional distribution \(p(x, t)\) and the corresponding Laplace transform \(\hat{p}(x, t) = \mathbb{E}[e^{-\xi x}]\). We have

\[
\begin{align*}
    \hat{p}(\xi, t) &= \frac{\delta}{\lambda(\xi)} + \left( \hat{p}_0(\xi) - \frac{\delta}{\lambda(\xi)} \right) e^{-\lambda(\xi)t} \\
    \lambda(\xi) &= \xi \mu - \xi^2 \sigma^2 + \frac{\varphi}{2} \left( \hat{f}(\xi) - 1 \right)
\end{align*}
\]

\(^{28}\)Alternatively, income innovations may themselves be drawn from a double-Pareto distribution.
We have that \( \hat{f}(0) = 0 \) and \( \hat{f}(\xi) > 1 \) for all \( \xi < 0 \). Hence the speed of convergence is actually lower than in the case \( \varphi = 0 \). This point can be made more precise by considering the case where the distribution \( f \) is symmetric. For this case, consider the Taylor series approximation of \( \hat{f}(\xi) \) around \( \xi = 0 \) and use the fact that derivatives of Laplace transforms of distributions evaluated at zero are moments \( (\hat{f}'(0) = -\mathbb{E}[g_u], \hat{f}''(0) = \mathbb{E}[g_u^2] \) and so on). Therefore

\[
\lambda(\xi) = \xi \mu - \xi^2 \frac{\sigma^2}{2} + \delta - \varphi \left( \frac{\xi^2}{2} \mathbb{E}[g_u^2] + \frac{\xi^4}{4} \mathbb{E}[g_u^4] + \ldots \right)
\]

The second order term means that the effective variance of income innovations is \( \sigma^2 + \mathbb{E}[g_u^2] \) which is precisely what an econometrician would measure in the data. Hence the only substantially new term is the fourth moment term and other terms corresponding to higher order moments. But this term affects the speed \( \lambda(\xi) \) negatively. 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.

### 5.5 Revisiting the Rise in Income Inequality

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). In the present, preliminary version of the paper, we do not have any microeconometric evidence suggesting that there has been such an increase in the dispersion of mean growth rates. We plan on exploring the related empirical evidence in more detail in future versions of the paper. That being said, our results suggest that any other shock that increases stationary tail inequality by a sufficiently large amount would work as well. The speed of convergence is always the same.

With this caveat in mind, we proceed as follows. We set the parameters \( \sigma_L \) and \( \mu_L \) to the same level as \( \mu \) and \( \sigma \) in the baseline random growth used in Section 4.3. We then parameterize the difference between the mean growth rates \( \mu_H \) and \( \mu_L \) as follows. We set \( \mu_H = \mu_L + 0.06 \) in the initial steady state corresponding to 1975. We then consider a once-and-for-all increase in \( \mu_H \) to \( \mu_H = \mu_L + 0.14 \). Such gaps between \( \mu_H \) and \( \mu_L \) are broadly
consistent with empirical evidence in Guvenen, Kaplan and Song (2014). In particular, their Figure 7 reveals difference in average growth rates of different population groups as large as 0.23 log points per year.\footnote{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 \textit{lifetime} earnings. Over a ten year period (ages 25 to 35), the earnings of the top 0.1% grow by 2.3 log points, those of the next 0.9% by 1.4 log points and those of the bottom 99% by 0.5 log points over 10 years.} We set the rate of switching from the high- to the low-growth regime to $\phi = 1/6$, corresponding to an expected duration of the high-growth regime of 6 years. We further set the death rate $\delta = 1/30$ as before and the fraction of individuals born in the high growth regime to $\theta = 0.3$. These parameters imply that 4.14% of the population are in the high growth regime. 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 get quite close to replicating the rapid rise in income inequality observed in the United States.

### 5.6 The Role of Startups – an Interpretation

A possible interpretation of the “generalized random growth model” of section 5.1 is as follows.

- There are two regimes $H$ and $L$. The high growth regime corresponds to entrepreneurship/owning a startup. The low growth regime corresponds to being a worker.
• Entrepreneurs have higher earnings growth than workers $\mu_H > \mu_L$ but they are also exposed to more idiosyncratic risk $\sigma_H > \sigma_L$.

• Entrepreneurs also have $b_H > b_L$, that is they are more exposed to aggregate risk. The latter point is consistent with empirical evidence. See Guvenen (2015).

• No Silicon Valley in France $\Rightarrow$ no increase in top inequality. Discuss relation to Jones and Kim (2014).

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 use these results to shed light on the question 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 Piketty (2014).

6.1 Stationary Wealth Inequality

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

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

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 is the marginal propensity to consume $\theta$ and the higher is the after-tax wealth volatility $\sigma$. 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, 2015). One can also see from (38) that top wealth inequality $\eta$ is a convex function of this gap, and therefore for some levels of $r - g$ additional increases may have large effects on steady state wealth inequality. However, this leaves unanswered the question whether such increases in top wealth inequality will come about quickly or take many hundreds of years to materialize, i.e. whether the transition to the new steady state is slow or fast.
6.2 Other Theories of Top Wealth Inequality

As discussed in section 3.4 for the case of income inequality, we explicitly confine ourselves to studying theories of top wealth inequality that can generate power laws. This is because we consider the fractal feature of top inequality discussed in Section 2 an important empirical regularity that deserves special attention.

Before proceeding we briefly discuss one example of a theory that does not generate a power law, and that we therefore do not consider. In particular, some readers may conjecture that wealth is simply accumulated labor income and that therefore top labor income inequality is a main determinant of top wealth inequality. Therefore, if we could explain the rise in top income inequality, we would get the rise in top wealth inequality “for free.” While it is certainly true that top labor income inequality affects top wealth inequality, theories of labor income alone cannot explain the observed Pareto tail of the wealth distribution, i.e. the fractal feature of top wealth inequality. This point is due to Benhabib, Bisin and Zhu (2011) who argue that what is needed instead is randomness in capital incomes or saving rates. One qualification needs to be made here: their result is derived under the assumption of a bounded or thin-tailed labor income distributions, whereas in the data, the labor income distribution itself has a Pareto tail. Therefore, in principle, labor income could generate a Pareto tail. However, in this case, the wealth distribution would inherit the tail exponent of the income distribution exactly. And this is inconsistent with the observation that, in the data, the wealth distribution has a fatter tail than the income distribution (Figure 2).  

Consistent with these theoretical findings, a large fraction of the wealthiest individuals in the United States and other advanced countries are “entrepreneurs” (in some form or another) whose wealth is invested primarily in private firms. Related, the threshold for membership in the top 0.1% of the wealth distribution was roughly $20,000,000 in 2012 which is an order of magnitude higher than the corresponding threshold for the labor income distribution of roughly $1,000,000, suggesting that large fortunes cannot be solely due to accumulated labor income.  

---

30 One can further show that, as long as the tail of the income distribution is thinner than that of the wealth distribution, the tail of the wealth distribution is completely independent of the properties of the labor income process. This follows from the property of Pareto distributions that the sum or product of two fat-tailed random variables inherits the fatter of the two tails, i.e. one tail always completely dominates.  

31 The corresponding thresholds for membership in the top 1% were roughly $4,000,000 for wealth and roughly $300,000 for income. The income numbers are from (Guvenen, Kaplan and Song, 2014) and the wealth numbers are from (Saez and Zucman, 2014). To examine this point in a more satisfactory manner would require reliable data on the joint distribution of income and wealth, which is unfortunately not available.
6.3 Dynamics of Wealth Inequality: Theoretical Results

We now show how our theoretical results can be extended to the case of the wealth dynamics (2) or equivalently (3) in combination friction 4. The logarithm of wealth \( x_{it} = \log w_{it} \) satisfies

\[
dx_{it} = (ye^{-x_{it}} + \mu)dt + \sigma dZ_{it}
\]

(39)

where the reader should recall that \( y \) denotes labor income which enters additively into (2). The addition of the labor income term \( y \) introduces some difficulties for extending Proposition 1 to the case of the wealth dynamics (39). However, note that for large wealth levels this term becomes negligible. It is therefore still possible to derive a tight upper bound on the speed of convergence of the cross-sectional distribution.

**Proposition 6** Consider the wealth process (39). If a stationary distribution \( p_\infty(x) \) exists, then 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 \( \lambda \). With friction 4, the rate of convergence

\[
\lambda = -\lim_{t \to \infty} \frac{1}{t} \ln ||p(x, t) - p_\infty(x)||
\]

is given by

\[
\lambda \leq \frac{1}{2} \frac{\mu^2}{\sigma^2} 1_{\{\mu < 0\}} + \delta
\]

where \( 1_{\{\cdot\}} \) is the indicator function, and with equality for \(|\mu| \) below a threshold \( |\mu^*| \). The corresponding half life is given by \( t_{1/2} = \ln(2)/\lambda \).

In the case of friction 4, it is not possible to obtain an exact formula for the speed of convergence for all parameter values. However, we show that the formula holds exactly if the drift of the process is “close enough” to zero. And more importantly, the speed of convergence is bounded above and, in particular, is equal to or less than the speed with friction 1.

6.4 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 for the United States.
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 inputs: data on the average pre-tax rate of return, on capital income taxes, and on a measure of the economy’s growth rate. We use the series of top marginal capital gains taxes from Saez, Slemrod and Giertz (2012, Table A1) and data on the GDP growth rate of the United States from the Penn World Tables. Good measures of the average before-tax rate of return are harder to come by. For now, we simply assume a constant pre-tax return of five percent. The results are identical for other constant returns, e.g. ten percent. In future versions of this paper, we plan on exploring more carefully constructed measures of \( r - g \), e.g. those computed by Auerbach and Hassett (2015). 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 increasing (Figure 1).\(^{32}\) The Figure therefore suggests that, a priori, the

\[ \text{Figure 7: Our measure of } r - g \text{ for the United States} \]

theory that \( r - g \) is a potential candidate for explaining increasing wealth inequality.

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

\(^{32}\)In the Figure and our experiment we smooth the time series for GDP growth rates by using four-year moving averages. Results are virtually unchanged when instead using the raw data.
is on the upper end of values estimated or used in the existing literature. Second, given \( \sigma \) 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 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 (14). A difficulty in operationalizing this formula is that it requires 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.

With these numbers in hand we obtain a half life of

\[
\frac{1}{2} \geq \frac{\ln(2) \times 8 \times \eta^2}{\sigma^2} = \frac{\ln(2) \times 8 \times (0.65)^2}{0.3^2} \approx 26 \text{ years}
\]

That is, on average, the distribution takes 26 years to cover half the distance to the new steady state. Consider now the speed of convergence in the tail, for example consider \( \lambda(\xi) \) from (21) for the 1.4th moment of the wealth distribution, \( \xi = -1.4 \). Using that \( \mu = -\sigma^2/(2\eta) \), we have that for \( \xi = -1.4 \), the half life \( \frac{1}{2}(\xi) = \ln(2)/\lambda(\xi) \) satisfies

\[
\frac{1}{2}(\xi) \geq \ln(2) \left( \frac{-\sigma^2}{2\eta} - \frac{\sigma^2}{2} \xi^2 \right)^{-1} = \ln(2) \left( \frac{(0.3)^2}{2 \times 0.65} 1.4 - \frac{(0.3)^2}{2} (1.4)^2 \right)^{-1} \approx 79.5 \text{ years}.
\]

Figure 8 displays the results of our experiment. 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).

---

33 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. They estimate an average \( \sigma \) of 0.18. Moskowitz and Vissing-Jorgensen (2002) argue for \( \sigma \) of 0.3. Angeletos (2007) and Angeletos and Calvet (2006) use \( \sigma = 0.2 \). These relatively high estimates for \( \sigma \) are consistent with the finding of Carroll (2000) that the portfolios of the rich are heavily skewed toward risky assets, particularly their own privately held businesses.

34 We compute \( \eta \) from the relative wealth shares in Figure 2 as \( \eta(p) = \log S(p/10)/S(p) \). We here use \( \eta(1) = S(0.1)/S(1) \).
6.5 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 (24) and (25) that feature heterogeneity in mean growth rates or deviations from Gibrat’s law have the potential to deliver fast transitions.\(^{35}\) We view both as potentially relevant for the case of wealth dynamics.

- Rise in rate of returns of super wealthy relative to wealthy (top 0.01 vs. top 1%). Better investment advice? Better at taking advantage of “tax loopholes”?
- Rise in saving rates of super wealthy relative to wealthy. Saez and Zucman (2014) provide some evidence.

7 Conclusion

TO BE COMPLETED.

\(^{35}\)The processes now have an additional term \(ye^{-\mu t} dt\).
Appendix

A Proofs

A.1 Complements to the Proof of Proposition 1

Proof of Lemma 5
Define operators $L$ and $B$ by $Lq = -q_t + Aq$ and $Bq = \phi E [q (x - g) - q (x)]$. Also define $s (x, t) := \text{sign} (q (x, t))$. Using $q (x, t) = |q (x, t)| s (x, t)$, we have:

$$L |q| = L (qs) = s \left[ -q_t - \delta q - \mu q_x + \frac{\sigma^2}{2} q_{xx} \right] + q \left[ -s_t - \mu s_x + \frac{\sigma^2}{2} s_{xx} \right]$$

$$+ \sigma^2 q_x s_x + \phi E \left[ |q (x - g)| - |q (x)| \right]$$

Note that $q_t = Aq$ implies:

$$s \left[ -q_t - \delta q - \mu q_x + \frac{\sigma^2}{2} q_{xx} \right] = -sBq$$

Also,

$$q \left[ -s_t - \mu s_x + \frac{\sigma^2}{2} s_{xx} \right] = 0$$

because if on an open connected set, $q (x, t) \neq 0$, then $s (x, t)$ is locally constant on that open set, and $-s_t - \mu s_x + \frac{\sigma^2}{2} s_{xx} = 0$.

In addition

$$\sigma^2 q_x s_x = \sigma^2 q_x \partial_x \text{sign} (q (x, t)) = \sigma^2 q_x^2 \text{sign}' (q (x, t)) \geq 0$$

Finally Jensen’s inequality implies:

$$E [||q (x - g)|| - |q (x)|] \geq |E [q (x - g)]| - |q (x)|$$

so

$$L |q| \geq -sBq + \phi \left\{ |E [q (x - g)]| - |q (x)| \right\}$$

$$= +\phi \left\{ -sE [q (x - g)] - sq (x) + |E [q (x - g)]| - |q (x)| \right\}$$

$$= +\phi \left\{ -sE [q (x - g)] + |E [q (x - g)]| \right\} \text{ as } sq (x) = |q (x)|$$

$$\geq 0 \text{ as } |s| \leq 1$$

We conclude: $L |q| \geq 0$, i.e. $-|f|_t + A|f| \geq 0$, and $|f|_t \leq A|f|$. □
Proof of Lemma 7  Proof: Write \( q(x, t) = e^{-\delta t}Q(x, t) \). We have, by an elementary calculation:

\[
Q_t = CQ = -\mu Q_x + \frac{\sigma^2}{2} Q_{xx} + \phi E [Q(x-g) - Q(x)]
\]

Also, call \( Y_t \) the process \( dY_t = \mu dt + \sigma dZ_t + dJ_t \). Call also \( A(t) := \int xQ(x, t) \, dx \), so that \( a(t) = e^{-\delta t} A(t) \). Then,

\[
A(t) = \int xQ(x, t) \, dx = \int Q(x, 0) E[x+Y_t] \, dx
\]

as \( \int Q(x, 0) \, dx = 0 \). This implies: \( a(t) = e^{-\delta t} a(0) \). This

This shows that the decay \( \lambda \leq \delta \).

A.1.1  Case with the reflecting barrier

First consider the case \( \delta = 0, \mu < 0, \sigma^2 > 0 \) and a reflecting barrier \( x = 0 \) (friction 1). In this case, the dynamics of \( p \) are described by

\[
p_t = \mathcal{A}^* p, \quad \mathcal{A}^* = -\mu \frac{\partial}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2}
\]  (40)

with boundary condition

\[
0 = -\mu p(0, t) + \frac{\sigma^2}{2} \frac{\partial p(0, t)}{\partial x}.
\]  (41)

As explained in the main text, the key insight is that the speed of convergence of \( p \) is governed by the second eigenvalue of \( \mathcal{A}^* \), and the key step is to obtain an analytic formula for this second eigenvalue which is given by \( |\lambda_2| = \frac{1}{2} \frac{\mu^2}{\sigma^2} \).

Mathematical Preliminaries.  The following definitions will be useful in what follows below.

Definition 1:  the inner product of two continuous functions \( u \) and \( v \) is

\[
<u, v> = \int_{-\infty}^{\infty} u(x)v(x) \, dx.
\]

Definition 2:  consider an operator \( \mathcal{A} \). The adjoint of \( \mathcal{A} \) is the operator \( \mathcal{A}^* \) satisfying

\[
<\mathcal{A}u, p> = <u, \mathcal{A}^* p>
\]
Definition 3: An operator $\mathcal{B}$ is *self-adjoint* if $\mathcal{B}^* = \mathcal{B}$.

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

Note that the adjoint is the infinite-dimensional analogue of a matrix transpose. To see this note that for any $N \times N$ matrix $A$ and $N \times 1$ vectors $u$ and $v$, the transpose $A^T$ satisfies $Au \cdot v = u \cdot A^Tv$ where $\cdot$ is the inner product $u \cdot v = \sum_{i=1}^{N} u_i v_i$. Similarly, a self-adjoint operator is the infinite-dimensional analogue of a symmetric matrix. Theorem A.1 is the analogue of the result that all eigenvalues of a symmetric matrix are real and can be found in any book on linear operators.

Lemma 8 The adjoint of $\mathcal{A}^*$ defined in (40) is

$$u_t = Au, \quad \mathcal{A} = \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} + \mu \frac{\partial}{\partial x}$$

with boundary condition

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

Proof of Lemma: Using Definition 2, we have

$$< Au, p > = \int_{0}^{\infty} \left( \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2} + \mu \frac{\partial u}{\partial x} \right) p dx$$

$$= \frac{\sigma^2}{2} \int_{0}^{\infty} \frac{\partial^2 u}{\partial x^2} p dx + \mu \int_{0}^{\infty} \frac{\partial u}{\partial x} p dx$$

$$= \frac{\sigma^2}{2} \frac{\partial u}{\partial x} \bigg|_{0}^{\infty} - \frac{\sigma^2}{2} u \bigg|_{0}^{\infty} + \mu u \bigg|_{0}^{\infty} + \int_{0}^{\infty} u \left( \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} - \mu \frac{\partial p}{\partial x} \right) dx$$

$$= \int_{0}^{\infty} u \left( \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} - \mu \frac{\partial p}{\partial x} \right) dx$$

$$= < u, A^* p >$$

where the fourth equality follows from the boundary conditions (41) and (43).

Main Proof. With these preliminaries in hand, we proceed with the proof of the theorem. The goal is to analyze the eigenvalues of $\mathcal{A}$ 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 9 Consider \( u \) satisfying (42) and the corresponding stationary distribution, \( p_\infty = e^{(2\mu/\sigma^2)x} \). Then \( v = up_\infty = ue^{(\mu/\sigma^2)x} \) satisfies
\[
v_t = \mathcal{B}v, \quad \mathcal{B} = \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} - \frac{1}{2} \frac{\mu^2}{\sigma^2}
\] (44)
with boundary condition \( \frac{\partial v(0,t)}{\partial x} = \frac{\mu}{\sigma} v(0,t) \). Furthermore, \( \mathcal{B} \) is self-adjoint.

Proof: \( (44) \) follows from differentiating \( v = ue^{(\mu/\sigma^2)x} \). To see that \( \mathcal{B} \) is self-adjoint, follow the same steps as in Lemma 8 to conclude that for any \( u, p \),
\[
\langle B u, p \rangle = \langle u, B p \rangle.
\]

Lemma 10 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 \(|\lambda| > |\lambda_2|\).

Proof of Lemma: Consider the eigenvalue problem \( \mathcal{B} \varphi = \lambda \varphi \) or equivalently
\[
\frac{\sigma^2}{2} \varphi''(x) - \frac{1}{2} \frac{\mu^2}{\sigma^2} \varphi(x) = \lambda \varphi(x) \quad (45)
\]
\[
\varphi'(0) = \frac{\mu}{\sigma^2} \varphi(0) \quad (46)
\]
We are looking for non-positive eigenvalues \( \lambda \) and the question is: for what values of \( \lambda \leq 0 \) does \( (45) \) have a solution \( \varphi(x) \) that satisfies the boundary condition \( (46) \) and stays bounded as \( x \to \infty \)? To answer this question, note that for a given \( \lambda \leq 0 \), the general solution to \( (45) \) is \( \varphi(x) = c_1 e^{ax} + c_2 e^{-ax} \) where \( a \) satisfies \( \frac{\sigma^2}{2} a^2 = \frac{1}{2} \frac{\mu^2}{\sigma^2} + \lambda \). Consider four different cases:

1. \( \lambda = 0 \). In this case, \( a = \frac{\mu}{\sigma^2} \), i.e. \( \varphi(x) = e^{\frac{\mu}{\sigma^2}x} \) which satisfies \( (46) \) and stays bounded as \( x \to \infty \) (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, \( a \) is real and positive. We therefore need \( c_1 = 0 \) so that \( \varphi \) stays bounded as \( x \to \infty \). But then \( (46) \) 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, \( a = 0 \) i.e. \( \varphi(x) = c \) which satisfies \( (46) \) with \( c = 0 \) and stays bounded as \( x \to \infty \). 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, \( a \) is complex. 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 \( x \to \infty \). We can therefore choose \( c_1, c_2 \neq 0 \) to satisfy \( (46) \). 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 \( \lambda = 0 \)
This concludes the proof of Proposition 1 for the case of Friction 1. The proof for the case of Friction 3 (reflecting barrier, \( \delta > 0 \)) follows analogous steps. \( \square \)
A.2 Proof of Proposition 5

The distribution \( p(x, t) \) satisfies the Kolmogorov Forward equation

\[
p_t = -\mu p_x + \frac{\sigma^2}{2} p_{xx} - \delta p + \varphi (p * f - p) + \delta \cdot \delta_0
\]

where \(*\) is the convolution operator. The Laplace transform is

\[
\hat{p}_t = -\xi \mu \hat{p} + \xi \frac{\sigma^2}{2} \hat{p} - \delta \hat{p} + \varphi (\hat{f} - 1) \hat{p} + \delta
\]

Integrating we obtain (37). □


We paraphrase the summary in Edmans, Gabaix and Landier (2009) of Gabaix and Landier (2008). A continuum of firms and potential managers are matched together. Firm \( n \in [0, N] \) has size \( S(n) \) and manager \( m \in [0, N] \) has talent \( T(m) \). Low \( n \) denotes a larger firm and low \( m \) a more talented manager: \( S'(n) < 0, \ T'(m) < 0 \). \( n \ (m) \) can be thought of as the rank of the manager (firm), or a number proportional to it, such as its quantile of rank.

We consider the problem faced by one particular firm. The firm has a “baseline” value of \( S \). At \( t = 0 \), it hires a manager of talent \( T \) for one period. The manager’s talent increases the firm’s value according to

\[
S' = S + CT S^\gamma,
\]

where \( C \) parameterizes the productivity of talent. If large firms are more difficult to change than small firms, then \( \gamma < 1 \). If \( \gamma = 1 \), the model exhibits constant returns to scale (CRS) with respect to firm size.

We now determine equilibrium wages, which requires us to allocate one CEO to each firm. Let \( w(m) \) denote the equilibrium compensation of a CEO with index \( m \). Firm \( n \), taking the market compensation of CEOs as given, selects manager \( m \) to maximize its value net of wages:

\[
\max_m C S(n)^\gamma T(m) - w(m).
\]

The competitive equilibrium involves positive assortative matching, i.e. \( m = n \), and so \( w'(n) = C S(n)^\gamma T'(n) \). Let \( \underline{w}_N \) denote the reservation wage of the least talented CEO \( (n = N) \). Hence we obtain the classic assignment equation (Sattinger (1993), Tervio (2007)):

\[
w(n) = -\int_n^N C S(u)^\gamma T'(u) du + \underline{w}_N.
\]

41
Specific functional forms are required to proceed further. We assume a Pareto firm size distribution with exponent $1/\alpha$: $S(n) = An^{-\alpha}$. Using results from extreme value theory, GL use the following asymptotic value for the spacings of the talent distribution: $T'(n) = -Bn^{\beta-1}$. These functional forms give the wage equation in closed form, taking the limit as $n/N \to 0$:

$$w(n) = \int_n^N A^\gamma BC u^{-\alpha\gamma+\beta-1} du + \frac{A^\gamma BC}{\alpha\gamma - \beta} [n^{-\alpha\gamma - \beta} - N^{-\alpha\gamma - \beta}] + w_{n} \sim \frac{A^\gamma BC}{\alpha\gamma - \beta} n^{-\alpha\gamma - \beta}.$$  

(49)

To interpret equation (49), we consider a reference firm, for instance firm number 250 – the median firm in the universe of the top 500 firms. Denote its index $n_*$, and its size $S(n_*)$. We obtain Proposition 2 from GL, which we repeat here. In equilibrium, manager $n$ runs a firm of size $S(n)$, and is paid according to the “dual scaling” equation

$$w(n) = D(n_*) S(n_*)^{\beta/\alpha} S(n)^{\gamma - \beta/\alpha}$$  

(50)

where $S(n_*)$ is the size of the reference firm and $D(n_*) = -Cn_* T'(n_*) / (\alpha\gamma - \beta)$ is a constant independent of firm size.\textsuperscript{36}

\section{B Stationary Distributions for Frictions 2 to 4}

[TO BE COMPLETED]

\textbf{References}


\textsuperscript{36}The derivation is as follows. Since $S = An^{-\alpha}$, $S(n_*) = An_*^{-\alpha}$, $n_* T'(n_*) = -Bn_*^\beta$, we can rewrite equation (49) as follows:

$$(\alpha\gamma - \beta) w(n) = A^\gamma BC n^{-(\alpha\gamma - \beta)} = CBn_*^\beta \cdot (An_*^{-\alpha})^{\beta/\alpha} \cdot (An^{-\alpha})^{(\gamma - \beta/\alpha)}$$

$$= -Cn_* T'(n_*) S(n_*)^{\beta/\alpha} S(n)^{\gamma - \beta/\alpha}.$$


DeBacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Rammath, and Ivan Vidangos. 2013. “Rising Inequality: Transitory or Persistent? New Evidence from a


Pareto, Vilfredo. 1896. Cours d’économie politique.


