The Dynamics of Inequality*

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

December 24, 2015

Abstract

The past forty years have seen a rapid rise in top income inequality in the United States. While there is a large number of existing theories of the Pareto tail of the long-run income distributions, almost none of these address the fast rise in top inequality observed in the data. We show that standard theories, which build on a random growth mechanism, generate transition dynamics that are an order of magnitude too slow relative to those observed in the data. We then suggest two parsimonious deviations from the canonical model that can explain such changes: "scale dependence" that may arise from changes in skill prices, and "type dependence," i.e. the presence of some "high-growth types." These deviations are consistent with theories in which the increase in top income inequality is driven by the rise of "superstar" entrepreneurs or managers.

*NYU Stern, Dauphine, Collège de France, Princeton. We thank the editor and referees for many constructive suggestions, Fernando Alvarez, Roland Bénabou, Jess Benhabib, Fatih Guvenen, Chad Jones, Greg Kaplan, Erzo Luttmer, Wojciech Kopczuk, Makoto Nirei, Ezra Oberfield, Jonathan Parker, John Shea, Joe Sullivan and Gabriel Zucman for their insights, and seminar participants at UQAM, Queen's University, Cornell, the Chicago Fed, Princeton, the University of Maryland, the University of Chicago, the 2015 NBER Summer Institute on Income Distribution and Macroeconomics, the NBER EFG meeting, and 2015 SAET conference for useful comments, and Moll thanks the Kauffman Foundation for financial support. We also thank Cristian Alonso, Joshua Bernstein, Nik Engbom, Mildred Hager and Jason Ravit for excellent research assistance.

1 Introduction

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

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

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

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

[1] See e.g. Piketty (2014), Saez and Zucman (2015), Bricker, Henriques, Krimmel, and Sabelhaus (2015) and Kopczuk (2015).

its upper tail. We argue that, in standard theories, transitions are even slower in the tail and, additionally, that our low measure of the average speed of convergence overestimates the speed of convergence in the upper tail. We also show that allowing for jumps in the income process, while useful for descriptively matching micro-level data, does not help with generating fast transitions.

Given this negative result, we are confronted with a puzzle: what, then, explains the observed rise in top income inequality? We develop an "augmented random growth model" that features two parsimonious departures from the canonical model that do generate fast transitions. Both departures are deviations from Gibrat's law, the assumption that the distribution of income growth rates is independent of the income level. The first departure is *type dependence* of the growth rate distribution and, in particular, the presence of some "high-growth types".[2] For instance, some highly skilled entrepreneurs or managers may experience much higher average earnings growth rates than other individuals over short to medium horizons. We argue analytically and quantitatively that this first departure can explain the observed fast rise in income inequality. The second departure consists of *scale dependence* of the growth rate distribution which arises from shocks that disproportionately affect high incomes, e.g. changing skill prices in assignment models.[3] Scale dependence can generate infinitely fast transitions in inequality.

To obtain our analytic formulas for the speed of convergence, we employ tools from ergodic theory and the theory of partial differential equations. Our measure of the average speed of convergence is the first non-trivial eigenvalue or "spectral gap" of the differential operator governing the stochastic process for income. One of the main contributions of this paper is to derive an analytic formula for this first non-trivial eigenvalue (i.e. the second eigenvalue) for a large variety of random growth processes.[4] We obtain our measure of the speed of convergence in the tail of the distribution by making use of the fact that the solution to the Kolmogorov Forward equation for random growth processes can be characterized tightly by calculating the Laplace transform of this equation. Our clean results, which a discrete-time analysis would be unable to deliver, constitute an example of the usefulness of continuous-time methods in economics.

A large theoretical literature builds on random growth processes to theorize about the

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

[3]Technically, shocks that generate scale dependence affect log income multiplicatively, rather than additively, as in the usual random growth model.

[4]See Hansen and Scheinkman (2009) for a related application of operator methods in economics.

upper tails of income and wealth distributions. Early theories of the income distribution include Champernowne (1953) and Simon (1955), with more recent contributions by Nirei (2009), Toda and Walsh (2015), Kim (2013), Jones and Kim (2014) and Luttmer (2015). Similarly, random growth theories of the wealth distribution include Wold and Whittle (1957) and more recently Benhabib, Bisin, and Zhu (2011, 2015a,b), Jones (2015), and Acemoglu and Robinson (2015). All of these papers focus on the income or wealth distribution *at a given point in time* by studying stationary distributions, and none of them analyze transition dynamics. Aoki and Nirei (2015) are a notable exception, who examine the dynamics of the income distribution and ask whether tax changes can account for the rise in top income inequality observed in the United States. Our paper differs from theirs in that we obtain a number of analytic results providing a tight characterization of transition dynamics in random growth models whereas their analysis of transition dynamics is purely numerical.[5]

Our finding that type dependence delivers fast dynamics of top inequality is also related to Guvenen (2007), who has argued that an income process with heterogeneous income profiles provides a better fit to the micro data than a model in which all individuals face the same income profile. In our model variant with multiple "growth types," 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, Karahan, Ozkan, and Song (2015). One key difference between our model with multiple growth types and the standard random growth model is that, in the standard model, the key determinant for an individual's place in the income distribution is her age. In contrast, in a model with type dependence another important determinant is the individual's growth type which may represent her occupation or her talent as an entrepreneur. This is consistent with salient patterns of the tail of the income distribution in the United States (Guvenen, Kaplan, and Song, 2014).[6]

One of the most ubiquitous regularities in economics and finance is that the empirical distribution of many variables is well approximated by a power law. For this reason, theories of random growth are an integral part of many different strands of the literature beside those studying the distributions of income and wealth.[7] For example, they have been used to study the distribution of city sizes (Gabaix, 1999) and firm sizes (Luttmer, 2007), the shape of the

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

[6] Luttmer (2011) makes a similar observation about the relationship between a firm's age and its place in the firm size distribution. As Luttmer puts it succinctly: "Gibrat implies 750-year-old firms."

[7] Our focus is on the dynamics of income inequality. However, our criticism and suggested fixes apply without change to random growth models of the wealth distribution. In Online Appendix H we work out in detail the implications of our theoretical results for the dynamics of wealth inequality.

production function (Jones, 2005), and in many other contexts (see the review by Gabaix, 2009). The tools and results presented in this paper should therefore also prove useful in other applications.[8]

The paper is organized as follows. Section 2 states the main motivating facts for our analysis, and Section 3 reviews random growth theories of the income 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 fast transitions, and shows that these have the potential to account for the fast transitions observed in the data. Section 6 concludes.

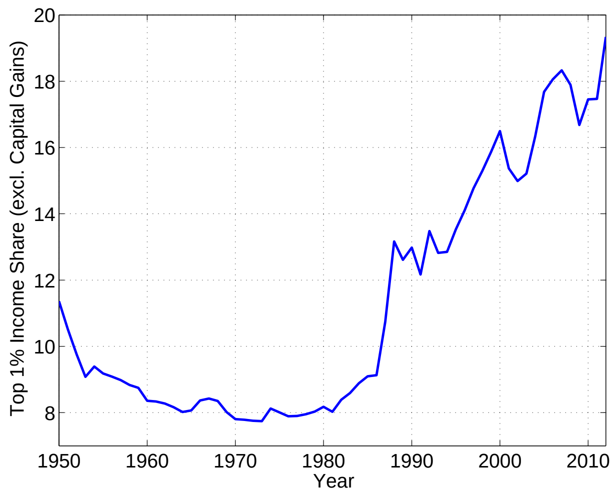
2 Motivating Facts

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

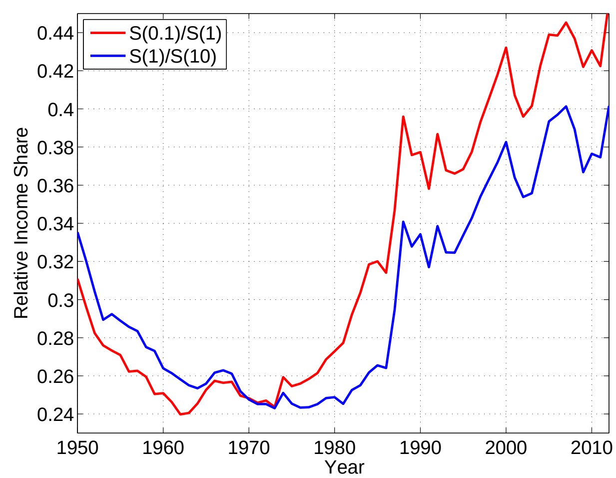
Panel (a) of Figure 1 displays the evolution of measures of the top 1% income share. It shows the large and rapid increase in the top 1% income share that has been extensively documented by Piketty and Saez (2003), Atkinson, Piketty, and Saez (2011) and others.[9] As already noted, the upper tail of the income distribution follows a power law, 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 are, in turn, X times richer than the top 10%, where X is a fixed number. Equivalently, the top 0.1% income share is a fraction Y of the top 1% income share, which, in turn, is a fraction Y of the top 10% income share, and so on. We now explore this fractal pattern in the data using a strategy borrowed from Jones and Kim (2014). Panel (b) of Figure 1 plots the income share of the top 0.1% relative to that of the

[8]Other useful tools are in the following works. Bouchaud and Mézard (2000) calculate the decay rate of the autocorrelation of an individual's wealth, and find that it depends on the tail exponent (when this tail exponent is smaller than two so that the variance ceases to exist, the expression for this decay rate coincides with the speed of convergence in a special case of our model). Saichev, Malevergne, and Sornette (2009) and Malevergne, Saichev, and Sornette (2013) calculate a number of probability densities and hazard rates at finite times. Those works study the dynamics of individuals in an economy already at the steady state – while we study the entire economy off its steady state, but transitioning towards it.

[9]The series is from the "World Top Incomes Database." Here, we plot total income (salaries plus business income plus capital income) excluding capital gains. The series display a similar trend when we include capital gains or focus on salaries only (though the levels are different). Note also that a significant part of the increase in top inequality is concentrated in 1987 and 1988 just after the Tax Reform Act of 1986 which sharply reduced top marginal income tax rates. Part of this increase may therefore be due to changes in tax reporting and realizations rather than actual changes in inequality. We discuss this in detail in Appendix G.



(a) Top Income Inequality



(b) Relative Income Shares

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

top 1% and the income share of the top 1% relative to that of the top 10%. As expected, the two lines track each other relatively closely. More importantly, there is an upward trend in both lines. That is, there has been a *relative* increase in top income shares. As we explain in more detail below, this increase in "fractal inequality" implies equivalently a "fattening" of the Pareto tail of the income distribution.

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

3 Random Growth Theories of Income Inequality

Our starting point is the existing theories that can explain top income inequality at a point in time, meaning that they can generate stationary income 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 present a relatively general random growth model of income dynamics and characterize its stationary distribution. This framework will also 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 indexed by i. Workers are heterogeneous in their income or wage w_{it}. For brevity, we here only spell out the reduced-form dynamics of income. We discuss possible microfoundations below and provide one example in Appendix B. We will later find it useful to conduct much of the analysis in terms of the logarithm of income, $x_{it} = \log w_{it}$, whose dynamics are:

$$dx_{it} = \mu dt + \sigma dZ_{it} + g_{it} dN_{it}, \tag{1}$$

where Z_{it} is a standard Brownian motion and where N_{it} is a jump process with intensity ϕ.[10] The innovations g_{it} are drawn from an exogenous distribution f. The distribution f can have arbitrary support and it may be either thin-tailed (e.g. a normal distribution) or fat-tailed.

All theories of top inequality add some "stabilizing force" to the pure random growth process (1) to ensure the existence of a stationary distribution (Gabaix, 2009). In the absence of such a stabilizing force, the cross-sectional variance of x_{it} grows without bound. We consider two possibilities. First, workers may die (retire) at rate δ, in which case they are replaced by a young worker with wage x_{it} drawn from a distribution $\psi(x)$. Second, there may be a lower bound $\underline{x}$ on income. The simplest possibility is that this lower bound takes the form of a reflecting barrier. More generally, we consider exit at $\underline{x}$ with entry (i.e. reinjection) at a point $x > \underline{x}$ drawn from a distribution $\rho(x)$. For instance, Luttmer (2007) analyzes the case of a "return process" where the reinjection occurs at a point x_*, which is the special case in which ρ is a Dirac distribution at x_*, $\rho(x) = \delta_{x_*}(x)$.[11] A natural interpretation for a lower bound on income is that workers exit the labor force if their income falls below some threshold. For simplicity, we normalize $\underline{x} = 0$ throughout the remainder of the paper, i.e. the corresponding threshold for income is $\underline{w} = 1$. When the process (1) features jumps $\phi \neq 0$, we only consider death as a stabilizing force.[12]

The income dynamics (1) can be microfounded in a variety of ways. Appendix B provides one such microfoundation: workers optimally invest to accumulate human capital, a process that also involves some luck. But other microfoundations are possible as well and a large number of theories of the upper tail of the income distribution ends up with a similar reduced

[10] That is, the innovations dZ_{it} are normally distributed: approximately, $dZ_{it} \simeq \varepsilon_{it}\sqrt{dt}, \varepsilon_{it} \sim \mathcal{N}(0,1)$, for a small dt. Similarly, there is a jump in $(t - dt, t]$ (i.e., $dN_{it} = 1$) with probability ϕdt and no jump (i.e., $dN_{it} = 0$) with probability $1 - \phi dt$; if there is a jump, it is a random $\tilde{g}$.

[11] Luttmer (2007) shows that the stationary distribution of the process with exit and entry converges to one associated with a reflecting barrier at $\underline{x}$ as $x_* \downarrow \underline{x}$.

[12] For instance, it is messy to define a reflecting barrier in the presence of jumps.

form.[13]

Because the process (1) allows for jumps, it is considerably more general than the more commonly used specification in which income innovations are log-normally distributed (a geometric Brownian motion for income). Recent research suggests that the standard specification is a quite imperfect description of the data. For instance, Guvenen, Karahan, Ozkan, and Song (2015) document, using administrative data, that earnings innovations are very fat-tailed and much more so than a normally distributed random variable. In our continuous time setup, the most natural way of generating such kurtosis is to allow for jumps.[14] At the same time, the process (1) makes the strong assumption that the parameters μ and σ as well as the distribution f do not depend on the level of income, a strict form of Gibrat's law. We show below that this assumption can be relaxed considerably to the case when the drift and diffusion are arbitrary functions $\mu(x)$ and $\sigma(x)$ of the income level that are constant for large x.

A large literature estimates reduced-form labor income processes similar to (1) using panel data.[15] In particular, (1) is the special case of the widespread "permanent-transitory model" of income dynamics, but with only a permanent component. As a result, good estimates are available for its parameter values. The process could easily be extended to feature a transitory component, e.g. by introducing jumps that are distributed i.i.d. over time and across individuals.

3.2 Stationary Income Distribution

The properties of the stationary distribution of the process (1) for the logarithm of income $x_{it} = \log w_{it}$ are well understood. In particular, under certain parameter restrictions, this stationary distribution has a Pareto tail[16]

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

where C is a constant and $\zeta > 0$ is a simple function of the parameters μ, σ and the distribution of jumps f (see e.g. Gabaix (2009)). Equivalently, the distribution of log

[13]See e.g. Champernowne (1953), Simon (1955), Nirei (2009), Toda and Walsh (2015), Aoki and Nirei (2015), Kim (2013), Jones and Kim (2014) and Luttmer (2015) for models with similar reduced forms. Some of these are derived from individual optimization, but others are not.

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

[15]See e.g. MaCurdy (1982), Heathcote, Perri, and Violante (2010) and Meghir and Pistaferri (2011).

[16]Here and elsewhere "$f(x) \sim g(x)$" for two functions f and g means $\lim_{x\to\infty} f(x)/g(x) = 1$.

income has an exponential tail, $\mathbb{P}(x_{it} > x) \sim Ce^{-\zeta x}$. Without jumps $\phi = 0$, ζ is the positive root of[17]

$$0 = \frac{\sigma^2}{2}\zeta^2 + \zeta\mu - \delta, \tag{2}$$

which equals

$$\zeta = \frac{-\mu + \sqrt{\mu^2 + 2\sigma^2\delta}}{\sigma^2}. \tag{3}$$

The constant ζ is called the "power law exponent,"with a smaller ζ corresponding to a fatter tail. We find it useful to refer to the inverse of the power law exponent $\eta = 1/\zeta$ as "top inequality". Intuitively, tail inequality is increasing in μ and σ and decreasing in the death rate δ. In Appendix C, we provide a complete characterization of the stationary distributions for different "stabilizing forces." In particular, we spell out the assumptions under which there exists a unique stationary distribution. For the remainder of the paper we assume that these assumptions are satisfied.

To make the connection to the empirical evidence in the introduction, note that if the distribution of w has a Pareto tail above the pth percentile, then the share of the top $p/10$th percentile relative to that of the pth percentile is given by $\frac{S(p/10)}{S(p)} = 10^{\eta-1}$. There is, therefore, a one-to-one mapping between the relative income shares in panel (b) of Figure 1 and the top inequality parameter $\eta = 1/\zeta$.[18] Most existing contributions focus on the stationary distribution of the process (1) and completely ignore the corresponding transition dynamics. It is unclear whether these theories can explain the observed dynamics of the tail parameter η. This is what we turn to next.

4 The Baseline Random Growth Model Generates Slow Transitions

Changes in the parameters of the income process (1) lead to changes in the fatness of the right tail of its stationary distribution. For example, an increase in the innovation variance σ^2 leads to an increase in stationary tail inequality η in (3). 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 standard random growth model (1) gives rise to very slow transition dynamics.

Throughout this section, we conduct the following thought experiment. Initially at time

[17] The proof is standard: we plug $p(x) = Ce^{-\zeta x}$ into (5), which leads to (2).

[18] In particular $\eta = 1 + \log_{10} \frac{S(p/10)}{S(p)}$. See Jones and Kim (2014) and Jones (2015) for two papers that use this fact extensively.

$t = 0$, the economy is in a Pareto steady state corresponding to some initial parameters μ_0, σ_0^2 and so on. At time $t = 0$, a parameter changes; for example, the innovation variance σ^2 may increase. Asymptotically as $t \to \infty$, the distribution converges to its new stationary distribution. The question is: what can we say about the speed of this transition? We present two sets of results corresponding to different notions of the speed of convergence. The first notion measures an "average" speed of convergence throughout the distribution. The second notion captures differential speeds of convergence across the distribution, allowing us in particular to put the spotlight on its upper tail.

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

$$||p(x,t) - p_\infty(x)|| := \int_{-\infty}^{\infty} |p(x,t) - p_\infty(x)| dx. \tag{4}$$

The cross-sectional distribution $p(x,t)$ satisfies a Kolmogorov Forward equation. Without jumps ($\phi = 0$) this equation is:

$$p_t = -\mu p_x + \frac{\sigma^2}{2} p_{xx} - \delta p + \delta \psi \tag{5}$$

with initial condition $p(x,0) = p_0(x)$, where we use the compact notation $p_t := \frac{\partial p(x,t)}{\partial t}$, $p_x := \frac{\partial p(x,t)}{\partial x}$, $p_{xx} := \frac{\partial^2 p(x,t)}{\partial x^2}$. The first two terms on the right hand side capture the evolution of x due to diffusion with drift μ and variance σ^2. The third term captures death and, hence, an outflow of individuals at rate δ, and the fourth term captures birth, namely that every "dying" individual is replaced with a newborn drawn from the distribution $\psi(x)$.

When there is a reflecting barrier, p must additionally satisfy the boundary condition:[19]

$$0 = -\mu p + \frac{\sigma^2}{2} p_x, \quad \text{at} \quad x = 0, \quad \text{for all } t. \tag{6}$$

When there is exit at $x = 0$ with reinjection at points strictly above $x = 0$, i.e. $\rho(0) = 0$, the boundary condition is

$$p(0,t) = 0 \quad \text{for all } t. \tag{7}$$

and an additional term $\gamma(t)\rho(x)$ is added to the right-hand side of (5), with $\gamma(t) = \frac{\sigma^2}{2} p_x(0,t)$:

[19] This boundary condition comes from integrating (5) from $x = 0$ to ∞.

$p_t = -\mu p_x + \frac{\sigma^2}{2} p_{xx} - \delta p + \delta\psi + \gamma\rho$. This term captures reinjection after exit: a density $\gamma(t) = \frac{\sigma^2}{2} p_x(0,t)$ of agents touch the barrier at time t^-, and they are reinjected at the random location drawn from the distribution $\rho(x)$.[20]

When there are jumps, the Kolmogorov Forward equation (5) becomes

$$p_t = -\mu p_x + \frac{\sigma^2}{2} p_{xx} - \delta p + \delta\psi + \phi\mathbb{E}\left[p(x-g) - p(x)\right]. \tag{8}$$

Relative to (5), the new term is the expectation $\mathbb{E}\left[p(x-g) - p(x)\right]$, which is taken over the random jump g and is multiplied by ϕ, the arrival rate of jumps.[21]

It is often convenient to write these partial differential equations in terms of a differential operator. For instance (5) is

$$p_t = \mathcal{A}^* p + \delta\psi, \qquad \mathcal{A}^* p := -\mu p_x + \frac{\sigma^2}{2} p_{xx} - \delta p. \tag{9}$$

This formulation is quite flexible and can be extended in a number of ways, in particular to the case with jumps or with exit and reinjection.

4.1 Average Speed of Convergence

We now state Proposition 1, one of the two main theoretical results of our paper. For now, we assume that the process (1) does not feature jumps ($\phi = 0$) and do not allow for exit and reinjection. We extend the results to the case with jumps in Proposition 3 and to exit and reinjection in Proposition 2. As mentioned above, we assume that the process (1) satisfies the assumptions in Appendix C that guarantee the existence of a unique stationary distribution $p_\infty(x)$. We additionally make the following assumption.

Assumption 1 *The initial distribution* $p_0(x)$ *satisfies* $\int_{-\infty}^{\infty} \frac{(p_0(x))^2}{e^{-\bar{\zeta}x}} dx < \infty$ *where* $\bar{\zeta} := \frac{-2\mu}{\sigma^2} \leq \zeta$*, and* μ, σ *are the parameters of the new steady state process.*

Note that Assumption 1 is a relatively weak restriction. For instance, assume that p_0 has a Pareto tail $p_0(x) \sim c_0 e^{-\zeta_0 x}$ for large x. Then Assumption 1 is equivalent to $\zeta_0 > \bar{\zeta}/2$, and

[20] To see why the rate at which people exit is given by $\gamma(t) = \frac{\sigma^2}{2} p_x(0,t)$, integrate the Kolmogorov equation (5) from $x = 0$ to ∞, which gives $0 = \mu p(0,t) - \frac{\sigma^2}{2} p_x(0,t) + \gamma$ (using $\int_0^\infty \rho(x)\,dx = 1$). Given $p(0,t) = 0$, we obtain $\gamma(t) = \frac{\sigma^2}{2} p_x(0,t)$.

[21] A jump of g at $x - g$ will transport $p(x-g)$ individuals to location x, hence the term $\phi\mathbb{E}\left[p(x-g)\right]$. Jumps at x make $\phi p(x)$ people leave location x, hence the term $-\phi p(x)$. The net effect is $\phi\mathbb{E}\left[p(x-g) - p(x)\right]$.

a sufficient condition is $\zeta_0 > \zeta/2$,[22] or in terms of top inequality $\eta = 1/\zeta$: $\eta_0 < 2\eta$. That is, Assumption 1 rules out cases in which top inequality in the initial steady state is more than twice as large as that in the new steady state. In particular, it is satisfied in all cases where top inequality in the new steady state is larger than that in the initial steady state, $\eta_0 < \eta$, the case we are interested in.[23]

Proposition 1 (Average speed of convergence) *Consider the income process* (1) *with death and/or a reflecting barrier as a stabilizing force but without jumps ($\phi = 0$). The cross-sectional distribution $p(x, t)$ converges to its stationary distribution $p_\infty(x)$ in the total variation norm for any initial distribution $p_0(x)$. The rate of convergence*

$$\lambda := -\lim_{t\to\infty} \frac{1}{t} \log ||p(x,t) - p_\infty(x)|| \tag{10}$$

depends on whether there is a reflecting barrier at $x = 0$. Without a reflecting barrier

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

With a reflecting barrier, under Assumption 1 and for generic initial conditions,

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

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

The interpretation of the rate of convergence (10) is that, asymptotically as $t \to \infty$, the distribution converges exponentially at rate λ: $||p(x,t) - p_\infty(x)|| \sim ke^{-\lambda t}$. We shall see that Proposition 1 implies that the traditional canonical model delivers convergence that is far too slow: λ is too low compared to empirical estimates.

The intuition for formulas (11) and (12) is as follows. Without a reflecting barrier, the speed is simply given by the death intensity δ. This is intuitive: the higher δ is, the more churning there is in the cross-sectional distribution and the faster the distribution settles down to its invariant distribution. In the extreme case where $\delta \to \infty$, the distribution jumps to its steady state immediately. Next, consider the case with a reflecting barrier, $\mu < 0$, and no death, $\delta = 0$. From (3), stationary tail inequality for this case is $\eta = 1/\zeta = -\frac{\sigma^2}{2\mu}$ and

[22] Indeed, if $\delta = 0$, $\zeta = \bar{\zeta}$. If $\delta > 0$, call $P(r) = -\delta + \mu r + \frac{\sigma^2}{2} r^2$, so that $P(\zeta) = 0$. Given $P(\bar{\zeta}) = -\delta < 0$, we have $\bar{\zeta} < \zeta$.

[23] Proposition 1 can also be extended to the case where p_0 does not decay fast enough, i.e. if $\eta_0 > 2\eta$. In particular, one can bound the speed of convergence, which becomes lower.

therefore the speed of convergence can also be written as

$$\lambda = \frac{\sigma^2}{8\eta^2}. \tag{13}$$

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

In section 4.3 we show that when the parameters μ, σ and δ 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.

As mentioned above, the process (1) is a bit restrictive because it assumes that Gibrat's law holds everywhere in the state space. In fact, it is possible to relax this assumption and still obtain an *upper bound* on the speed of convergence. To this end, consider the more general process:

$$dx_{it} = \mu(x_{it})dt + \sigma(x_{it})dZ_{it}, \quad \mu(x) = \bar{\mu},\ \sigma(x) = \bar{\sigma},\ \text{for } x \geq x^*. \tag{14}$$

where the growth and standard deviation of income depend on income itself. Here $\mu(x)$ and $\sigma(x)$ are quite arbitrary functions that satisfy one condition: there is a threshold income level x^* such that the process is a strict random growth process above this threshold.[24] Furthermore, we now allow for exit with reinjection in addition to a reflecting barrier.

Proposition 2 (Upper bound on average speed of convergence with general process (14)) *Consider the income process* (14) *with a stabilizing force. The cross-sectional distribution* $p(x,t)$ *converges to its stationary distribution* $p_\infty(x)$ *in the total variation norm. The rate of convergence* $\lambda := -\lim_{t\to\infty} \frac{1}{t} \log ||p(x,t) - p_\infty(x)||$ *is* at most as large as *that with a strict random growth process* (1) *from Proposition 1. Without a lower bound on income,* $\lambda \leq \delta$. *With a lower bound on income (either a reflecting barrier or exit with reinjection),* $\lambda \leq \frac{1}{2}\frac{\bar{\mu}^2}{\bar{\sigma}^2}\mathbf{1}_{\{\bar{\mu}<0\}} + \delta$.

As most readers will be more interested in the message of Propositions 1 and 2 than their proofs, we only sketch here the intuition for the proofs. The proof of Proposition 1 without a reflecting barrier analyzes directly the L^1-norm (4) by means of a differential equation

[24] In addition, we assume that the functions μ and σ satisfy simple sufficient conditions ensuring the existence and uniqueness of a steady state. See Appendix E.3.1.

for $|p(x,t) - p_\infty(x)|$. The rough idea of the proof of Proposition 1 with a lower bound is as follows. The entire dynamics of the process for x_{it} are summarized by the operator $\mathcal{A}^*$ defined in (9). This operator is the appropriate generalization of a transition matrix for a finite-state process to processes with a continuum of states such as (1), and it can be analyzed in an exactly analogous way. In particular, the critical property of $\mathcal{A}^*$ governing the speed of convergence of p is its largest non-trivial eigenvalue.[25] The key contribution of Proposition 1 and the main step of the proof is then to obtain an explicit formula for the second eigenvalue of $\mathcal{A}^*$ in the form of (12).[26] The proof of Proposition 2 is similar.

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.

We also ask whether departing from the standard log-normal framework by introducing jumps can help resolve the puzzle raised in the preceding section that random growth processes cannot explain the fast rise of income inequality observed in the data. We find that they cannot: while jumps are useful descriptively for capturing certain features of the data, they do not increase the speed of convergence of the cross-sectional income distribution.

Because we use somewhat different arguments depending on whether there is a lower bound on income or not, we present the results for the two cases separately.

4.2.1 Speed of Convergence in the Tail Without a Lower Bound on Income

Without a lower bound on income, the distribution $p(x,t)$ satisfies the Kolmogorov Forward equation (8), which potentially allows for jumps. One can show (see e.g. Gabaix, 2009, and Appendix C) that in this case the stationary distribution has Pareto tails both as $x \to \infty$

[25] The largest non-trivial eigenvalue λ is the relevant speed of convergence for *generic* initial conditions p_0. As we explain in more detail in Appendix E.2.1, "generic" here means that, for any given p_0, we can find an arbitrarily close $\tilde{p}_0$ that converges at the rate λ. The logic is exactly the same as in the finite-dimensional case analyzed in Appendix E.1: there could in principle be initial conditions that are exactly orthogonal to the eigenvector corresponding to the largest non-trivial eigenvalue. But such initial conditions are knife-edge and the second eigenvalue governs the speed of convergence for any perturbations of such initial conditions.

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

and as $x \to -\infty$:

$$p_\infty(x) \sim \begin{cases} e^{-\zeta_+ x}, & x \to \infty, \\ e^{-\zeta_- x}, & x \to -\infty, \end{cases} \tag{15}$$

with $\zeta_- < 0 < \zeta_+$.[27] Apart from the stationary distribution, the solution to the Kolmogorov Forward equation is cumbersome.

Without a lower bound on income, the entire time path of the solution to the Kolmogorov Forward equation can be characterized conveniently in terms of the "Laplace transform" of p

$$\widehat{p}\,(\xi, t) := \int_{-\infty}^{\infty} e^{-\xi x} p\,(x, t)\, dx = \mathbb{E}\left[e^{-\xi x_{it}}\right], \tag{16}$$

where ξ is a real number and x_{it} represents the random variable (log income) with distribution $p\,(x, t)$.[28] For $\xi \le 0$, the Laplace transform has the natural interpretation of the $-\xi$th moment of the distribution of income, that is $\widehat{p}\,(\xi, t) = \mathbb{E}[w_{it}^{-\xi}]$, where $w_{it} = e^{x_{it}}$ is income. Similarly note that, up to a minus, the Laplace transform is the moment generating function corresponding to the distribution $p(x, t)$, and one can therefore also calculate all moments of *log* income.[29] We show momentarily that we can obtain a clean analytic formula for the entire time path of this object for all t. This is useful because a complete characterization of a function's Laplace transform is equivalent to a complete characterization of the function itself. This is because by varying the variable ξ, we can trace out the behavior of different parts of the distribution. In particular, the more negative ξ is, the more we know about the distribution's tail behavior. In a similar vein, our analysis using Laplace transforms will allow us to characterize tightly the behavior of a weighted version of the L^1-norm in (4):

$$||p(x, t) - p_\infty(x)||_\xi := \int_{-\infty}^{\infty} |p(x, t) - p_\infty(x)| e^{-\xi x} dx. \tag{17}$$

In the special case $\xi = 0$, this distance measure coincides with the L^1-norm defined in (4). But by taking $\xi < 0$, (17) puts more weight on the behavior of the distribution's tail, the main focus of the current section. Note that the Laplace transform (16) ceases to exist if ξ is too negative or too positive. To ensure that the Laplace transform exists we impose $\max\{\zeta_{0,-}, \zeta_-\} < -\xi < \min\{\zeta_{0,+}, \zeta_+\}$ where ζ_+, ζ_- are the tail parameters of the stationary

[27] For instance, without jumps and with reinjection at $x = 0$, the stationary distribution is a double Pareto distribution $p_\infty(x) = c \min\{e^{-\zeta_- x}, e^{-\zeta_+ x}\}$ where $c = -\zeta_- \zeta_+ / (\zeta_+ - \zeta_-)$ and where $\zeta_- < 0 < \zeta_+$ are the two roots of (2).

[28] Note that we here work with the "bilateral" or "two-sided" Laplace transform which integrates over the entire real line. This is in contrast to the one-sided Laplace transform defined as $\int_0^\infty e^{-\xi x} p\,(x, t)\, dx$.

[29] The first moment of log income can be calculated as $-\widehat{p}_\xi(0, t) = \int_{-\infty}^{\infty} x p(x, t) dx = \mathbb{E}[x_{it}]$, the second moment from the second derivative, and so on.

distribution (15) and $\zeta_{0,+}, \zeta_{0,-}$ those of the initial distribution.

We apply the Laplace transform to the Kolmogorov Forward equation (8). For the first two terms we use the rules $\widehat{p_x} = \xi\widehat{p}$ and $\widehat{p_{xx}} = \xi^2\widehat{p}$. Next consider the term capturing jumps, which can be written as

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

where $*$ is the convolution operator. Conveniently, integral transforms like the Laplace transform are the ideal tool for handling convolutions. In particular, the Laplace transform of a convolution of two functions is the product of the Laplace transforms of the two functions: $\widehat{(p*f)}(\xi) = \widehat{p}(\xi)\widehat{f}(\xi)$. Note that the Laplace transform can handle *arbitrary* jump distributions f. Applying these rules to (8), we obtain

$$\widehat{p}_t(\xi,t) = -\lambda(\xi)\widehat{p}(\xi,t) + \delta\widehat{\psi}(\xi) \quad \text{where} \quad \lambda(\xi) := \mu\xi - \frac{\sigma^2}{2}\xi^2 + \delta - \phi(\widehat{f}(\xi)-1) \tag{18}$$

with initial condition $\widehat{p}(\xi,0) = \widehat{p}_0(\xi)$, the Laplace transform of $p_0(x)$ and where $\widehat{\psi}(\xi)$ and $\widehat{f}(\xi)$ are the Laplace transforms of $\psi(x)$ and $f(x)$. Importantly, note that for fixed ξ, (18) is a simple ordinary differential equation for $\widehat{p}$ that can be solved analytically. Note that this strategy would work even if the coefficients μ, σ, δ and ϕ were arbitrary functions of time t. However, it would not work if μ, σ, δ and ϕ depended on income x.

Proposition 3 (Speed of convergence in the tail) *Consider the Laplace transform of the income distribution* $\widehat{p}(\xi,t)$ *defined in* (16)*. Its time path is*

$$\widehat{p}(\xi,t) = \widehat{p}_\infty(\xi) + \left(\widehat{p}_0(\xi) - \widehat{p}_\infty(\xi)\right)e^{-\lambda(\xi)t}, \tag{19}$$

$$\lambda(\xi) := \xi\mu - \xi^2\frac{\sigma^2}{2} + \delta - \phi(\widehat{f}(\xi)-1), \tag{20}$$

$$\widehat{p}_\infty(\xi) := \frac{\delta\widehat{\psi}(\xi)}{\mu\xi - \frac{\sigma^2}{2}\xi^2 + \delta - \phi(\widehat{f}(\xi)-1)}. \tag{21}$$

The Laplace transform of the distribution of jumps f*,* $\widehat{f}(\xi) = \int_{-\infty}^{\infty} e^{-\xi g} f(g)dg$ *satisfies* $\widehat{f}(0) = 1$ *and (if* $\mathbb{E}[g] \geq 0$*)* $\widehat{f}(\xi) \geq 1$ *for all* $\xi < 0$*. Furthermore,* $\lambda(\xi)$ *is also the rate of convergence of the weighted* L^1*-norm* (17)*:* $-\lim_{t\to\infty}\frac{1}{t}\log||p(x,t)-p_\infty(x)||_\xi = \lambda(\xi)$*.*

Consider first the formula for the speed of convergence of the weighted distance measure without jumps $\phi = 0$. For the special case $\xi = 0$, we have $\lambda(\xi) = \delta$; when the weighted L^1-norm places no additional weight on the behavior of the distribution's tail, we recover our original result from Proposition 1, as expected. As we take ξ to be more and more negative, the weighted norm places more and more weight on the behavior of the distribution's upper tail, and the corresponding speed of convergence is given by $\lambda(\xi)$. Note that for $\mu > 0$, the speed of convergence $\lambda(\xi)$ is always lower the lower ξ is, for all $\xi \leq 0$. If $\mu < 0$, the same is true for all ξ less than some critical value. The formula for $\lambda(\xi)$ therefore indicates that convergence is slower the more weight we put on observations in the distribution's tail.

Next, consider the case with jumps $\phi > 0$. First note that the average speed of convergence as measured by the unweighted L^1-norm is entirely unaffected by the presence of jumps: explicitly spelling out the dependence of the speed of convergence $\lambda(\xi;\phi)$ on ϕ, we have $\lambda(0;\phi) = \delta$ for all ϕ. With $\xi < 0$, jumps make the speed of convergence *lower* than in the absence of jumps: $\lambda(\xi;\phi) \leq \lambda(\xi;0), \phi > 0$ (since $\widehat{f}(\xi) \geq 1$ for $\xi < 0$). Furthermore, for $\xi < 0$, $\lambda(\xi;\phi)$ is *decreasing* in ϕ, that is the higher is the jump intensity, the lower is the rate of convergence. Summarizing, if we confine attention to the average speed of convergence $||p(x,t) - p_\infty(x)||$ jumps have no effect whatsoever. If instead we put more weight on observations in the distribution's tail, $\xi < 0$, then the rate of convergence becomes worse, not better. 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.

Next consider (19) which provides a closed form solution for the evolution of the Laplace transform or equivalently for the evolution of all moments of the cross-sectional income distribution. These moments converge at the same rate $\lambda(\xi)$ as the weighted norm in (17). Hence, the closed form solution for the Laplace transform in (19) shows that high moments converge more slowly than low moments. We illustrate these results graphically below. Also note that all moments of the income distribution converge exponentially and hence monotonically.

Finally, note that one can identify the Pareto tail of the distribution p from knowledge of its Laplace transform only: the tail parameter is simply the critical value $\zeta > 0$ such that $\widehat{p}(\xi)$ ceases to exist for all $\xi \leq -\zeta$.[30] This strategy is useful because it also works in some cases in which the tail parameter cannot be computed using standard methods, e.g. with jumps $\phi > 0$.

4.2.2 Speed of Convergence in the Tail With Reflecting Barrier

Proposition 3 can also be extended to an income process with a reflecting barrier. The Kolmogorov Forward equation for the distribution can then no longer be solved by means of the Laplace transform. The proof therefore uses a different strategy, closely related to

[30]For any distribution p with a Pareto tail, that is $p(x) \sim ce^{-\zeta x}$ $x \to \infty$ for constants c and ζ, the Laplace transform (16) satisfies $\widehat{p}(\xi) \sim \frac{c}{\zeta+\xi}$ as $\xi \downarrow -\zeta$. Therefore $\zeta = -\inf\{\xi : \widehat{p}(\xi) < \infty\}$.

that in Proposition 1. While this strategy applies with a reflecting barrier, we can no longer handle jumps.

Proposition 4 *Consider the income process* (1) *without jumps $\phi = 0$ but with a reflecting barrier. Under Assumption 1, the rate of convergence $\lambda(\xi) := -\lim_{t\to\infty} \frac{1}{t} \log ||p(x,t) - p_\infty(x)||_\xi$ of the weighted L^1-norm* (17) *is*

$$\lambda(\xi) = \begin{cases} \frac{1}{2}\frac{\mu^2}{\sigma^2} + \delta, & \xi \geq \frac{\mu}{\sigma^2} \\ \mu\xi - \frac{\sigma^2}{2}\xi^2 + \delta, & \xi < \frac{\mu}{\sigma^2} \end{cases} \tag{22}$$

The speed of transition $\lambda(\xi)$ weakly decreases as the weight $-\xi$ on the right tail increases.

4.2.3 An Instructive Special Case: the Steindl Model

We briefly illustrate the result of Propositions 3 and 4 in an instructive special case originally due to Steindl (1965) with an analytic solution for the time path of the cross-sectional income distribution: $\sigma = 0, \mu, \delta > 0$ and ψ is the Dirac delta function at $x = 0$. In this model, the logarithm of income x_{it} grows at rate μ and gets reset to $x_{i0} = 0$ at rate δ. The Steindl model has recently also been examined by Jones (2015). The distribution $p(x,t)$ then satisfies the Kolmogorov Forward equation (5) with $\sigma = 0$ for $x > 0$. The corresponding stationary distribution is a Pareto distribution $p_\infty(x) = \zeta e^{-\zeta x}$ with $\zeta = \frac{\delta}{\mu}$. For concreteness, consider an economy starting in a steady state with some growth rate μ_0 (and death rate δ_0). At $t = 0$ the growth rate changes permanently to $\mu > \mu_0$ (and death rate δ). Then, the new steady state distribution is more fat-tailed, $\zeta < \zeta_0$. The following Lemma derives the path (it is valid for any ζ_0, not necessarily greater than ζ).[31]

Lemma 1 (Closed form solution for the transition in the Steindl model) *The time path of $p(x,t)$ is the solution to* (5) *with $\sigma = 0$ and initial condition $p_0(x) = \zeta_0 e^{-\zeta_0 x}, \zeta_0 = \delta_0/\mu_0$ and is given by*

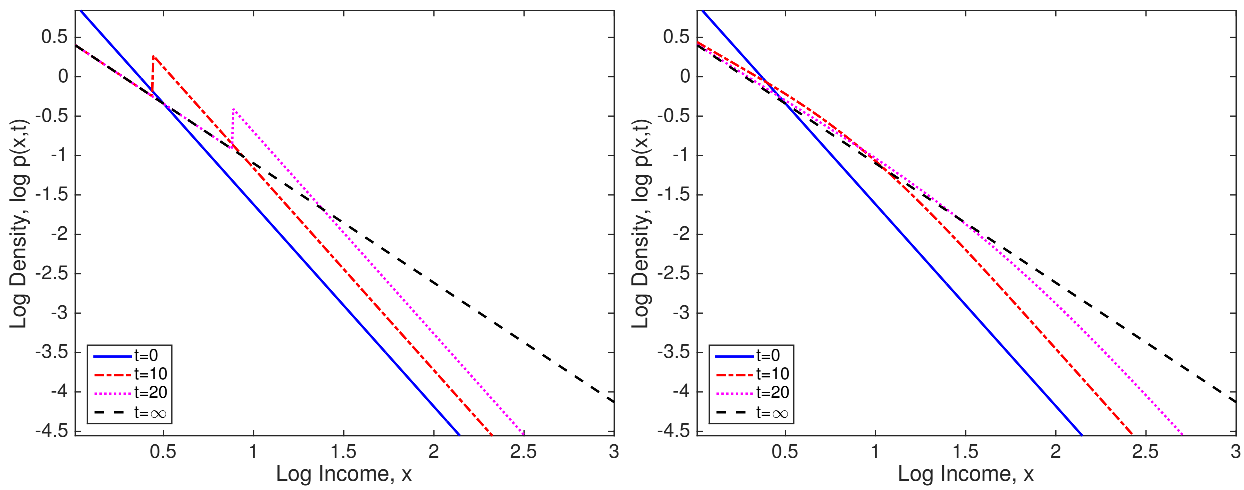
$$p(x,t) = \zeta e^{-\zeta x}\mathbf{1}_{\{x\leq\mu t\}} + \zeta_0 e^{-\zeta_0 x + (\zeta_0 - \zeta)\mu t}\mathbf{1}_{\{x>\mu t\}} \tag{23}$$

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

The solution is depicted in Figure 2 (a).[32] Consider, in particular, the local power law exponent $\zeta(x,t) = -\partial \log p(x,t)/\partial x$. Since the figure plots the log density, $\log p(x,t)$,

[31] Section D of the Online Appendix gives more closed forms, e.g. with $\sigma > 0$.

[32] The Steindl model is too stylized for a systematic calibration, an exercise we pursue in Section 4.3. Figure 2 uses comparable parameter values: we set $\delta = 1/30$, $\zeta_0 = 1/0.39, \zeta = 1/0.66$ and choose $\mu_0 = \delta/\zeta_0 = 0.013$ and $\mu = \delta/\zeta = 0.022$. In panel (b) we set $\sigma = 0.1$ and recalibrate μ_0 and μ to deliver the same ζ and ζ_0.



(a) Steindl Model, $\sigma = 0$

(b) More General Model, $\sigma > 0$

Figure 2: Transition of Cross-Sectional Income Distribution

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 exponent does not move at all when $t < \tau(x) = x/\mu$, then fully jumps to its steady state value at $t = \tau(x)$. In the Steindl model, the convergence of the distribution is slower the further out in the tail we look. In particular, note from the Figure that the *asymptotic* (for large x) power law exponent $\zeta(t) = -\lim_{x\to\infty} \partial \log p(x,t)/\partial x$ 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. Figure 2 (b) shows that also in the more general case with $\sigma > 0$ the power law exponent ζ (equivalently top inequality η) does not change at first and the distribution instead shifts out in parallel.[33]

Consider the behavior of top income shares in response to the permanent increase in μ considered above. Lemma 1 implies that the relative income of the 0.1% versus 1% income quantiles is *constant* for a while; it budges only when the "traveling discontinuity" hits the top 1% quantile. In contrast, the levels of the top 1% income quantile and the 0.1% income quantile increase quickly after the shock (to be more precise, after any time $t > 0$, they have

[33] This is more than a numerical result. Defining the local power law exponent $\zeta(x,t) := -P_x(x,t)/P(x,t)$ where P is the CDF corresponding to p, one can show using (5) that this local power law exponent does not move on impact following a shock, $\zeta_t(x,t)|_{t=0} = 0$ for all $x > 0$.

moved, in parallel). Hence, the ratio of the 0.1% to 1% share moves slowly (indeed, not at all for a while), though the top 1% share moves fairly fast.

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

We now revisit Figure 1 from Section 2 and ask: can standard random growth models generate the observed increase in income inequality? We find that they cannot. In particular, the transition dynamics generated by the model are too slow relative to the dynamics observed in the data. This operationalizes, by means of a simple calibration exercise using estimates from the micro data, the theoretical results in the preceding two sections.

More precisely, we ask whether an increase in the variance of the permanent component of wages σ^2 can explain the increase in income inequality observed in the data. That an increase in the variance of permanent earnings has contributed to the rise of inequality observed in the data has been argued by Moffitt and Gottschalk (1995), Haider (2001), Kopczuk, Saez, and Song (2010) and DeBacker, Heim, Panousi, Ramnath, and Vidangos (2013) (however, Guvenen, Ozkan, and Song (2014) examine administrative data and dispute that there has been such a trend – either way our argument is that an increase in σ cannot explain the rise in top inequality). The particular experiment we consider below is an increase in the variance of permanent earnings σ^2 from 0.01 in 1973 to 0.025 today. This implies that the standard deviation σ increases from 0.1 to 0.158, broadly consistent with evidence in Heathcote, Perri, and Violante (2010).

Before proceeding to the calibration exercise, we first use our theoretical results for some simple back-of-the-envelope calculations that illustrate our main point that transition dynamics of standard random growth models are extremely slow. We here focus on the case $\mu \geq 0$, i.e. that individuals' incomes grow at least as fast on average as the aggregate economy.[34] Proposition 1 then implies that the average speed of convergence is simply $\lambda = \delta$ and the corresponding half-life is $t_{1/2} = \log(2)/\delta$.[35] As shown in Propositions 3 and 4, the speed of convergence in the tail can be much slower. In particular, consider the formula (20) for

[34] Our model is stationary whereas the U.S. economy features long-run growth. The parameter μ should therefore be interpreted as the growth rate of individual incomes over the lifecycle *relative* to the growth rate of the aggregate economy (this can also be seen from the fact that in the model the distribution of starting wages $\psi(x)$ is stationary). Parameterizations with $\mu < 0$ are therefore relatively natural as well. In that case and with a reflecting barrier, convergence may be marginally faster – see (12).

[35] Because we are dealing with exponential decay in multiple (in fact, infinite) dimensions, $t_{1/2}$ only equals half the time it takes for $||p - p_\infty||$ to converge for the particular initial conditions p_0 for which $||p - p_\infty|| = ||p_0 - p_\infty||e^{-\lambda t}$ (so that $||p_0 - p_\infty||e^{-\lambda t_{1/2}} = \frac{1}{2}||p_0 - p_\infty||$ implies $t_{1/2} = \log 2/\lambda$). For other initial conditions this equation only holds asymptotically – see (10). It is nevertheless standard to refer to $t_{1/2}$ as "half-life."

the speed of convergence without jumps $\phi = 0$

$$\lambda(\xi) = \xi\mu - \xi^2 \frac{\sigma^2}{2} + \delta. \tag{24}$$

Here the reader should recall that by varying ξ, we can trace out the speed of convergence of all moments of the distribution and $\lambda(\xi)$ is the speed of convergence of the -ξth moment. Equivalently, $-\xi$ is the weight on the tail in the weighted L^1-norm (17). For our calculations, it is convenient to express (24) in terms of tail inequality $\eta = 1/\zeta$ which is directly measurable from cross-sectional data. From (2) we have $\mu = \delta\eta - \sigma^2/(2\eta)$ and therefore

$$\lambda(\xi) = \xi\left(\delta\eta - \frac{\sigma^2}{2\eta}\right) - \xi^2\frac{\sigma^2}{2} + \delta = \left(\delta\eta - \frac{\sigma^2}{2}\xi\right)\left(\frac{1}{\eta} + \xi\right). \tag{25}$$

In the relevant range $-1/\eta < \xi < 0$, the speed of convergence is strictly decreasing in tail inequality η, i.e. higher inequality goes hand in hand with a slower transition. It is also strictly increasing in the innovation variance σ^2.

Using this formula, we can now examine how the parameters η, δ and σ^2 affect the speed of convergence. To get a "quantitative feel" for (25), consider first the "Steindl" case $\sigma^2 = 0$ so that $\lambda(\xi) = \delta(1 + \eta\xi)$. While unrealistic, this simple case has the advantages that computations are particularly easy and only require estimates for two parameters, η and δ (the implied speed also turns out to be similar for the more realistic case where $\sigma^2 > 0$). We use $\delta = 1/30$ corresponding to an expected work life of thirty years. A slight difficulty arises because η in (25) is tail inequality in the new stationary equilibrium. We use observed tail inequality in 2012 which equals $\eta_{2012} = 0.66$, a conservative estimate because $\lambda(\xi)$ is decreasing in η (and η is increasing in the data).[36] The resulting half-life of the $-\xi$th moment is given by $t_{1/2}(\xi) = \log 2/\lambda(\xi) = 0.69 \times 30 \times \frac{1}{1+0.66\xi}$. For example, the half-life of convergence of the first moment ($\xi = -1$) is around 60 years. Note that this calibration is conservative. In particular, a longer expected work life or higher estimate of tail inequality would result in even slower transitions.

We use (25) to perform similar calculations for the more general case where $\sigma^2 > 0$. Figure 3 plots the corresponding half-life $t_{1/2}(\xi) = \log(2)/\lambda(\xi)$ for the parameter values used in our experiment as a function of the moment under consideration $-\xi$. Consider first the solid line which plots the half life $t_{1/2}(\xi)$ for $\sigma^2 = 0.025$, the variance of the permanent component of wages used in our experiment. There are two main takeaways from the figure.

[36] We compute η from the relative income shares in panel (b) of Figure 1. If the distribution is Pareto, relative income shares satisfy $\frac{S(p/10)}{S(p)} = 10^{\eta-1}$ and we therefore compute $\eta(p) = 1 + \log_{10} S(p/10)/S(p)$. We here use $\eta(1) = 1 + \log_{10} S(0.1)/S(1)$.

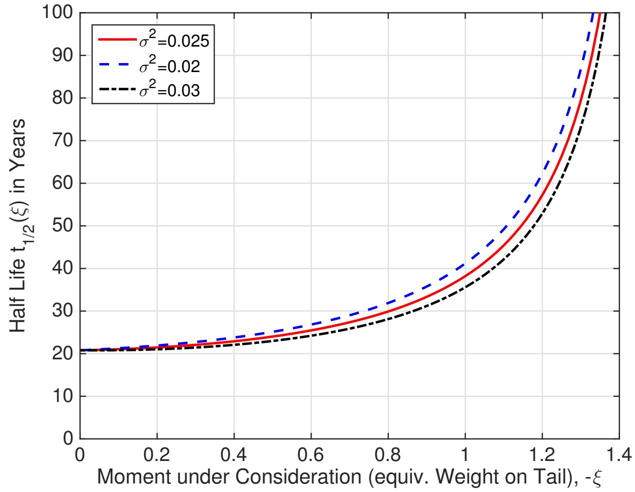


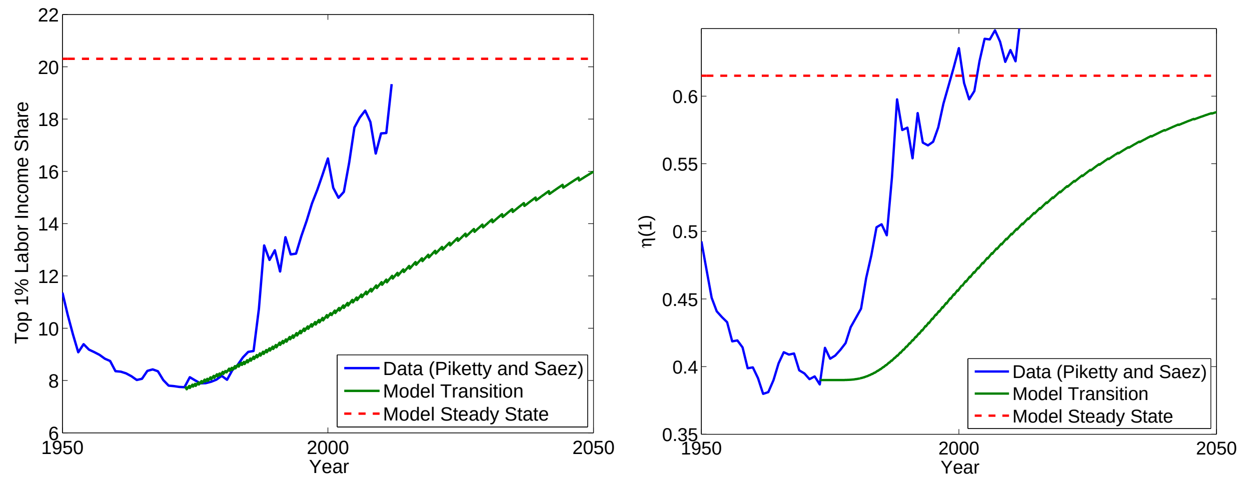
Figure 3: Theoretical Speed of Convergence of Different Moments of Income Distribution

First, even for relatively low moments the speed of convergence is considerably lower. For example, the half-life of convergence of the first moment ($\xi = -1$) is around 40 years, i.e. twice as much as the average speed of roughly 20 years. Second, the speed of convergence becomes slower and slower the higher the moment under consideration, with half lives of 100 years close to the highest admissible moment $1/\eta = 1.52$. The figure also shows that the speed of convergence is not particularly sensitive to the value of the variance σ^2.

We next consider the effects of an increase in σ^2 from 0.01 in 1973 to 0.025 today in the baseline random growth model and how they compare to the evolution of inequality in the data. We set $\delta = 1/30$ as above and set μ to match the observed tail inequality in 1973, $\eta_{1973} = 0.39$ which yields $\mu = \delta\eta - \sigma^2/(2\eta) = 0.002$, i.e. individual income growth 0.2% above the economy's long-run growth rate. Figure 4 plots the time paths for the top 1% income share (panel (a)) and the empirical power law exponent (panel (b)) following the increase in σ^2 in the baseline random growth model and compares them to the data.[37] Not surprisingly given our analytical results, the model fails spectacularly.[38] An increase in the variance in the permanent component of income σ^2 in the standard random growth model is therefore not a promising candidate for explaining the observed increase in top income inequality. It is also worth emphasizing again that allowing for jumps ($\phi > 0$) in the income process would only slow down the speed of convergence even more (Proposition 3).

[37]We solve the Kolmogorov Forward equation (5) numerically using a finite difference method.

[38]Note that the power law exponent in panel (b) is completely flat on impact, consistent with Figure 2 and footnote 33.



(a) Top 1% Labor Income Share (b) Empirical Inverse Power Law Exponent

Figure 4: Dynamics of Income Inequality in the Baseline Model

5 Models that Generate Fast Transitions

Given the negative results of the preceding section, it is natural to ask: what then explains the observed fast rise in top income inequality? We argue that fast transitions require very specific departures from the standard random growth model. We extend the model along two dimensions, both of which constitute deviations from Gibrat's law. First, we allow for *type dependence* in the growth rate distribution. Second, we consider *scale dependence*. We discuss the role of these two additions in turn in Sections 5.2 and 5.3. In Section 5.4, we then revisit the rise in income inequality and argue that our augmented random growth model can generate transitions that are as fast as those observed in the data.

5.1 The Augmented Random Growth Model

In its most general form, we consider a random growth model with *type dependence* in the form of distinct "growth types" indexed by $j = 1, ..., J$, and *scale dependence* captured by a process χ_t. The dynamics of log income x_{it} of individual i of type j are given by

$$\begin{aligned} x_{it} &= \chi_t^{b_j} y_{it}, \\ dy_{it} &= \mu_j dt + \sigma_j dZ_{it} + g_{jit} dN_{jit} + \text{Injection} - \text{Death}, \end{aligned} \tag{26}$$

where dN_{jit} is a Poisson process with intensity ϕ_j and g_{jit} is a random variable with distribution f_j. The latent variable y_{it} can be interpreted as a worker's skill. As before, workers retire at rate δ and get replaced by labor entrants with income drawn from a distribution ψ.

A fraction θ_j of labor force entrants are born as type j and workers switch from being type j to type k at rate $\alpha_{j,k}$. Our baseline model is the special case with $J = 1$ and $\chi_t = 1$.

The process features type dependence in that μ_j, σ_j, ϕ_j and f_j differ across types. Guvenen (2007) has argued that an income process with heterogeneous income profiles provides a better fit to the micro data than a model in which all individuals face the same income profile, 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 σ_j, similar to the mixture specification advocated by Guvenen, Karahan, Ozkan, and Song (2015). We also build on Luttmer (2011), who studies a related framework applied to firm dynamics and argues that persistent heterogeneity in mean firm growth rates is needed to account for the relatively young age of very large firms at a given point in time (a statement about the stationary distribution rather than transition dynamics as in our paper). Aoki and Nirei (2015) present a related and more complex economic model with entrepreneurs and workers that are subject to different income growth rates, and Jones and Kim (2014) examine a model with different types of entrepreneurs.

Scale dependence is captured by χ_t, an arbitrary stochastic process satisfying $\chi_t > 0$ and $\lim_{t\to\infty} \mathbb{E}[\log \chi_t] < \infty$. Basically, with $b_j > 0$, an increase in χ_t means that the income growth rate is higher for higher incomes: hence it violates Gibrat's law. To see this, write (26) as

$$dx_{it} = \tilde{\mu}_{jt} dt + \tilde{\sigma}_{jt} dZ_{it} + b_j x_{it} d \log \chi_t + g_{jit} d\tilde{N}_{jit} + \text{Injection} - \text{Death} \tag{27}$$

where $\tilde{\mu}_{jt} = \mu_j \chi_t^{b_j}, \tilde{\sigma}_{jt} = \sigma_j \chi_t^{b_j}$ and $d\tilde{N}_{jit} = dN_{jit}\chi_t^{b_j}$. If $b_j d \log \chi_t > 0$, the growth rate of income x_{it} is increasing in income, i.e. a deviation from Gibrat's law.[39] Appendix F.1 provides conditions under which the process (26) features a unique stationary distribution with a Pareto tail, and we assume that these conditions hold throughout this section.[40]

5.2 The Role of Type Dependence

First, consider the special case of (26) with type dependence but without scale dependence $\chi_t = 1$ (or jumps $\phi = 0$). Here we focus on a simple case with two types, a high-growth type and a low-growth type, but our results can be extended to more types (Appendix F.2).

[39] Also note that Z_{it} is an idiosyncratic stochastic process whereas S_t is an aggregate or common shock that hits all individuals simultaneously.

[40] More precisely, we provide conditions under which the process for skill y_{it} has a fat-tailed stationary distribution. A necessary and sufficient condition for income x_{it} to also have a fat-tailed stationary distribution is that χ_t is constant. More generally though, we want to allow for time-variation in χ_t, thereby capturing secular changes in skill prices or shocks disproportionately affecting high incomes at business-cycle frequencies.

Denote the density of individuals who are currently in the high and low growth states by $p^H(x,t)$ and $p^L(x,t)$ and the cross-sectional wage distribution by $p(x,t) = p^H(x,t)+p^L(x,t)$. We assume that a fraction θ of individuals start their career as high-growth types and the remainder as low-growth types, and that individuals switch from high to low growth with intensity α. Low growth is an absorbing state that is only left upon retirement. Then, the densities satisfy the following system of Kolmogorov Forward equations

$$\begin{aligned} p_t^H &= -\mu_H p_x^H + \frac{\sigma_H^2}{2} p_{xx}^H - \alpha p^H - \delta p^H + \beta_H \delta_0, \\ p_t^L &= -\mu_L p_x^L + \frac{\sigma_L^2}{2} p_{xx}^L + \alpha p^H - \delta p^L + \beta_L \delta_0, \end{aligned} \tag{28}$$

with initial conditions $p^H(x,0) = p_0^H(x), p^L(x,0) = p_0^L(x)$ and where $\beta_H = \theta\delta$ and $\beta_L = (1-\theta)\delta$ are the birth rates of the two types.[41]

While we are not aware of an analytic solution method for the system of partial differential equations (28), this system can be conveniently analyzed by means of Laplace transforms as in Section 4.2. In particular, $\widehat{p}^H(\xi,t)$ and $\widehat{p}^L(\xi,t)$ satisfy

$$\widehat{p}_t^H(\xi,t) = -\lambda_H(\xi)\widehat{p}^H(\xi,t) + \beta_H, \qquad \lambda_H(\xi) := \xi\mu_H - \xi^2\frac{\sigma_H^2}{2} + \alpha + \delta, \tag{29}$$

$$\widehat{p}_t^L(\xi,t) = -\lambda_L(\xi)\widehat{p}^L(\xi,t) + \alpha\widehat{p}^H(\xi,t) + \beta_L, \qquad \lambda_L(\xi) := \xi\mu_L - \xi^2\frac{\sigma_L^2}{2} + \delta, \tag{30}$$

with initial conditions $\widehat{p}^H(\xi,0) = \widehat{p}_0^H(\xi), \widehat{p}^L(\xi,0) = \widehat{p}_0^L(\xi)$. Importantly, for fixed ξ, this is again simply a system of ordinary (rather than partial) differential equations which can be solved analytically. Note that the system is triangular so that one can first solve the equation for $\widehat{p}^H(\xi,t)$ and then the one for $\widehat{p}^L(\xi,t)$.[42]

Proposition 5 (Speed of convergence with type dependence) *Consider the cross-sectional distribution $p(x,t) := p^H(x,t)+p^L(x,t)$. The stationary distribution $p_\infty(x) = p_\infty^H(x)+p_\infty^L(x)$ has a Pareto tail with tail exponent $\zeta = \min\{\zeta_L,\zeta_H\}$ where ζ_H is the positive root of $0 = \zeta^2\frac{\sigma_H^2}{2} + \zeta\mu_H - \alpha - \delta$ and ζ_L is the positive root of $0 = \zeta^2\frac{\sigma_L^2}{2} + \zeta\mu_L - \delta$. The time paths of the*

[41] Assuming that the fraction of high types is stationary, it equals $\theta\delta/(\alpha+\delta)$.

[42] Proposition 5 can easily be extended to a non-triangular system, i.e. if the low state is not an absorbing state and low types can switch to being high types. See Appendix F.2. This is achieved by writing the analogue of (29) and (30) in matrix form. The speed of convergence is then governed by the eigenvalues of that matrix. In the triangular case, these eigenvalues are simply $-\lambda_L(\xi)$ and $-\lambda_H(\xi)$. Therefore, while triangularity yields simple formulae, all results can be extended to the more general case.

Laplace transforms of $p^H(x,t)$ *and* $p(x,t)$ *are*

$$\widehat{p}^H(\xi,t) - \widehat{p}^H_\infty(\xi) = e^{-\lambda_H(\xi)t}(\widehat{p}^H_0(\xi) - \widehat{p}^H_\infty(\xi)), \tag{31}$$

$$\widehat{p}(\xi,t) - \widehat{p}_\infty(\xi) = c_H(\xi)e^{-\lambda_H(\xi)t} + c_L(\xi)e^{-\lambda_L(\xi)t}, \tag{32}$$

where $\lambda_H(\xi)$ *and* $\lambda_L(\xi)$ *are defined in* (29) *and* (30)*,* $\widehat{p}^H_\infty(\xi)$ *and* $\widehat{p}_\infty(\xi)$ *are the Laplace transforms of the stationary distributions and* $c_H(\xi)$ *and* $c_L(\xi)$ *are constants of integration. Finally, the weighted* L^1*-norm of the distribution of high types converges at rate* $-\lim_{t\to\infty}\frac{1}{t}\log||p^H(x,t) - p^H_\infty(x)||_\xi = \lambda_H(\xi)$*.*

The transition dynamics of the income distribution therefore take place on two different time scales: part of the transition happens at rate $\lambda_H(\xi)$ and another part at rate $\lambda_L(\xi)$.[43] The model then has the theoretical potential to explain fast short-run dynamics and, as we argue in Section 5.4, the observed rise in income inequality.

5.3 The Role of Scale Dependence

Next consider the special case of (26) with scale dependence $d\log\chi_t \neq 0$ but without type dependence $J = 1$ (only one growth type). The logarithm of income then satisfies $x_{it} = \chi_t y_{it}$ and the level of income is $w_{it} = (e^{y_{it}})^{\chi_t}$, with χ_t disciplining the convexity of income as a function of skill $e^{y_{it}}$. We therefore interpret changes in χ_t as changes in skill prices. Appendix F.3 shows that such changing skill prices naturally arise in assignment models with "superstars" effects.[44] Note that the distribution of skills (and hence income) can have a Pareto tail even if the support of the distribution of workers' true underlying "talent" is bounded above (Gabaix and Landier, 2008). Since income is $w_{it} = (e^{y_{it}})^{\chi_t}$, it is easy to see that an increase in χ_t (which generates scale dependence) leads to an instantaneous fattening of the tail of the income distribution.

Proposition 6 (Infinitely fast adjustment in models with scale dependence) *Consider the special case of* (26) $x_{it} = \chi_t y_{it}$ *where the distribution of* y_{it} *is stationary and where* χ_t *is an aggregate shock. This process has an infinitely fast speed of adjustment:* $\lambda = \infty$*. Denoting by* ζ^x_t *and* ζ^y *the power law exponents of log income and skill* x_{it} *and* y_{it}*, we have* $\zeta^x_t = \zeta^y/\chi_t$*.*

[43] A natural assumption is that the switching rate α is large enough to swamp any differences between the μ's and σ's in the two states and so $\lambda_H(\xi) > \lambda_L(\xi)$ in (29) and (30). In contrast to the baseline random growth model of section 4, transition dynamics following a parameter change 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.

[44] As in Rosen (1981), Garicano and Rossi-Hansberg (2006), Gabaix and Landier (2008), Tervio (2008) and Geerolf (2014).

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

$$\mathbb{P}(x_{it} > x) = \mathbb{P}(\chi_t y_{it} > x) = \mathbb{P}(y_{it} > x/\chi_t) = ce^{-\zeta^y x/\chi_t} \quad \Rightarrow \quad \zeta_t^x = \zeta^y/\chi_t. \ \square$$

Hence, the process is extremely fast – it features instantaneous transitions in the power law exponent. Therefore, if χ_t has a secular trend, the power law exponent inherits this trend. Fast transitions are therefore consistent with theories in which the increase in top income inequality is driven by changing skill prices, e.g. due to the rise of "superstars."

Parker and Vissing-Jorgensen (2010) provide supportive evidence for scale dependence at high frequencies. They find that in good (respectively bad) times, the incomes of top earners increase (respectively decrease), in a manner consistent with (27): the sensitivity to the shock at time t is proportional to x_{it}, as in

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

with $S_t := d \log \chi_t$. Note that the shock $x_{it} dS_t$ to log income is multiplicative in log income, as opposed to additive as in the traditional random growth model. This finding is broadly confirmed by Guvenen (2015, p.40). We conclude that scale dependence is an empirically grounded source of fast transitions.

5.4 Fast Transitions in the Augmented Model

We now use the framework of this section to revisit the rise in income inequality in the United States. We argue that, in contrast to the spectacular failure of the standard random growth model, the model with type dependence 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 high types μ_H (while μ_L is unchanged). This is motivated in part by casual evidence of very rapid income growth rates since the 1980s, for instance for Bill Gates, Mark Zuckerberg, hedge fund managers and the like – their growth is very high for a while, then tails off. This impression is confirmed by Jones and Kim (2014), who find that there has been a substantial increase in the average growth rate in the upper tail of the growth rate distribution since the late 1970s.[45] We

[45] Jones and Kim (2014) proxy μ_H with the median of the upper decile, i.e. the 95th percentile, of the distribution of income growth rates. Combining evidence from the IRS public use panel of tax returns and from Guvenen, Ozkan, and Song (2014), they show that this measure of μ_H has increased substantially from 1979-81 to 1988-90 to 1995-96. Jones and Kim note that this evidence should be viewed as suggestive due to limited sample sizes in the IRS data and comparability of the IRS and the Social Security Administration

follow a similar calibration strategy as in Section 4.3. First, note from Proposition 3 that, if μ_H is sufficiently bigger than μ_L, the Pareto tail of the stationary income distribution is determined only by the dynamics of high-growth types and given by

$$\zeta = \min\{\zeta_L, \zeta_H\} = \frac{-\mu_H + \sqrt{\mu_H^2 + 2\sigma_H^2(\delta + \alpha)}}{\sigma_H^2}, \tag{33}$$

and the parameters σ_L and μ_L do not affect top inequality. As before, we set $\delta = 1/30$ and impose that the economy is initially in a Pareto steady state with $\eta_{1973} = 0.39$. We set $\sigma_H = 0.15$, which is a conservative estimate.[46] We do not have precise estimates for α, the rate of switching from high to low growth. For our baseline results, we set $\alpha = 1/6$, corresponding to an expected duration of being a high-growth type of 6 years, and we report results under alternative parameter values. Given values for σ_H, δ and α, we calibrate the initial μ_H so that (33) yields $\eta_{1973} = 0.39$. In the initial steady state, the difference in mean growth rates between high- and low-growth types is $\mu_H - \mu_L = 0.06$.

Our baseline exercise considers a once-and-for-all increase in μ_H by 8 percentage points. The resulting gap of $\mu_H - \mu_L = 0.14$ is broadly consistent with empirical evidence in Guvenen, Kaplan, and Song (2014).[47] Figure 5 plots the corresponding results. The difference to the earlier experiment in Figure 4 is striking. The model with type dependence can replicate the rapid rise in income inequality observed in the United States.

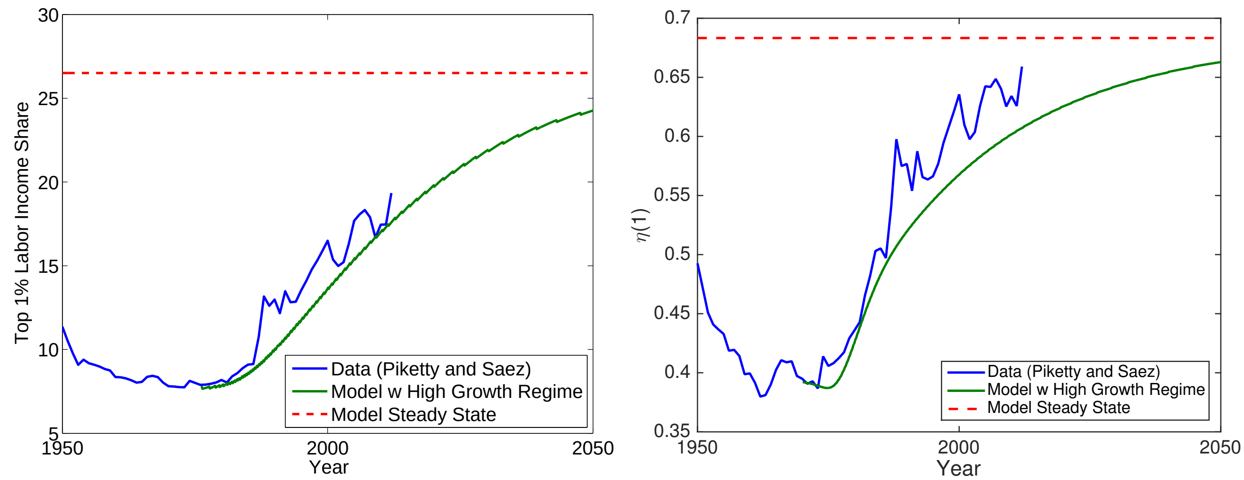
The key parameters that govern the speed of transition are μ_H and α, the growth rate of high types and the probability of leaving it. In the Online Appendix, we report results from alternative parameterizations and experiments. As expected given our theoretical results, transitions are fastest when α and μ_H are high, i.e. when individuals can experience very short-lived, very high-growth spurts, what one may call "live-fast-die-young dynamics".[48] In summary, the model with type dependence is capable of generating fast transition dynamics of top inequality for a number of alternative parameterizations that are broadly consistent with the micro data. The common feature of these parameterizations is a combination of

data used by Guvenen, Ozkan, and Song (2014). Below we discuss ongoing work and directions for future work that could improve on these estimates. In the meantime, Jones and Kim provide the best available evidence documenting potential drivers of the increase in top income inequality.

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

[47] Guvenen, Kaplan, and Song (2014) document differences in average growth rates of different population groups as large as 0.23 log points per year. See in particular their Figure 7. Reader may also wonder how the model with type dependence compares to the baseline model when subjected to the same shock. Appendix F.4 reports results from such an experiment. As expected, transitions are faster.

[48] In their ongoing work using a very similar model, Jones and Kim (2014) propose such a "live-fast-die-young" calibration with very high α and μ_H.



(a) Top 1% Labor Income Share

(b) Empirical Inverse Power Law Exponent

Figure 5: Transition Dynamics in Model with Type Dependence

relatively high growth rates for part of the population (high enough μ_H) over relatively short time horizons (high enough α). The absence of better micro estimates for these critical parameters and the stylized nature of our model mean that the quantitative explorations in this section should be viewed as suggestive. Future research should explore these mechanisms in richer, more fully-fledged quantitative models. Similarly, better empirical evidence is a clear priority.

6 Conclusion

This paper makes two contributions. First, it finds that standard random growth models cannot explain rapid changes in tail inequality, for robust analytical reasons. This required developing new tools to analyze transition dynamics, as most previous literature could analyze only separate steady states, without being able to assess analytically the speed of transition between them and without identifying the above-mentioned important defect of the standard model. Second, it suggests two parsimonious deviations from the basic model that can explain such fast changes: (i) type dependence and (ii) scale dependence. We view them as promising, because they have some support in the data (as we argued above, see especially Jones and Kim (2014), Parker and Vissing-Jorgensen (2010) and Guvenen (2015)). We hope that future research explores their importance in more detail. A clear priority for future research is empirical evidence, in combination with quantitative theory, that allows for an assessment of various concrete economic mechanisms put forth in the public debate ("Is the rise in top inequality due to: technical change, superstars, rent-seeking, globalization, and so on?") The forces we have analyzed in this paper may serve to guide future empirical and theoretical work on the determinants of fast changes in inequality.

Appendix

A Proof of Proposition 1

Proposition 1 is concerned with two different cases. The first case involves models with death and reinjection where the dynamics without those terms are not ergodic. The second one concerns ergodic models (with or without death and reinjection). The strategy of the proof in both cases is different and we therefore present the two cases separately. In the first case ("non-ergodic case"), the rate of convergence is obtained by directly analyzing the dynamics of the L^1 norm (4) of the cross-sectional distribution. In the second case ("ergodic case"), the convergence is to a real invariant measure and the rate of convergence is obtained by a spectral analysis (in particular it is given by the "spectral gap").[49]

A.1 Proof of Proposition 1: Non-ergodic Case

We here study the non-ergodic case, starting with a generally useful lemma.

Lemma 2 *Suppose that a function $q(x,t)$ solves $q_t = \mathcal{A}q$ with $\mathcal{A}q = a(x,t)\,q + b(x,t)q_x + c(x,t)q_{xx}$ with $c(x,t) \geq 0$ for all x. Then $|q(x,t)|$ is a "subsolution" of the same equation, that is*

$$|q|_t \leq \mathcal{A}\,|q|\,. \tag{34}$$

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

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

[49]More precisely, in the "ergodic case" there is a reflecting barrier and $\delta \geq 0, \mu < 0, \sigma^2 > 0$. In the "non-ergodic case," there is no reflecting barrier and no restriction on μ but $\delta > 0, \sigma^2 \geq 0$. In the Proposition we distinguish between the case "with a reflecting barrier" and the one "without a reflecting barrier." Note that this distinction is related to but somewhat different from "ergodic" and "non-ergodic."

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

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

Lemma 3 *The decay rate of the L^1 norm $d(t) := \|q(\cdot,t)\|$ is at least δ: $\lambda \geq \delta$.*

Proof of Lemma 3: We have $d(t) := \|q(\cdot,t)\| = \int |q(x,t)|\, dx$ and hence

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

where the inequality follows from Lemma 2 and the last equality from the boundary conditions corresponding to p. Hence $d'(t) \leq -\delta \int |q|\, dx = -\delta d(t)$ and therefore $d(t) \leq e^{-\delta t}d(0)$ by Grönwall's Lemma. □

We next prove the opposite inequality (the overly technical proof is in Appendix E.2.2):

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

Gathering the arguments and putting together Lemmas 3 and 4, we obtain that $\lambda = \delta$.

A.2 Proof of Proposition 1: Ergodic Case

We next study the "ergodic case": there is a reflecting barrier on income and additionally $\mu < 0$. Then the process (1) is ergodic even with $\delta = 0$. In this case, the cross-sectional distribution satisfies (9) with boundary condition (6). The key insight is that the speed of convergence of p is governed by the second eigenvalue of the operator $\mathcal{A}^*$ ($\mathcal{A}^*p := -\mu p_x + \frac{\sigma^2}{2}p_{xx} - \delta p$), and the key step is to obtain an analytic formula for this second eigenvalue given by $|\lambda_2| = \frac{1}{2}\frac{\mu^2}{\sigma^2} + \delta$.

A.2.1 Preparation: Boundary Conditions

We first review some mathematical concepts that will be useful.[50] First, the *inner product* of two continuous functions u and v is $\langle u, v\rangle = \int_{-\infty}^{\infty} u(x)v(x)dx$. Second, for an operator $\mathcal{A}$, the (formal) *adjoint* of $\mathcal{A}$ is the operator $\mathcal{A}^*$ satisfying $\langle \mathcal{A}u, p\rangle = \langle u, \mathcal{A}^*p\rangle$. Third, an operator

[50] A more systematic treatment can be found in many textbooks on functional analysis or partial differential equations, particularly applications to physics. See e.g. Weidmann (1980) and the more accessible Hunter and Nachtergaele (2001) and Stone and Goldbart (2009, Ch.4).

$\mathcal{B}$ is *self-adjoint* if $\mathcal{B}^* = \mathcal{B}$.[51] It is well-known that eigenvalues of a self-adjoint operator are real. Fourth, the *infinitesimal generator* of a Brownian motion with death at Poisson rate δ is the operator $\mathcal{A}$ defined by

$$\mathcal{A}u = \mu u_x + \frac{\sigma^2}{2} u_{xx} - \delta u \tag{35}$$

Some care is needed with the boundary condition. As we shall see, the boundary condition is:

$$u_x(0) = 0. \tag{36}$$

The domain of $\mathcal{A}$ here is the set of functions u in L^2 (i.e. square-integrable functions) such that $\mathcal{A}u$ is also in L^2, i.e. the u's such that u, u_x, u_{xx} are in L^2.

We next state a lemma. Its proof is instructive, because it shows where the boundary condition (36) comes from.

Lemma 5 *The Kolmogorov Forward operator* $\mathcal{A}^*$ *in* (9) *with boundary condition* (6) *in the reflecting case is the adjoint of the infinitesimal generator* $\mathcal{A}$ *in* (35) *with boundary condition* (36)*.*

Proof of Lemma 5: The boundary for $p(x)$ is: $\mu p(0) - \frac{\sigma^2}{2} p_x(0) = 0$ (this comes from integrating the Forward Kolmogorov equation from $x = 0$ to ∞). We have

$$\begin{aligned}
\langle u, \mathcal{A}^* p \rangle &= \int_0^\infty u \left(-\mu p_x + \frac{\sigma^2}{2} p_{xx} - \delta p \right) dx \\
&= \left[-u\mu p + \frac{\sigma^2}{2} u p_x \right]_0^\infty - \int_0^\infty \left(-\mu u_x p + \frac{\sigma^2}{2} u_x p_x \right) dx - \int_0^\infty \delta u p dx \\
&= \left[-u\mu p + \frac{\sigma^2}{2} u p_x + \frac{\sigma^2}{2} u_x p \right]_0^\infty + \int_0^\infty \left(\mu u_x p + \frac{\sigma^2}{2} u_{xx} p - \delta u p \right) dx \\
&= u(0) \left(\mu p(0) - \frac{\sigma^2}{2} p_x(0) \right) - \frac{\sigma^2}{2} u_x(0) p(0) + \langle \mathcal{A}u, p \rangle \\
&= -\frac{\sigma^2}{2} u_x(0) p(0) + \langle \mathcal{A}u, p \rangle \qquad \text{from (6)} \\
&= \langle \mathcal{A}u, p \rangle .
\end{aligned}$$

For the last equality we need $\frac{\sigma^2}{2} u_x(0) p(0) = 0$, which leads to the boundary condition (36).

A.2.2 Main Proof

With these preliminaries in hand, we proceed with the proof of the proposition.

[51] Note that the adjoint is the infinite-dimensional analogue of a matrix transpose.

We first show how the case $\delta \geq 0$ can be derived from the case $\delta = 0$. Suppose an initial condition $p_0(x)$. Given that $p_t = \mathcal{A}^* p + \delta\psi$ (equation (9)), we have $0 = \mathcal{A}^* p_\infty + \delta\psi$, and by subtraction $\tilde{q} := p - p_\infty$ satisfies $\tilde{q}_t = \mathcal{A}^* \tilde{q}$. Next, define $q(x,t) := e^{\delta t}\tilde{q}(x,t) = e^{\delta t}(p(x,t) - p_\infty(x))$. Then, a simple calculation gives: $q_t = \mathcal{C}^* q := -\mu q_x + \frac{\sigma^2}{2} q_{xx}$. Operator $\mathcal{C}^*$ has no "death", so that the case $\delta = 0$ applies to q. If we have shown (as we will shortly) that $\|q(x,t)\|$ decays in $e^{-\lambda t}$ (more precisely, that $\lambda = -\lim_{t\to\infty} \frac{1}{t} \log ||q(x,t)||$), that will show that $\|p(x,t) - p_\infty(x)\| = e^{-\delta t}\|q(x,t)\|$ decays in $e^{-\delta t - \lambda t}$ (more precisely, that $\delta + \lambda = -\lim_{t\to\infty} \frac{1}{t} \log \|p(x,t) - p_\infty(x)\|$). Hence, the case $\delta > 0$ follows easily from the case $\delta = 0$.

We next proceed to the case $\delta = 0$. The goal is to analyze the eigenvalues of the infinitesimal generator $\mathcal{A}$ or equivalently its adjoint $\mathcal{A}^*$. The difficulty is that $\mathcal{A}$ is not self-adjoint, $\mathcal{A}^* \neq \mathcal{A}$, and therefore its eigenvalues could, in principle, be anywhere in the complex plane. We therefore construct a self-adjoint transformation $\mathcal{B}$ of $\mathcal{A}$ as follows.

Lemma 6 *Consider u satisfying $u_t = \mathcal{A}u$ with $\delta = 0$ and boundary condition* (36) *and the corresponding stationary distribution $\bar{p}_\infty(x) = \frac{2\mu}{\sigma^2} e^{(2\mu/\sigma^2)x}$. Then $v := u\bar{p}_\infty^{-1/2} = \frac{2\mu}{\sigma^2} u e^{(\mu/\sigma^2)x}$ satisfies*

$$v_t = \mathcal{B}v := \frac{\sigma^2}{2} v_{xx} - \frac{1}{2}\frac{\mu^2}{\sigma^2} v \tag{37}$$

with boundary condition $v_x(0) = \frac{\mu}{\sigma^2} v(0)$ and where the domain of $\mathcal{B}$ is the set of functions v in L^2 such that $\mathcal{B}v$ is also in L^2. Furthermore, $\mathcal{B}$ is self-adjoint.

Proof: (37) follows from differentiating $v = \frac{2\mu}{\sigma^2} u e^{(\mu/\sigma^2)x}$. To see that $\mathcal{B}$ is self-adjoint, we integrate by parts as in Lemma 5 to conclude that for any v, q in the domain of $\mathcal{B}$, $\langle \mathcal{B}v, q\rangle = \langle v, \mathcal{B}q\rangle$. $\square$

Lemma 7 *The spectrum of $\mathcal{B}$ consists of an isolated first eigenvalue $\Lambda_1 = 0$, $\Lambda_2 = -\frac{1}{2}\frac{\mu^2}{\sigma^2}$, and all other points in the spectrum satisfy $|\Lambda| > |\Lambda_2|$. Hence the spectral gap of $\mathcal{B}$ equals $\lambda := |\Lambda_2| = \frac{1}{2}\frac{\mu^2}{\sigma^2}$.*

Proof of Lemma 7: Consider the eigenvalue problem $\mathcal{B}\varphi = \Lambda\varphi$ or equivalently

$$\frac{\sigma^2}{2}\varphi''(x) - \frac{1}{2}\frac{\mu^2}{\sigma^2}\varphi(x) = \Lambda\varphi(x), \tag{38}$$

with boundary condition

$$\varphi'(0) = \frac{\mu}{\sigma^2}\varphi(0). \tag{39}$$

The question is: for what values of $\Lambda \leq 0$ does (38) have a solution $\varphi(x)$ that satisfies the boundary condition (39) and is either in the domain of $\mathcal{B}$ (i.e. v, v_x, v_{xx} are in L^2) or has at most polynomial growth. If so, φ is an eigenfunction of $\mathcal{B}$ and Λ is in the spectrum of $\mathcal{B}$ (essentially meaning that Λ is an eigenvalue of $\mathcal{B}$).[52]

To answer this question, note that for a given $\Lambda \leq 0$, the general solution to (38) is $\varphi(x) = c_1 e^{ax} + c_2 e^{-ax}$ where a satisfies

$$\frac{\sigma^2}{2} a^2 = \frac{1}{2} \frac{\mu^2}{\sigma^2} + \Lambda. \tag{40}$$

Consider four different cases:

1. $\Lambda = 0$. In this case the solution to (40) is $a = \frac{\mu}{\sigma^2}$, i.e. $\varphi(x) = e^{\frac{\mu}{\sigma^2} x}$ which satisfies (39) and stays bounded as $x \to \infty$ (since $\mu < 0$). Hence $\Lambda = 0$ is an eigenvalue of $\mathcal{B}$ and is therefore in the spectrum of $\mathcal{B}$.

2. $-\frac{1}{2}\frac{\mu^2}{\sigma^2} < \Lambda < 0$. In this case, a solving (40) is real and positive. We therefore need $c_1 = 0$ so that φ does not explode exponentially as $x \to \infty$. But then the boundary condition (39) implies $-a = \frac{\mu}{\sigma^2}$, which is a contradiction. Hence points in $\left(-\frac{1}{2}\frac{\mu^2}{\sigma^2}, 0\right)$ are not in the spectrum of $\mathcal{B}$.

3. $\Lambda = -\frac{1}{2}\frac{\mu^2}{\sigma^2}$. In this case, (38) becomes $\varphi''(x) = 0$. A solution is $\varphi(x) = \alpha x + b$ where we can take $\alpha > 1$ and b is adjusted to satisfy the boundary condition (39). Since φ is polynomially bounded, $\Lambda = -\frac{1}{2}\frac{\mu^2}{\sigma^2}$ is in the spectrum of $\mathcal{B}$.

4. $\Lambda < -\frac{1}{2}\frac{\mu^2}{\sigma^2}$. In this case, a solving (40) is a purely imaginary number. We have $e^{ix} = \cos x + i \sin x$, so $\varphi(x) = c_1 e^{ax} + c_2 e^{-ax}$ oscillates but stays bounded as $x \to \infty$. We can therefore choose $c_1, c_2 \neq 0$ to satisfy the boundary condition (39). Hence any $\Lambda < -\frac{1}{2}\frac{\mu^2}{\sigma^2}$ is also in the spectrum of $\mathcal{B}$.

Summarizing, the spectrum of $\mathcal{B}$ consists of an isolated first eigenvalue $\Lambda = 0$ and all $\Lambda \in \left(-\infty, -\frac{1}{2}\frac{\mu^2}{\sigma^2}\right]$. $\square$

A.3 Proof of Proposition 2

The proof is in the Online Appendix; here we just give some intuition for its strategy.

[52] There is a subtle distinction between the eigenvalues of $\mathcal{B}$ and the spectrum of $\mathcal{B}$: Λ is only an eigenvalue if φ is in the domain of $\mathcal{B}$. If φ is not in the domain of $\mathcal{B}$ but has at most polynomial growth, Λ is not an eigenvalue but still in the spectrum of $\mathcal{B}$. Similarly, in this case φ is not an eigenfunction but a "generalized eigenfunction." Intuitively, φ is "almost in the domain of $\mathcal{B}$." See Simon (1981) for a proof that a polynomially bounded solution φ implies that Λ is in the domain of $\mathcal{B}$.

"Non-ergodic case." We take an initial distribution that is essentially completely in the "upper tail." There, the process is basically a constant-coefficient process. Then, as in Proposition 1, the speed of convergence is at most δ.

"Ergodic case." As in the proof of Proposition 1, we convert the situation into a self-adjoint operator $\mathcal{B}$ (detrended for death). Then, we find that there is some (generalized) eigenfunction of $\mathcal{B}$ corresponding to the spectral value $-\frac{1}{2}\frac{\bar{\mu}^2}{\bar{\sigma}^2}$. That means that the spectral gap and hence the speed of mean-reversion is at most $\frac{1}{2}\frac{\bar{\mu}^2}{\bar{\sigma}^2}$ or $\frac{1}{2}\frac{\bar{\mu}^2}{\bar{\sigma}^2} + \delta$ with death.

References

Acemoglu, D., and J. A. Robinson (2015): "The Rise and Decline of General Laws of Capitalism," *Journal of Economic Perspectives*, 29(1), 3–28.

Aoki, S., and M. Nirei (2015): "Zipf's Law, Pareto's Law, and the Evolution of Top Incomes in the U.S.," Working paper, Hitotsubashi University.

Atkinson, A. B., T. Piketty, and E. Saez (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49(1), 3–71.

Benhabib, J., A. Bisin, and S. Zhu (2011): "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents," *Econometrica*, 79(1), 123–157.

——— (2015a): "The Distribution of Wealth in the Blanchard–Yaari Model," *Macroeconomic Dynamics*, FirstView, 1–16.

——— (2015b): "The Wealth Distribution in Bewley Economies with Capital Income Risk," *Journal of Economic Theory*, 159, 489–515.

Bouchaud, J.-P., and M. Mézard (2000): "Wealth Condensation in a Simple Model of Economy," *Physica A*, 282(3), 536–545.

Bricker, J., A. M. Henriques, J. Krimmel, and J. Sabelhaus (2015): "Measuring Income and Wealth at the Top Using Administrative and Survey Data," Finance and Economics Discussion Series 2015-30, Board of Governors of the Federal Reserve System.

Champernowne, D. G. (1953): "A Model of Income Distribution," *The Economic Journal*, 63(250), 318–351.

DeBacker, J., B. Heim, V. Panousi, S. Ramnath, and I. Vidangos (2013): "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns," *Brookings Papers on Economic Activity*, 46(1), 67–142.

Feenberg, D. R., and J. M. Poterba (1993): "Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns," in *Tax Policy and the Economy, Volume 7*, NBER Chapters, pp. 145–177. National Bureau of Economic Research, Inc.

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

——— (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1(1), 255–293.

Gabaix, X., and A. Landier (2008): "Why Has CEO Pay Increased So Much?," *The Quarterly Journal of Economics*, 123(1), 49–100.

Garicano, L., and E. Rossi-Hansberg (2006): "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, 121(4), 1383–1435.

Geerolf, F. (2014): "A Static and Microfounded Theory of Zipf's Law for Firms and of the Top Labor Income Distribution," Working paper, UCLA.

Guvenen, F. (2007): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *American Economic Review*, 97(3), 687–712.

——— (2015): "Income Inequality and Income Risk: Old Myths vs. New Facts," Lecture slides available at `https://fguvenendotcom.files.wordpress.com/2014/04/handout_inequality_risk_webversion_may20151.pdf`.

Guvenen, F., G. Kaplan, and J. Song (2014): "The Glass Ceiling and The Paper Floor: Gender Differences among Top Earners, 1981-2012," NBER Working Paper 20560.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2015): "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?," NBER Working Paper 20913.

Guvenen, F., S. Ozkan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122(3), 621–660.

Haider, S. J. (2001): "Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991," *Journal of Labor Economics*, 19(4), 799–836.

Hansen, L. P., and J. A. Scheinkman (2009): "Long-Term Risk: An Operator Approach," *Econometrica*, 77(1), 177–234.

Harrison, M. (1985): *Brownian Motion and Stochastic Flow Systems*. Wiley & Sons.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006," *Review of Economic Dynamics*, 13(1), 15–51.

Hunter, J., and B. Nachtergaele (2001): *Applied Analysis*. World Scientific.

Jones, C. I. (2005): "The Shape of Production Functions and the Direction of Technical Change," *The Quarterly Journal of Economics*, 120(2), 517–549.

——— (2015): "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality," *Journal of Economic Perspectives*, 29(1), 29–46.

Jones, C. I., and J. Kim (2014): "A Schumpeterian Model of Top Income Inequality," Working paper, Stanford.

Kim, J. (2013): "The Effect of the Top Marginal Tax Rate on Top Income Inequality," Working paper, KAIST.

Kopczuk, W. (2015): "What Do We Know about the Evolution of Top Wealth Shares in the United States?," *Journal of Economic Perspectives*, 29(1), 47–66.

Kopczuk, W., E. Saez, and J. Song (2010): "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937," *The Quarterly Journal of Economics*, 125(1), 91–128.

Linetsky, V. (2005): "On the Transition Densities for Reflected Diffusions," *Advances in Applied Probability*, 37(2), 435–460.

Luttmer, E. G. (2012): "Slow Convergence in Economies with Firm Heterogeneity," Discussion paper.

Luttmer, E. G. J. (2007): "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, 122(3), 1103–1144.

——— (2011): "On the Mechanics of Firm Growth," *Review of Economic Studies*, 78(3), 1042–

1068.

——— (2015): "An Assignment Model of Knowledge Diffusion and Income Inequality," Staff Report 509, Federal Reserve Bank of Minneapolis.

MaCurdy, T. E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18(1), 83–114.

Malevergne, Y., A. Saichev, and D. Sornette (2013): "Zipf's Law and Maximum Sustainable Growth," *Journal of Economic Dynamics and Control*, 37(6), 1195–1212.

Meghir, C., and L. Pistaferri (2011): *Earnings, Consumption and Life Cycle Choices* vol. 4 of *Handbook of Labor Economics*, chap. 9, pp. 773–854. Elsevier.

Moffitt, R. A., and P. Gottschalk (1995): "Trends in the Covariance Structure of Male Earnings in the U.S., 1969-1987," Working paper, Johns Hopkins University.

Nirei, M. (2009): "Pareto Distributions in Economic Growth Models," IIR Working Paper 09-05, Institute of Innovation Research, Hitotsubashi University.

Pareto, V. (1896): *Cours d'Économie politique*. Librairie Droz. Reprinted in 1964.

Parker, J. A., and A. Vissing-Jorgensen (2010): "The Increase in Income Cyclicality of High-Income Households and Its Relation to the Rise in Top Income Shares," *Brookings Papers on Economic Activity*, 41(2), 1–70.

Piketty, T. (2014): *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T., and E. Saez (2003): "Income Inequality In The United States, 1913-1998," *The Quarterly Journal of Economics*, 118(1), 1–39.

Rosen, S. (1981): "The Economics of Superstars," *American Economic Review*, 71(5), 845–58.

Saez, E., and G. Zucman (2015): "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," *The Quarterly Journal of Economics*, forthcoming.

Saichev, A., Y. Malevergne, and D. Sornette (2009): *Theory of Zipf's Law and Beyond*, vol. 632. Springer.

Simon, B. (1981): "Spectrum and Continuum Eigenfunctions of Schrödinger Operators," *Journal of Functional Analysis*, 42.

Simon, H. A. (1955): "On a Class of Skew Distribution Functions," *Biometrika*, 42(3/4), 425–440.

Steindl, J. (1965): *Random Processes and the Growth of Firms: a Study of the Pareto Law*, Economic theory and applied statistics. Hafner Publishing Company.

Stone, M., and P. Goldbart (2009): *Mathematics for Physics: A Guided Tour for Graduate Students*. Cambridge University Press.

Tervio, M. (2008): "The Difference That CEOs Make: An Assignment Model Approach," *American Economic Review*, 98(3), 642–68.

Toda, A. A., and K. Walsh (2015): "The Double Power Law in Consumption and Implications for Testing Euler Equations," *Journal of Political Economy*, 123(5), 1177 – 1200.

Weidmann, J. (1980): *Linear Operators in Hilbert Spaces*. Springer.

Wold, H. O. A., and P. Whittle (1957): "A Model Explaining the Pareto Distribution of Wealth," *Econometrica*, 25(4), 591–595.