Online Appendix to "The Dynamics of Inequality"

Xavier Gabaix, Jean-Michel Lasry, Pierre-Louis Lions, Benjamin Moll
Appendix G with Zhaonan Qu
August 9, 2016

C Micro-foundation of Income Process (1)

Time is continuous, and there is a continuum of workers with preferences

$$\mathbb{E}_0 \int_0^\infty e^{-\rho t} \log c_{it} dt \tag{41}$$

A worker's consumption c_{it} equals her wage w_{it}, which is given by

$$w_{it} = \bar{w} h_{it}(1 - e_{it}). \tag{42}$$

Here $\bar{w}$ is an exogenous skill price that we normalize to one, h_{it} is her human capital or her skills, e_{it} is the fraction of time spent accumulating human capital. Workers die (retire) at rate δ, in which case they are replaced by a young worker with human capital h_{i0} drawn from a distribution $\tilde{\psi}(h)$. A worker's human capital evolves as

$$dh_{it} = \tilde{\gamma}(e_{it})h_{it} + \sigma h_{it} dZ_{it} + g_{it} h_{it} dN_{it} \tag{43}$$

where Z_{it} is a standard Brownian motion, N_{it} is a Poisson process with intensity ϕ, and g_{it} is drawn from a distribution f. The function $\tilde{\gamma}$ is increasing and concave. Workers maximize (41) subject to (42) and (43).

Lemma 8 *Worker's optimal investment decision e^* is independent of her human capital h_{it}. The resulting wage dynamics are given by*

$$dw_{it} = \gamma w_{it} dt + \sigma w_{it} dZ_{it} + g_{it} w_{it} dN_{it}, \tag{44}$$

where $\gamma := \tilde{\gamma}(e^*)$. By Ito's formula therefore $x_{it} = \log w_{it}$ satisfies (1) with $\mu = \gamma - \sigma^2/2$.

Proof of Lemma: Define the value function $v(h) = \max_e \mathbb{E}_0 \int_0^\infty e^{-\rho t} \log c_{it} dt$ where the expectation $\mathbb{E}_0$ is conditional on $h_{i0} = h$. A worker's HJB equation is

$$\rho v(h) = \max_e \; \log((1-e)h) + v'(h)\tilde{\gamma}(e)h + \frac{1}{2}v''(h)\sigma^2 h^2 + \phi \int_0^\infty (v(gh) - v(h)) f(g) dg \tag{45}$$

with corresponding first-order condition $\frac{1}{1-e} = v'(h)\tilde{\gamma}'(e)h$. We proceed with a guess-and-verify strategy. We guess that the value function takes the form

$$v(h) = A + B \log h \tag{46}$$

for constants A and B to be determined. Then the FOC implies e is independent of h and satisfies $\frac{1}{1-e} = Bg'(e)$. Plugging this and (46) into (45) we have

$$\rho(A + B \log h) = \log((1-e)h) + B\tilde{\gamma}(e) - \frac{1}{2}B\sigma^2 + \phi B \int_0^\infty \log g f(g) dg.$$

Collecting the terms involving h, we have $\rho B \log h = \log h$ or $B = 1/\rho$. Hence the optimal choice e^* solves $\frac{1}{1-e} = \tilde{\gamma}'(e)/\rho$, which is independent of h as asserted in the Lemma. Multiplying (43) by $1 - e^*$ and using $w_{it} = (1-e^*)h_{it}$ yields (44). For completeness, A satisfies $\rho A = \log(1-e^*) + \frac{1}{\rho}\tilde{\gamma}(e^*) - \frac{1}{2}\frac{1}{\rho}\sigma^2 + \phi\frac{1}{\rho}\int_0^\infty \log g f(g) dg.\square$

Labor Force Participation: To motivate exit with reinjection (or a reflecting barrier), extend the model above by assuming that workers have an early retirement option: if they choose to retire, they receive an income stream b instead of w_{it} each period. Once retired, they can never re-enter the labor force and are instead replaced by a new entrant with human capital drawn from a distribution $\tilde{\rho}(h)$. Workers therefore solve a stopping time problem, and one can show that they optimally retire when their wage drops below some threshold $\underline{w}$.

D Stationary Distributions of the Standard Random Growth Process

Consider the process for income (44) or equivalently the process for log income (1) with $\mu = \gamma - \sigma^2/2$. We here provide a complete characterization of the stationary distributions of the process for the various "stabilizing forces" discussed in Section 3.1. We first consider the case without jumps $\phi = 0$ and then turn to the case with jumps $\phi > 0$. The results for the former case are well-known (see e.g. Gabaix, 2009) and we state them without proof.

D.1 Stationary Distributions without Jumps $\phi = 0$

The following two assumptions will be used multiple times throughout this section and we collect them here to avoid unnecessary repetition.

Assumption 3 *The distribution of starting wages following death, ψ, satisfies:*

$$\lim_{x\to\infty}\frac{\psi(x)}{e^{-\zeta_+ x}}=\lim_{x\to-\infty}\frac{\psi(x)}{e^{-\zeta_- x}}=0$$

where $\zeta_-<0<\zeta_+$ are the two roots of (2).

Assumption 4 *The distribution of starting wages following exit, ρ, satisfies:*

$$\lim_{x\to\infty}\frac{\rho(x)}{e^{-\zeta_+ x}}=\lim_{x\to-\infty}\frac{\rho(x)}{e^{-\zeta_- x}}=0$$

where $\zeta_-<0<\zeta_+$ are the two roots of (2). *Furthermore, $\rho(x)=0$ for all $x\leq 0$.*

Stabilizing Force 1: Poisson Death, No Lower Bound on Income. Consider the process (44) or equivalently (1) with death at rate $\delta>0$ and no lower bound on income. We first analyze the case with rebirth at $x=0$ (equivalently $w=1$), i.e. ψ is the Dirac delta function at $x=0$, and then turn to general ψ.

With rebirth at $x=0$, the process has a stationary distribution given by the double Pareto distribution

$$f_\infty(w)=c\min\{w^{-\zeta_--1},w^{-\zeta_+-1}\},$$

where $c=-\zeta_-\zeta_+/(\zeta_+-\zeta_-)$ and where $\zeta_-<0<\zeta_+$ are the two roots of (2). Equivalently, the stationary distribution of the logarithm of income or wealth $x=\log w$ is given by

$$p_\infty(x)=c\min\{e^{-\zeta_- x},e^{-\zeta_+ x}\}. \tag{47}$$

Note that $f_\infty(w)$ has a Pareto tail, that is $\mathbb{P}(w_{it}>w)\sim Cw^{-\zeta_+}$ as $w\to\infty$ and $p_\infty(x)$ has an exponential tail, $\mathbb{P}(x_{it}>x)\sim Ce^{-\zeta_+ x}$. The expression for ζ in (3) is the expression for the positive root ζ_+.

The derivation is standard (see e.g. Gabaix, 2009) and can be carried out either in terms of $f_\infty(w)$ or $p_\infty(x)$. We here briefly restate the latter derivation. For $x\neq 0$ (outside of the point of rebirth), the stationary version of the Kolmogorov Forward equation (5) is

$$0=-\mu p_x+\frac{\sigma^2}{2}p_{xx}-\delta p. \tag{48}$$

Guess that $p(x)=ce^{-\zeta x}$ and hence $p_x(x)=-\zeta ce^{-\zeta x}$ and $p_{xx}(x)=\zeta^2 ce^{-\zeta x}$. Substituting this guess into (48), we get the quadratic (2) which has two roots $\zeta_-<0<\zeta_+$. Hence the general solution of (48) for $x \neq 0$ is

$$p_\infty(x) = \begin{cases} c_- e^{-\zeta_- x} + c_+ e^{-\zeta_+ x}, & x < 0 \\ C_- e^{-\zeta_- x} + C_+ e^{-\zeta_+ x}, & x > 0 \end{cases}$$

The two different cases come from the fact that (48) does not hold at $x = 0$ and hence $p_\infty(x)$ may not be differentiable (though it does have to be continuous). Because $p_\infty(x)$ has to be integrable as $x \to \infty$ and $\zeta_- < 0$, we require $C_- = 0$. Similarly, that $p_\infty(x)$ has to be integrable as $x \to -\infty$ and $\zeta_+ > 0$ imposes $c_+ = 0$. Given that $p_\infty(x)$ has to be continuous at $x = 0$, we finally have $c_- = C_+ = c$ and so the solution can be written as (47). Finally, the constant c is pinned down by the requirement that the distribution integrates to one.

Now consider the case of a general distribution of starting wages $\psi(x)$. One can no longer obtain an analytic solution of the form (47). But one can show that there exists a unique stationary distribution under Assumption 3, and that this stationary distribution satisfies

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

where $\zeta_- < 0 < \zeta_+$ are the two roots of (2), i.e. the *asymptotic* tail behavior of the distribution is *the same* as (47). Intuitively, Assumption 3 ensures that the endogenously generated part of the tail dominates any exogenous tail of ψ.

Stabilizing Force 2: Reflecting Barrier. Consider the process (44) with a reflecting barrier at $w = 1$ or equivalently (1) with a reflecting barrier at $x = 0$. If $\mu = \gamma - \sigma^2/2 < 0$, the process has a stationary distribution given by

$$f_\infty(w) = \zeta w^{-\zeta-1}, \quad w \geq 1, \qquad \zeta = -\frac{\mu}{\sigma^2/2} = 1 - \frac{\gamma}{\sigma^2/2}.$$

Equivalently, the stationary distribution of the logarithm of income or wealth $x = \log w$ is given by $p_\infty(x) = \zeta e^{-\zeta x}, x \geq 0$. Note that again $f_\infty(w)$ has a Pareto tail and $p_\infty(x)$ has an exponential tail.

Stabilizing Force 3: Exit and Reinjection. Consider first the case where reinjection occurs at a point $x_* > 0$, i.e. ρ is the Dirac delta function at x_*. If $\mu < 0$, then there exists a unique stationary distribution equal to

$$f_\infty(w) = cw^{-\zeta-1}\min\left\{w^{\zeta+1}-1, w_*^{\zeta+1}-1\right\}, \quad \zeta = -\frac{\mu}{\sigma^2/2}$$

or equivalently

$$p_\infty(x) = ce^{-\zeta x}\min\left\{e^{\zeta x}-1, e^{\zeta x_*}-1\right\}, \quad \zeta = -\frac{\mu}{\sigma^2/2}, \tag{49}$$

where c is pinned down by the requirement that the distribution integrates to one. This is a simplified version of equation (20) in Luttmer (2007). See his Figure II for a graphical representation. Finally note that as $x_* \downarrow 0$, (49) converges to $p_\infty(x) = \zeta e^{-\zeta x}$, i.e. the stationary distribution of the process with exit and entry converges to one associated with a reflecting barrier at $\underline{x} = 0$.

The derivation of (49) is straightforward. The stationary Kolmogorov Forward equation outside of the point of reinjection, i.e. for $x \neq x_*$, is

$$0 = -\mu p_x + \frac{\sigma^2}{2}p_{xx} \tag{50}$$

with boundary condition $p(0) = 0$. We solve this equation using a guess-and-verify strategy as before. Guess that

$$p(x) = \begin{cases} c\left(1-e^{-\zeta x}\right), & x < x_* \\ Ce^{-\zeta x}, & x > x_* \end{cases} \tag{51}$$

for $c, C > 0$. Note that this guess satisfies $p(0) = 0$ and $p(x) \to 0$ as $x \to \infty$. Take $x < x_*$. We have $p_x = c\zeta e^{-\zeta x}$ and $p_{xx} = -c\zeta^2 e^{-\zeta x}$. Plugging into (50) and canceling terms we have $\zeta = -2\mu/\sigma^2$ as asserted. One can check that this ζ can also be obtained by solving the branch of p for $x > x_*$. Since p has to be continuous at x_*, $c\left(1-e^{-\zeta x_*}\right) = Ce^{-\zeta x_*}$ or equivalently $C = c\left(e^{\zeta x_*}-1\right)$. Plugging this into (51), we have (49).

Next, one can relax the assumption that reinjection is at $x = x_*$ to allow for reinjection with a wage drawn from an arbitrary distribution $\rho(x)$ satisfying Assumption 4. While one can no longer obtain an analytic solution, one can then show that p has an asymptotic Pareto tail: $p(x) \sim e^{-\zeta x}$ for large x with the same tail exponent ζ as in (49).

Stabilizing Force 4: Reflecting Barrier and Poisson Death. Consider the process (44) with Poisson death at rate δ and both rebirth and a reflecting barrier at $w = 1$ or equivalently (1) with both rebirth and a reflecting barrier at $x = 0$. If either $\delta > 0$ or

$\mu = \gamma - \sigma^2/2 < 0$, the process has a stationary distribution given by

$$f_\infty(w) = \zeta w^{-\zeta-1}, \quad w \geq 1,$$

where ζ is the positive root of $0 = \frac{\sigma^2}{2}\zeta^2 + \zeta\mu - \delta$. Equivalently, the stationary distribution of the logarithm of income or wealth $x = \log w$ is given by $p_\infty(x) = \zeta e^{-\zeta x}, x \geq 0$. Note that again $f_\infty(w)$ has a Pareto tail and $p_\infty(x)$ has an exponential tail. As above, the assumption that rebirth occurs at $x = 0$ can be relaxed to rebirth from any function ψ satisfying Assumption 3. In this case again $p_\infty(x) \sim e^{-\zeta x}$ for large x.

Stabilizing Force 5: Additive Income Term ydt. Consider the process (54), i.e. a standard random growth process with the addition of an additive income term ydt, or equivalently (55). If $y > 0$ and $\mu = \gamma - \sigma^2/2 < 0$, the process has a stationary distribution, i.e. this additive term acts as a stabilizing force. No closed form solution exists but the distribution has a Pareto tail: for $w \to \infty$ and a constant c

$$f_\infty(w) \sim cw^{-\zeta-1}, \qquad \zeta = -\frac{\mu}{\sigma^2/2} = 1 - \frac{\gamma}{\sigma^2/2}. \tag{52}$$

Equivalently, the stationary distribution of the logarithm of income or wealth $x = \log w$ has an exponential tail: $p_\infty(x) \sim ce^{-\zeta x}$ for $x \to \infty$.

D.2 Stationary Distributions with Jumps $\phi > 0$

Before characterizing the stationary distribution of the process with jumps, we report a useful result that allows one to conclude that a distribution has a Pareto tail from a characterization of its Laplace transform alone and to characterize the corresponding tail exponent.

D.2.1 From Laplace Transform to Pareto Tail: A Tauberian Result

As we noted in the main text, if a distribution p has a Pareto tail, that is $p(x) \sim ce^{-\zeta x}$ $x \to \infty$ for constants c and ζ, then the Laplace transform (16) satisfies $\widehat{p}(\xi) \sim \frac{c}{\zeta+\xi}$ as $\xi \downarrow -\zeta$. Therefore $\zeta = -\inf\{\xi : \widehat{p}(\xi) < \infty\}$. Note that this characterization only applies *if* we already know that p has a Pareto tail. However, there are situations in which it is not known whether p has a Pareto tail but in which we still have a characterization of the Laplace transform $\widehat{p}$. It is therefore natural to ask whether there also is a converse result, i.e. whether we can say anything about a distribution's tail from its Laplace transform alone? The answer is "yes, we can."

There are a number of results along these lines, all of which are typically referred to as "Tauberian" results. These in turn are applications of Karamata's theory of regular variation. The most well-known example is Karamata's Tauberian Theorem. We here cite one Tauberian result that applies to our setup in which we would like to say something about the exponential decay of p given knowledge of the Laplace transform $\widehat{p}$. The result is Corollary 1.4 from Mimica (2013). It makes use of the concept of a function's "abscissa of convergence" which we define before stating the result.

Definition 1 (Negative Abscissa of Convergence of $\widehat{p}$) *Since p is a density on $(-\infty, \infty)$, there exists a real number $\xi^* \in (-\infty, 0]$ such that the integral defining the Laplace transform $\widehat{p}$ in* (16) *converges for $\xi \in (\xi^*, 0]$, diverges for $\xi < \xi^*$ and has a singularity at ξ^*. The number ξ^* is known as the* negative abscissa of convergence.

The following result is from Mimica (2013).[60] In addition to the "abscissa of convergence" we just defined, it also uses the concept of the "pole" of a function. Roughly, a pole of a function $f(\xi)$ is a singularity of f at ξ^* that behaves like $1/(\xi - \xi^*)^n$ for some positive integer n. For a complete definition, see any book on Complex Analysis.

Proposition 7 (Mimica (2013)) *Consider the Laplace transform $\widehat{p}(\xi, t)$ defined in* (16) *with the negative abscissa of convergence $\xi^* \in (-\infty, 0]$, and assume that ξ^* is a pole of $\widehat{p}$. Then*

$$\lim_{x \to \infty} \frac{1}{x} \log R(x, t) = \xi^* < 0,$$

where $R(x, t) := \int_x^\infty p(y, t) dy$ is the right-CDF corresponding to the density p. That is, $R(x, t) \sim Ce^{-\zeta x}$ or equivalently $p(x, t) \sim ce^{-\zeta x}$ as $x \to \infty$ with $\zeta = -\xi^*$. This means that p, the density of log income, has an exponential tail with parameter $\zeta = -\xi^*$ or equivalently the density of the level of income $w = e^x$ has a Pareto tail with that same tail parameter.

With this useful result in hand, we now go on to characterize the stationary distribution of the process with jumps.

D.2.2 Stationary Distributions with Jumps $\phi > 0$

As we will see, a sufficient condition for ensuring the existence of a unique stationary distribution with a Pareto tail is:

[60] Note that Mimica's result is stated in terms of the one-sided Laplace transform (which is the more commonly used version). It is easy to extend the result to the two-sided Laplace transform by using the fact that the two-sided Laplace transform is the sum of two one-sided Laplace transforms.

Assumption 5 *The distribution of jumps in the income growth rate satisfies:*

$$\lim_{g\to\infty}\frac{f(g)}{e^{-\beta_+ g}}=\lim_{g\to-\infty}\frac{f(g)}{e^{-\beta_- g}}=0$$

where $\beta_- < 0 < \beta_+$ *are the two roots of* (2), *i.e. the tail coefficients in the absence of jumps.*

We here denote the tail coefficients *in the absence of jumps* by β_-, β_+ so as to distinguish them from the tail coefficients with jumps ζ_-, ζ_+. We characterize the stationary distribution.

Proposition 8 *Under Assumptions 3 and 5, there is a unique stationary distribution with an asymptotic Pareto tail* $p_\infty(x) \sim e^{-\zeta x}$ *as* $x \to \infty$ *where* ζ *is the unique solution of*

$$P(\zeta) := \mu\zeta + \frac{\sigma^2}{2}\zeta^2 - \delta + \phi(\widehat{f}(-\zeta) - 1) = 0 \tag{53}$$

and satisfies $0 < \zeta \leq \beta_+$, *i.e. the tail is at least as fat as that in the model without jumps.*

Proof: Consider the Laplace transform of the stationary distribution $\widehat{p}_\infty(\xi)$ in equation (21) in Proposition 3. Assumptions 3 and 5 guarantee that $\widehat{\psi}(\xi) < \infty$ and $\widehat{f}(\xi) < \infty$ for all $\beta_- \leq -\xi \leq \beta_+$. Therefore the stationary Laplace transform $\widehat{p}_\infty(\xi)$ in (21) exists for $\beta_- \leq -\xi \leq \beta_+$ (at least). By the Tauberian result in Proposition 7, the underlying stationary distribution $p_\infty(x)$ has an exponential tail (and the distribution of $w = e^x$ a Pareto tail) if the Laplace transform $\widehat{p}_\infty(\xi)$ has a finite negative abscissa of convergence that also is a pole. We now show that this is the case.

Consider a point $\xi^* < 0$ at which the denominator in (21) equals zero, i.e. $\mu\xi^* - \frac{\sigma^2}{2}(\xi^*)^2 + \delta - \phi(\widehat{f}(\xi^*) - 1) = 0$. Equivalently, $\zeta = -\xi^* > 0$ satisfies (53). We first show that there is a unique such point and that $\zeta \leq \beta_+$. To this end, note that $\widehat{f}(-\zeta) := \int_{-\infty}^{\infty} e^{\zeta g} f(g)dg$ is strictly increasing in ζ with $\widehat{f}(0) = 1$ and $\widehat{f}(-\zeta) \geq 1$ for $\zeta > 0$. The function P is then the sum of the quadratic function (2), which has a positive root β_+, and a strictly increasing function $\phi(\widehat{f}(-\zeta) - 1)$. Therefore, it can only have one positive root. Next, note that

$$\begin{aligned}
P(0) &= -\delta < 0,\\
P(\beta_+) &= \mu\beta_+ + \frac{\sigma^2}{2}\beta_+^2 - \delta + \phi(\widehat{f}(-\beta_+) - 1) = \phi(\widehat{f}(-\beta_+) - 1) \geq 0,
\end{aligned}$$

where the last equality follows from the definition of β_+. Therefore, the unique positive root ζ of (53) satisfies $\delta < \zeta \leq \beta_+$.

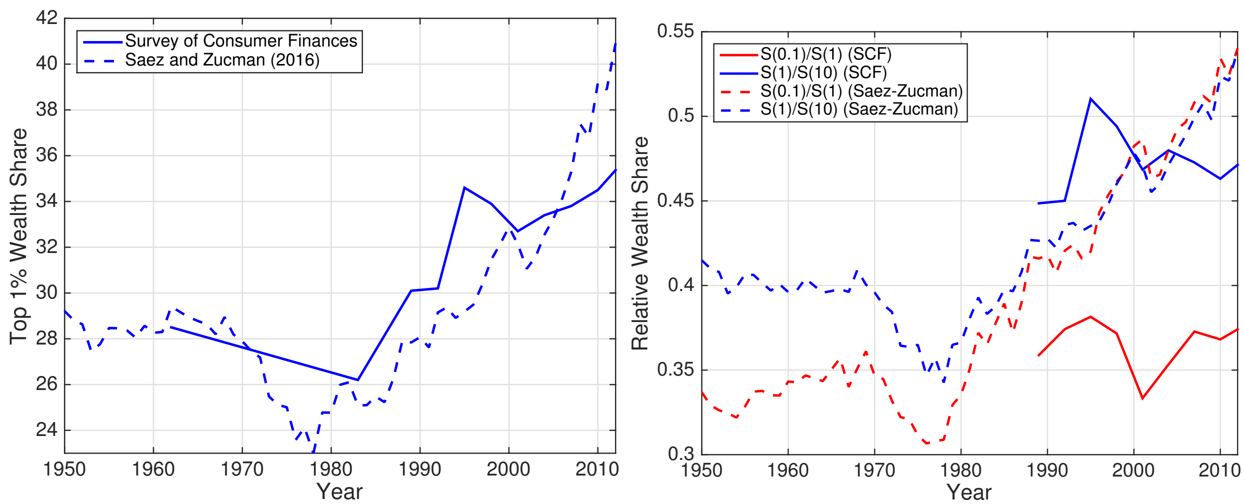
Summarizing, $\xi^* = -\zeta$ is the negative abscissa of convergence of $\widehat{p}_\infty(\xi)$. Furthermore, assuming that $\widehat{f}$ is an analytic function (in the complex analysis sense) for ξ in a neighborhood of ξ^* in the complex plane, this abscissa of convergence ξ^* is also a pole of $\widehat{p}_\infty$. Therefore by the Tauberian result in Proposition 7, p has an exponential tail (and the distribution of $w = e^x$ a Pareto tail) with parameter $\zeta = -\xi^*$ which satisfies (53).□

E The Dynamics of Wealth Inequality

In this appendix we explore the implications of our results for the dynamics of wealth inequality. We first provide a brief overview of the facts, and then show how our theoretical results can be extended to a simple model of top wealth inequality. We then ask whether an increase in $r - g$, the gap between the after-tax average rate of return and the growth rate, can explain the increase in top wealth inequality observed in some datasets as suggested by Piketty (2014).

E.1 Motivating Facts: the Evolution of Top Wealth Inequality

Figure 7 presents facts about the evolution of top wealth inequality, analogous to those about top income inequality in Figure 1. Panel (a) shows the time path of the top 1%



(a) Top Wealth Inequality

(b) Relative Wealth Shares

Figure 7: Evolution of Top Wealth Inequality

wealth share from two different data sources. The first is the Survey of Consumer Finances (SCF) and the second is a series constructed by Saez and Zucman (2015) by capitalizing capital income data.[61] The two series suggest quite different conclusions. In particular, data

[61]The SCF data for 1989 to 2013 is from the Online Appendix of Saez and Zucman (2015) available at `http://gabriel-zucman.eu/files/SaezZucman2014MainData.xlsx` in Sheet `DataFig1-6-7-11-12`. The

from the SCF suggest a relatively gradual rise in the top 1% wealth share, whereas Saez and Zucman's estimates suggest a much more dramatic rise, a discrepancy that has generated some controversy (see e.g. Kopczuk, 2015; Bricker, Henriques, Krimmel, and Sabelhaus, 2015).[62] Finally, comparing Figures 1 and 7 one can also see that wealth is much more unequally distributed than income.

Panel (b) plots the evolution of relative wealth shares which are informative about the fatness of the Pareto tail of the wealth distribution as discussed in Section 2. The finding depends again on the underlying data source, with the SCF showing no clear pattern and the capitalization method suggesting a large thickening of the tail of the wealth distribution.

There are three main takeaways from this section. First, top wealth shares appear to have increased though it is unclear by how much. Second, it is ambiguous whether the thickness of the tail of the wealth distribution has increased over time. And finally, wealth is more unequally distributed than income and, relatedly, the wealth distribution has a fatter Pareto tail than the income distribution.

E.2 A Simple Model of Top Wealth Inequality

The following simple model captures the main features of a large number of models of the upper tail of the wealth distribution.[63] Time is continuous and there is a continuum of individuals that are heterogeneous in their wealth $\tilde{w}_{it}$. At the individual level, wealth evolves as

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

where τ is the capital income tax rate, dR_{it} is the rate of return on wealth which is stochastic, y_t is labor income and c_{it} is consumption. To keep things simple, we make the following assumptions. First, capital income is i.i.d. over time and, in particular, $dR_{it} = \tilde{r}dt + \tilde{\nu}dZ_{it}$, where $\tilde{r}$ and $\tilde{\nu}$ are parameters, and Z_{it} is a standard Brownian motion, which reflects idiosyncratic returns to human capital or to financial capital (this idiosyncratic shock captures

SCF data for 1962 and 1983 is from Wolff (1987, Table 3). The 1962 dataset is a precursor of the SCF called the "Survey of Financial Characteristics of Consumers" or SFCC. Note that the pre- and post-1989 data use different wealth definitions and may therefore not be directly comparable. See the discussion in Kopczuk (2015) and Roine and Waldenström (2015), and the data appendix of Roine and Waldenström (2015) for alternative series that extend the SCF back in time.

[62]Kopczuk (2015) notes that a third method of measuring top wealth shares, the estate-tax multiplier technique, suggests an even smaller increase in the top one percent wealth share than the SCF. Also see critique of Piketty (2014) by Auerbach and Hassett (2015).

[63]See e.g. Wold and Whittle (1957), Benhabib, Bisin, and Zhu (2011, 2015a,b), Piketty and Zucman (2014b), Jones (2015) and Acemoglu and Robinson (2015).

the undiversified ownership of an entrepreneur, for instance).[64] Second, we assume that individuals consume an exogenous fraction θ of their wealth at every point in time, $c_{it} = \theta \tilde{w}_{it}$.[65] Third, we assume that all individuals earn the same labor income y_t, which grows deterministically at a rate g, $y_t = ye^{gt}$. Given these assumptions, it is easy to show that detrended wealth $w_{it} = \tilde{w}_{it} e^{-gt}$ follows the stochastic process

$$dw_{it} = (y + \gamma w_{it})dt + \sigma w_{it} dZ_{it}, \quad \gamma := r - g - \theta \tag{54}$$

where $r = (1 - \tau)\tilde{r}$ is the after-tax average rate of return on wealth and $\sigma = (1 - \tau)\tilde{\nu}$ is the after-tax wealth volatility. Note that (54) is a standard random growth process with the addition of an additive income term ydt. This income term acts as a stabilizing force. Many other shocks (e.g. demographic shocks or shocks to saving rates) result in a similar reduced form. From Ito's formula, the logarithm of wealth $x_{it} = \log w_{it}$ satisfies

$$dx_{it} = (ye^{-x_{it}} + \mu)dt + \sigma dZ_{it}, \quad \mu := r - g - \theta - \frac{\sigma^2}{2}. \tag{55}$$

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

$$\eta = \frac{1}{\zeta} = \frac{\sigma^2/2}{\sigma^2/2 - (r - g - \theta)} \tag{56}$$

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

The model can easily be extended to the case where labor income is stochastic, i.e. y

[64] Benhabib, Bisin, and Luo (2015) argue that, in the data, such uninsured capital income risk is the main determinant of the wealth distribution's right tail. In Section 5.3 we additionally consider common shocks.

[65] A consumption rule with such a constant marginal propensity to consume can also be derived from optimizing behavior, at least for large wealth levels w_{it}.

in (54) follows some stochastic process. As long as the income process does not itself have a fat-tailed stationary distribution, this does not affect the tail parameter of the wealth distribution (56). Intuitively, as wealth $w_{it} \to \infty$, labor income becomes irrelevant as an income source.

E.3 Dynamics of Wealth Inequality: Theoretical Results

We now show how our theoretical results can be extended to wealth dynamics. The addition of the labor income term y in (55) introduces some difficulties for extending Proposition 1. However, note that for large wealth levels this term becomes negligible, which makes it possible to derive a tight *upper bound* on the speed of convergence of the cross-sectional distribution.

Proposition 9 (Speed of convergence for wealth dynamics) *Consider the wealth process (55). Under Assumption 1, and if $\mu < 0$, the cross-sectional distribution $p(x,t)$ converges to its stationary distribution. The rate of convergence $\lambda := -\lim_{t\to\infty} \frac{1}{t} \log ||p(x,t) - p_\infty(x)||$ satisfies*

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

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

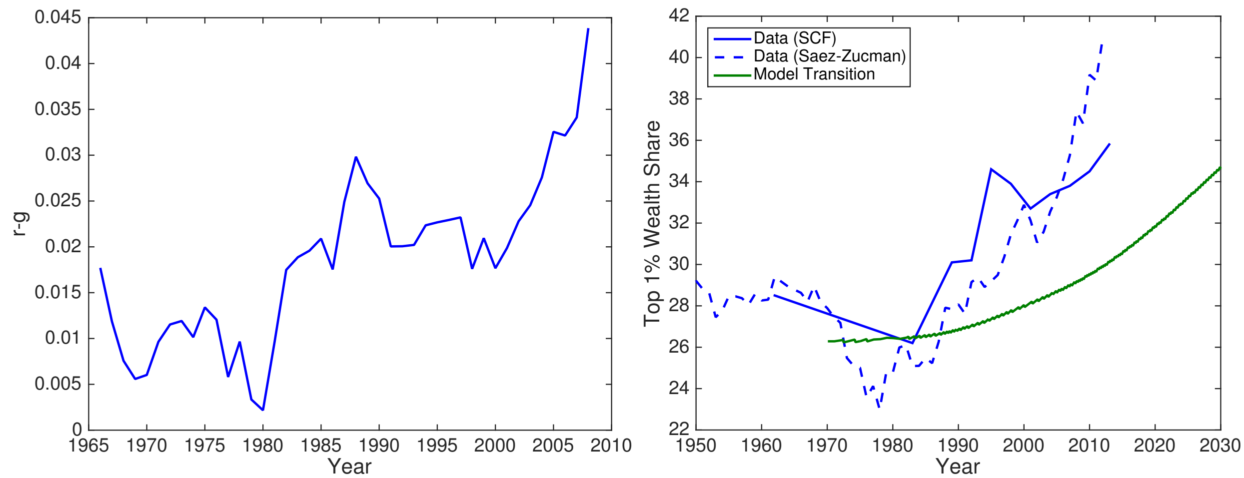
We conjecture that with $\mu > 0$, $\lambda = \delta$, as in Proposition 1.

With the process (55), it is not possible to obtain an exact formula for the speed of convergence. However, the speed of convergence is bounded above and, in particular, is equal to or less than the speed with a reflecting barrier from Proposition 1. It is also no longer possible to characterize the corresponding Kolmogorov Forward equation by using Laplace transforms (due to the presence of the term ye^{-xt}). Numerical experiments nevertheless confirm our results from section 4.2 that the speed of convergence in the tail can be substantially lower than the average speed of convergence characterized in Proposition 9.

E.4 Wealth Inequality and Capital Taxes

In this section we ask whether an increase in $r - g$, the gap between the (average) after-tax rate of return on wealth and the economy's growth rate, can explain the increase in wealth inequality observed in some data sets, as suggested by Piketty (2014). To do so, we first construct a measure of the time series of $r - g$. This requires three data inputs: on the average pre-tax rate of return, on capital income taxes, and on a measure of the economy's growth rate. We use data on the average before-tax rate of return from Piketty and Zucman

(2014a), the series of top marginal capital income tax rates from Auerbach and Hassett (2015) and data on the growth rate of per capita GDP of the United States from the Penn World Tables. Panel (a) of Figure 8 plots our time series for $r-g$, displaying a strong upward trend starting in the late 1970s, which coincides with the time when top wealth inequality started to increase (Figure 1).[66] The figure therefore suggests that, a priori, the theory using variations in $r-g$ is a potential candidate for explaining increasing wealth inequality.



(a) Evolution of of $r-g$ (b) Top 1% Wealth Share

Figure 8: Dynamics of Wealth Inequality in the Baseline Model

We now ask whether the simple model of wealth accumulation from Section 3 has the potential to explain the different data series for wealth inequality in Figure 1. To this end, recall equation (54) and note that the dynamics of this parsimonious model are described by two parameter combinations only, $r-g-\theta$, where θ is the marginal propensity to consume out of wealth, and the cross-sectional standard deviation of the return to capital, σ. Our exercise proceeds in three steps. First, we obtain an estimate for σ. We use $\sigma = 0.3$, which is on the upper end of values estimated or used in the existing literature.[67] Second, given σ and our data for $r-g$ in 1970, we calibrate the marginal propensity to consume θ so as to match the tail inequality observed in the data in 1970, $\eta = 0.6$. Third, we feed the time path for $r-g$ from panel (a) of Figure 8 into the calibrated model.

Before comparing the model's prediction to the evolution of top wealth inequality in the

[66] We have tried a number of alternative exercises with different data series for the return on capital and taxes, e.g. we set the pre-tax r equal to the yields of 10-year government bonds as in Auerbach and Hassett (2015) and Piketty and Zucman (2014a). Results are very similar.

[67] Overall, good estimates of σ are quite hard to come by and relatively dispersed. Campbell (2001) provides the only estimates for an exactly analogous parameter using Swedish wealth tax statistics on asset returns. He estimates an average σ of 0.18. Moskowitz and Vissing-Jorgensen (2002) argue for a σ of 0.3.

data, we make use of our analytic formulas from Section 4 to calculate measures of the speed of convergence. To this end, revisit the average speed of convergence in Proposition 1, and in particular the formula in terms of inequality (13). To operationalize this formula, we use the tail exponent observed in 2010 in the SCF of $\eta = 0.65$ together with our other parameter values.[68] With these numbers in hand, we obtain a half-life of

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

That is, on average, the distribution takes 26 years to cover half the distance to the new steady state. Panel (b) of Figure 8 displays the results of our experiment using the parameter values just discussed. The main takeaway is that the baseline random growth model cannot even explain the gradual rise in top wealth inequality found in the SCF. It fails even more obviously in explaining the rise in top wealth inequality found by Saez and Zucman (2015).

E.5 Fast Dynamics of Wealth Inequality

What, then, explains the dynamics of wealth inequality observed in the data? The lessons from Section 5 still apply. In particular, processes of the form (26) that feature deviations from Gibrat's law in the form of "type dependence" or "scale dependence" have the potential to deliver fast transitions. We view both as potentially relevant for the case of wealth dynamics. Wealth dynamics at the individual level depend on both rates of returns and saving rates, and heterogeneity or wealth-dependence in either would result in such deviations from Gibrat's law.

With regard to rates of returns, Fagereng, Guiso, Malacrino, and Pistaferri (2016), using high-quality Norwegian administrative data, find evidence for type dependence across the entire support of the wealth distribution. Additionally, they find some evidence for scale dependence particularly above the 95th percentile of the wealth distribution, mostly because wealthier people take more risk and compensated in the form of higher returns. This is also consistent with evidence in Bach, Calvet, and Sodini (2015) using Swedish administrative data, as well as Kacperczyk, Nosal, and Stevens (2014). With regard to saving rates, scale dependence may arise because the saving rates of the super wealthy relative to those of the wealthy may change over time (Saez and Zucman, 2015).[69]

[68] Ideally, one would use an estimate of tail inequality in the new stationary distribution η. Since λ is decreasing in inequality, we use the tail exponent observed in 2010 in the SCF of $\eta = 0.65$, which provides an upper bound on the speed of convergence λ. Since inequality in the new stationary distribution may be even higher, true convergence may be even slower.

[69] Finally, it is natural to ask whether the extension to multiple distinct growth regimes of Section 5.2 can generate fast transition dynamics in response to the increase in $r - g$ from Section E.4. Numerical

E.6 Proof of Proposition 9

The proof follows steps exactly analogous to the proof of Proposition 1, particularly the "ergodic case" in Appendix A.2. The only difference to the earlier proof is that some of the arguments need to be adjusted to account for the presence of the term $ye^{-x_{it}}dt$ in the wealth process (55). We again present the proof for the case $\delta = 0$. Denote the drift of wealth by

$$b(x) = \mu + ye^{-x}.$$

Consider the Kolmogorov Forward equation corresponding to (55)

$$p_t = \mathcal{A}^* p, \quad \mathcal{A}^* p = -(b(x)p)_x + \frac{\sigma^2}{2} p_{xx} \tag{57}$$

and its adjoint

$$\mathcal{A}u = b(x)u_x + \frac{\sigma^2}{2} u_{xx}.$$

As in the proof of Proposition 1, the strategy is again to construct a self-adjoint transformation $\mathcal{B}$ of $\mathcal{A}$, which is again found as the operator corresponding to $v = up_\infty^{1/2}$ where p_∞ is the stationary distribution corresponding to (57). To find p_∞, define $B(x) := -ye^{-x} + \mu x$ such that $B'(x) = b(x)$ and write

$$0 = -(B'p)' + \frac{\sigma^2}{2} p'' \quad \Rightarrow \quad \frac{p'}{p} = \frac{2B'}{\sigma^2} \quad \Rightarrow \quad p_\infty(x) \propto e^{2B(x)/\sigma^2}.$$

Since $\mu < 0$ and $y > 0$, $B(x) \to -\infty$ as $x \to \pm\infty$. Hence $p_\infty(x) \to 0$ for $x \to \pm\infty$ and there is a well-defined stationary distribution.

The rest of the proof establishes analogous versions of Lemmas 6 and 7.

Lemma 9 *Consider u satisfying $u_t = \mathcal{A}u$ and the corresponding stationary distribution, $p_\infty(x) = ce^{2B(x)/\sigma^2}$. Then $v = up_\infty^{1/2} := ue^{B(x)/\sigma^2}$ satisfies*

$$v_t = \mathcal{B}v, \qquad \mathcal{B}v = \frac{\sigma^2}{2} v_{xx} - \frac{1}{2\sigma^2}\left(\mu^2 + y^2 e^{-2x} + 2ye^{-x}\mu - \sigma^2 ye^{-x}\right) v. \tag{58}$$

Furthermore, $\mathcal{B}$ is self-adjoint.

Proof.

We have $v_t = e^{B(x)/\sigma^2} b(x) u_x + \frac{\sigma^2}{2} e^{B(x)/\sigma^2} u_{xx}$. We need to check that the right hand side

experiments suggest that the answer is no.

is equal to $\frac{\sigma^2}{2}v_{xx} - \frac{1}{2\sigma^2}\left(\mu^2 + y^2 e^{-2x} + 2ye^{-x}\mu - \sigma^2 y e^{-x}\right)v$. For this, we need to calculate

$$\begin{aligned} v_x &= u_x e^{B(x)/\sigma^2} + \frac{b(x)}{\sigma^2}e^{B(x)/\sigma^2}u \\ v_{xx} &= u_{xx}e^{B(x)/\sigma^2} + 2\frac{b(x)}{\sigma^2}u_x e^{B(x)/\sigma^2} + \frac{-ye^{-x}}{\sigma^2}e^{B(x)/\sigma^2}u + (\frac{b(x)}{\sigma^2})^2 e^{B(x)/\sigma^2}u \\ b(x)^2 &= \mu^2 + 2\mu y e^{-x} + y^2 e^{-2x} \end{aligned}$$

and so

$$\frac{\sigma^2}{2}v_{xx} = \frac{\sigma^2}{2}u_{xx}e^{B(x)/\sigma^2} + b(x)u_x e^{B(x)/\sigma^2} + \frac{-ye^{-x}}{2}e^{B(x)/\sigma^2}u + \frac{\mu^2 + 2\mu y e^{-x} + y^2 e^{-2x}}{2\sigma^2}e^{B(x)/\sigma^2}u$$

which gives the desired result. □

Lemma 10 *The first eigenvalue of $\mathcal{B}$ is $\lambda_1 = 0$. The second eigenvalue satisfies the following properties: there exists $-\infty < \mu^* < 0$ such that $\lambda_2 = -\frac{1}{2}\frac{\mu^2}{\sigma^2}$ for $|\mu| \le |\mu^*|$ and $\lambda_2 \le -\frac{1}{2}\frac{\mu^2}{\sigma^2}$ for all $\mu < 0$. All remaining eigenvalues satisfy $|\lambda| > |\lambda_2|$. Put differently, the spectral gap of $\mathcal{B}$ satisfies $|\lambda_2| \le \frac{1}{2}\frac{\mu^2}{\sigma^2}$.*

The conclusion of the proof (from spectral gap to L^1-norm) is unchanged from that of Proposition 1.

F Proof Complements

F.1 Intuition Why Second Eigenvalue Matters

With a lower bound, the key step in the proof of Proposition 1 was to derive an expression for the second eigenvalue of the operator $\mathcal{B}$ defined in (37) (see section A.2). We here provide some additional intuition for why it is the second eigenvalue that determines the speed of convergence. Rather than considering the infinite-dimensional case studied in the paper, we consider a finite-state Markov chain. One can then study the speed of convergence using standard linear algebra tools. See Lawler (2006, Chapter 7.2) for an excellent introductory treatment.[70]

Consider a finite-state Markov chain for $x_{it} \in \{x_1, ..., x_n\}$ with some $n \times n$ transition matrix $\mathbf{B}$. The distribution of x_{it} is a vector $p(t) = (p_1(t), ..., p_n(t))$ which satisfies

$$\dot{p} = \mathbf{B}'p, \qquad p(0) = p_0$$

[70]See also `http://en.wikipedia.org/wiki/Markov_chain#Convergence_speed_to_the_stationary_distribution`.

Assume that $\mathbf{B}$ is symmetric so that all its eigenvalues are real. Further assume that $\mathbf{B}$ is diagonalizable, and denote its eigenvalues by $(\Lambda_1, ..., \Lambda_n)$ and the corresponding eigenvectors by $(v_1, ..., v_N)$. One can always decompose the initial density as $p_0 = \sum_{i=1}^n c_i v_i$ for some weights c_i. Then

$$p(t) = \sum_{i=1}^{n} c_i e^{\Lambda_i t} v_i.$$

If the process has a stationary distribution, the first (principal) eigenvalue is zero and the stationary distribution is given by the first eigenvector v_1 that has only positive entries (Perron-Frobenius Theorem). It can be seen that $p(t) \to c_1 v_1$ as $t \to \infty$ (the stationary distribution is proportional to the first eigenvector). The *speed* of convergence is instead governed by the second eigenvalue. To see this note that for large t (assuming that the initial distribution p_0 is not orthogonal to the second eigenvector v_2 and so $c_2 \neq 0$)

$$p(t) \approx c_1 v_1 + c_2 e^{\Lambda_2 t} v_2$$

because all remaining terms go to zero faster due to the fact that $|\Lambda_i| > |\Lambda_2|$ for $i > 2$.[71]

For example, consider a two-dimensional Poisson process with symmetric intensity ϕ. In that case the transition matrix is

$$\mathbf{B} = \begin{bmatrix} -\phi & \phi \\ \phi & -\phi \end{bmatrix}$$

which has eigenvalues $\Lambda_1 = 0, \Lambda_2 = -2\phi$ (the two roots of $0 = \det(\mathbf{B} - \Lambda I) = (\phi + \Lambda)^2 - \phi^2$). Intuitively, the speed of convergence Λ_2 is larger the higher is the Poisson intensity ϕ.

Proposition 1 generalizes this argument to the infinite-dimensional operator $\mathcal{B}$ in (37).

F.2 Complements to the Proof of Proposition 1

F.2.1 Last part of the proof: from spectral gap to L^1 norm

Denoting $\lambda = \frac{\mu^2}{2\sigma^2}$, it remains to show that under Assumption 1, the following two statements are true. First,

$$\int_0^\infty |p(x,t) - p_\infty(x)| \, dx \leq k e^{-\lambda t}. \tag{59}$$

This inequality proves that the cross-sectional distribution $p(x,t)$ converges to its stationary distribution $p_\infty(x)$ in the total variation norm for any initial distribution $p_0(x)$. Second, we

[71] If the initial condition is orthogonal to the second eigenvector, the asymptotic speed of convergence is faster than $|\Lambda_2|$. For instance, if the initial condition is proportional to the third eigenvector, $p_0 = c_3 v_3$, the asymptotic speed of convergence is $|\Lambda_3|$. But cases like this are knife-edge and for generic initial conditions, the speed of convergence is governed by the second eigenvalue.

need to show that the inequality in (59) is tight, in the sense that the upper bound in (59) (with speed λ) is attained asymptotically for generic initial conditions p_0.[72] As in the main proof, we continue to consider the case $\delta = 0$, which contains all the difficulty. The general case is an easy extension.

The difficult part is to show the first part, the inequality in (59), and we show it below. The easier part is to show that the upper bound (with speed λ) is indeed generically attained. The logic is exactly the same as in the finite-dimensional case analyzed in Appendix F.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. The same phenomenon holds for continuous spectra (e.g. Kato 1995).

Proof of (59): The key difficulty is contained in the case $\delta = 0$, so we start with it.

First (and central) case: $\delta = 0$. The following applies both to the plain constant-coefficient case (μ, σ constant) and to the more general case with varying $\mu(x), \sigma(x)$ of Proposition 2. First note that Assumption 1 implies the condition we truly need, which is:[73]

$$\int_0^\infty \frac{(p_0(x) - p_\infty(x))^2}{\bar{p}_\infty(x)} dx < \infty, \tag{60}$$

where $\bar{p}_\infty(x) := \bar{\zeta} e^{-\bar{\zeta} x}$ with $\bar{\zeta} := \frac{-2\mu}{\sigma^2}$ in the constant coefficient case, and in the more general case with varying coefficients,

$$\bar{p}_\infty(x) = \frac{K}{\sigma^2(x)} e^{\int_1^x \frac{2\mu(y)}{\sigma^2(y)} dy} \tag{61}$$

with K a normalization constant. Here $\bar{p}_\infty$ is a surrogate steady state distribution, for a process with $\delta = 0$.

Note that in (60), the numerator features p_∞, the true steady state density (which depends on δ, and the location of births $\psi(x)$), while the denominator features $\bar{p}_\infty$ (which does not depend on δ and the location of births). We have $p_\infty = \bar{p}_\infty$ with reflected barrier and no death, but not in general. As we will see, the power 2 in the numerator shows up because the argument converts a statement involving the L^2-norm (63) into a statement involving the L^1-norm (59).

[72] To be precise, note that in (59) the constant k depends on the initial condition p_0. In contrast, the rate of decay λ is independent of this initial condition.

[73] We have chosen to use Assumption 1, rather than the more complex (60), to have a simple formulation.

We now move to the core statement of this proof. Consider $p(x,t)$ with $p_t = \mathcal{A}^* p$ and initial condition $p_0(x)$. Our main goal here is to prove that there is a constant C_0 such that:

$$\int |p(t) - J\bar{p}_\infty| \, dx \le C_0 e^{-\lambda t} \tag{62}$$

where $J := \int p_0(y)\, dy$ is not necessarily 1. Indeed, when we later apply this result to the case $\delta > 0$, we shall take p_0 to be the difference between two densities, so that p_0 can take negative values and J can be 0.

Next, consider $v = u\bar{p}_\infty^{1/2}$ defined in Lemma 6 satisfying $v_t = \mathcal{B}v$, i.e. (37). We appeal to the standard spectral decomposition of a self-adjoint operator.[74] We use the fact that the eigenfunction corresponding to the 0 eigenvalue is $m = \bar{p}_\infty^{1/2} = \left(-\frac{2\mu}{\sigma^2}\right)^{1/2} e^{(\mu/\sigma^2)x}$ from Lemma 7 in the constant coefficient case.[75] Hence:

$$\left(\int |v(x,t) - cm(x)|^2 \, dx\right)^{1/2} \le \left(\int |v_0(x) - cm(x)|^2 \, dx\right)^{1/2} e^{-\lambda t}, \tag{63}$$

where we suppress the limits of integration which are always 0 and ∞ for the remainder of the proof and where the constant c is given by (using $v = um$)

$$c := \frac{\langle v_0, m\rangle}{\langle m, m\rangle} = \frac{\int v_0(y)m(y)dy}{\left(\int (m(y))^2 dy\right)^{1/2}} = \frac{\int u_0(y)m(y)^2 dy}{\left(\int (m(y))^2 dy\right)^{1/2}} = \int u_0(y)\, \bar{p}_\infty(y) dy. \tag{64}$$

To see why (63) is true, consider the finite-dimensional case as an analogy, as in Section F.1. Calling $0 = \Lambda_1 \ge \Lambda_2 \ge ... \ge \Lambda_n$ the eigenvalues of $\mathbf{B}$ and $e_1, ..., e_n$ the corresponding eigenvectors (normalized to have unit norm), we decompose $v(0) = \sum_{i=1}^n c_i e_i$. Then, $\dot{v} = \mathbf{B}v$ implies $v(t) = \sum_{i=1}^n c_i e^{\Lambda_i t} e_i$. Note also that m is the eigenvector corresponding to $\Lambda_1 = 0$: $e_1 = m$. Also, $\Lambda_2 = -\lambda$ is the first non-trivial eigenvalue. Then, $c := \frac{\langle v(0), m\rangle}{\langle m, m\rangle} = c_1$ is the projection of $v(0)$ on the top eigenvector, e_1. Next, we have

$$v(t) - cm = v(t) - c_1 e_1 = \sum_{i=2}^n c_i e^{\Lambda_i t} e_i.$$

Denoting by $||\cdot||_2$ the L^2-norm (for any $x \in \mathbb{R}^n$, $||x||_2 := \left(\sum_{i=1}^n x_i^2\right)^{1/2}$), using the fact that

[74] This is the infinite-dimensional analogue of the diagonalization of a symmetric matrix.

[75] Unfortunately, when μ and σ vary spatially, the generalization of $\mathcal{B}$ is no longer self-adjoint. However, it is still true that $\mathcal{B}m = 0$ with $m = \sqrt{\bar{p}_\infty(x)} = \sqrt{\frac{K}{\sigma^2(x)} e^{\int_1^x \frac{2\mu(y)}{\sigma^2(y)} dy}}$ and $\mathcal{B}v = m\mathcal{A}(m^{-1}v)$.

the eigenvectors are orthogonal (because $\mathbf{B}$ is self-adjoint):

$$\begin{aligned}\|v(t) - cm\|_2^2 &= \left\| \sum_{i=2}^n c_i e^{\Lambda_i t} e_i \right\|_2^2 = \sum_{i=2}^n \left\| c_i e^{\Lambda_i t} e_i \right\|_2^2 = \sum_{i=2}^n c_i^2 e^{2\Lambda_i t} \\ &\leq \sum_{i=2}^n c_i^2 e^{2\Lambda_2 t} = e^{-2\lambda t} \sum_{i=2}^n c_i^2 = e^{-2\lambda t} \|v(0) - cm\|_2^2\end{aligned}$$

i.e.

$$\|v(t) - cm\|_2 \leq \|v(0) - cm\|_2 e^{-\lambda t}. \tag{65}$$

Equation (63) is simply the infinite-dimensional analogue of (65).[76]

Next, we note that if u and p satisfy $u_t = \mathcal{A}u$ and $p_t = \mathcal{A}^* p$, then

$$\int u(x,t) p_0(x) dx = \int u_0(x) p(x,t) dx. \tag{66}$$

[76]Complications from boundary conditions and unboundedness of $\mathcal{B}$ renders the extension technically involved. Instead, we can prove the infinite-dimensional analogue of this equation (up to a constant) using an energy argument. More precisely, note that $\int v(t) m dx = \int v_0 m dx$, so that $v(t) - cm$ is always orthogonal to m. Note that $\tilde{v} = v - cm$ also satisfies $\tilde{v}_t = \mathcal{B}\tilde{v}$ if $v_t = \mathcal{B}v$. Denote $\tilde{v}$ by v and assume $\int v(t) m dx = 0$ for all t. Multiply the equation $v_t = \mathcal{B}v = \frac{\sigma^2}{2} v_{xx} - \frac{\mu^2}{2\sigma^2} v$ by $2v$ and integrate from 0 to ∞, we get

$$\begin{aligned}\frac{d}{dt} \int v^2 dx &= \sigma^2 \int v_{xx} v dx - \frac{\mu^2}{\sigma^2} \int v^2 dx \\ &= -\sigma^2 \int v_x^2 dx - \frac{\mu^2}{\sigma^2} \int v^2 dx - \mu v^2(0) \\ &= -\sigma^2 \int v_x^2 dx - \frac{\mu^2}{\sigma^2} \int v^2 dx + 2\mu \int v_x v dx \\ &= -\sigma^2 \int (v_x - \frac{\mu}{\sigma^2} v)^2 dx\end{aligned}$$

using boundary condition $v_x(0) = \frac{\mu}{\sigma^2} v(0)$ for $\mathcal{B}$. The term $\int (v_x - \frac{\mu}{\sigma^2} v)^2 dx = 0$ only if $v_x = \frac{\mu}{\sigma^2} v$ for all x, i.e. $v = m$ or $v = 0$. However, since $\int vm dx = 0$, $\int (v_x - \frac{\mu}{\sigma^2} v)^2 dx > 0$ for all t. It follows that $\int v^2 dx \to 0$. Moreover, letting $w(x) = v_x(x) - \frac{\mu}{\sigma^2} v(x)$ we have $\mathcal{B}w = w_t$ and $w(0) = 0$ for all t. Running the same argument, we see that $\frac{d}{dt} \int w^2 dx = -\sigma^2 \int w_x^2 dx - \frac{\mu^2}{\sigma^2} \int w^2 dx$, so that $\int w^2 dx = \Theta(e^{-2\lambda t})$ where $\lambda \geq \frac{\mu^2}{2\sigma^2}$. Now $w^2(x) = v_x^2 + \frac{\mu^2}{\sigma^4} v^2 - \frac{2\mu}{\sigma^2} v v_x$, so that

$$\frac{d}{dt} \int v^2 dx = -\sigma^2 \int w^2(x) dx$$

and so $\int v^2 dx \leq C \int v_0^2 dx e^{-2\lambda t} \leq C \int v_0^2 dx e^{-2(\frac{\mu^2}{2\sigma^2})t}$, as was to be shown.

To prove this "dual equation", let $I(s) = \int u(x,t-s)p(x,s)dx$. Then

$$\begin{aligned}\frac{d}{ds}I(s) &= -\int \partial_t u(x,t-s)p(x,s)dx + \int u(x,t-s)\partial_s p(x,s)dx \\ &= -\int \mathcal{A}u(x,t-s)p(x,s)dx + \int u(x,t-s)\mathcal{A}^* p(x,s)dx = 0\end{aligned}$$

Setting $s = 0$ and t gives the result. Therefore

$$\int \left(u - \int u_0 \bar{p}_\infty dy\right) p_0 dx = \int u p_0 dx - J \int u_0 \bar{p}_\infty dy = \int u_0 \left(p - J\bar{p}_\infty\right) dx. \qquad (67)$$

Substituting (64) into (63),

$$\begin{aligned}L &:= \int |v(x,t) - cm(x)|^2 \, dx = \int |m(x)\, u(x,t) - cm(x)|^2 \, dx = \int m(x)^2 \, |u(x,t) - c|^2 \, dx \\ &= \int \bar{p}_\infty(x) \, |u(x,t) - c|^2 \, dx, \\ L &\le e^{-2\lambda t} \int |v_0(x) - cm(x)|^2 \, dx = e^{-2\lambda t} \int m(x)^2 \, |u_0(x) - c|^2 \, dx \\ &= e^{-2\lambda t} \int \bar{p}_\infty \, |u_0(x) - c|^2 \, dx,\end{aligned}$$

hence, using $c = \int u_0 \bar{p}_\infty dy$,

$$\left(\int \bar{p}_\infty \left(u - \int u_0 \bar{p}_\infty dy\right)^2 dx\right)^{1/2} \le e^{-\lambda t} \left(\int \bar{p}_\infty \left(u_0 - \int u_0 \bar{p}_\infty dy\right)^2 dx\right)^{1/2}. \qquad (68)$$

Next, define $J := \int p_0 dx$ (as mentioned above, $J = 1$ is typical but J may also equal zero if p_0 is the difference between two densities as in the proof of the case $\delta > 0$). Using

(67), we have:

$$\left|\int u_0\left(p(t)-J\bar{p}_\infty\right)dx\right| = \left|\int\left(u(t)-\int u_0\bar{p}_\infty dy\right)p_0 dx\right| = \left|\int\left(\left(u(t)-\int u_0\bar{p}_\infty dy\right)\bar{p}_\infty^{1/2}\right)\frac{p_0}{\bar{p}_\infty^{1/2}}dx\right|$$
$$\leq\left(\int\frac{(p_0)^2}{\bar{p}_\infty}dx\right)^{1/2}\left(\int\left(u(t)-\int u_0\bar{p}_\infty dy\right)^2\bar{p}_\infty dx\right)^{1/2}$$
$$\leq C_0 e^{-\lambda t}\left(\int(u_0-\int u_0\bar{p}_\infty)^2\bar{p}_\infty dx\right)^{1/2}$$
$$= C_0 e^{-\lambda t}\left(\int(u_0)^2\bar{p}_\infty dx-(\int u_0\bar{p}_\infty dx)^2\right)^{1/2}\leq C_0 e^{-\lambda t}\left(\int(u_0)^2\bar{p}_\infty dx\right)^{1/2}$$

where the inequality in the second line follows from the Cauchy-Schwarz inequality and the inequality in the third line follows from (60) and (68).

Dividing by $\left(\int u_0^2\bar{p}_\infty dx\right)^{1/2}$,

$$\frac{\left|\int u_0\left(p(t)-J\bar{p}_\infty\right)dx\right|}{\left(\int u_0^2\bar{p}_\infty dx\right)^{1/2}}\leq C_0 e^{-\lambda t}.$$

We next optimize on u_0 to extract, so to speak, the maximum information from this inequality. The maximum of the left-hand side is attained at $u_0=(p(t)-J\bar{p}_\infty)/\bar{p}_\infty$.[77] Hence, applying the above inequality to $u_0=(p(t)-J\bar{p}_\infty)/\bar{p}_\infty$,

$$\left(\int\frac{(p(t)-J\bar{p}_\infty)^2}{\bar{p}_\infty}dx\right)^{1/2}=\frac{\left|\int u_0\left(p(t)-J\bar{p}_\infty\right)dx\right|}{\left(\int u_0^2\bar{p}_\infty dx\right)^{1/2}}\leq C_0 e^{-\lambda t}.$$

[77] To see this, note that the optimal u_0 is only determined up to a constant. Therefore, the problem is equivalent to

$$\max_{u_0}\int u_0\left(p(t)-J\bar{p}_\infty\right)dx\quad\text{s.t.}\quad\left(\int u_0^2\bar{p}_\infty dx\right)^{1/2}=1$$

The Lagrangian is $\mathcal{L}=\int u_0\left(p(t)-\bar{p}_\infty\right)dx+\nu\left(1-\left(\int u_0^2\bar{p}_\infty dx\right)^{1/2}\right)$ with first-order condition $p(t)-\bar{p}_\infty=\nu\left(\int u_0^2\bar{p}_\infty dx\right)^{-1/2}u_0\bar{p}_\infty$ and so u_0 is proportional to $(p(t)-\bar{p}_\infty)/\bar{p}_\infty$. Technically, we still need to show that for our choice of u_0, $\int u_0^2\bar{p}_\infty dx<\infty$. This is delayed till when we prove the more general case of Proposition 2.

Finally, using the Cauchy-Schwarz inequality

$$\int |p(t) - J\bar{p}_\infty|\, dx \le \left(\int \frac{(p(t) - J\bar{p}_\infty)^2}{\bar{p}_\infty} dx \right)^{1/2} \underbrace{\left(\int \bar{p}_\infty dx \right)^{1/2}}_{=1} \le C_0 e^{-\lambda t},$$

which is the desired result.

To show that the convergence rate λ is generically attained, we note that given an initial p_0, we can form $\tilde{p}_0$ by $\tilde{p}_0(x) = \partial_x p_0 - \frac{2\mu}{\sigma^2} p_0$ so that $\tilde{p}_0(0) = 0$. Then extend $\tilde{p}_0(x)$ to the real line by reflecting $\tilde{p}_0$ with respect to the origin and using the boundary condition $\tilde{p}_0(0) = 0$. Then we find the solution $\tilde{p}$ to the equation $\tilde{p}_t = \frac{\sigma^2}{2} \frac{\partial^2 \tilde{p}}{\partial x^2} - \mu \frac{\partial \tilde{p}}{\partial x}$ on the real line with initial condition $\tilde{p}_0$.

Now it can be verified that if $\tilde{p}_t = \frac{\sigma^2}{2} \frac{\partial^2 \tilde{p}}{\partial x^2} - \mu \frac{\partial \tilde{p}}{\partial x}$ and $\frac{\sigma^2}{2} \frac{\partial^2 p_\infty}{\partial x^2} - \mu \frac{\partial p_\infty}{\partial x} = 0$, then $q(x,t) = \tilde{p}(x,t) p_\infty^{-1/2}$ (with $p_\infty^{-1/2}$ extended to $\mathbb{R}$ as an even function) satisfies

$$q_t - \frac{\sigma^2}{2} q_{xx} + \frac{\mu^2}{2\sigma^2} q = 0$$

the Kolmogorov Forward equation for a Brownian motion with death rate $\frac{\mu^2}{2\sigma^2}$. The solution to the above equation on $\mathbb{R}$ is given by $q(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-y)^2}{2\sigma^2 t} - \frac{\mu^2}{2\sigma^2} t} q_0(y) dy$, where $q_0 = \tilde{p}_0(x) p_\infty^{-1/2}$, which is an odd function by construction.

The crucial property of $q(x,t)$ is that $q(x,t) = -q(-x,t)$, i.e. q is an odd function. Thus in fact $\tilde{p}(x,t)$ restricted to $x \ge 0$ gives the solves $\mathcal{A}^* \tilde{p} = \tilde{p}_t$ with boundary condition $\tilde{p}(0,t) = 0$ and initial condition $\tilde{p}_0$. Now it is a standard result that the L^1 norm of solution to the heat equation is conserved. Thus

$$\begin{aligned}
\int_0^\infty |q(x,t)| dx &= \frac{1}{2} \int_{-\infty}^{\infty} |q(x,t)| dx \\
&= \frac{1}{2} \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-y)^2}{2\sigma^2 t} - \frac{\mu^2}{2\sigma^2} t} q_0(y) dy \right| dx \\
&= \frac{1}{2} e^{-\frac{\mu^2}{2\sigma^2} t} \int_{-\infty}^{\infty} \left| \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-y)^2}{2\sigma^2 t}} q_0(y) dy \right| dx \\
&= \frac{1}{2} e^{-\frac{\mu^2}{2\sigma^2} t} \int_{-\infty}^{\infty} |q_0(x)| dx = \frac{1}{4} e^{-\frac{\mu^2}{2\sigma^2} t} \int_0^\infty q_0(x) dx
\end{aligned}$$

and we can conclude that $\lim_{t\to\infty} -\frac{1}{t} \log \int_0^\infty |q(x,t)| dx = \frac{\mu^2}{2\sigma^2}$. Finally, $\int_0^\infty |\tilde{p}(x,t)| p_\infty^{-1/2} dx = \int_0^\infty |q(x,t)| dx$, and if we choose p_0 with mass concentrated near the origin and uniformly bounded above by 1, then $\int_0^\infty |\tilde{p}(x,t)| dx \ge C \int_0^\infty |\tilde{p}(x,t)| p_\infty^{-1/2} dx$, so that $\lim_{t\to\infty} -\frac{1}{t} \log \int_0^\infty |\tilde{p}(x,t)| dx \le$

$\frac{\mu^2}{2\sigma^2}$.

Thus given f uniformly bounded by 1 with mass concentrated near the origin, we can form $\eta(x) = (\partial_x f - \frac{2\mu}{\sigma^2} f)$, and given any initial distribution p_0, we can perturb p_0 by $\varepsilon\eta(x)$, and $p_0 + \varepsilon\eta(x)$ will have rate of convergence at most $\frac{\mu^2}{2\sigma^2}$.

This concludes the proof of (62), and of Proposition 1 for the case $\delta = 0$.

Case with $\delta > 0$. Define an operator: $\mathcal{C}^* p := -\mu p_x + \frac{\sigma^2}{2} p_{xx}$, so that $\mathcal{A}^* = \mathcal{C}^* - \delta$. Suppose an initial condition $p_0(x)$. Given $p_t = \mathcal{A}^* p + \delta\psi$ and $0 = \mathcal{A}^* p_\infty + \delta\psi$, the difference $q := p - p_\infty$ satisfies $q_t = \mathcal{A}^* q = \mathcal{C}^* q - \delta q$. Next, define $Q(x) := q(x) e^{\delta t}$. This gives: $Q_t = \mathcal{C}^* Q$. But operator $\mathcal{C}^*$ has no "death" rate, and Q has the right boundary condition, so that the previous case ($\delta = 0$) applies to Q. Note also that $\int Q(x)\, dx = 0$. Then, equation (62) gives: $\int |Q(t)|\, dx \le C_0 e^{-\lambda t}$, i.e.

$$\int |p(x,t) - p_\infty(x)|\, dx \le C_0 e^{-(\lambda+\delta)t}.$$

This concludes the proof of Proposition 1 in the case $\delta > 0$.

Comment on the use of Assumption 1: Note that Assumption 1 is only used for one particular step, namely to go from the spectral decomposition of the operator $\mathcal{B}$ to the asymptotic behavior of the L^1-norm $||p - p_\infty||$. This really comes from the fact that a spectral decomposition is an "L^2-statement" which we then convert into an "L^1-statement" (this is also why a power 2 shows up). Put differently, the main step in the proof – that the spectral gap of $\mathcal{B}$ is λ – is true regardless of whether Assumption 1 holds. The only step that relies on Assumption 1 is converting this into a statement about the rate of convergence of the L^1-norm.

F.2.2 Proof of Lemma 4

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

$$r_t = -\mu r_x + \frac{\sigma^2}{2} r_{xx}. \tag{69}$$

Note further that

$$\int_{-\infty}^{+\infty} |r(x,t)| dx = \int_{-\infty}^{+\infty} |\tilde{r}(x,t)| dx \tag{70}$$

where $\tilde{r}(x,t) = r(x+\mu t, t)$. Note that this works only because the limits of integration are $\pm\infty$ and hence one can simply "translate everything" by μt, i.e. it does not work with a lower bound. From (69), and using that $\tilde{r}_t = r_t + \mu r_x$, $\tilde{r}$ solves the standard heat equation

$$\tilde{r}_t = \frac{\sigma^2}{2}\tilde{r}_{xx}. \tag{71}$$

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

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

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

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

i.e. $e^{\beta t}\tilde{r}(x,t) \in L^1(\mathbb{R}_+ \times \mathbb{R})$. Hence $R(x) := \int_0^\infty e^{\beta t}\tilde{r}(x,t)\, dt$ is defined and $R(\cdot) \in L^1$. Using the heat equation (71) (normalizing without loss of generality $\sigma^2 = 2$), we have

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

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

$$\hat{R}(\xi)\left(-|\xi|^2 + \beta\right) = -\widehat{\tilde{r}}_0(\xi). \tag{72}$$

This implies that $\widehat{\tilde{r}}_0(\xi) = 0$ for $\xi = \beta^{1/2}$ (we just plug the value $\xi = \beta^{1/2}$ in (72)).

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

G Proof of Proposition 2

Xavier Gabaix, Jean-Michel Lasry, Pierre-Louis Lions, Benjamin Moll and Zhaonan Qu
August 9, 2016

Due to space constraints, we here only outline a sketch of the proof of Proposition 2. For the complete proof, see Appendix J on the authors' websites.

G.1 Existence and uniqueness of the steady state with non-constant coefficients

We expand on the indications given in the text for the existence and uniqueness of a steady state for the process (14) with non-constant coefficients: $dx_{it} = \mu(x_{it}, t)dt + \sigma(x_{it}, t)dZ_{it}$, with rebirth after death. Here are simple sufficient conditions. Assumption 2 imposes that $\mu(x,t) \to \tilde{\mu}(x)$ and $\sigma(x,t) \to \tilde{\sigma}(x)$ uniformly in x as $t \to \infty$ and that $\tilde{\mu}(x) \to \bar{\mu}$ and $\tilde{\sigma}(x) \to \bar{\sigma}$ as $x \to \infty$. Naturally, only the asymptotic coefficients $\tilde{\mu}(x)$ and $\tilde{\sigma}(x)$ are relevant for the stationary distribution. We assume that $\tilde{\sigma}(x)$ and $\tilde{\mu}(x)$ are continuous, bounded, and that $\tilde{\sigma}(x) \geq \underline{\sigma}$ for some $\underline{\sigma} > 0$ (this ensures that the process is always remixed, and does not get stuck in a zone of 0 volatility), and (i) in the ergodic case: $\bar{\mu} < 0$ (ii) in the non-ergodic case: there are positive constants C_0, C_1 such that $|\tilde{\mu}(x)| \leq C_0 + C_1 |x|$ for $x \to \infty$, $\delta > 0$ (this is so that the large incomes get killed faster than they "escape to infinity").

Ergodic case. A sufficient condition is that $\bar{\mu} < 0$ (in the general case, that $\tilde{\mu}(x) \leq C_0$ with $C_0 < 0$). A necessary and sufficient condition is that, starting from a point $x_0 > 1$, the expected time to reach $x = 1$ is finite. This is developed in Meyn and Tweedie (2009) and Lions (2014).

Non-ergodic case. A sufficient condition is that there are positive constants C_0, C_1 $|\tilde{\mu}(x)| \leq C_0 + C_1 |x|$ for $x \to \infty$. This implies that operator $-\mathcal{A}^*$ (with $\mathcal{A}^* p = \delta p - (\tilde{\mu} p)' + \left(\frac{\tilde{\sigma}^2}{2} p\right)''$) is "accretive" (which is a form of generalization of monotonicity): for all p, all $\lambda > 0$, $\|(\lambda I - \mathcal{A}^*) p\| \geq \lambda \|p\|$. See Kato (1970), https://en.wikipedia.org/wiki/Dissipative_operator, and https://fr.wikipedia.org/wiki/Opérateur_accrétif.

G.2 Sketch of Proof of Proposition 2: Case without lower bound ("non-ergodic")

In Section J.1 we prove Proposition 2 in the case without lower bound, i.e. non-ergodic. We use a translation-at-infinity argument to ensure that the process stays at large spatial domain, so that the coefficients are essentially constant. Then we can apply results from

Proposition 1 to conclude that there exists a particular q_* whose rate of convergence is exactly δ. Finally, we conclude the proof by showing that a small perturbation of any initial condition by q_* yields a speed of convergence that is exactly δ.

G.3 Sketch of Proof of Proposition 2: Case with lower bound ("ergodic")

Section J.2 is dedicated to the proof in the ergodic case with a lower bound. The techniques employed there are based on energy methods that are different from the proof of Proposition 1. Thus for simplicity, we first demonstrate the proof in the case of constant coefficients. The generalization to space dependent but time-independent coefficients is then through a "translation at infinity argument", similar to the one employed in Section J.1, to essentially reduce the problem to the constant-coefficient case. The generalization to time- and space-dependent coefficients requires some new ideas, but the essence is still approximation by the time-independent coefficient case and bounding the error effectively. This we carry out in Section J.2.6.

G.3.1 Setting the stage: a unified one-parameter model with a lower bound

Section J.2.1 sets the stage for the energy method for constant coefficients by establishing a unified one-parameter model that encompasses both the model with exit and reinjection and the model with a reflecting barrier. More precisely, we let $\mathcal{A}^*$ be the operator

$$\mathcal{A}^* p := \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2} - \mu\frac{\partial p}{\partial x} + \frac{\sigma^2}{2}\frac{1-\theta}{\theta}p(0)\rho(x)$$

with $\mu < 0$ and boundary condition $\frac{\sigma^2}{2}p_x(0) - \mu p(0) = \frac{\sigma^2}{2}\frac{1-\theta}{\theta}p(0)$ where we recall $\rho(x)$ is the distribution of the reinjection point following exit, subject to Assumption 4. When $\theta = 0$ (taking $\theta \to 0$ first) we recover the model with exit and reinjection, and when $\theta = 1$ we obtain the model with reflecting barrier. We show that $\mathcal{A}^*$ has a unique invariant distribution p_∞: $\mathcal{A}^* p_\infty = 0$ and p_∞ can be rescaled to be a probability distribution. Moreover, $p_\infty(x) \sim C_\theta e^{\frac{2\mu}{\sigma^2}x}$.

Corresponding to $\mathcal{A}^*$ we have its adjoint operator $\mathcal{A}$:

$$\mathcal{A}u := \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}u + \mu\frac{\partial}{\partial x}u$$

where $\mu < 0$ with boundary condition $-\theta u_x(0) + (1-\theta)\left(u(0) - \int u(x)\rho(x)dx\right) = 0$. Intuitively, this boundary condition describes the following behavior: if the process ever reaches

$x = 0$, then, with probability θ, the process is reflected; and with probability $1 - \theta$, the process jumps to some $x > 0$, drawn from the distribution $\rho(x)$.

G.3.2 Proof Strategy for Proposition 2

When $\theta < 1$, i.e. when we depart from the pure reflection case, one can no longer construct a self-adjoint transformation $\mathcal{B}$ of $\mathcal{A}$ as in the proof of Proposition 1. Therefore, it is no longer possible to obtain an explicit formula for the spectral gap of the operator $\mathcal{A}$. We instead follow an alternative approach that works directly with the operator $\mathcal{A}$ using "energy methods" (i.e. techniques techniques involving L^2-norms of various expressions – see Evans (1998) for their usefulness in other applications).

The proof of Proposition 2 has three parts. The first part proves that the cross-sectional income distribution converges to its stationary distribution exponentially at *some* rate $\lambda > 0$. The second part is to prove that this rate λ satisfies $\lambda \leq \frac{\mu^2}{2\sigma^2}$. The third part simply concludes the proof by combining the two previous parts.

G.3.3 Part 1: exponential convergence to stationary distribution

The first parts proves that the cross-sectional income distribution converges to its stationary distribution exponentially at some rate $\lambda > 0$, i.e.

$$\int_0^\infty |p(x,t) - p_\infty(x)| dx \leq e^{-\lambda t} \int_0^\infty |p_0(x,t) - p_\infty(x)| dx$$

for some $\lambda > 0$. This is proved in Lemmas 14 and 15 in Section J.2.3.

Lemma 14, which is the heart of the matter, establishes a Poincaré-like energy inequality. More precisely, let $\partial_t u = \mathcal{A}u$, and $\int p_\infty u_0 dx = 0$. For $\theta \in (0,1]$, let

$$\lambda := \frac{1}{2} \inf_u \Big\{ \sigma^2 \int u_x^2 p_\infty dx + \frac{\sigma^2}{2} \frac{1-\theta}{\theta} p_\infty(0) \left[\left(u(0) - \int u\rho dy \right)^2 + \int \left(u - \int u \rho dy \right)^2 \rho dx \right]$$
$$\text{s.t.} \quad \int u^2 p_\infty dx = 1, \qquad \int u p_\infty dx = 0 \Big\}.$$

and when $\theta = 0$, replace $\frac{\sigma^2}{2}\frac{1-\theta}{\theta} p_\infty(0)$ with $\frac{\sigma^2}{2}(p_\infty)_x$.

Then
$$\int u(x,t)^2 p_\infty(x) dx \leq e^{-2\lambda t} \int u_0(x)^2 p_\infty(x) dx \ .$$

Next, Lemma 15 establishes the bridge from this energy inequality to the exponential convergence of the cross-sectional income distribution:

$$\int_0^\infty |p(x,t) - p_\infty(x)|dx \leq e^{-\lambda t} \int_0^\infty |p_0(x,t) - p_\infty(x)|dx.$$

G.3.4 Part 2: the rate of convergence cannot be larger than $\frac{\mu^2}{2\sigma^2}$

The second part of the proof of Proposition 2 is to show that this λ defined in Lemma 14 satisfies $\lambda \leq \frac{\mu^2}{2\sigma^2}$, and that for arbitrarily small perturbations of any initial distribution, the bound $\frac{\mu^2}{2\sigma^2}$ is indeed achieved. This is done in Lemma 16 in Section J.2.4. The proof of Lemma 16 features the translation-at-infinity argument that brings the PDE to arbitrarily large spatial domain to obtain a Kolmogorov Forward equation for a Brownian motion, whose solution can be explicitly represented by a convolution with the heat kernel multiplied by $e^{-\frac{\mu^2}{2\sigma^2}t}$. Then through explicit calculations a contradiction is obtained when we assume $\lambda > \frac{\mu^2}{2\sigma^2}$.

G.3.5 Part 3: conclusion of proof of Prop. 2

Combining Lemmas 15 and 16, we see that when $\delta = 0$ and $\mu < 0$,

$$-\lim_{t\to\infty} \frac{1}{t} \log \int_0^\infty |p(x,t) - p_\infty(x)|dx \geq \lambda \quad \text{with} \quad \lambda \leq \frac{\mu^2}{2\sigma^2}.$$

Appendix J.2.5 further shows that the rate of convergence $\lambda = \frac{\mu^2}{2\sigma^2}$ to p_∞ is generically attained.

G.3.6 Extension of Prop. 2 to income-dependent coefficients $\mu(x,t)$ and $\sigma(x,t)$

Section J.2.6 extends the proof to the case of time and space dependent coefficients $\mu(x,t)$ and $\sigma(x,t)$. This is done by extending Lemmas 14, 15, and 16. In this case the operators $\mathcal{A}$ and $\mathcal{A}^*$ are time dependent:

$$\mathcal{A}^*(t)p(x) = \left(\frac{\sigma^2(x,t)}{2}p\right)_{xx} - (\mu(x,t)p)_x + \frac{\sigma^2(0,t)}{2}\frac{1-\theta}{\theta}p(0)\rho(x)$$

with the time-dependent boundary condition $\left(\frac{\sigma^2(x,t)}{2}p\right)_x(0) - \mu(0,t)p(0) = \frac{\sigma^2(0,t)}{2}\frac{1-\theta}{\theta}p(0)$, and

$$\mathcal{A}(t)u(x) = \frac{\sigma^2(x,t)}{2}\frac{\partial^2}{\partial x^2}u + \mu(x,t)\frac{\partial}{\partial x}u$$

with boundary condition $-\theta u_x(0) + (1-\theta)\left(u(0) - \int u(x)\rho(x)dx\right) = 0$.

H Proofs of Propositions 3 –5

H.1 Proof of Proposition 3

Equations (19) to (21) follow from integrating (18). The speed of convergence

$$-\lim_{t\to\infty} \frac{1}{t} \log ||p(x,t) - p_\infty(x)||_\xi = \lambda(\xi) \tag{73}$$

is obtained in two steps. The first is to show that the rate of convergence of the weighted L^1-norm is at least $\lambda(\xi)$. The second is to show that it is at most $\lambda(\xi)$. For the first step, define $q(x,t) := p(x,t) - p_\infty(x)$ and note that from Lemma 2[78], $|q(x,t)|$ is a subsolution of the same equation as $q(x,t)$:

$$|q|_t \le -\mu|q|_x + \frac{\sigma^2}{2}|q|_{xx} - \delta|q| + \phi\left[|q| * f - |q|\right]$$

Applying the Laplace transform $\widehat{|q|}(\xi,t) := \int_{-\infty}^{\infty} e^{-\xi x}|q(x,t)|dx$:

$$\frac{\partial\widehat{|q|}(\xi,t)}{\partial t} \le -\lambda(\xi)\widehat{|q|}(\xi,t), \qquad \lambda(\xi) = \xi\mu - \xi^2\frac{\sigma^2}{2} + \delta - \phi(\widehat{f}(\xi) - 1).$$

Therefore, by Grönwall's lemma, $\widehat{|q|}(\xi,t) \le e^{-\lambda(\xi)t}\widehat{|q_0|}(\xi)$ or equivalently

$$\int_{-\infty}^{\infty} e^{-\xi x}|p(x,t) - p_\infty(x)|dx \le e^{-\lambda(\xi)t}\widehat{|q_0|}(\xi) = Ae^{-\lambda(\xi)t} \tag{74}$$

where $A > 0$. This proves that the rate of convergence is at least $\lambda(\xi)$. For the second step

$$\begin{aligned}\int_{-\infty}^{\infty} |p(x,t) - p_\infty(x)|e^{-\xi x}dx \ge \left|\int_{-\infty}^{\infty} (p(x,t) - p_\infty(x))e^{-\xi x}dx\right| = |\widehat{p}(\xi,t) - \widehat{p}_\infty(\xi)| \\ = e^{-\lambda(\xi)t}\left|\widehat{p}_0(\xi) - \widehat{p}_\infty(\xi)\right| = ae^{-\lambda(\xi)t},\end{aligned} \tag{75}$$

where $a > 0$, i.e. the rate of convergence is at most $\lambda(\xi)$. (74) and (75) imply (73). Finally the Laplace transform of f satisfies $\widehat{f}(\xi) = \mathbb{E}[e^{-\xi g}] \ge e^{-\xi\mathbb{E}[g]} \ge 1$ for $\xi < 0$ where the weak inequality follows from Jensen's inequality and the assumption that $\mathbb{E}[g] \ge 0$.□

[78] There is a difference here since in the equation satisfied by q, there is an extra term $\phi\left[q * f - q\right]$ involving jump, but $\phi(|q| * f) \ge 0$, so the proof of Lemma 2 goes through.

H.2 Proof of Proposition 4

We have $||p(x,t) - p_\infty(x)||_\xi = ||\tilde{p}(x,t) - \tilde{p}_\infty(x)||$ where $\tilde{p}\,(x,t) := e^{-\xi x} p\,(x,t)$. Next, since $p(x,t)$ satisfies (5), simple calculations show that $\tilde{p}\,(x,t)$ satisfies

$$\tilde{p}_t = -\tilde{\mu}(\xi)\tilde{p}_x + \frac{\sigma^2}{2}\tilde{p}_{xx} - \tilde{\delta}(\xi)\tilde{p} + \delta\tilde{\psi}(x), \qquad \tilde{\mu}(\xi) := \mu - \xi\sigma^2, \quad \tilde{\delta}(\xi) := \mu\xi - \frac{\sigma^2}{2}\xi^2 + \delta$$

and where $\tilde{\psi}(x) := e^{-\xi x}\psi(x)$. Similarly, the new surrogate state $\tilde{\bar{p}}_\infty(x) \sim e^{-\tilde{\zeta}x}$ where $\tilde{\zeta} = \bar{\zeta} + 2\xi$. Hence Assumption 1 implies

$$\int_{-\infty}^{\infty} \frac{(\tilde{p}_0(x))^2}{e^{-\tilde{\zeta}x}}dx = \int_{-\infty}^{\infty} \frac{(p_0(x))^2 e^{-2\xi x}}{e^{-\bar{\zeta}x}e^{-2\xi x}}dx < \infty,$$

i.e. an identical assumption in terms of $\tilde{p}$. But from Proposition 1 we know that the speed of convergence of $||\tilde{p}(x,t) - \tilde{p}_\infty(x)||$ is $\lambda\,(\xi) = \frac{1}{2}\frac{\tilde{\mu}(\xi)^2}{\sigma^2}\mathbf{1}_{\{\tilde{\mu}(\xi)\leq 0\}} + \tilde{\delta}\,(\xi)$. Substituting in for $\tilde{\delta}(\xi)$ and $\tilde{\mu}(\xi)$, we obtain (22).

H.3 Proof of Proposition 5

First consider the stationary Laplace transforms which satisfy the stationary version of (29) and (30)

$$0 = -\lambda_H(\xi)\widehat{p}^H(\xi) + \beta_H, \qquad 0 = -\lambda_L(\xi)\widehat{p}^L(\xi) + \alpha\widehat{p}^H(\xi) + \beta_L$$

with solution

$$\widehat{p}^H_\infty(\xi) = \frac{\beta_H}{\lambda_H(\xi)}, \qquad \widehat{p}^L_\infty(\xi) = \frac{\beta_L + \alpha\widehat{p}^H(\xi)}{\lambda_L(\xi)} = \frac{\beta_L}{\lambda_L(\xi)} + \alpha\frac{\beta_H}{\lambda_L(\xi)\lambda_H(\xi)}.$$

Adding these two expressions we obtain

$$\widehat{p}_\infty(\xi) = \frac{\beta_H}{\lambda_H(\xi)} + \frac{\beta_L}{\lambda_L(\xi)} + \alpha\frac{\beta_H}{\lambda_L(\xi)\lambda_H(\xi)}. \tag{76}$$

As explained in section 4.2 (particularly footnote 37), the tail exponent of the distribution $p_\infty(x)$ equals $\zeta = -\inf\{\xi : \widehat{p}_\infty(\xi) < \infty\}$, i.e. it is the critical value of ζ such that $\widehat{p}_\infty(\xi)$ ceases to exist for $\xi < -\zeta$. Examining (76) we see that ζ is as asserted in the Proposition.

Next consider the dynamics of the Laplace transform (29) and (30). Defining $\widehat{q}^j = \widehat{p}^j - \widehat{p}^j_\infty$

$$\begin{aligned} \widehat{q}^H_t(\xi,t) &= -\lambda_H(\xi)\widehat{q}^H(\xi,t), \\ \widehat{q}^L_t(\xi,t) &= -\lambda_L(\xi)\widehat{q}^L(\xi,t) + \alpha\widehat{q}^H(\xi,t), \end{aligned} \tag{77}$$

with solution

$$\widehat{q}^H(\xi,t) = e^{-\lambda_H(\xi)t}\widehat{q}_0^H(\xi),$$
$$\widehat{q}^L(\xi,t) = e^{-\lambda_L(\xi)t}\widehat{q}_0^L(\xi) + \frac{\alpha}{\lambda_H(\xi)-\lambda_L(\xi)}\left(e^{-\lambda_L(\xi)t} - e^{-\lambda_H(\xi)t}\right)\widehat{q}_0^H(\xi). \tag{78}$$

Summing these two equations and collecting terms, we obtain the expression for the cross-sectional distribution $\widehat{p}(\xi,t) = \widehat{p}^L(\xi,t)+\widehat{p}^H(\xi,t)$ in (32). The constants of integration referred to in the Proposition are given by:

$$c_H(\xi) := \frac{\lambda_H(\xi)-\lambda_L(\xi)-\alpha}{\lambda_H(\xi)-\lambda_L(\xi)}(\widehat{p}_0^H(\xi) - \widehat{p}_\infty^H(\xi)),$$
$$c_L(\xi) := (\widehat{p}_0^L(\xi) - \widehat{p}_\infty^L(\xi)) + \frac{\alpha}{\lambda_H(\xi)-\lambda_L(\xi)}(\widehat{p}_0^H(\xi) - \widehat{p}_\infty^H(\xi)).\square$$

I Complements to Section 5

I.1 Stationary Distribution of the Augmented Random Growth Model

We now provide sufficient conditions under which (26) has a unique stationary distribution. More precisely, we allow for time variation in the stochastic process capturing scale-dependence χ_t, and we provide conditions under which there exists a stationary distribution for the process for y_{it}, which we restate here for the reader's convenience:

$$dy_{it} = \mu_j dt + \sigma_j dZ_{it} + g_{jit}dN_{jit} + \text{Injection} - \text{Death}. \tag{79}$$

In this case, there will be a unique stationary distribution for $x_{it} = \chi_t^{b_j} y_{it}$ if 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.

Proposition 10 *Assume that $\delta > 0, \alpha_{j,k} \geq 0$ for all $j \neq k$, there is no lower bound on income and that Assumption 3 holds. Then there exists a unique vector of stationary type-specific distributions $p_\infty^j(y), j = 1, ..., J$ and therefore also an overall stationary distribution $p_\infty(y) = \sum_{j=1}^J p_\infty^j(y)$. This stationary distribution has a Pareto tail $p_\infty(y) \sim e^{-\zeta y}$ for large y with $\zeta > 0$. If $\alpha_{j,k} = 0, j > k$ or $\alpha_{j,k} = 0, j < k$ (analogously to the triangular case in Section 5.2), then $\zeta = \min\{\zeta_1, ..., \zeta_J\}$ and where for each $j = 1, ..., J$, ζ_j is the unique positive root* *of*

$$0 = \zeta^2 \frac{\sigma_j^2}{2} + \zeta\mu_j + \phi_j(\widehat{f}_j(-\zeta) - 1) - \sum_{k\neq j} \alpha_{j,k} - \delta. \tag{80}$$

Proof: The Kolmogorov Forward equation corresponding to (79) is

$$p_t^j = -\mu_j p_y^j + \frac{\sigma_j^2}{2} p_{yy}^j + \phi_j \mathbb{E}_j[p^j(y-g) - p^j(y)] - \sum_{k\neq j} \alpha_{j,k} p^j + \sum_{k\neq j} \alpha_{k,j} p^k - \delta p^j + \delta\theta_j \psi \tag{81}$$

for $j = 1, ..., J$. The Laplace transform of (81) is given by

$$\widehat{p}_t^j = -\lambda_j(\xi)\widehat{p}^j + \sum_{k\neq j} \alpha_{k,j} \widehat{p}^k + \delta\theta_j \widehat{\psi} \tag{82}$$

$$\lambda_j(\xi) := \xi\mu_j - \xi^2 \frac{\sigma_j^2}{2} - \phi_j(\widehat{f}_j(\xi) - 1) + \sum_{k\neq j} \alpha_{j,k} + \delta \tag{83}$$

or in matrix notation

$$\widehat{\mathbf{p}}_t = -\mathbf{M}(\xi)\widehat{\mathbf{p}} + \delta\widehat{\psi}\theta \tag{84}$$

where $\widehat{\mathbf{p}} = (\widehat{p}_1(\xi, t), ..., \widehat{p}_J(\xi, t))^T$, $\mathbf{M}(\xi)$ is a $J \times J$ with diagonal entries $\lambda_1(\xi), ..., \lambda_J(\xi)$ and off-diagonal entries $-\alpha_{j,k}$ for $j \neq k$, and $\theta = (\theta_1, ..., \theta_J)^T$. When $\xi = 0$, $\lambda_j(0) = \sum_{k\neq j} \alpha_{j,k} + \delta$. Therefore, for ξ sufficiently close to zero, $\mathbf{M}(\xi)$ is strictly diagonally dominant. Strictly diagonally dominant matrices are invertible. Therefore, for ξ sufficiently close to zero, there is a unique stationary Laplace transform that solves

$$\widehat{\mathbf{p}}_\infty = \mathbf{M}(\xi)^{-1}(\delta\widehat{\psi}\theta).$$

Since there is a unique vector of stationary type-specific Laplace transforms $\widehat{p}_\infty^j, j = 1, ..., J$ for ξ sufficiently close to zero, there is also a unique vector of type-specific stationary distributions $p_\infty^j(y), j = 1, ..., J$.

As we take ξ more and more negative, the matrix $\mathbf{M}(\xi)$ ceases to be invertible and hence the Laplace transform $\widehat{\mathbf{p}}_\infty$ ceases to exist. That is, $\widehat{\mathbf{p}}_\infty$ has a finite negative abscissa of convergence ξ^*. If additionally ξ^* is a pole of $\widehat{\mathbf{p}}_\infty$, then it follows from the Tauberian result in Proposition 7 that the stationary distribution $\mathbf{p}_\infty$ has a Pareto tail with tail parameter $\zeta = -\xi^*$. In this case, $\zeta = -\inf\{\xi : \det(\mathbf{M}(\xi)) \neq 0\}$. More can be said in the case where $\alpha_{j,k} = 0, j > k$ or $\alpha_{j,k} = 0, j < k$ and hence $\mathbf{M}(\xi)$ is triangular. In this case

$$\det(\mathbf{M}(\xi)) = \lambda_1(\xi) \times \lambda_2(\xi) \times ... \times \lambda_J(\xi).$$

Therefore let $\xi^* = \inf\{\xi : \lambda_j(\xi) > 0, \text{all } j\}$. Then ξ^* is indeed a pole and hence the $\mathbf{p}_\infty$ has a Pareto tail with tail parameter $\zeta = -\xi^*$. Equivalently, $\zeta = \min\{\zeta_1, ..., \zeta_J\}$ and where for each $j = 1, ..., J$, ζ_j is the unique positive root of (80). That this equation has a unique positive root can be shown in exactly the same way as in the proof of Proposition 8 in Appendix D.□

I.2 Speed of Convergence with Type Dependence and J Types

We here provide an informal discussion how the model with type dependence and $J \geq 2$ types can be analyzed. Following exactly the same steps as in the proof of Proposition 10, one can show that the vector $\widehat{\mathbf{p}} = (\widehat{p}_1(\xi, t), ..., \widehat{p}_J(\xi, t))^T$ satisfies (84). Equivalently, $\widehat{\mathbf{q}} := \widehat{\mathbf{p}} - \widehat{\mathbf{p}}_\infty$ satisfies

$$\widehat{\mathbf{q}}_t = -\mathbf{M}(\xi)\widehat{\mathbf{q}} \tag{85}$$

This is a simple system of J ordinary differential equations that can be analyzed using standard methods, in particular by diagonalizing the matrix $\mathbf{M}(\xi)$. More can be said in special cases. In the triangular case $\alpha_{j,k} = 0, j > k$ or $\alpha_{j,k} = 0, j < k$, the eigenvalues are simply given by $\lambda_1(\xi), ..., \lambda_J(\xi)$. Since all of these are strictly positive, the steady state Laplace transform $\widehat{\mathbf{p}}_\infty$ is globally stable and the asymptotic speed of convergence is given by the eigenvalue that is closest to zero. In fact, we have assumed in the previous section that $\mathbf{M}(\xi)$ is invertible for the stationary Laplace transform $\widehat{\mathbf{p}}_\infty$ to exist. Since $\mathbf{M}(\xi)$ is invertible it must also be positive definite, as none of its eigenvalues can cross 0 and thus must remain positive for the range of ξ we consider.

I.3 A Microfoundation for "Scale Dependence"

Here we provide an example of a micro foundation for scale dependence. It is based on a dynamic generalization of the Gabaix and Landier (2008) model.

I.3.1 Summary of the Static Gabaix and Landier (2008) Model

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

We consider the problem faced by one particular firm. The firm has a "baseline" value of S. At $t = 0$, it hires a manager of talent T for one period. The manager's talent increases the firm's value according to

$$S' = S + CTS^{\gamma}, \tag{86}$$

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

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

$$\max_m CS(n)^{\gamma} T(m) - w(m).$$

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

$$w(n) = -\int_n^N CS(u)^{\gamma} T'(u)\, du + \underline{w}_N. \tag{87}$$

Specific functional forms are required to proceed further. We assume a Pareto firm size distribution with exponent $1/\alpha$:

$$S(n) = An^{-\alpha} \tag{88}$$

Using results from extreme value theory, GL use the following asymptotic value for the spacings of the talent distribution: $T'(n) = -Bn^{\beta-1}$. These functional forms give the wage equation in closed form, taking the limit as $n/N \to 0$:

$$w(n) = \int_n^N A^{\gamma} BCu^{-\alpha\gamma+\beta-1} du + \underline{w} = \frac{A^{\gamma}BC}{\alpha\gamma - \beta}\left[n^{-(\alpha\gamma-\beta)} - N^{-(\alpha\gamma-\beta)}\right] + \underline{w}_N \sim \frac{A^{\gamma}BC}{\alpha\gamma-\beta} n^{-(\alpha\gamma-\beta)}. \tag{89}$$

Therefore using (88), the equilibrium pay of a CEO of talent rank n is

$$w(n) = Gn^{-\chi}, \qquad \chi := \alpha\gamma - \beta, \qquad G = \frac{A^{\gamma}BC}{\alpha\gamma - \beta} \tag{90}$$

To interpret equation (89), we consider a reference firm, for instance firm number 250 – the median firm in the universe of the top 500 firms. Denote its index n_*, and its size $S(n_*)$.

We obtain Proposition 2 from GL, which we repeat here. In equilibrium, manager n runs a firm of size $S(n)$, and is paid according to the "dual scaling" equation

$$w(n) = D(n_*) S(n_*)^{\beta/\alpha} S(n)^{\gamma-\beta/\alpha} \tag{91}$$

where $S(n_*)$ is the size of the reference firm and $D(n_*) = -Cn_* T'(n_*) / (\alpha\gamma - \beta)$ is a constant independent of firm size.[79]

I.3.2 Dynamic Version

In the dynamic model, a CEO of talent T_{it} increases the value of a firm of size S_{it} by $C_t S_{it}^{\gamma_t} T_{it}$.

We define $y_{it} := -\log n_{it}$ to be the relative talent (so that a high y_{it} corresponds to a high talent, and a fraction $e^{-y_{it}}$ of managers are better than manager i). So, CEO i's relative talent y_{it} evolves stochastically as

$$dy_{it} = \mu_t dt + \sigma_t dZ_{it}$$

with death rate δ_t.[80] From equation (89), the wage of a CEO with talent rank n_{it} is given by $w_{it} \propto n_{it}^{-(\alpha\gamma-\beta)}$. Therefore, using the definition of the log talent quantile $y_{it} := -\log n_{it}$, the log wage $x_{it} = \log w_{it}$ of individual i is given by

$$x_{it} = \chi_t y_{it} + a_t. \tag{92}$$

with

$$\chi_t = \alpha\gamma_t - \beta. \tag{93}$$

and where a_t is common across all individuals. Note that χ_t and a_t depend on economy-wide forces, while x_{it}, y_{it} are specific to each CEO.[81] Hence this is a process exactly like $x_{it} = \chi_t y_{it}$

[79] The derivation is as follows. Since $S(n) = An^{-\alpha}$ from (88), $S(n_*) = An_*^{-\alpha}$, $n_* T'(n_*) = -Bn_*^{\beta}$, we can rewrite equation (89) as follows:

$$\begin{aligned}(\alpha\gamma - \beta) w(n) = A^{\gamma} BCn^{-(\alpha\gamma-\beta)} &= CBn_*^{\beta} \cdot \left(An_*^{-\alpha}\right)^{\beta/\alpha} \cdot \left(An^{-\alpha}\right)^{(\gamma-\beta/\alpha)} \\ &= -Cn_* T'(n_*) S(n_*)^{\beta/\alpha} S(n)^{\gamma-\beta/\alpha}.\end{aligned}$$

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

[81] We have $a_t = \ln\left(\frac{A^{\gamma_t} BC_t}{\alpha\gamma_t - \beta}\right)$. Under the appropriate parametrization of C_t, we can have a_t a constant. Otherwise, equation (92) suggests an enrichment of the model with both "multiplicative" forces (γ_t) and "additive" forces (a_t). However, only the multiplicative force γ_t changes the local Pareto exponent.

and changes in $\chi_t = \alpha\gamma_t - \beta$ exactly generate scale dependence. We took parameter γ_t to be a technology parameter, which changes over time. That gives a time-varying χ_t. (We could also vary α and β, so that $\chi_t = \alpha_t\gamma_t - \beta_t$, but we keep the simplest formulation here).

Recall that a CEO of talent T_{it} increases the value of a firm of size S_{it} by $C_t S_{it}^{\gamma_t} T_{it}$. Hence, when the "reach" of the CEO is larger, γ_t increases. Hence, when γ_t increases, "effective scale" of the CEO impact is large. In the context of this model, it is just a technological parameter.[82]

The bottomline is the "scale dependence" model can be microfounded by a dynamic version of existing static models. It would be interesting to analyze this in a quantitative papers with micro data, but this is outside the scope of the present study.

I.4 Additional Parameterizations/Experiments in Section 5.4

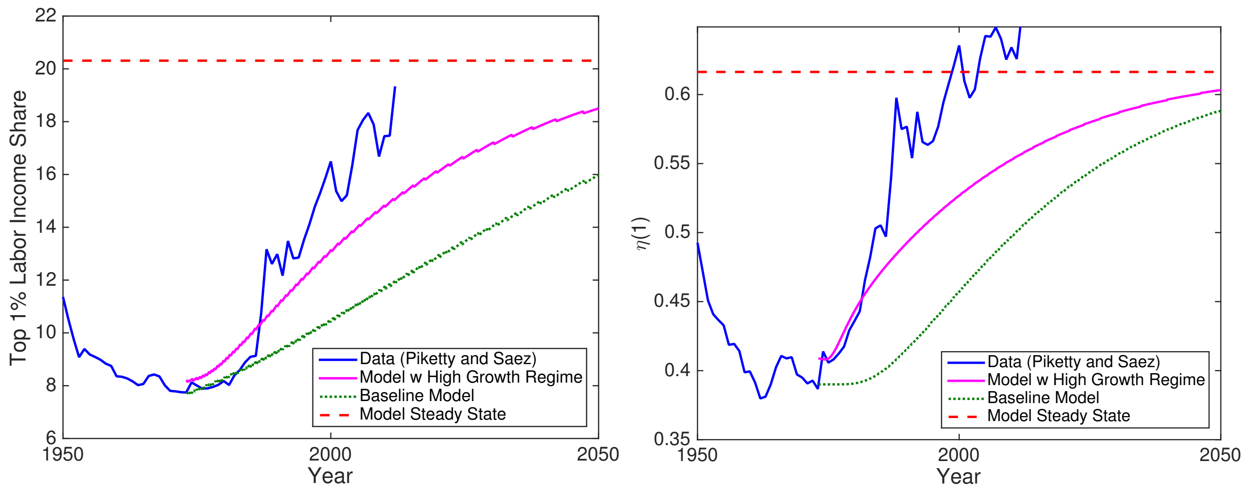
Since we do not have precise estimates for the parameters ψ and μ_H, we have explored a number of alternative parameterizations. We computed results for both higher and lower switching rates $\psi = 1/3$ and $\psi = 1/12$. In each case, we use (33) to recalibrate the initial μ_H so as to match the tail inequality observed in the data. As expected given our theoretical results, transitions are fastest when ψ 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". On the other hand, the speed of convergence becomes close to that in the benchmark model as ψ and μ_H become small. Indeed, as $\psi \to 0$ the model collapses to the one-type model of Section 4.3. This is because in this case we need $\mu_H \to \mu_L$ so as to still match data on the tail exponent ζ (see (33)). In their ongoing work using a very similar model, Jones and Kim (2014) propose such a "live-fast-die-young" calibration with very high ψ and μ_H.

Second, we have also conducted experiments in which we feed in a gradual increase in μ_H rather than a once-and-for-all increase. We find that in this case a higher switching rate ψ is needed than the one used in our baseline experiment. A parameter combination that generates time paths quite similar to those in Figure 5 is a gradual increase over a time period of 20 years (1973 to 1993) of μ_H by 11%, together with $\psi = 1/4$. This is still more conservative than the calibration of Jones and Kim who feature a larger increase of μ_H and a higher ψ.

Third, in the quantitative discussion in section 4.3 we studied the speed of change in inequality after an increase in σ (Figure 4). However, in section 5.4 which studies a model with type dependence we examine the dynamics of inequality after an increase in the gap

[82] See Garicano and Rossi-Hansberg (2006) and Geerolf (2014) for more microfoundation of this type of "scope" of CEO talent.

$\mu_H - \mu_L$ (Figure 5). Readers may therefore wonder how type dependence affects the speed of changes in inequality if we subject the two models to the same shocks. To answer this question, Figure 9 examines the dynamics following an increase in σ_H in the model with type dependence (note that only σ_H matters for the fatness of the right tail as discussed in Section 5.4). To focus on the speed of transition only, we choose the increase in σ_H so as to



(a) Top Income Inequality (b) Relative Income Shares

Figure 9: Transition Dynamics in Model with Type Dependence: Alternative Experiment

match the level of top inequality in the new steady state (the vertical dashed line labelled "Model Steady State" in Figure 9 is the same as that in Figure 4). It can be seen that, as expected, the transition in the model with type dependence is considerably faster than the transition in the baseline model. The model does not match the time paths of the two measures of top inequality as well as in Figure 5. This is mainly because in this experiment inequality in the new steady state is smaller than after an increase in $\mu_H - \mu_L$.

Appendix 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.

Auerbach, A. J., and K. Hassett (2015): "Capital Taxation in the Twenty-First Century," *American Economic Review*, 105(5), 38–42.

Bach, L., L. E. Calvet, and P. Sodini (2015): "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy," Les Cahiers de Recherche 1126, HEC Paris.

Benhabib, J., A. Bisin, and M. Luo (2015): "Wealth Distribution and Social Mobility in the US: A Quantitative Approach," Working paper, NYU.

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.

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.

Campbell, J. Y. (2001): "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 56(1), 1–43.

Edmans, A., X. Gabaix, and A. Landier (2009): "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium," *Review of Financial Studies*, 22(12), 4881–4917.

Evans, L. C. (1998): *Partial Differential Equations*, vol. 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2016): "Heterogeneity and Persistence in Returns to Wealth," Working paper.

Gabaix, X. (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.

Jones, C. I. (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.

Kacperczyk, M., J. B. Nosal, and L. Stevens (2014): "Investor Sophistication and Capital Income Inequality," NBER Working Papers 20246, National Bureau of Economic Research, Inc.

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.

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

Mimica, A. (2013): "Exponential decay of measures and Tauberian theorems," *ArXiv e-prints*.

Moskowitz, T. J., and A. Vissing-Jorgensen (2002): "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?," *American Economic Review*, 92(4), 745–778.

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

Piketty, T., and G. Zucman (2014a): "Capital is Back: Wealth-Income Ratios in Rich Countries, 1700-2010," 129(3), 1255–1310.

——— (2014b): "Wealth and Inheritance in the Long Run," in *Handbook of Income Distribution*, Handbook of Income Distribution. Forthcoming.

Roine, J., and D. Waldenström (2015): "Chapter 7 - Long-Run Trends in the Distribution of Income and Wealth," in *Handbook of Income Distribution*, ed. by A. B. Atkinson, and F. Bourguignon, vol. 2 of *Handbook of Income Distribution*, pp. 469 – 592. Elsevier.

Saez, E., and G. Zucman (2016): "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," 131(2), 519–578.

Sattinger, M. (1993): "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31(2), 831–80.

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