Online Appendix
for "Rapid Dynamics of Top Wealth Shares and Self-Made Fortunes: What is the Role of Family Firms?"
by Andrew G. Atkeson and Magnus Irie

Appendix A: Definition of $\mathbb{T}$

The evolution of the distribution of wealth by type from t to $t+1$ can be described by an operator $\mathbb{T}$ that maps pairs of densities of wealth by type at t to pairs of densities of wealth by type at $t+1$. The definition of this operator $\mathbb{T}$ is given by two linked second-order differences equations characterizing the updating of the distribution of wealth for each type. For $n \geq 1$ the difference equations are for each type j

$$g_{j,t+1}(n) = \phi_j \left[p_{u,j} g_{j,t}(n-1) + p_{d,j} g_{j,t}(n+1) + (1 - p_{u,j} - p_{d,j}) g_{j,t}(n)\right] + (1-\phi_j)\left[p_{u,-j} g_{-j,t}(n-1) + p_{d,-j} g_{-j,t}(n+1) + (1 - p_{u,-j} - p_{d,-j}) g_{-j,t}(n)\right], \tag{A1}$$

where $-j$ denotes the type opposite to j. For $n = 0$, this evolution is given by

$$g_{j,t+1}(0) = \phi_j \left[p_{d,j} g_{j,t}(1) + (1 - p_{u,j}) g_{j,t}(0)\right] + (1-\phi_j)\left[p_{d,-j} g_{-j,t}(1) + (1 - p_{u,-j}) g_{-j,t}(0)\right]. \tag{A2}$$

When the size of the grid is finite, we have the following additional equation describing the evolution at $n = N$

$$g_{j,t+1}(N) = \phi_j \left[p_{u,j} g_{j,t}(N-1) + (1 - p_{d,j}) g_{j,t}(N)\right] + (1-\phi_j)\left[p_{u,-j} g_{-j,t}(N-1) + (1 - p_{d,-j}) g_{-j,t}(N)\right]. \tag{A3}$$

Appendix B: Setting parameters as $\Delta_t \to 0$

To compare results in our discrete time model with closely related results in continuous time versions of the model as presented in Luttmer (2016), Gabaix et al. (2016) and elsewhere, we use the following procedure to adjust the parameters of our model as we change the length of the time period Δ_t. This is done to consider the limiting implications of our model as the time period gets short. We set $p_{d,j}$ and $\frac{p_{u,j}}{p_{d,j}}$ to match annualized means μ_j and variances σ_j^2 of innovations to the logarithm of the idiosyncratic component of assets. Specifically, we set the grid step size Δ as a function of the length of a time period Δ_t as

$$\Delta = \sigma_{max}\sqrt{2\Delta_t},$$

where σ_{max} is the largest annualized standard deviation of innovations to the logarithm of assets that we consider.

Under the model assumptions regarding the evolution of wealth for each type, the expected value at t of the innovations to the logarithm of wealth for all dynasties of type j, except those at the lowest node on the grid, is given by

$$\mathbb{E}_t \left[\log W_{i,t+1} - \log W_{i,t}\right] = (p_{u,j} - p_{d,j})\Delta. \tag{B1}$$

The uncentered second moment of these innovations to the logarithm of the idiosyncratic component of assets is given by

$$\mathbb{E}_t \left[\log W_{i,t+1} - \log W_{i,t}\right]^2 = (p_{u,j} + p_{d,j})\Delta^2. \tag{B2}$$

We then choose the parameters $p_{d,j}$ and $\frac{p_{u,j}}{p_{d,j}}$ so that the expression in equation (B1) is equal to the target per period mean $\Delta_t \mu_j$, and the expression in equation (B2) is equal to the target per period uncentered second moment $\Delta_t \sigma_j^2 + \Delta_t^2 \mu_j^2$. We set the transition probabilities over types as $1 - \phi_j = \kappa_j \Delta_t$ for fixed target values of κ_j.

In the case in which dynasties do not switch type, as we shrink the time interval to zero the tail coefficients for wealth for each type of dynasty approaches the standard formulas when log wealth follows a Brownian motion with a reflecting barrier at the bottom, namely $\zeta_{ss,j} = -2\mu_j/\sigma_j^2$ for $j = F, D$. To see this, we use that the tail coefficient is $\zeta_{ss,j} = \log\left(\frac{p_{u,j}}{p_{d,j}}\right)/\Delta$ when the types do not switch. Moreover, equations B1 and B2 together with $\Delta = \sigma_{max}\sqrt{2\Delta_t}$, imply that

(B3)

$$\begin{aligned}
\frac{\log\left(\frac{p_{u,j}}{p_{d,j}}\right)}{\Delta} &= \frac{1}{\Delta}\log\left(\frac{\sigma_j^2 + \mu_j^2\Delta_t + \mu_j\Delta}{\sigma_j^2 + \mu_j^2\Delta_t - \mu_j\Delta}\right) \\
&= \frac{1}{\Delta}\log\left(\frac{\sigma_j^2 + \mu_j^2\frac{1}{2}\frac{\Delta^2}{\sigma_{max}^2} + \mu_j\Delta}{\sigma_j^2 + \mu_j^2\frac{1}{2}\frac{\Delta^2}{\sigma_{max}^2} - \mu_j\Delta}\right) \\
&= \frac{1}{\Delta}\log\left(\sigma_j^2 + \mu_j^2\frac{1}{2}\frac{\Delta^2}{\sigma_{max}^2} + \mu_j\Delta\right) - \frac{1}{\Delta}\log\left(\sigma_j^2 + \mu_j^2\frac{1}{2}\frac{\Delta^2}{\sigma_{max}^2} - \mu_j\Delta\right)
\end{aligned}$$

Taking $\Delta_t \to 0$ implies taking $\Delta \to 0$, and applying L'Hôpital's rule to the above two terms separately gives us

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\log\left(\sigma_j^2+\mu_j^2\frac{1}{2}\frac{\Delta^2}{\sigma_{max}^2}+\mu_j\Delta\right)=\lim_{\Delta\to 0}\frac{\Delta\frac{\mu_j^2}{\sigma_{max}^2}+\mu_j}{\sigma_j^2+\mu_j^2\frac{1}{2}\frac{\Delta^2}{\sigma_{max}^2}+\mu_j\Delta}=\frac{\mu_j}{\sigma_j^2}$$

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\log\left(\sigma_j^2+\mu_j^2\frac{1}{2}\frac{\Delta^2}{\sigma_{max}^2}-\mu_j\Delta\right)=\lim_{\Delta\to 0}\frac{\Delta\frac{\mu_j^2}{\sigma_{max}^2}-\mu_j}{\sigma_j^2+\mu_j^2\frac{1}{2}\frac{\Delta^2}{\sigma_{max}^2}+\mu_j\Delta}=-\frac{\mu_j}{\sigma_j^2}$$

and hence

$$\lim_{\Delta_t\to 0}\zeta=\frac{\mu_j}{\sigma_j^2}-\left(-\frac{\mu_j}{\sigma_j^2}\right)=\frac{2\mu_j}{\sigma_j^2} \tag{B4}$$

Appendix C:
Analytical Solution for the Evolution of the Density of Wealth

In our main proposition, we provide an analytical solution for the evolution of the density of wealth in the transition to steady-state. We prove that proposition here.

C1. One-type model

We begin by providing an analytical expression for the evolution of the distribution of wealth in the context of the model with only one type, or, equivalently, as in the model in which dynasties do not switch types. In the one type model, the equations (A1) and (A2) can be written as

$$g_{t+1}(n)=p_u g_t(n-1)+p_d g_t(n+1)+(1-p_u-p_d)g_t(n) \tag{C1}$$

$$g_{t+1}(0)=p_d g_t(1)+(1-p_u)g_t(0). \tag{C2}$$

Champernowne (1953) showed that the stationary distribution implied by these equations is

$$g_{ss}(n)=(1-\lambda_{ss})\lambda_{ss}^n$$

where $\lambda_{ss}=\frac{p_u}{p_d}$. The stationary distribution exists provided that $p_u < p_d$. The proposition presented in this paper establishes an analytical expression for the distribution of wealth at each time period of the transitions from one steady state to another. Specifically, we consider initial distributions of wealth across dynasties that are of the same form as the steady-state distribution but with a different parameter, $\lambda_0 \neq \lambda_{ss}$. That is, we assume that the initial distribution is of the form

$$g_0(n) = (1 - \lambda_0)\lambda_0^n.$$

To develop our analytical formula in this case, we use the following notation. Let $\mathbb{T}$ be the operator mapping distributions over nodes n of our grid to new distributions defined by equations (C1) and (C2). Let Λ_0 be a vector corresponding to the initial distribution $g_0(n) = (1-\lambda_0)\lambda_0^n$. Let Λ_{ss} be the distribution to which the economy converges, $g_{ss}(n) = (1 - \lambda_{ss})\lambda_{ss}^n$. Let $\mathbf{1}$ denote a distribution that places weight 1 on the node $n = 0$ and weight 0 on every node $n \geq 1$. That is, $\mathbf{1}$ corresponds to the distribution of assets for a cohort of dynasties all starting with the minimum level of assets. With this notation, we have the following result stated as a Corollary of our main proposition in the text.

Corollary Assume that the initial distribution at $t = 0$ of the idiosyncratic component of assets across dynasties is given by Λ_0 and that the transition probabilities in equations (C1) and (C2) are constant at p_d and $p_u = \lambda_{ss}p_d$ so that the stationary distribution of the idiosyncratic component of assets across dynasties is given by Λ_{ss}. Then the distribution at date t implied by equations (C1) and (C2) is given recursively by

$$(g_{t+1} - \Lambda_{ss}) = A\,(g_t - \Lambda_{ss}) + (1 - A)\left(\mathbb{T}^t(\mathbf{1}) - \Lambda_{ss}\right), \tag{C3}$$

so that the distribution at time t is given by

$$g_t = A^t\Lambda_0 + (1 - A)\sum_{k=0}^{t-1} A^{t-1-k}\mathbb{T}^k(\mathbf{1}) \tag{C4}$$

where A is a scalar given by

$$A \equiv \left(p_d(1 - \lambda_0)(\frac{\lambda_{ss}}{\lambda_0} - 1) + 1\right),$$

Proof: Direct calculation gives that

$$\mathbb{T}(\Lambda_0) = A\Lambda_0 + (1 - A)\mathbf{1}.$$

The operator $\mathbb{T}$ is linear, and $\mathbb{T}(\Lambda_{ss}) = \Lambda_{ss}$. Repeated application of this operator to $g_{t+1} = \mathbb{T}(g_t)$ starting from $g_0 = \Lambda_0$ then gives the result (C3). Solving (C3) forward then implies (C4).

C2. Continuous-time analogue

Aleh Tsyvinski generously provided the continuous-time result presented in this section. This result is analogous in the sense that it gives an analytical expression for the density of the logarithm of wealth at all times during the course of a transition to steady state from an initial distribution in which the logarithm of wealth is exponentially distributed (and hence, wealth is Pareto distributed).

In particular, let X_t be a Brownian motion with drift $\mu = -r < 0$ and diffusion σ, with a reflecting barrier at zero. The transition density $p_t(x, y)$ of the process X_t satisfies the following Kolmogorov backward equation

$$\frac{\partial p_t(x,y)}{\partial t} = -r\frac{\partial p_t(x,y)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p_t(x,y)}{\partial x^2} \tag{C5}$$

and the Neumann boundary condition

$$\frac{\partial p_t(x,y)}{\partial x}\Big|_{x=0} = 0. \tag{C6}$$

The stationary distribution for transition densities p_t is exponential with rate $2\frac{r}{\sigma^2}$. To see this, note that with $g(x) = 2\frac{r}{\sigma^2}e^{-2\frac{r}{\sigma^2}x}, x > 0$, we have

$$\frac{\partial}{\partial t}\int_0^\infty 2\frac{r}{\sigma^2}e^{-2\frac{r}{\sigma^2}x}p_t(x,y)dx = \int_0^\infty 2\frac{r}{\sigma^2}e^{-2\frac{r}{\sigma^2}x}\left(-r\frac{\partial p_t(x,y)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p_t(x,y)}{\partial x^2}\right)dx =$$

$$\int_0^\infty \left(-\frac{\partial\left(re^{-2\frac{r}{\sigma^2}x}\right)}{\partial x} - 2\frac{r^2}{\sigma^2}e^{-2\frac{r}{\sigma^2}x}\right)\frac{\partial p_t(x,y)}{\partial x}dx = 0.$$

Where the second equality follows from integrating by parts. Consider now a transition experiment analogous to the one considered in our Corollary above. In other words, suppose that the initial distribution of the logarithm of wealth is given by $g_0(y) = \lambda e^{-\lambda y}, y > 0$. The distribution at time t, which we denote by $g_t(y)$ is then given by

$$g_t(y) = \int_0^\infty g_0(x)p_t(x,y)dx$$

Differentiating this (and using integration by parts) we obtain

$$\frac{\partial g_t(y)}{\partial t} = \int_0^\infty \lambda e^{-\lambda x}\frac{\partial p_t(x,y)}{\partial t}dx = \int_0^\infty \lambda e^{-\lambda x}\left(-r\frac{\partial p_t(x,y)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p_t(x,y)}{\partial x^2}\right)dx =$$

$$\int_0^\infty \left(-\frac{\partial\left(\frac{\lambda\sigma^2}{2}e^{-\lambda x}\right)}{\partial x} - r\lambda e^{-\lambda x}\right)\frac{\partial p_t(x,y)}{\partial x}dx = \left(\frac{\lambda^2\sigma^2}{2} - r\lambda\right)\int_0^\infty e^{-\lambda x}\frac{\partial p_t(x,y)}{\partial x}dx$$

$$= \left(\frac{\lambda^2\sigma^2}{2} - r\lambda\right)\left(-p_t(0,y) + \int_0^\infty \lambda e^{-\lambda x} p_t(x,y)dx\right) =$$

$$\left(\frac{\lambda^2\sigma^2}{2} - r\lambda\right)(g_t(y) - p_t(0,y))$$

In other words, the distribution at time t satisfies the following non-homogeneous ordinary differential equation

$$\frac{\partial g_t(y)}{\partial t} = \left(\frac{\lambda^2\sigma^2}{2} - r\lambda\right)(g_t(y) - p_t(0,y))$$

Using the initial condition $g_0(y) = \lambda e^{-\lambda y}$ we obtain the solution

$$g_t(y) = e^{\left(\frac{\lambda^2\sigma^2}{2} - r\lambda\right)t}\lambda e^{-\lambda y} - \left(\frac{\lambda^2\sigma^2}{2} - r\lambda\right)\int_0^t e^{\left(\frac{\lambda^2\sigma^2}{2} - r\lambda\right)(t-s)} p_s(0,y)ds$$

This is analoguous to equation C4 in that it shows that the distribution at time t is a linear combination of the initial distribution $\lambda e^{-\lambda y}$, and the distribution of agents coming up from the bottom, $\int_0^t e^{\left(\frac{\lambda^2\sigma^2}{2} - r\lambda\right)(t-s)} p_s(0,y)ds$.

C3. Two-type model

In this section, we prove our main Proposition in the model with switching between the two types. We denote by Λ_i the distribution over nodes given by $\Lambda_i(n) = (1-\lambda_i)\lambda_i^n$ for any $\lambda_i \in (0,1)$ and for $n \geq 0$. We use $\mathbf{1}$ to denote a distribution that puts weight one on the node $n = 0$ and zero on every other node.

In the two-type model, the operator $\mathbb{T}$ defined by equations (A1) and (A2) maps a pair of distributions by type at t, $[g_{F,t}, g_{D,t}]'$ to a pair of distributions by type at $t+1$, $[g_{F,t+1}, g_{D,t+1}]'$. Define $\mathbb{T}_j$ to be the operator which maps pairs of distributions at t, $[g_{F,t}, g_{D,t}]'$ to the distribution for type j at $t+1$. With these definitions

$$[g_{F,t+1}, g_{D,t+1}]' = \mathbb{T}[g_{F,t}, g_{D,t}]' = \left[\mathbb{T}_F [g_{F,t}, g_{D,t}]', \mathbb{T}_D [g_{F,t}, g_{D,t}]'\right]'$$

Our main proposition provides an analytical expression for the distribution of wealth at each time period in the transition between one steady state distribution and another. Specifically, fix the parameters of the operator $\mathbb{T}$ given by $\{p_{u,j}, p_{d,j}, \phi_j\}$. Let the initial distribution of assets by type be given by

$$g_{j,0} = a_{j,0}\Lambda_a + b_{j,0}\Lambda_b$$

with $a_{j,0}+b_{j,0} = 1$ for arbitrary non-negative weights $a_{j,0}, b_{j,0}$ and arbitrary Λ_a, Λ_b defined by $\lambda_a, \lambda_b \in [0,1)$. Then the following holds:

Main Proposition In the transition experiment described above the distributions of wealth by type at date t are given by

$$\text{(C7)} \qquad \begin{bmatrix} g_{F,t} \\ g_{D,t} \end{bmatrix} = \begin{bmatrix} a_{F,t}\Lambda_a \\ a_{D,t}\Lambda_a \end{bmatrix} + \begin{bmatrix} b_{F,t}\Lambda_b \\ b_{D,t}\Lambda_b \end{bmatrix} + \sum_{k=0}^{t-1} \mathbb{T}^k \begin{bmatrix} c_{F,t-k}\mathbf{1} \\ c_{D,t-k}\mathbf{1} \end{bmatrix}.$$

where $a_{j,0}, b_{j,0}$ are given by the initial distributions at $t = 0$,

$$\text{(C8)} \qquad \begin{bmatrix} a_{F,t+1} \\ a_{D,t+1} \end{bmatrix} = \begin{bmatrix} \phi_F A_F & (1-\phi_F)A_D \\ (1-\phi_D)A_F & \phi_D A_D \end{bmatrix} \begin{bmatrix} a_{F,t} \\ a_{D,t} \end{bmatrix}$$

and

$$\text{(C9)} \qquad \begin{bmatrix} b_{F,t+1} \\ b_{D,t+1} \end{bmatrix} = \begin{bmatrix} \phi_F B_F & (1-\phi_F)B_D \\ (1-\phi_D)B_F & \phi_D B_D \end{bmatrix} \begin{bmatrix} b_{F,t} \\ b_{D,t} \end{bmatrix}$$

where

$$\text{(C10)} \qquad A_j = \left[1 + p_{u,j}\frac{1-\lambda_a}{\lambda_a} - p_{d,j}(1-\lambda_a)\right]$$

$$\text{(C11)} \qquad B_j = \left[1 + p_{u,j}\frac{1-\lambda_b}{\lambda_b} - p_{d,j}(1-\lambda_b)\right]$$

and $c_{F,0} = c_{D,0} = 0$ and

$$c_{F,t+1} = \phi_F(a_{F,t} + b_{F,t}) + (1-\phi_F)(a_{D,t} + b_{D,t}) - (a_{F,t+1} + b_{F,t+1})$$

$$c_{D,t+1} = \phi_D(a_{D,t} + b_{D,t}) + (1-\phi_D)(a_{F,t} + b_{F,t}) - (a_{D,t+1} + b_{D,t+1})$$

Proof: Note that the operator $\mathbb{T}$ is linear in acting on pairs of distributions. Direct calculation gives that

$$\mathbb{T}_F \begin{bmatrix} a_{F,t}\Lambda_a \\ a_{D,t}\Lambda_a \end{bmatrix} = [\phi_F A_F a_{F,t} + (1-\phi_F)A_D a_{D,t}]\,\Lambda_a +$$

$$[\phi_F(1-A_F)a_{F,t} + (1-\phi_F)(1-A_D)a_{D,t}]\,\mathbf{1} =$$

$$a_{F,t+1}\Lambda_a + [\phi_F a_{F,t} + (1-\phi_F)a_{D,t} - a_{F,t+1}]\,\mathbf{1}$$

$$\mathbb{T}_F \begin{bmatrix} b_{F,t}\Lambda_b \\ b_{D,t}\Lambda_b \end{bmatrix} = b_{F,t+1}\Lambda_b + [\phi_F b_{F,t} + (1-\phi_F)b_{D,t} - b_{F,t+1}]\,\mathbf{1}$$

$$\mathbb{T}_D \begin{bmatrix} a_{F,t}\Lambda_a \\ a_{D,t}\Lambda_a \end{bmatrix} = a_{D,t+1}\Lambda_a + [\phi_D a_{D,t} + (1-\phi_D)a_{F,t} - a_{D,t+1}]\,\mathbf{1}$$

$$\mathbb{T}_D \begin{bmatrix} b_{F,t}\Lambda_b \\ b_{D,t}\Lambda_b \end{bmatrix} = b_{D,t+1}\Lambda_b + [\phi_D b_{D,t} + (1-\phi_D)b_{F,t} - b_{D,t+1}]\,\mathbf{1}$$

These results imply that when the operator $\mathbb{T}$ is applied to the initial distribution at $t = 0$, the pair of distributions that results at $t = 1$ is given by

$$\begin{bmatrix} g_{F,1} \\ g_{D,1} \end{bmatrix} = \begin{bmatrix} a_{F,1}\Lambda_a \\ a_{D,1}\Lambda_a \end{bmatrix} + \begin{bmatrix} b_{F,1}\Lambda_b \\ b_{D,1}\Lambda_b \end{bmatrix} + \begin{bmatrix} c_{F,1}\mathbf{1} \\ c_{D,1}\mathbf{1} \end{bmatrix}$$

Now consider applying the operator $\mathbb{T}$ to a pair of distributions at t of the form in equation (7). We get

$$\begin{bmatrix} g_{F,t+1} \\ g_{D,t+1} \end{bmatrix} = \mathbb{T}\begin{bmatrix} g_{F,t} \\ g_{D,t} \end{bmatrix} = \begin{bmatrix} a_{F,t+1}\Lambda_a \\ a_{D,t+1}\Lambda_a \end{bmatrix} + \begin{bmatrix} b_{F,t+1}\Lambda_b \\ b_{D,t+1}\Lambda_b \end{bmatrix} +$$

$$\begin{bmatrix} c_{F,t+1}\mathbf{1} \\ c_{D,t+1}\mathbf{1} \end{bmatrix} + \sum_{k=0}^{t-1} \mathbb{T}^{k+1} \begin{bmatrix} c_{F,t-k}\mathbf{1} \\ c_{D,t-k}\mathbf{1} \end{bmatrix} =$$

$$\begin{bmatrix} a_{F,t+1}\Lambda_a \\ a_{D,t+1}\Lambda_a \end{bmatrix} + \begin{bmatrix} b_{F,t+1}\Lambda_b \\ b_{D,t+1}\Lambda_b \end{bmatrix} + \sum_{k=0}^{t} \mathbb{T}^{k} \begin{bmatrix} c_{F,t+1-k}\mathbf{1} \\ c_{D,t+1-k}\mathbf{1} \end{bmatrix}$$

which proves the result.

C4. Conditions that the Steady-State Distribution must satisfy

The following are necessary conditions of Steady-State that are useful in our calibration of the model.

We take as given the parameters of the two-type model $\phi_F, \phi_D, p_{u,F}, p_{d,F}, p_{u,D}, p_{d,D}$. Provided that these parameters are such that equation (7) converges to a steady state of the form

$$\begin{bmatrix} g_F \\ g_D \end{bmatrix} = \begin{bmatrix} a_F\Lambda_a + b_F\Lambda_b \\ a_D\Lambda_a + b_D\Lambda_b \end{bmatrix}$$

we can characterize the steady state as follows. The steady state distribution is given by six parameters: $\lambda_a, \lambda_b \in (0, 1)$ and $a_F, a_D, b_F, b_D \in [0, 1]$. These six parameters have to satisfy the following conditions. The weights a_F, a_D, b_F, b_D have to satisfy

$$a_F + b_F = 1$$

$$a_D + b_D = 1$$

and be a stationary solution to equations (C8) and (C9) with the coefficients A_j and B_j given by equations (C10) and (C11). These equations imply that

$$\frac{a_F}{a_D} = \frac{(1-\phi_F)A_D}{(1-\phi_F A_F)} = \frac{(1-\phi_D A_D)}{(1-\phi_D)A_F} \tag{C12}$$

The second of these equations implies

$$0 = 1 - (1 - \phi_F - \phi_D)A_F A_D - \phi_D A_D - \phi_F A_F \tag{C13}$$

Since λA_F and λA_D are both quadratic in λ, we can multiply the left hand side of (C13) and obtain a fourth order polynomial in λ when $(1 - \phi_F - \phi_D) \neq 0$. To have a unique stationary distribution, one must check that only two of the roots of this polynomial lie in the interval $(0, 1)$. By convention, λ_a is the largest root of this polynomial that lies in the interval $(0, 1)$ and λ_b is the smaller of the two roots in this interval. We have that b_F and b_D solve the analogous equation to (C12) with λ_b being the smaller root in $(0, 1)$ of the analog to equation (C13) defined by B_F and B_D in place of A_F and A_D.

Appendix D: The Steady State Distribution of Wealth

We previously provided necessary conditions that the steady-state distribution must satisfy if it is of a particular form. This appendix shows that the steady state of the two type model is of the form $g_j(n) = a_j(1 - \lambda_a)\lambda_a^n + b_j(1 - \lambda_b)\lambda_a^n$ for $j \in \{F, D\}$ provided that a steady state exists. We begin by writing the equations (A1) and (A2), that define the operator $\mathbb{T}$ in the form of matrix equations.

$$\begin{aligned} x_{t+1}(n+1) &= \Psi x_t(n+2) + \Gamma x_t(n+1) + \Theta x_t(n) \\ x_{t+1}(0) &= \Psi x_t(1) + \Xi x_t(0) \end{aligned} \tag{D1}$$

where $x_t(n) = \begin{bmatrix} g_{t,F}(n) \\ g_{t,D}(n) \end{bmatrix}$ and the following matrices

$$\Psi = \begin{bmatrix} \phi_F p_{d,F} & (1-\phi_F)p_{d,D} \\ (1-\phi_D)p_{d,F} & \phi_D p_{d,D} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \phi_F p_{s,F} & (1-\phi_F)p_{s,D} \\ (1-\phi_D)p_{s,F} & \phi_D p_{s,D} \end{bmatrix}$$
$$\Theta = \begin{bmatrix} \phi_F p_{u,F} & (1-\phi_F)p_{u,D} \\ (1-\phi_D)p_{u,F} & \phi_D p_{u,D} \end{bmatrix}, \quad \Xi = \begin{bmatrix} \phi_F(1-p_{u,F}) & (1-\phi_F)(1-p_{u,D}) \\ (1-\phi_D)(1-p_{u,F}) & \phi_D(1-p_{u,D}) \end{bmatrix}$$

Since our goal is to find the stationary distribution, we consider these equations with time-subscripts removed. In particular, we want to solve the following second-order matrix difference equation

$$x(n+1) = \Psi x(n+2) + \Gamma x(n+1) + \Theta x(n) \tag{D2}$$

with the initial condition $x(0) = \Psi x(1) + \Xi x(0)$. To solve this equation, we rewrite it as a first-order difference equation by letting $z(n) = \begin{bmatrix} x(n+1) \\ x(n) \end{bmatrix}$ and write the system as follows

$$z(n+1) = Lz(n), \quad \text{for} \quad n \geq 1$$
$$z(0) = \begin{bmatrix} \Psi^{-1}(I_{2\times 2} - \Xi)x(0) \\ x(0) \end{bmatrix}$$

with

(D3) $$L = \begin{bmatrix} \Psi^{-1}(I_{2\times 2} - \Gamma) & -\Psi^{-1}\Theta \\ I_{2\times 2} & 0_{2\times 2} \end{bmatrix}$$

The inverse of the matrix Ψ is given by $\Psi^{-1} = \frac{1}{\phi_D + \phi_F - 1} \begin{bmatrix} \frac{\phi_D}{p_{d,F}} & \frac{\phi_F - 1}{p_{d,F}} \\ \frac{\phi_D - 1}{p_{d,D}} & \frac{\phi_F}{p_{d,D}} \end{bmatrix}$. The inverse exists provided that $\phi_F + \phi_D \neq 1$ and the probability of moving down is positive for each type. Provided that L has four distinct eigenvalues we can diagonalize it and write

(D4) $$z(n) = L^n z(0) = P\Lambda^n P^{-1} z(0)$$

where P is the matrix with the eigenvectors of L as columns, and Λ is the diagonal matrix of eigenvalues. Moreover, we have

(D5) $$P^{-1}z(n) = \Lambda^n P^{-1} z(0)$$

so that to ensure that the $\sum_{n=0}^{\infty} z(n) < \infty$ holds we need to impose the condition that

(D6) $$\tilde{p}_i z(0) = 0, \text{ for every eigenvalue } |\lambda_i| \geq 1$$

where $\tilde{p}_i$ is a row vector from $P^{-1} = [\tilde{p}_1, ..., \tilde{p}_4]'$. Let $\lambda_a, \lambda_b, \lambda_c$ and λ_d be the eigenvalues of L. It turns out that L has two eigenvalues that are stable, i.e., less than 1 in absolute value. Without loss of generality let λ_a and λ_b be the stable eigenvalues. Hence, for $i = c, d$ $|\lambda_i| \geq 1$. With $\tilde{p}_3 z(0) = \tilde{p}_4 z(0) = 0$ we can write equation (D4) as

$$z(n) = P\Lambda^n \begin{bmatrix} \tilde{p}_1 \\ \tilde{p}_2 \\ \tilde{p}_3 \\ \tilde{p}_4 \end{bmatrix} z(0) = P \begin{bmatrix} \lambda_a^n \tilde{p}_1 z(0) \\ \lambda_b^n \tilde{p}_2 z(0) \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} p_{11}\lambda_a^n \tilde{p}_1 z(0) + p_{12}\lambda_b^n \tilde{p}_2 z(0) \\ p_{21}\lambda_a^n \tilde{p}_1 z(0) + p_{22}\lambda_b^n \tilde{p}_2 z(0) \\ p_{31}\lambda_a^n \tilde{p}_1 z(0) + p_{32}\lambda_b^n \tilde{p}_2 z(0) \\ p_{41}\lambda_a^n \tilde{p}_1 z(0) + p_{42}\lambda_b^n \tilde{p}_2 z(0) \end{bmatrix}$$

In other words, the pair of densities can be written on the form

$$g_F(n) = p_{31}\tilde{p}_1 z(0)\lambda_a^n + p_{32}\tilde{p}_2 z(0)\lambda_b^n$$
$$g_D(n) = p_{41}\tilde{p}_1 z(0)\lambda_a^n + p_{42}\tilde{p}_2 z(0)\lambda_b^n$$

By defining the weights a_F, b_F and a_D, b_D to solve the following system of equations

$$\text{(D7)} \qquad (1-\lambda_a)a_F = p_{31}\tilde{p}_1 \begin{bmatrix} (1-\lambda_a)a_F\lambda_a + (1-\lambda_b)b_F\lambda_b \\ (1-\lambda_a)a_F + (1-\lambda_b)b_F \end{bmatrix}$$

$$\text{(D8)} \qquad (1-\lambda_a)a_D = p_{41}\tilde{p}_1 \begin{bmatrix} (1-\lambda_a)a_F\lambda_a + (1-\lambda_b)b_F\lambda_b \\ (1-\lambda_a)a_F + (1-\lambda_b)b_F \end{bmatrix}$$

$$\text{(D9)} \qquad (1-\lambda_b)b_F = p_{32}\tilde{p}_2 \begin{bmatrix} (1-\lambda_a)a_F\lambda_a + (1-\lambda_b)b_F\lambda_b \\ (1-\lambda_a)a_F + (1-\lambda_b)b_F \end{bmatrix}$$

$$\text{(D10)} \qquad (1-\lambda_b)b_D = p_{42}\tilde{p}_2 \begin{bmatrix} (1-\lambda_a)a_F\lambda_a + (1-\lambda_b)b_F\lambda_b \\ (1-\lambda_a)a_F + (1-\lambda_b)b_F \end{bmatrix}$$

we have shown that the stationary distributions can be written as

$$g_F(n) = (1-\lambda_a)a_F\lambda_a^n + (1-\lambda_b)b_F\lambda_b^n$$
$$g_D(n) = (1-\lambda_a)a_D\lambda_a^n + (1-\lambda_b)b_D\lambda_b^n$$

which is what we wanted to show.

Appendix E: Calibration Details

This appendix details the procedure for implementing the baseline calibration of the model as well as the calibrations considered in various transition experiments.

The time step size Δ_t, the grid step size Δ, the size of of the grid N, the maximum standard deviation accomodated by the grid σ_{max} and the fraction of dynasties in the overall population that belong to the different types, ν_F and ν_D, as well as the rate at which family firms diversify, κ_F, are common to all calibrations. In particular, $\Delta_t = 1/15000$, $\sigma_{max} = 0.70$, $\Delta = \sigma_{max}\sqrt{2\Delta_t}$, $N = \frac{50}{\sqrt{\Delta_t}}$, $\nu_F = 0.05$, $\nu_D = 1 - \nu_F = 0.95$, and $\kappa_F = 1/15$. The relationship between Δ_t, σ_{max} and Δ ensures that the model is well behaved when $\Delta_t \to 0$, analogous to when one considers the continuous time limit of a binomial option pricing model.

With these parameters set directly, we set the remaining four parameters governing the first two moments of the innovations to log wealth for the two types of families as described next.

E1. Baseline calibration

The baseline calibration targets four data moments. These are (a) the steady state tail coefficient of top wealth which is set to a target of ζ =1.43, (b) the difference in expected growth rates in the level of wealth of families at the top 0.01% and the bottom of the wealth distribution which is set to a target of 5.69%, (c) the cross-sectional dispersion of innovations to log wealth for families at the bottom of the wealth distribution which is set to a target of 8.13%, and (d) the cross-sectional dispersion of innovations to log wealth for families at the top 0.01% of the wealth distribution which is set to a target of 35.79%.

The moment (a) is estimated using equation 4 and data on ratios of wealth shares for the top 0.01% and 0.1% in 2016. This tail coefficient corresponds to a ratio of these top shares of 0.5. This lies in between the ratio estimated by Smith, Zidar and Zwick (2021) and Piketty, Saez and Zucman (2018) that report ratios of 0.47 and 0.51 in the year 2016, respectively. To illustrate the ranges of values of that one could use for the ratios of wealth shares, which in turn imply a tail coefficients through equation 4, Figure E1 displays the ratio of the top 0.01% to the top 0.1% wealth shares as well as the top 0.1% to the top 1%. The data comes from both Piketty, Saez and Zucman (2018) and Smith, Zidar and Zwick (2021). Note that their findings in each paper that the ratio of the wealth shares of the top 0.01% to the 0.1% and of the top 0.1% to the top 1% are similar is consistent with the maintained assumption that the top of the wealth distribution above the top 1% has a Pareto density with a constant tail coefficient.

The moment (b) is taken from Bach, Calvet and Sodini (2020) Table 1 column 1. The moments (c) and (d) are taken from Bach, Calvet and Sodini (2020) Table 8, column 1.

Equation 4 can be derived as follows. Assume that the density of log wealth is geometric with parameter λ above some node $\bar{n}$ on our grid of wealth levels. That is, let $g(n) = \bar{g}\lambda^n$ for $n > \bar{n}$ for some constant $\bar{g}$. Let $H(n)$ be the complementary CDF corresponding to this density. Then $H(n) = \bar{H}\lambda^n$ for $n > \bar{n}$ for some constant $\bar{H}$. With these assumptions, we have that the tail coefficient of wealth at nodes $n > \bar{n}$ is given by $\zeta(n) = \zeta_{top} = -\log(\lambda)/\Delta$.

Let $x > y$ be two top wealth percentiles (for example, the top 0.1% and 0.01%). Let $n(y) > n(x) > \bar{n}$ be the cutoff nodes for those percentiles. That is, let $n(x)$ solve

$$x = \bar{H}\lambda^{n(x)}$$

and likewise for $n(y)$. Assume that $\exp(\Delta)\lambda < 1$ so that top wealth shares are defined. Then the aggregate wealth held by the top x percentile is given by $W(x) = \bar{W}(\exp(\Delta)\lambda)^{n(x)}$ for some constant $\bar{W}$ and ratio of the share of wealth held by the top y to top x percentiles is given by

$$\frac{S(y)}{S(x)} = (\exp(\Delta)\lambda)^{n(y)-n(x)}$$

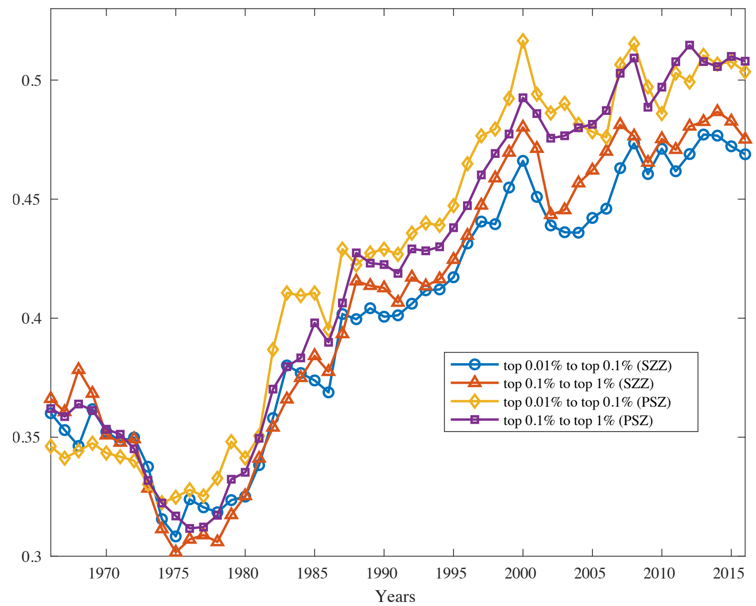


Figure E1. : Ratios of top wealth shares for the years 1966-2016.

Source: Estimates of top wealth shares are from Piketty, Saez and Zucman (2018) (PSZ) and Smith, Zidar and Zwick (2021) (SZZ).

This implies that

$$\log S(y) - \log S(x) = (n(y) - n(x))(\Delta + \log \lambda) = \Delta(n(y) - n(x))(1 - \zeta)$$

Since

$$n(x) = (\log(x) - \log(H))/\log(\lambda)$$

and likewise for $n(y)$, we have

$$\log S(y) - \log S(x) = (\log(y) - \log(x))(1 - \frac{1}{\zeta})$$

which gives equation 4.

E2. Calibration procedure

To hit these four moments, we have 4 parameters: μ_F, σ_F, μ_D and σ_D. The subsequent steps of the calibration are the following

1) Guess values for μ_j and σ_j, $j \in \{F, D\}$.
2) Compute the corresponding probabilities $p_{u,j}$ and $p_{d,j}$.
3) Compute the stationary distribution implied by these probabilities

4) Check what the implied tail coefficient of the resulting stationary distribution and check if the difference in average growth rates between the top and the bottom as well as the target values for dispersion of wealth growth are obtained.

5) Update guess until targets are hit.

In step 2, we must translate the annualized moments μ_j and σ_j, $j \in \{F, D\}$ in to probabilities of moving up and down on the grid. The annualized moments of the innovations to log wealth for each type are related to the probabilities through the following equations for the first and second moments of growth in log wealth

$$\mu_j \Delta_t = (p_{u,j} - p_{d,j})\Delta$$

$$\sigma_j^2 \Delta_t + \mu_j^2 \Delta_t^2 = (p_{u,j} + p_{d,j})\Delta^2$$

Solving these equations for the probabilities, using $\Delta = \sigma_{max}\sqrt{2\Delta_t}$, gives

$$p_{u,j} = \frac{1}{2}\left[\sigma_j^2 \frac{\Delta_t}{\Delta^2} + \mu_j^2 \frac{\Delta_t^2}{\Delta^2} + \mu_j \frac{\Delta_t}{\Delta}\right] = \frac{1}{4\sigma_{max}^2}\left[\sigma_j^2 + \mu_j^2 \Delta_t + \mu_j \Delta\right]$$

$$p_{d,j} = \frac{1}{2}\left[\sigma_j^2 \frac{\Delta_t}{\Delta^2} + \mu_j^2 \frac{\Delta_t^2}{\Delta^2} - \mu_j \frac{\Delta_t}{\Delta}\right]$$

Therefore

$$p_{u,j} = \frac{1}{4\sigma_{max}^2}\left[\sigma_j^2 + \mu_j^2 \Delta_t + \mu_j \Delta\right] \tag{E1}$$

$$p_{d,j} = \frac{1}{4\sigma_{max}^2}\left[\sigma_j^2 + \mu_j^2 \Delta_t - \mu_j \Delta\right] \tag{E2}$$

In step 3, we must compute the stationary distribution. We do this by finding the two stable eigenvalues of the matrix L defined in equation (D3) in Appendix D. We know that the steady-state distribution for each type is of the form

$$g_j(n) = a_j(1 - \lambda_a)\lambda_a^n + b_j(1 - \lambda_b)\lambda_b^n$$

so for high levels of wealth, the tail coefficient is given by $\zeta_{ss} = \frac{1}{\Delta}\log(\lambda_a)$, where λ_a is the larger of the two eigenvalues. This is the first of our targets. To fully specify the stationary distribution we also need to compute the weights a_j and b_j. Steady state implies that

$$\text{(E3)} \qquad \frac{a_F}{a_D} = \frac{(1-\phi_F)A_D}{(1-\phi_F A_F)}$$

$$\text{(E4)} \qquad \frac{b_F}{b_D} = \frac{(1-\phi_F)B_D}{(1-\phi_F B_F)}$$

where

$$\text{(E5)} \qquad A_j = \left[1 + p_{u,j}\frac{1-\lambda_a}{\lambda_a} - p_{d,j}(1-\lambda_a)\right]$$

$$\text{(E6)} \qquad B_j = \left[1 + p_{u,j}\frac{1-\lambda_b}{\lambda_b} - p_{d,j}(1-\lambda_b)\right]$$

Combining this with the fact that the steady state densities must sum to 1 also implies that $a_j + b_j = 1$, we obtain the system of equations

$$\text{(E7)} \qquad \frac{a_F}{a_D} = \frac{(1-\phi_F)A_D}{(1-\phi_F A_F)}$$

$$\text{(E8)} \qquad \frac{1-a_F}{1-a_D} = \frac{(1-\phi_F)B_D}{(1-\phi_F B_F)}$$

which implies

$$\text{(E9)} \qquad a_F = \frac{(1-\phi_F)A_D}{(1-\phi_F A_F)}$$

$$\text{(E10)} \qquad \frac{1-a_F}{1-a_D} = \frac{(1-\phi_F)B_D}{(1-\phi_F B_F)}$$

which can be solved for a_F and a_D, which in turn imply values for $b_F = 1 - a_F$ and $b_D = 1 - a_D$. The overall steady-state distribution of dynasties over nodes is therefore

$$\text{(E11)} \quad \nu_F g_F(n) + \nu_D g_D(n) = (\nu_F a_F + \nu_D a_D)(1-\lambda_a)\lambda_a^n + (\nu_F b_F + \nu_D b_D)(1-\lambda_b)\lambda_b^n$$

and the fraction of family firm dynasties at node n is given by

$$\text{(E12)} \qquad \nu_F(n) = \frac{\nu_F(a_F(1-\lambda_a)\lambda_a^n + (1-a_F)(1-\lambda_b)\lambda_b^n)}{(\nu_F a_F + \nu_D a_D)(1-\lambda_a)\lambda_a^n + (1-\nu_F a_F - \nu_D a_D)(1-\lambda_b)\lambda_b^n}$$

which can be used to calculate node-specific moments. In particular, the average growth rate of wealth and the dispersion of log wealth growth at node n is given by

(E13)
$$\bar{g}_n = \nu_F(n)(\mu_F + 0.5\sigma_F^2) + (1 - \nu_F(n))(\mu_D + 0.5\sigma_D^2)$$
(E14)
$$\bar{\sigma}_n^2 = ((\mu_F^2 + \sigma_F^2)\nu_F(n) + (1 - \nu_F(n))(\mu_D^2 + \sigma_D^2) - (\mu_F\nu_F(n) + \mu_D(1 - \nu_F(n)))^2$$

These formulas are the formulas for the moments of a mixture of two normal distributions. Recall that target (b) is $\bar{g}_N - \underline{g}_0 = 0.0569$, target (c) is $\bar{\sigma}_0 = 0.0813$ and target (d) is $\bar{\sigma}_{n_{top0.01\%}} = 0.3579$. The node $n_{top0.01\%}$ is defined through the relationship

$$G(n_{top0.01\%}) \equiv \sum_{n_{top0.01\%}}^{N} g(n) = 0.0001$$

We use the MATLAB function 'fsolve' to find values of μ_j and σ_j that hit these targets. The resulting parameters are reported in row A of Table 1. We can compute the excess kurtosis at node n implied by this calibration using the following formula

(E15)
$$\text{ex kurtosis}(n) = \frac{\nu_F(n)\left(\mu_F^4 + 6\mu_F^2\sigma_F^2 + 3\sigma_F^4\right) + \nu_D(n)\left(\mu_D^4 + 6\mu_D^2\sigma_D^2 + 3\sigma_D^4\right)}{\left(\nu_F(n)\left(\mu_F^2 + \sigma_F^2\right) + \nu_D(n)\left(\mu_D^2 + \sigma_D^2\right)\right)^2} - 3$$

Table E1 reports the excess kurtosis implied by the baseline calibration and compares it to the excess kurtosis reported by Gomez (2021).

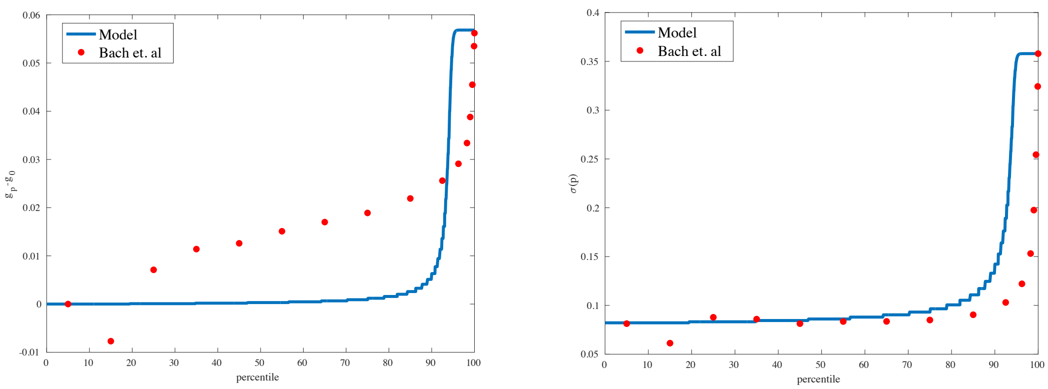
Table E1—: Excess Kurtosis of Innovations to Top Wealth

Data	6.58
Baseline	1.31

Note: The data on excess kurtosis for the Forbes 400 is from panel b) of Table 2 of Gomez (2021) for the period 1983-2017. The percentile used for the Forbes 400 in our model is the top 0.0003 percentile.

Gomez (2021) also reports that less than 10% of the members of the 1983 cohort of the Forbes 400 list were still members in 2017. When we compute this measure of persistence in the membership of the Forbes 400 in the context of the steady state of our baseline calibration we obtain that about 7% of the members of the Forbes 400 are still members over a 34 year period.

We calibrated our model to match the differences in the expected growth rate of wealth and cross section dispersion of innovations to wealth at the top and the bottom of the wealth distribution. To evaluate how well our model fits the data at intermediate levels of wealth, in Figure E2, in the left panel (E2a), we show the expected growth in the level of wealth for dynasties at each wealth percentile, and in the right panel (E2b), we show the corresponding cross section dispersion of growth rates of the logarithm of wealth at each wealth percentile implied by these changing fractions of dynasties of each type by wealth level. The red dots in these figures correspond to the data in Tables 1 and 8 of Bach, Calvet and Sodini (2020).



(a) Expected Growth

(b) Standard Deviation

Figure E2. : **Moments of innovations to wealth across the distribution:**

Note: The left panel (E2a) shows the difference in the expected growth rate of the level of wealth for dynasties at different percentiles of the steady-state distribution of wealth relative to the bottom of the distribution. The right panel (E2b) shows the corresponding dispersion of innovations to the logarithm of wealth. These moments of innovations to wealth differ across families at different percentiles of the wealth distribution because the mix of dynasties with family firms and with diversified portfolios varies with the level of wealth.

E3. Transition experiments

Once we have the parameters governing the growth of wealth, we can compute the evolution over time of a given distribution of wealth. This is done by applying the $\mathbb{T}$ operator repeatedly to a given initial distribution. The $\mathbb{T}$ operator is defined by equation D1. For instance, to compute the tail coefficient at node n at time $t+1$ given a vector of distributions of wealth by type at time t, x_t, we apply equations D1 to obtain x_{t+1}. We then obtain the overal distribution of wealth as $g_{t+1}(n) = [\nu_F, \nu_D] \cdot x_{t+1}(n)$, which we use to compute the negative of the slope of the CCDF at the given node n. There is a question about what to do about at the last node of the grid. We impose a reflecting barrier at the top of the grid analogous to the one at the bottom. However, the grid size is so large that the mass at the top of the grid is very close to zero. In the numerical examples we compute, it does not seem to matter if one puts a reflecting barrier at the top or not. To understand this, consider the version of the model when types do not switch. As long as $p_{u,j}/p_{d,j} < 1$, the mass at the top of the grid is going to be negligible if the grid size is large enough since the mass is proportional to $\lambda_{j,ss}^n = (p_{u,j}/p_{d,j})^n$.

E4. Calibration of alternative experiments presented in Section III

We conduct a series of quantitative experiments. This appendix describes the calibration procedures of those experiments.

The first two counterfactual experiments are presented in Section III of the paper. Relative to the calibration procedure for the baseline, these two experiments replace the target for the dispersion of wealth growth at the top with directly setting the volatility of the F type. In particular, the first experiment sets $\sigma_F = 0.3306$, while the second sets $\sigma_F = 0.2204$. Recall that the baseline calibration does not set σ_F directly, but the the implied value for this parameter in the baseline calibration is $\sigma_F = 0.4409$. The values for the calibrated parameters are presented in rows B and C of Table 1 of the paper.

In addition, when computing the persistence of membership in the Forbes 400, these alternative calibrations feature higher persistence than in the baseline and in the data reported by Gomez (2021). In particular, Gomez (2021) reports that less than 10% of the Forbes 400 cohort of 1983 were still on the list 33 years later. The corresponding number in the baseline calibration is around 7% while it is closer to 13% and 21% in the two alternative calibrations discussed here.

E5. Additional quantitative experiments

As robustness checks, we also consider three additional quantitative experiments in this appendix. In the first two additional experiments, we examine the results of calibrations wherein the volatility of the F type is reduced in the same manner as the two alternative calibrations presented in Section III of the paper, while the target for the difference in mean growth rates across the wealth distribution is simultaneously doubled. In other words, relative to the alternative calibrations considered in Section III, we now also change the calibration target b) to $\bar{g}_N - \bar{g}_0 = 0.1138$. Increasing the target difference in mean growth rates is meant to gauge the extent to which larger differences in mean growth rates between the types rather than the very high volatility of the F type can account for the prevalence of new large fortunes and rapid transitions of top wealth inequality. The following Table E2 presents the values of the calibrated parameters. Figure E3 compares the fraction of the Forbes 400 members that where at the bottom within the last k years and the transition of ratios of top wealth shares between these alternative calibrations and the baseline calibration. We see that the presence of a substantially larger difference in mean growth rates across the wealth distribution is not enough to compensate for the absence of the high volatility of the F type that is characteristic of the baseline calibration.

Table E2—: Calibrated Parameters in the Baseline and Alternative Calibrations

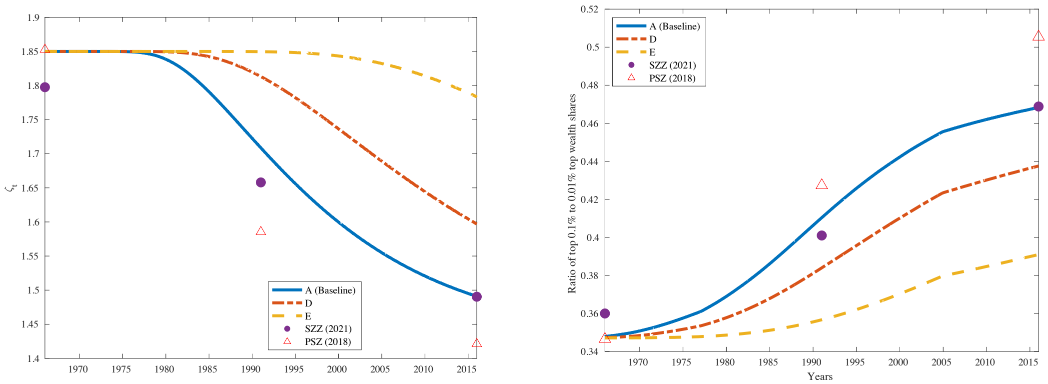
Parameters	g_F	g_D	σ_F	σ_D	ν_F	κ_F
A (Baseline)	0.0035	-0.0847	0.4409	0.0779	0.05	0.067
D	0.0223	-0.1319	0.3306	0.0798	0.05	0.067
E	0.0352	-0.1221	0.2204	0.0805	0.05	0.067
F	0.0068	-0.0828	0.4441	0.0778	0.05	0.067
Targets	ζ_{ss}	$\bar{g} - \underline{g}$	$\bar{\sigma}_{bottom}$	$\bar{\sigma}_{n_{top0.01\%}}$	σ_F	
A (Baseline)	1.43	0.0569	0.0813	0.3579	N/A	
D	1.43	0.1138	0.0813	N/A	0.3306	
E	1.43	0.1138	0.0813	N/A	0.2204	
F	1.4	0.0569	0.0813	N/A	0.3579	

Note: Calibrated parameters in the baseline as well as the alternative calibrations D and E where the volatility of the F type is reduced to 75% and 50% of its baseline value, respectively, while the targeted difference in growth rates between the top and the bottom of the wealth distribution is doubled relative to the baseline. Alternative calibration F instead features a lower target for the steady state tail coefficient.

The final alternative calibration we consider is one in which the target steady state wealth coefficient is set to $\zeta_{ss} = 1.4$ instead of the baseline value of $\zeta_{ss} = 1.43$. This is motivated by the following two reasons. First, there is some discrepancy between the ratios of top wealth shares reported by Piketty, Saez and Zucman (2018) and Smith, Zidar and Zwick (2021). Second, the mapping between ratios of top wealth shares in equation 4 is a steady state relationship. It is entirely possible that the parameters governing wealth growth at a specific point in time are associated with a different steady state than what the current ratio of top wealth shares would imply. The resulting parameter values are reported in row F of Table E2. Figure E4 plots the transition of the tail coefficient as well as the evolution of the ratio of the top 0.01% to the top 0.1% wealth shares with this alternative target together with data from Piketty, Saez and Zucman (2018) and Smith, Zidar and Zwick (2021). We see that the lower target value for the steady state distribution implies that the transition is somewhat faster.

Appendix F: A Spectral Analysis of the Dynamics of the Distribution

In this paper, we provide an analytical expression for the dynamics of the distribution of wealth over time as it converges to steady-state if the initial distribution of wealth is in a certain class of distributions. Gabaix et al. (2016) use an alternative approach to analyze the dynamics of the distribution of wealth to steady-state based on a spectral analysis of these dynamics in continuous time. In this appendix, we provide direct analogs of their spectral analysis in our discrete



(a) Tail coefficient.

(b) Ratio of top wealth shares

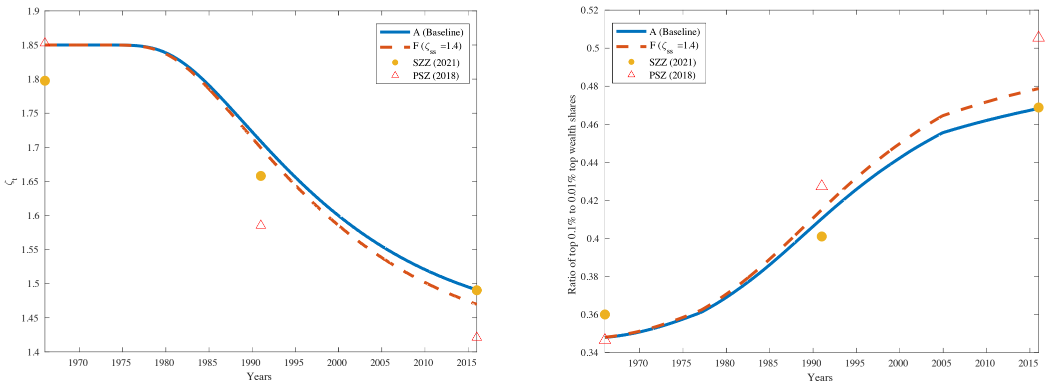
Figure E3. : Transition Results from Baseline and Alternative Calibrations

Note: This figure displays comparisons along two dimensions of the baseline calibration A with the alternative calibrations, D and E. In calibrations D and E the value of σ_F is set to 75% and 50% of its baseline value, respectively, while the target difference in mean growth rates across the wealth distribution is doubled. Figure (E3a) compares the transition of the tail coefficient, and Figure (E3b) considers the transition of the ratio of the top 0.01% to the top 0.1% wealth shares. Along both dimensions, the presence of a minority of dynasties with very high idiosyncratic volatility is important for obtaining rapid transitions. The transition is computed for the years 1966-2016. Marked are also the data from Smith, Zidar and Zwick (2021) (circles) and Piketty, Saez and Zucman (2018) (triangles). These are the ratios of the top 0.01% to the top 0.1% wealth shares and the implied tail coefficient using equation 4.

time - discrete state setting with the model restricted to have only one type by analyzing the eigenvalues and eigenvectors of our operator $\mathbb{T}$ in the version of our model with only one type of dynasty.

To prove their results, Gabaix et al. (2016) impose a boundedness assumption on tail coefficients of the distributions of wealth under consideration that is described in their Assumption 1. Here, we consider a related bound by computing the eigenvalues and eigenvectors of our operator $\mathbb{T}$ when the grid of wealth levels is finite (so $N < \infty$). In this case, this operator $\mathbb{T}$ is simply a square Markov transition matrix of size $(N+1) \times (N+1)$, so the calculation of eigenvalues and eigenvectors is standard. As is the case with finite Markov transition matrices, the largest eigenvalue is equal to one, and the speed of convergence of the distribution to steady-state is related to the size of the second largest eigenvalue, which is less than one. We are able to compute this second largest eigenvalue and consider its limiting value as $N \to \infty$. We find that this limiting value of the second largest eigenvalues of our finite Markov transition matrix $\mathbb{T}$ as N grows large corresponds to the formula found in Gabaix et al. (2016) Proposition 1.

We present this analysis for two reasons. First, it may be of interest to readers wishing to better understand spectral methods for analyzing dynamics of distributions. Second, it allows us to highlight two differences between the analytical characterization of the dynamics of the distribution of wealth that we present in



(a) Tail coefficient

(b) Ratio of top wealth shares

Figure E4. : Transition Results from Final Alternative Calibration

Note: This figure displays the transition of the tail coefficient as well as the ratio of the top 0.01% wealth share to the top 0.1% wealth share when the target steady state tail coefficient is $\zeta_{ss} = 1.4$ instead of the baseline value $\zeta_{ss} = 1.43$. The transition is computed for the years 1966-2016. Marked are also the data from Smith, Zidar and Zwick (2021) (circles) and Piketty, Saez and Zucman (2018) (triangles).

our paper and those obtained using spectral methods. These are, first, that our analysis does not require that we impose a bound on the tail coefficient of the initial distribution under consideration. Second, and more important, our analysis directly highlights the connection between the speed of wealth mobility from the bottom of the wealth distribution to the top and the dynamics of the shape of the top of the wealth distribution as it converges to steady state.

F1. The eigenvalue problem of $\mathbb{T}$

In the version of the model with one type, the operator $\mathbb{T}$ that maps a distribution g at time t to a distribution $\mathbb{T}(g)$ at time $t+1$ can be defined through the following equations

For $0 < n < N$,

$$\mathbb{T}(g)(n) = p_u g(n-1) + (1 - p_u - p_d) g(n) + p_d g(n+1) \tag{F1}$$

for $n = 0$

$$\mathbb{T}(g)(0) = (1 - p_u) g(0) + p_d g(1) \tag{F2}$$

and, if $N < \infty$, for $n = N$

$$\mathbb{T}(g)(N) = (1 - p_d) g(N) + p_u g(N-1) \tag{F3}$$

The eigenvalue problem $\lambda g = \mathbb{T}(g)$ is therefore defined by the following equations:

For $0 < n < N$,

$$\lambda g(n) = p_u g(n-1) + (1 - p_u - p_d) g(n) + p_d g(n+1), \tag{F4}$$

for $n = 0$

$$\lambda g(0) = (1 - p_u) g(0) + p_d g(1) \tag{F5}$$

and, if $N < \infty$, for $n = N$

$$\lambda g(N) = (1 - p_d) g(N) + p_u g(N-1) \tag{F6}$$

Note that when $N < \infty$, $\mathbb{T}$ can be represented by an $N+1 \times N+1$ matrix P of the form

$$P = \begin{pmatrix} 1-p_u & p_d & 0 & \dots & \dots & \dots & 0 \\ p_u & 1-p_u-p_d & p_d & 0 & \dots & \dots & 0 \\ 0 & p_u & 1-p_u-p_d & p_d & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & \dots & 0 & p_u & 1-p_d-p_u & p_d \\ 0 & \dots & \dots & 0 & 0 & p_u & 1-p_d \end{pmatrix} \tag{F7}$$

So that for all g given by vectors of size $N+1 \times 1$,

$$\mathbb{T}(g) = Pg$$

Thus, the eigenvalue problem for $\mathbb{T}$ corresponds to finding the eigenvalues of P.

Note that the matrix P is not symmetric. Similarly, when $N = \infty$, $\mathbb{T}$ is not self-adjoint. This prevents a direct application of the Spectral Theorem for analyzing the eigenvalue problem presented above.

Following Lemma 6 in Gabaix et al. (2016), we analyze a related operator $\mathbb{S}$ that is self-adjoint and which, under certain conditions discussed below, has the same eigenvalues as $\mathbb{T}$.

We define this related self-adjoint operator $\mathbb{S}$ as follows. For each n, scale the equations (F1)-(F3) that define the operator $\mathbb{T}$ by a factor $\left(\sqrt{p_d/p_u}\right)^n$. This gives the equations

$$\left(\frac{p_d}{p_u}\right)^{n/2}\mathbb{T}(g)(n) = p_u\left(\frac{p_d}{p_u}\right)^{n/2} g(n-1) + (1-p_u-p_d)\left(\frac{p_d}{p_u}\right)^{n/2} g(n) + p_d\left(\frac{p_d}{p_u}\right)^{n/2} g(n+1)$$

$$\mathbb{T}(g)(0) = (1-p_u)g(0) + p_d g(1)$$

$$\left(\frac{p_d}{p_u}\right)^{N/2}\mathbb{T}(g)(N) = (1-p_d)\left(\frac{p_d}{p_u}\right)^{N/2} g(N) + p_u\left(\frac{p_d}{p_u}\right)^{N/2} g(N-1)$$

For any vector g, let $h(n) = g(n)\left(\sqrt{p_d/p_u}\right)^n$. We will use the notation h_g refer to this vector. For $N < \infty$, define the operator $\mathbb{S}$ by

$$\mathbb{S}(h)(n) = \left(\frac{p_d}{p_u}\right)^{n/2}\mathbb{T}(g)(n) \tag{F8}$$

In other words, $\mathbb{S}$ is defined by the following set of equations:

For $0 < n < N$,

$$\mathbb{S}(h)(n) = (\sqrt{p_u p_d})h(n-1) + (1-p_u-p_d)h(n) + (\sqrt{p_u p_d})h(n+1) \tag{F9}$$

for $n = 0$,

$$\mathbb{S}(h)(0) = (1-p_u)h(0) + (\sqrt{p_u p_d})h(1) \tag{F10}$$

and, if $N < \infty$

$$\mathbb{S}(h)(N) = (1-p_d)h(N) + (\sqrt{p_u p_d})h(N-1) \tag{F11}$$

As with the operator $\mathbb{T}$, for fixed $N < \infty$, the operator $\mathbb{S}$ can be represented as an $N+1 \times N+1$ matrix Q:

(F12)

$$Q = \begin{pmatrix} 1-p_u & \sqrt{p_u p_d} & 0 & \dots & \dots & \dots & 0 \\ \sqrt{p_d p_u} & 1-p_u-p_d & \sqrt{p_u p_d} & 0 & \dots & \dots & 0 \\ 0 & \sqrt{p_u p_d} & 1-p_u-p_d & \sqrt{p_u p_d} & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & \dots & 0 & \sqrt{p_u p_d} & 1-p_u-p_d & \sqrt{p_u p_d} \\ 0 & \dots & \dots & 0 & 0 & \sqrt{p_u p_d} & 1-p_d \end{pmatrix}$$

That is, for all h given by vectors of size $N+1 \times 1$,

$$\mathbb{S}(h) = Qh$$

Note that for any fixed $N \leq \infty$, we can recover the dynamics of g from the dynamics of h. That is, if we start from g_0, we construct $h_0(n) = g_0(n)\left(\frac{p_d}{p_u}\right)^{n/2}$. We then construct h_t by applying the operator $\mathbb{S}$, t times, or, equivalently, when $N < \infty$

$$h_t = Q^t h_0$$

We then can construct g_t from $g_t(n) = h_t(n)\left(\sqrt{p_u/p_d}\right)^n$.

Note as well that when $N < \infty$ the matrix Q is real valued and symmetric. That is

$$Q(i,j) = Q(j,i), \forall i,j$$

Thus, we have that when $N < \infty$ the eigenvalues of Q are real, that the eigenvectors are orthogonal, and that the Spectral Theorem for finite dimensional spaces applies. That is, we can diagonalize Q and use that eigenvalue-eigenvector decomposition to characterize the dynamics of h_t.

The eigenvalue problem $\lambda h = \mathbb{S}(h)$ can be written

$$\lambda h(n) = (\sqrt{p_u p_d})h(n-1) + (1 - p_u - p_d)h(n) + (\sqrt{p_u p_d})h(n+1) \tag{F13}$$

for $0 < n < N$ and for $n = 0$

$$\lambda h(0) = (1 - p_u)h(0) + (\sqrt{p_u p_d})h(1) \tag{F14}$$

and, if $N < \infty$, for $n = N$

$$\lambda h(N) = (1 - p_d)h(N) + (\sqrt{p_u p_d})h(N-1) \tag{F15}$$

Direct comparison of these two eigenvalue problems gives us our first proposition:

Proposition F1: When $N < \infty$, the set of $N+1$ eigenvalues $\{\lambda_k\}_{k=1}^{N+1}$ of the two operators $\mathbb{T}$ and $\mathbb{S}$ are the same, and the eigenvectors of the two eigenvalue problems are related by $h(n;\lambda_k) = (\sqrt{p_d/p_u})^n g(n;\lambda_k)$.

To prove this proposition, observe that the operators satisfy $\mathbb{S}(h)(n) = \left(\frac{p_d}{p_u}\right)^{n/2}\mathbb{T}(g)(n)$ for any two vectors h and g such that $h(n) = (\sqrt{p_d/p_u})^n g(n)$. Suppose that λ_k is an eigenvalue of $\mathbb{S}$, and that $h(n;\lambda_k)$ is the corresponding eigenvector, then for all $0 \le n \le N$:

$$
\begin{aligned}
\mathbb{S}(h)(n) &= \lambda_k h(n;\lambda_k) \\
&\Leftrightarrow \\
\left(\frac{p_d}{p_u}\right)^{n/2} \mathbb{T}(g)(n) &= \lambda_k h(n;\lambda_k) \\
&\Leftrightarrow \\
\mathbb{T}(g)(n) &= \lambda_k g(n;\lambda_k)
\end{aligned}
$$

So that λ_k is also an eigenvalue of $\mathbb{T}$, with $g(n;\lambda_k)$ being the corresponding eigenvector. By the same argument, if λ_k is an eigenvalue of $\mathbb{T}$, with $g(n;\lambda_k)$ being the corresponding eigenvector, then λ_k is an eigenvalue of $\mathbb{S}$, with $h(n;\lambda_k)$ as the corresponding eigenvector.

Note that equations (F4) and (F13) in the eigenvalue problems for the operators $\mathbb{T}$ and $\mathbb{S}$ are both regular homogeneous second-order difference equation with constant coefficients. For a given eigenvalue, λ, the characteristic equations for these two difference equations are as follows

$$
k(\lambda)^2 - \frac{(\lambda - 1 + p_u + p_d)}{p_d} k(\lambda) + \frac{p_u}{p_d} = 0 \tag{F16}
$$

$$
m(\lambda)^2 - \frac{(\lambda - 1 + p_u + p_d)}{\sqrt{p_u p_d}} m(\lambda) + 1 = 0 \tag{F17}
$$

The characteristic equation for the eigenvalue problem for the operator $\mathbb{T}$ has the two solutions

$$
k_1(\lambda) = \frac{(\lambda - 1 + p_u + p_d)}{2p_d} + \frac{\sqrt{(\lambda - 1 + p_u + p_d)^2 - 4p_u p_d}}{2p_d} \tag{F18}
$$

$$
k_2(\lambda) = \frac{(\lambda - 1 + p_u + p_d)}{2p_d} - \frac{\sqrt{(\lambda - 1 + p_u + p_d)^2 - 4p_u p_d}}{2p_d} \tag{F19}
$$

while that for the operator $\mathbb{S}$ has the two solutions

(F20)

$$
m_1(\lambda) = \frac{(\lambda - 1 + p_u + p_d)}{2\sqrt{p_u p_d}} + \frac{\sqrt{(\lambda - 1 + p_u + p_d)^2 - 4p_u p_d}}{2\sqrt{p_u p_d}} = \sqrt{p_d/p_u}\, k_1(\lambda)
$$

(F21)

$$
m_2(\lambda) = \frac{(\lambda - 1 + p_u + p_d)}{2\sqrt{p_u p_d}} - \frac{\sqrt{(\lambda - 1 + p_u + p_d)^2 - 4p_u p_d}}{2\sqrt{p_u p_d}} = \sqrt{p_d/p_u}\, k_2(\lambda)
$$

Note that these roots of these characteristic equations are both real whenever

$$(\lambda - 1 + p_u + p_d)^2 - 4p_u p_d \geq 0 \tag{F22}$$

and are complex conjugates of each other whenever

$$(\lambda - 1 + p_u + p_d)^2 - 4p_u p_d < 0 \tag{F23}$$

Define the cutoffs $\bar{\lambda}$ and $\underline{\lambda}$ as the two solutions to

$$(\lambda - 1 + p_u + p_d)^2 - 4p_u p_d = 0$$

These are given by

$$\bar{\lambda} = 1 - (p_u + p_d) + 2\sqrt{p_u p_d} \tag{F24}$$

and

$$\underline{\lambda} = 1 - (p_u + p_d) - 2\sqrt{p_u p_d} \tag{F25}$$

we have $1 > \bar{\lambda} > \underline{\lambda}$. We distinguish between three cases surrounding the larger cutoff point $\bar{\lambda}$:

1. In the interval $(\bar{\lambda}, 1)$, the characteristic equations corresponding to the difference equations (F4) and (F13) have two distinct real roots, and the solutions to the difference equations are of the form

   $$g(n; \lambda) = a_1(\lambda) k_1(\lambda)^n + a_2(\lambda) k_2(\lambda)^n \tag{F26}$$

   $$h(n; \lambda) = a_1(\lambda) m_1(\lambda)^n + a_2(\lambda) m_2(\lambda)^n \tag{F27}$$

   respectively. Here the parameters $a_1(\lambda)$ and $a_2(\lambda)$ are to be chosen to match boundary conditions.

2. At $\bar{\lambda}$, the characteristic equations have one real root and the solutions to the difference equations are of the form

   $$g(n; \lambda) = (a_1(\lambda) + n a_2(\lambda))\, k(\lambda)^n \tag{F28}$$

   $$h(n; \lambda) = (a_1(\lambda) + n a_2(\lambda))\, m_2(\lambda)^n \tag{F29}$$

3. When $\lambda \in (\underline{\lambda}, \bar{\lambda})$, the roots of the two characteristic equations are complex

and the solution to the difference equations can be written

$$g(n;\lambda) = \left(\sqrt{p_u/p_d}\right)^n a(\lambda)\cos\left(\theta(\lambda)n + \omega(\lambda)\right) \tag{F30}$$

$$h(n;\lambda) = a(\lambda)\cos\left(\theta(\lambda)n + \omega(\lambda)\right) \tag{F31}$$

where

$$\theta(\lambda) = \cos^{-1}\left(\frac{(\lambda - 1 + p_u + p_d)}{2\sqrt{p_d p_u}}\right) \tag{F32}$$

and $a(\lambda)$ and $\omega(\lambda)$ are to be chosen to match boundary conditions

$$[a_1(\lambda)m_1(\lambda) + a_2(\lambda)m_2(\lambda)]$$

$$(a_1(\lambda)m_1(\lambda)^N + a_2(\lambda)m_2(\lambda)^N)(\lambda - 1 + p_d) = (\sqrt{p_u p_d})\left[a_1(\lambda)m_1(\lambda)^{N-1} + a_2(\lambda)m_2(\lambda)^{N-1}\right]$$

these in turn imply

$$a_1(\lambda) = -\left(\frac{\lambda - 1 + p_u - (\sqrt{p_d p_u})m_2(\lambda)}{\lambda - 1 + p_u - (\sqrt{p_d p_u})m_1(\lambda)}\right) a_2(\lambda)$$

$$a_1(\lambda) = -\left(\frac{m_2(\lambda)}{m_1(\lambda)}\right)^N \left(\frac{\lambda - 1 + p_d - (\sqrt{p_d p_u})m_2(\lambda)^{-1}}{\lambda - 1 + p_d - (\sqrt{p_d p_u})m_1(\lambda)^{-1}}\right) a_2(\lambda)$$

Hence, λ is an eigenvalue if and only if

(F33)

$$\left(\frac{\lambda - 1 + p_u - (\sqrt{p_d p_u})m_2(\lambda)}{\lambda - 1 + p_u - (\sqrt{p_d p_u})m_1(\lambda)}\right) = \left(\frac{m_2(\lambda)}{m_1(\lambda)}\right)^N \left(\frac{\lambda - 1 + p_d - (\sqrt{p_d p_u})m_2(\lambda)^{-1}}{\lambda - 1 + p_d - (\sqrt{p_d p_u})m_1(\lambda)^{-1}}\right)$$

Recall that when $\lambda \in (\bar{\lambda}, 1)$, $m_1(\lambda) > 1 > m_2(\lambda)$. Hence, the left-hand side of (F33) is larger than 1. But, the right-hand side is smaller than 1 for the same reason. Hence, there are no eigenvalues in the interval $\lambda \in (\bar{\lambda}, 1)$, when N is finite. A similar argument can be used to rule out eigenvalues smaller than $\underline{\lambda}$. We thus have the following proposition

Proposition F2: For $N < \infty$, all eigenvalues of the operator $\mathbb{S}$ that are less than 1 lie in the interval $(\underline{\lambda}, \bar{\lambda})$.

Since $\mathbb{S}$ and $\mathbb{T}$ have the same eigenvalues when $N < \infty$, the proposition holds for the operator $\mathbb{T}$ as well.

Next, we show how to find the $N + 1$ eigenvalues. Since all eigenvalues lie in the interval $(\underline{\lambda}, \bar{\lambda})$, we know that for a given eigenvalue λ, the eigenvectors are of of the form (F31) corresponding to complex roots of the characteristic equation associated with the difference equation defining $\mathbb{S}$:

$$h(n;\lambda) = a(\lambda)\cos\left(\theta(\lambda)n + \omega(\lambda)\right)$$

To pin down $\omega(\lambda)$ for a given eigenvalue λ, we use the lower boundary condition (F14) and the fact that $\cos\left(x+y\right) = \cos(y)\cos(x) - \sin(y)\sin(x)$. This gives us

$$0 = (1-p_u-\lambda)h(0;\lambda) + (\sqrt{p_u p_d})h(1;\lambda)$$
$$\Leftrightarrow$$
$$0 = (1-p_u-\lambda)\cos(\omega(\lambda)) + (\sqrt{p_u p_d})\left[\cos(\omega(\lambda))\cos(\theta(\lambda)) - \sin(\omega(\lambda))\sin(\theta(\lambda))\right]$$

This condition can in turn be written

$$\frac{\lambda - 1 + p_u - p_d}{2\sqrt{p_u p_d}} = -\sin(\theta(\lambda))\tan(\omega(\lambda))$$

by using (F32). Moreover, note that

$$\sin(\theta(\lambda)) = \left(1 - \left(\frac{(\lambda - 1 + p_u + p_d)}{2\sqrt{p_d p_u}}\right)^2\right)^{1/2}$$

since $\theta(\lambda) = \cos^{-1}\left(\frac{(\lambda-1+p_u+p_d)}{2\sqrt{p_d p_u}}\right)$, so we can solve for $\omega(\lambda)$ as:

$$\omega(\lambda) = \arctan\left(-\frac{\left(\frac{\lambda-1+p_u-p_d}{2\sqrt{p_u p_d}}\right)}{\left(1-\left(\frac{(\lambda-1+p_u+p_d)}{2\sqrt{p_d p_u}}\right)^2\right)^{1/2}}\right) \tag{F34}$$

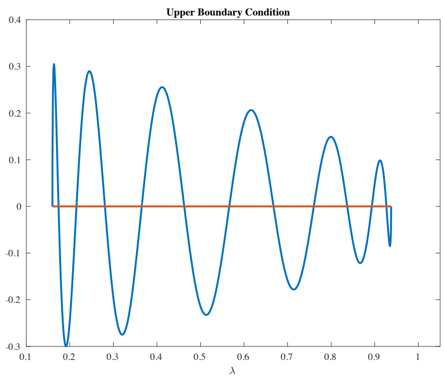
We can then find all eigenvalues as solutions to the upper boundary condition (F15) plugging in the above expressions for $\theta(\lambda)$ and $\omega(\lambda)$:

$$(\lambda - 1 + p_d)\cos\left(\theta(\lambda)N + \omega(\lambda)\right) - (\sqrt{p_u p_d})\cos\left(\theta(\lambda)(N-1) + \omega(\lambda)\right) = 0 \tag{F35}$$

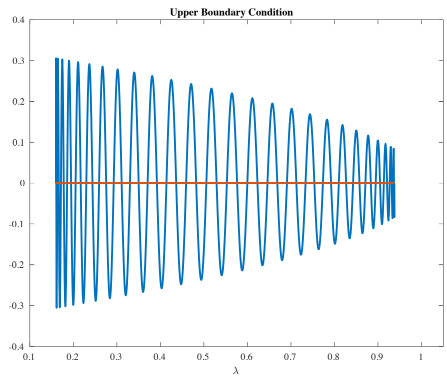
where $\theta(\lambda)$ and $\omega(\lambda)$ are given by (F32) and (F34), respectively. In figure F1, we plot the left-hand side of (F35) for increasing N. The eigenvalues are the points at which the left-hand side of (F35) is equal to zero.

We see that as N grows, the eigenvalues successively fill out the entire interval $(\underline{\lambda}, \bar{\lambda})$. This shows that the second largest eigenvalue in a model with a finite grid approaches $\bar{\lambda}$ as the size of the grid grows.

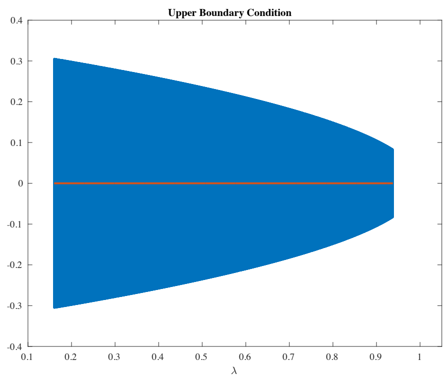
In conclusion, an upper bound on the second largest eigenvalue of the operator $\mathbb{T}$ is given by $\bar{\lambda}$. To relate this to Gabaix et al. (2016) we compute the continuous



(a) $N = 10$



(b) $N = 50$



(c) $N = 2000$

Figure F1. : Roots of the upper boundary condition

Note: Roots of the upper boundary condition for $N = 10$, $N = 50$, and $N = 2000$. The red line indicates the interval $(\underline{\lambda}, \bar{\lambda})$

time analogue of $\bar{\lambda}$:

$$\lim_{\Delta_t \to 0} -\frac{1}{\Delta_t} \log(\bar{\lambda})$$

and show that it is equal to $\frac{\mu^2}{2\sigma^2}$ which is the same value they obtain. To show this we first rewrite $\bar{\lambda}$ in terms of the annualized moments μ and σ using equations (E1) and (E2), and then apply L'Hôpital's rule. Specifically, we can rewrite $\bar{\lambda}$ as

$$
\begin{aligned}
\bar{\lambda} &= 1 - p_u - p_d + 2\sqrt{p_d p_u} \\
&= 1 - \left(\frac{\Delta_t}{\Delta^2}\sigma^2 + \frac{\Delta_t^2}{\Delta^2}\mu^2\right) + 2\sqrt{\frac{1}{4}\left(\left(\frac{\Delta_t}{\Delta^2}\sigma^2 + \frac{\Delta_t^2}{\Delta^2}\mu^2\right) + \frac{\Delta_t}{\Delta}\mu\right)\left(\left(\frac{\Delta_t}{\Delta^2}\sigma^2 + \frac{\Delta_t^2}{\Delta^2}\mu^2\right) - \frac{\Delta_t}{\Delta}\mu\right)} \\
&= 1 - \left(\frac{\Delta_t}{\Delta^2}\sigma^2 + \frac{\Delta_t^2}{\Delta^2}\mu^2\right) + \sqrt{\left(\frac{\Delta_t}{\Delta^2}\sigma^2 + \frac{\Delta_t^2}{\Delta^2}\mu^2\right)^2 - \frac{\Delta_t^2}{\Delta^2}\mu^2} \\
&= 1 - \left(c\sigma^2 + c\Delta_t\mu^2\right) + \sqrt{(c\sigma^2 + c\Delta_t\mu^2)^2 - c\Delta_t\mu^2}
\end{aligned}
$$

where $c = \frac{1}{2\sigma_{max}^2}$ is a constant. To use L'Hôpital's rule we need to compute $\frac{d}{d\Delta_t}\log(\bar{\lambda})$, which is given by

$$
\frac{d}{d\Delta_t}\log(\bar{\lambda}) = \frac{-c\mu^2 + \frac{2(c\sigma^2 + c\Delta_t\mu^2)c\mu^2 - c\mu^2}{2\sqrt{(c\sigma^2 + c\Delta_t\mu^2)^2 - c\Delta_t\mu^2}}}{1 - (c\sigma^2 + c\Delta_t\mu^2) + \sqrt{(c\sigma^2 + c\Delta_t\mu^2)^2 - c\Delta_t\mu^2}}
$$

Letting $\Delta_t \to 0$ we have

$$
\frac{d}{d\Delta_t}\log(\bar{\lambda}) \to \frac{-c\mu^2 + \frac{2c^2\sigma^2\mu^2 - c\mu^2}{2c\sigma^2}}{1} = \frac{-\mu^2}{2\sigma^2}
$$

So by L'Hôpital's

$$
\lim_{\Delta_t \to \infty} -\frac{1}{\Delta_t}\log(\bar{\lambda}) = \lim_{\Delta_t \to \infty} -\frac{\frac{d}{d\Delta_t}\log(\bar{\lambda})}{1} = \frac{\mu^2}{2\sigma^2}
$$

which is what we wanted to show.