

Persistent Heterogeneous Returns and Top End Wealth Inequality: Online Appendix

Dan Cao and Wenlan Luo

A Continuous Time Derivations

Proof of Lemma 1. By the Principle of Optimality and the fact that after an infinitesimal time interval Δt , the survival probability of an entrepreneur is $e^{-\bar{\lambda}\Delta t}$, we have

$$V(t, i_t, x_t, w_t) = \max_{c_t, k_t, b_t} u(c_t)\Delta t + e^{-(\rho + \bar{\lambda})\Delta t} \mathbb{E}_t [V(t + \Delta t, i_{t+\Delta t}, x_{t+\Delta t}, w_{t+\Delta t})] + o(\Delta t) \quad (\text{A.1})$$

subject to $k_t \geq 0$ and $mk_t + b_t + Q_t \geq 0$, and

$$w_{t+\Delta t} = w_t + ((1 - \tau_{i_t})R_{i_t}(e_t)k_t + r_t b_t + e_t x_t - c_t)\Delta t + o(\Delta t).$$

Furthermore, under the Poisson arrival rates of idiosyncratic shocks

$$\begin{aligned} \mathbb{E}_t [V(t + \Delta t, i_{t+\Delta t}, x_{t+\Delta t}, w_{t+\Delta t})] &= e^{-\lambda_{i_t, -i_t}\Delta t} V(t + \Delta t, i_t, x_{t+\Delta t}, w_{t+\Delta t}) \\ &\quad + (1 - e^{-\lambda_{i_t, -i_t}\Delta t}) V(t + \Delta t, -i_t, x_{t+\Delta t}, w_{t+\Delta t}) + o(\Delta t). \end{aligned}$$

Using the fact that $e^z = 1 + z + o(z)$ for z close to 0, we rewrite the previous expression as

$$\begin{aligned} \mathbb{E}_t [V(t + \Delta t, i_{t+\Delta t}, x_{t+\Delta t}, w_{t+\Delta t})] &= (1 - \lambda_{i_t, -i_t}\Delta t) V(t + \Delta t, i_t, x_{t+\Delta t}, w_{t+\Delta t}) \\ &\quad + \lambda_{i_t, -i_t}\Delta t V(t + \Delta t, -i_t, x_{t+\Delta t}, w_{t+\Delta t}) + o(\Delta t). \end{aligned}$$

Plugging this into equation (A.1), we obtain

$$\begin{aligned} V(t, i_t, x_t, w_t) &= \max_{c_t, k_t, b_t} \{u(c_t)\Delta t + V(t + \Delta t, i_t, x_{t+\Delta t}, w_{t+\Delta t}) - (\rho + \bar{\lambda})\Delta t V(t + \Delta t, i_t, x_{t+\Delta t}, w_{t+\Delta t}) \\ &\quad + \lambda_{i_t, -i_t}\Delta t (V(t + \Delta t, -i_t, x_{t+\Delta t}, w_{t+\Delta t}) - V(t + \Delta t, i_t, x_{t+\Delta t}, w_{t+\Delta t})) + o(\Delta t)\}. \end{aligned}$$

Subtracting both sides by $V(t + \Delta t, i_t, x_t, w_{t+\Delta t})$ and dividing both sides by Δt then take the limit $\Delta t \rightarrow 0$, noticing that

$$\lim_{\Delta t \rightarrow 0} \frac{V(t, i_t, x_t, w_t) - V(t + \Delta t, i_t, x_{t+\Delta t}, w_{t+\Delta t})}{\Delta t} = -\frac{\partial V}{\partial t} - \frac{\partial V}{\partial w} \frac{dw_t}{dt} - \frac{\partial V}{\partial x} g_x x_t,$$

we obtain (6). □

Proof of Lemma 2. The total wealth conditional on investment productivity type, $W_{i,t}$ is the sum of wealth across agents $h \in [0, N_t]$ whose type equal to i :

$$\begin{aligned} W_{i,t+\Delta t} &= \int_{h:i_t^h=i} w_{t+\Delta t}^h dt \\ &= \int_{h \leq N_t: i_{t+\Delta t}^h=i, i_t^h=i} w_{t+\Delta t}^h dt + \int_{h \leq N_t: i_{t+\Delta t}^h=i, i_t^h=-i} w_{t+\Delta t}^h dt \\ &\quad + \int_{h \geq N_t: i_{t+\Delta t}^h=i} w_{t+\Delta t}^h dt + o(\Delta t). \end{aligned}$$

We write the first term as

$$\begin{aligned} &\int_{h \leq N_t: i_{t+\Delta t}^h=i, i_t^h=i} w_{t+\Delta t}^h dt \\ &= \int_{h \leq N_t: i_{t+\Delta t}^h=i, i_t^h=i, j_t^h=1} \left((w_t^h + Q_t^h) (1 + \Delta t ((1 - \tau_i) R_i(e_t) k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^*)) - Q_{t+\Delta t}^h \right) dt \\ &\quad + \int_{h \leq N_t: i_{t+\Delta t}^h=i, i_t^h=i, j_t^h=0} w_{t+\Delta t}^h dt + o(\Delta t) \\ &= e^{-\lambda_{i,-i}\Delta t} e^{-\bar{\lambda}\Delta t} \int_{h \leq N_t: i_t^h=i} \left((w_t^h + Q_t^h) (1 + \Delta t ((1 - \tau_i) R_i(e_t) k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^*)) - Q_{t+\Delta t}^h \right) dt \\ &\quad + (1 - e^{-\bar{\lambda}\Delta t}) M_i N_t Z_t / ((\bar{\lambda} + n) N_t) + o(\Delta t), \end{aligned}$$

where

$$\begin{aligned} &\int_{h \leq N_t: i_t^h=i} \left((w_t^h + Q_t^h) (1 + \Delta t ((1 - \tau_i) R_i(e_t) k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^*)) - Q_{t+\Delta t}^h \right) dt \\ &= W_{i,t} + q_t M_i G_t N_t \\ &\quad + \Delta t ((1 - \tau_i) R_i(e_t) k_{i,t}^* + b_{i,t}^* r_t - c_{i,t}^*) (W_{i,t} + q_t M_i G_t N_t) - \left(q_t + \frac{dq_t}{dt} \Delta t \right) M_i G_t N_t (1 + g_x \Delta t) + o(\Delta t). \end{aligned}$$

Therefore,

$$\begin{aligned}
& \int_{h \leq N_t: i_{t+\Delta t}^h = i, i_t^h = i} w_{t+\Delta t}^h dt \\
&= W_{i,t} + \Delta t ((1 - \tau_i) R_i(e_t) k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^*) (W_{i,t} + q_t M_i G_t N_t) - \frac{dq_t}{dt} M_i G_t N_t \Delta t - q_t M_i G_t N_t g \Delta t \\
&- (\lambda_{i,-i} + \bar{\lambda}) \Delta t W_{i,t} + \bar{\lambda} \Delta t M_i \frac{Z_t}{\bar{\lambda} + n} + o(\Delta t).
\end{aligned}$$

We write the second term as

$$\begin{aligned}
& \int_{h \leq N_t: i_{t+\Delta t}^h = i, i_t^h = -i} w_{t+\Delta t}^h dt \\
&= \int_{h \leq N_t: i_{t+\Delta t}^h = i, i_t^h = -i, j_t^h = 1} \left((w_t^h + Q_t^h) (1 + \Delta t ((1 - \tau_{-i}) R_{-i}(e_t) k_{-i,t}^* + r_t b_{-i,t}^* - c_{-i,t}^*)) - Q_{t+\Delta t}^h \right) dt \\
&= (1 - e^{-\lambda_{-i,i} \Delta t}) e^{-\bar{\lambda} \Delta t} \\
& \int_{h \leq N_t: i_t^h = -i} \left((w_t^h + Q_t^h) (1 + \Delta t ((1 - \tau_{-i}) R_{-i}(e_t) k_{-i,t}^* + r_t b_{-i,t}^* - c_{-i,t}^*)) - Q_{t+\Delta t}^h \right) dt
\end{aligned}$$

where

$$\begin{aligned}
& \int_{h \leq N_t: i_t^h = -i} \left((w_t^h + Q_t^h) (1 + \Delta t ((1 - \tau_{-i}) R_{-i}(e_t) k_{-i,t}^* + r_t b_{-i,t}^* - c_{-i,t}^*)) - Q_{t+\Delta t}^h \right) dt \\
&= W_{-i,t} + o(1).
\end{aligned}$$

Therefore

$$\int_{h \leq N_t: i_{t+\Delta t}^h = i, i_t^h = -i} w_{t+\Delta t}^h dt = \lambda_{-i,i} \Delta t W_{-i,t} + o(\Delta t).$$

And lastly,

$$\begin{aligned}
\int_{h \geq N_t: i_{t+\Delta t}^h = i} w_{t+\Delta t}^h dt &= M_i (N_{t+\Delta t} - N_t) Z_t / ((\bar{\lambda} + n) N_t) \\
&= M_i n \Delta t N_t Z_t / ((\bar{\lambda} + n) N_t) + o(\Delta t).
\end{aligned}$$

Therefore

$$\begin{aligned}
W_{i,t+\Delta t} - W_{i,t} &= \Delta t((1 - \tau_i)R_i(e_t)k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^*) (W_{i,t} + q_t M_i G_t N_t) \\
&\quad - (\lambda_{i,-i} + \bar{\lambda}) \Delta t W_{i,t} - \frac{dq_t}{dt} M_i M_x G_t N_t \Delta t - q_t M_i G_t N_t g \Delta t \\
&\quad + N_t \bar{\lambda} \Delta t M_i Z_t / ((\bar{\lambda} + n) N_t) + \lambda_{-i,i} \Delta t W_{-i,t} \\
&\quad + M_i n \Delta t N_t Z_t / ((\bar{\lambda} + n) N_t) + o(\Delta t).
\end{aligned}$$

Dividing both sides by Δt and taking the limit $\Delta t \rightarrow 0$,

$$\begin{aligned}
\frac{dW_{i,t}}{dt} &= ((1 - \tau_i)R_i(e_t)k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^*) (W_{i,t} + q_t M_i G_t N_t) \\
&\quad - (\lambda_{i,-i} + \bar{\lambda}) W_{i,t} + \lambda_{-i,i} W_{-i,t} - \frac{dq_t}{dt} M_i G_t N_t - q_t M_i G_t N_t + M_i Z_t.
\end{aligned}$$

Rearranging, we arrive at (14). □

Proof of Lemma 3. To derive the partial differential equations for the wealth distributions, we use the following measure

$$\mathcal{M}_i(t + \Delta t, \omega, x) = \int_h \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega \text{ and } i_{t+\Delta t}^h = i \text{ and } \frac{x_{t+\Delta t}^h}{G_{t+\Delta t}} = x) dh$$

where $\mathbf{1}\{\cdot\}$ is the set indicator function. Since the population grows, we decompose the measure to two components, one with $h \in [0, N_t]$ and the other with $h \in (N_t, N_{t+\Delta t}]$:

$$\begin{aligned}
\int_h \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, x_{t+\Delta t}^h = x) dh &= \int_{h \leq N_t} \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, x_{t+\Delta t}^h = x) dh \\
&\quad + \int_{h > N_t} \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, x_{t+\Delta t}^h = x) dh
\end{aligned}$$

Conditioning on the type at t , we write the first term as

$$\begin{aligned}
&\int_{h \leq N_t} \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, x_{t+\Delta t}^h = x) dh \\
&= \int_{h \leq N_t} \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = i, \frac{x_t^h}{G_t} = x) dh \\
&\quad + \int_{h \leq N_t} \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = -i, \frac{x_t^h}{G_t} = x) dh
\end{aligned}$$

Conditional on whether the death shock hits between t and $t + \Delta t$, we obtain

$$\begin{aligned}
& \int_{h \leq N_t} \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = i, \frac{x_t^h}{G_t} = x) dh \\
&= \int_{h \leq N_t} \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = i, j_t^h = 0, \frac{x_t^h}{G_t} = x) dh \\
&+ \int_{h \leq N_t} \mathbf{1}(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = i, j_t^h = 1, \frac{x_{t+\Delta t}^h}{G_{t+\Delta t}} = x) dh
\end{aligned}$$

Let $\phi(\cdot)$ denote the pdf associated to the cdf $\Phi(\cdot)$. Expanding further, and we arrive at

$$\begin{aligned}
& \int_{h \leq N_t} \mathbf{1} \left(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = i, j_t^h = 0, \frac{x_t^h}{G_t} = x \right) dh \\
&= e^{-\lambda_{i,-i}\Delta t} e^{-\bar{\lambda}\Delta t} \int_{h \leq N_t} \mathbf{1} \left(\omega_t^h \left(\frac{1 + \Delta t \left((1 - \tau_i) R_i(e_t) k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^* \right)}{\frac{(W_t + Q_t)(1 + g_t \Delta t)}{N_t(1 + n \Delta t)}} \right) \geq \omega, i_t^h = i, \frac{x_t^h}{G_t} = x \right) dh \\
&= (1 - (\lambda_{i,-i} + \bar{\lambda}) \Delta t) \mathcal{M}_i \left(t, \frac{\omega}{1 + \Delta t \left((1 - \tau_i) R_i(e_t) k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^* - g_t + n \right)}, x \right) + o(\Delta t)
\end{aligned}$$

and

$$\begin{aligned}
& \int_{h \leq N_t} \mathbf{1} \left(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = i, j_t^h = 1, \frac{x_{t+\Delta t}^h}{G_{t+\Delta t}} = x \right) dh \\
&= e^{-\lambda_{i,-i}\Delta t} (1 - e^{-\bar{\lambda}\Delta t}) M_i N_t \int_{\psi} \mathbf{1} \left(\frac{\psi Z_t / ((\bar{\lambda} + n) N_t) + q_{t+\Delta t} x G_{t+\Delta t}}{Z_t / ((\bar{\lambda} + n) N_t)} \geq \omega \right) d\Gamma(\psi) \phi(x) \\
&= \bar{\lambda} \Delta t M_i N_t \Gamma \left(\frac{\omega \frac{(W_{t+\Delta t} + Q_{t+\Delta t})}{N_{t+\Delta t}} - q_{t+\Delta t} x G_{t+\Delta t}}{Z_t / ((\bar{\lambda} + n) N_t)} \right) \phi(x) + o(\Delta t) \\
&= \bar{\lambda} \Delta t M_i N_t \Gamma \left(\frac{\omega \frac{W_t + Q_t}{N_t} - q_t x G_t}{Z_t / ((\bar{\lambda} + n) N_t)} \right) \phi(x) + o(\Delta t)
\end{aligned}$$

Similarly

$$\begin{aligned}
& \int_{h \leq N_t} \mathbf{1} \left(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = -i, \frac{x_{t+\Delta t}^h}{G_{t+\Delta t}} = x \right) dh \\
&= \int_{h \leq N_t} \mathbf{1} \left(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = -i, j_t^h = 0, \frac{x_t^h}{G_t} = x \right) dh
\end{aligned}$$

Expanding further, we arrive at

$$\begin{aligned}
& \int_{h \leq N_t} \mathbf{1} \left(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, i_t^h = -i, j_t^h = 0, \frac{x_t^h}{G_t} = x \right) dh \\
&= \left(1 - e^{-\lambda_{-i,i}\Delta t} \right) e^{-\bar{\lambda}\Delta t}. \\
& \int_{h \leq N_t} \mathbf{1} \left(\omega_t^h (1 + \Delta t ((1 - \tau_{-i})R_{-i}(e_t)k_{-i,t}^* + r_t b_{-i,t}^* - c_{-i,t}^*) - g_t \Delta t + n \Delta t) \geq \omega, i_t^h = -i, \frac{x_t^h}{G_t} = x \right) dh \\
&= \lambda_{-i,i} \Delta t \mathcal{M}_{-i} \left(t, \frac{\omega}{1 + \Delta t ((1 - \tau_{-i})R_{-i}(e_t)k_{-i,t}^* + r_t b_{-i,t}^* - c_{-i,t}^* - g_t + n)}, x \right) + o(\Delta t)
\end{aligned}$$

Lastly

$$\begin{aligned}
& \int_{h > N_t} \mathbf{1} \left(\omega_{t+\Delta t}^h \geq \omega, i_{t+\Delta t}^h = i, \frac{x_{t+\Delta t}^h}{G_{t+\Delta t}} = x \right) dh \\
&= M_i (N_{t+\Delta t} - N_t) \int_{\psi} \mathbf{1} \left(\frac{\psi Z_t / ((\bar{\lambda} + n)N_t) + q_{t+\Delta t} x G_{t+\Delta t}}{\frac{W_{t+\Delta t} + Q_{t+\Delta t}}{N_{t+\Delta t}}} \geq \omega \right) d\Gamma(\psi) \phi(x) \\
&= M_i \Delta t N_t \Gamma \left(\frac{\omega \frac{(W_{t+\Delta t} + Q_{t+\Delta t})}{N_{t+\Delta t}} - q_{t+\Delta t} x G_{t+\Delta t}}{Z_t / ((\bar{\lambda} + n)N_t)} \right) \phi(x) + o(\Delta t) \\
&= M_i n \Delta t N_t \Gamma \left(\frac{\omega \frac{W_t + Q_t}{N_t} - q_t x G_t}{Z_t / ((\bar{\lambda} + n)N_t)} \right) \phi(x) + o(\Delta t)
\end{aligned}$$

Therefore

$$\begin{aligned}
& \mathcal{M}_i(t + \Delta t, \omega, x) - \mathcal{M}_i(t, \omega, x) \\
&= -\Delta t \frac{\partial \mathcal{M}_i}{\partial \omega} \omega ((1 - \tau_i)R_i(e_t)k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^* - g_t + n) - (\lambda_{i,-i} + \bar{\lambda}) \Delta t \mathcal{M}_i(t, \omega, x) \\
&+ \bar{\lambda} \Delta t M_i N_t \Gamma \left(\frac{\omega \frac{W_t + Q_t}{N_t} - q_t x G_t}{Z_t / ((\bar{\lambda} + n)N_t)} \right) \phi(x) + \lambda_{-i,i} \Delta t \mathcal{M}_{-i}(t, \omega, x) \\
&+ M_i n \Delta t N_t \Gamma \left(\frac{\omega \frac{W_t + Q_t}{N_t} - q_t x G_t}{Z_t / ((\bar{\lambda} + n)N_t)} \right) \phi(x) + o(\Delta t).
\end{aligned}$$

Let $\mathcal{M}_i^0(t, \omega, x) = \frac{\mathcal{M}_i(t, \omega, x)}{M_i N_t}$, the last equation implies

$$\begin{aligned}
& \frac{\mathcal{M}_i^0(t + \Delta t, \omega, x) - \mathcal{M}_i^0(t, \omega, x)}{\Delta t} \\
&= \frac{\mathcal{M}_i^0(t + \Delta t, \omega, x) - \mathcal{M}_i^0(t, \omega, x)}{M_i N_t} - n \mathcal{M}_i^0(t, \omega, x) + o(1) \\
&= -\frac{\partial \mathcal{M}_i^0}{\partial \omega} \omega \left((1 - \tau_i) R_i(e_t) k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^* - g_t + n \right) - (\lambda_{i,-i} + \bar{\lambda}) \mathcal{M}_i^0(t, \omega, x) \\
&+ \bar{\lambda} \Gamma \left(\frac{\omega \frac{W_t + Q_t}{N_t} - q_t x G_t}{Z_t / ((\bar{\lambda} + n) N_t)} \right) \phi(x) + \lambda_{-i,i} \frac{M_{-i}}{M_i} \mathcal{M}_{-i}^0(t, \omega, x) \\
&+ n \Gamma \left(\frac{\omega \frac{W_t + Q_t}{N_t} - q_t x G_t}{Z_t / ((\bar{\lambda} + n) N_t)} \right) \phi(x) + o(1)
\end{aligned}$$

Taking the limit $\Delta t \rightarrow 0$, we obtain

$$\begin{aligned}
\frac{\partial \mathcal{M}_i^0}{\partial t} &= -\frac{\partial \mathcal{M}_i^0}{\partial \omega} \omega \left((1 - \tau_i) R_i(e_t) k_{i,t}^* + r_t b_{i,t}^* - c_{i,t}^* - g_t + n \right) - (\lambda_{i,-i} + \bar{\lambda}) \mathcal{M}_i^0 + \lambda_{-i,i} \frac{M_{-i}}{M_i} \mathcal{M}_{-i}^0 \\
&+ (\bar{\lambda} + n) \Gamma \left(\frac{\omega \frac{W_t + Q_t}{N_t} - q_t x G_t}{Z_t / ((\bar{\lambda} + n) N_t)} \right) \phi(x) - n \mathcal{M}_i^0.
\end{aligned}$$

Now integrate both-side in x and let $p_i(t, \omega) = \int \mathcal{M}_i^0(t, \omega, x) dx$, we arrive at (15). $\square$

To facilitate the analyses of the dynamic equilibrium in this economy, let X_t denote the total wealth share of agents with high investment productivity:

$$X_t = \frac{W_{H,t} + M_H Q_t}{W_t + Q_t}.$$

The following result characterizes the dynamics of g_t and X_t as functions of the agents' policy functions.

Lemma A.1. *In a competitive equilibrium, the dynamics of the growth rate of total wealth is given by*

$$\begin{aligned}
g_t &= (R_H(e_t) k_{H,t}^* + r_t^* b_{H,t}^* - c_{H,t}^*) X_t \\
&+ (R_L(e_t) k_{L,t}^* + r_t^* b_{L,t}^* - c_{L,t}^*) (1 - X_t) \\
&+ n \frac{Q_t}{W_t + Q_t}.
\end{aligned} \tag{A.2}$$

In addition, X_t satisfies

$$\begin{aligned}\frac{dX_t}{dt} &= (g_{H,t}^* - g_{L,t}^*)X_t(1 - X_t) - (\lambda_{HL} + \bar{\lambda})X_t + \lambda_{LH}(1 - X_t) \\ &\quad + (M_H - X_t)(\tau_H k_{H,t}^* R_H(e_t)X_t + \tau_L k_{L,t}^* R_L(e_t)(1 - X_t)) \\ &\quad + M_H \bar{\lambda} + n \frac{Q_t}{W_t + Q_t} (M_H - X_t).\end{aligned}\tag{A.3}$$

Proof. From the definition of $W_{i,t}$ we have

$$\frac{dW_t}{dt} = \sum_i (k_{i,t}^* R_i(e_t) + b_{i,t}^* r_t - c_{i,t}^*) (W_{i,t} + q_t M_i G_t N_t) - \frac{dq_t}{dt} G_t N_t - g_x q_t G_t N_t$$

and

$$\frac{dQ_t}{dt} = \frac{dq_t}{dt} G_t N_t + g_x Q_t + n Q_t.$$

Therefore

$$\frac{dW_t}{dt} + \frac{dQ_t}{dt} = \sum_i (k_{i,t}^* R_i(e_t) + b_{i,t}^* r_t - c_{i,t}^*) (W_{i,t} + q_t M_i G_t N_t) + n Q_t.$$

Dividing both sides by $W_t + Q_t$, we obtain

$$g_t = \sum_i (k_{i,t}^* R_i(e_t) + b_{i,t}^* r_t - c_{i,t}^*) \frac{W_{i,t} + q_t M_i G_t N_t}{W_t + Q_t} + n \frac{Q_t}{W_t + Q_t},$$

which is equivalent to (A.2).

Now, we turn to the dynamics of X_t . From the definition of X_t , we have

$$dX_t = \frac{\frac{dW_{H,t}}{dt} + M_H \frac{dQ_t}{dt}}{W_t + Q_t} - X_t g_t.$$

From the dynamics of $W_{H,t}$

$$\begin{aligned}\frac{dW_{H,t}}{dt} \frac{1}{W_t + Q_t} &= ((1 - \tau_H) R_H(e_t) k_{H,t}^* + r_t b_{H,t}^* - c_{H,t}^*) X_t - (\lambda_{HL} + \bar{\lambda}) \frac{W_{H,t}}{W_t + Q_t} \\ &\quad + \lambda_{LH} \frac{W_{L,t}}{W_t + Q_t} + M_H \frac{Z_t}{W_t + Q_t} \\ &\quad - \frac{\frac{dq_t}{dt} M_H G_t N_t + g_x q_t M_H G_t N_t}{W_t + Q_t}.\end{aligned}$$

Therefore,

$$\begin{aligned} \frac{dW_{H,t}}{dt} \frac{1}{W_t + Q_t} &= ((1 - \tau_H)R_H(e_t)k_{H,t}^* + r_t b_{H,t}^* - c_{H,t}^*)X_t \\ &\quad - (\lambda_{HL} + \bar{\lambda}) \frac{W_{H,t} + M_H Q_t}{W_t + Q_t} + \lambda_{LH} \frac{W_{L,t} + M_L Q_t}{W_t + Q_t} \\ &\quad + M_H \frac{Z_t}{W_t + Q_t} - \frac{\frac{dq_t}{dt} M_H G_t N_t + g_x q_t M_H G_t N_t}{W_t + Q_t} + \bar{\lambda} \frac{M_H Q_t}{W_t + Q_t}. \end{aligned}$$

So

$$\begin{aligned} \frac{\frac{d}{dt} W_{H,t} + M_H \frac{dQ_t}{dt}}{W_t + Q_t} &= ((1 - \tau_H)R_H(e_t)k_{H,t}^* + r_t b_{H,t}^* - c_{H,t}^*)X_t - (\lambda_{HL} + \bar{\lambda})X_t + \lambda_{LH}(1 - X_t) \\ &\quad + M_H (\tau_H k_{H,t}^* R_H(e_t)X_t + \tau_L k_{L,t}^* R_L(e_t)(1 - X_t)) \\ &\quad + M_H \bar{\lambda} - \bar{\lambda} \frac{M_H Q_t}{W_t + Q_t} + (\bar{\lambda} + n) \frac{M_H Q_t}{W_t + Q_t}. \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{dX_t}{dt} &= g_{H,t}^* X_t - (\lambda_{HL} + \bar{\lambda}) X_t + \lambda_{LH}(1 - X_t) \\ &\quad + M_H (\tau_H k_{H,t}^* R_H(e_t)X_t + \tau_L k_{L,t}^* R_L(e_t)(1 - X_t)) \\ &\quad + M_H \bar{\lambda} - \bar{\lambda} \frac{M_H Q_t}{W_t + Q_t} + (\bar{\lambda} + n) \frac{M_H Q_t}{W_t + Q_t}. \end{aligned}$$

From the expression for g_t in (A.2) and the definition of $g_{i,t}^*$ above, we obtain

$$g_{H,t}^* = -\tau_H R_H(e_t)k_{H,t}^* X_t - \tau_L R_L(e_t)k_{L,t}^* (1 - X_t) + (g_{H,t}^* - g_{L,t}^*) (1 - X_t) - n \frac{Q_t}{W_t + Q_t}. \quad (\text{A.4})$$

Plugging this into the previous expression for $\frac{dX_t}{dt}$, we obtain (A.3). $\square$

B General CRRA Utility Functions

In this online appendix, we show that most of the results for log utility extend to general CRRA utility functions

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma},$$

where $\sigma \neq 1$. In this case, the solution to the HJB equation (6) has the form:

$$V^h(t, i_t^h, x_t^h, w_t^h) = v(t, i_t^h) \left(w_t^h + x_t^h q_t \right)^{1-\sigma}$$

and $v(.,.)$ satisfies:

$$\begin{aligned} (\rho + \bar{\lambda})v(t, i_t^h) - \frac{\partial v(t, i_t^h)}{\partial t} = \max_{c, k, b} \mathcal{H}_{i_t^h}(c, k, \tilde{b}; t, v(t, i_t^h)) \\ + \lambda_{i_t^h, -i_t^h}(v(t, -i_t^h) - v(t, i_t^h)) \end{aligned} \quad (\text{B.1})$$

where

$$\mathcal{H}_i(c, k, \tilde{b}; t, v) = u(c) + (1 - \sigma)v((1 - \tau_i)R_i(e_t)k + r_t \tilde{b} - c) \quad (\text{B.2})$$

and the maximization problem is subject to (10).

In a BGP, the value functions in (B.1) become $v(t, i) \equiv \bar{v}_i$, where $\bar{v}_i$ satisfies

$$(\rho + \bar{\lambda})\bar{v}_i = \max_{c, k, b} \mathcal{H}_i(c, k, \tilde{b}; \bar{v}_i) + \lambda_{i, -i}(\bar{v}_{-i} - \bar{v}_i) \quad (\text{B.3})$$

where

$$\tilde{\mathcal{H}}_i(c, k, \tilde{b}; v) = u(c) + (1 - \sigma)v((1 - \tau_i)R_i(\bar{e})k + \bar{r} \tilde{b} - c) \quad (\text{B.4})$$

and the maximization is subject to the constraints (10).

Theorem 1 extends to this model, including the last result on saving rate as shown in the following lemma.

Lemma B.1. *In Case 1 and Case 2 in the proof of Theorem 1,*

$$\frac{(1 - \tau_H)R_H(\bar{e}) - m\bar{r}}{1 - m} - \bar{c}_H^* > \bar{r} - \bar{c}_L^*. \quad (\text{B.5})$$

Proof. We show this inequality for two cases separately: $\sigma < 1$, and $\sigma > 1$.

Case A $\sigma < 1$: We first show that $\bar{v}_H > \bar{v}_L > 0$. Indeed, from the definition of value functions and from the fact that $u(c) > 0$ for all $c > 0$, we have $\bar{v}_H, \bar{v}_L > 0$. Assume by contradiction that $\bar{v}_H \leq \bar{v}_L$. From the first-order condition in c in (17), we obtain

$$(\bar{c}_i^*)^{-\sigma} = (1 - \sigma)\bar{v}_i. \quad (\text{B.6})$$

Plugging this into (B.3) and let $\bar{r}_i^* = (1 - \tau_i)R_i(\bar{e})\bar{k}_i^* + \bar{r}\bar{b}_i^*$, (since we are in Case 1 or Case

2, $\bar{r}_H^* > \bar{r}_L^*$) we rewrite (B.3) as

$$\rho + \bar{\lambda} + \lambda_{i,-i} \left(1 - \frac{\bar{v}_{-i}}{\bar{v}_i} \right) - \sigma \bar{c}_i^* = (1 - \sigma) \bar{r}_i^* \quad (\text{B.7})$$

Because $0 < \bar{v}_H \leq \bar{v}_L$, $\frac{\bar{v}_L}{\bar{v}_H} \geq 1 \geq \frac{\bar{v}_H}{\bar{v}_L}$. In addition, (B.6) implies that $\bar{c}_H^* \geq \bar{c}_L^*$. Therefore,

$$\rho + \bar{\lambda} + \lambda_{HL} \left(1 - \frac{\bar{v}_L}{\bar{v}_H} \right) - \sigma \bar{c}_H^* < \rho + \bar{\lambda} + \lambda_{LH} \left(1 - \frac{\bar{v}_H}{\bar{v}_L} \right) - \sigma \bar{c}_L^*,$$

which, together with (B.7) and $\sigma < 1$, contradicts the fact that $\bar{r}_H^* > \bar{r}_L^*$. Therefore by contradiction $\bar{v}_H > \bar{v}_L$. (B.6) implies that $\bar{c}_H^* < \bar{c}_L^*$. (B.5) then immediately follows from $\bar{r}_H^* > \bar{r}_L^*$.

Case B $\sigma > 1$: As for the case $\sigma < 1$, we first show that $\bar{v}_L < \bar{v}_H < 0$. From the definition of value functions and from the fact that $u(c) < 0$ for all $c > 0$, we have $\bar{v}_H, \bar{v}_L < 0$. Assume by contradiction that $\bar{v}_H \leq \bar{v}_L$. Because $\bar{v}_H \leq \bar{v}_L < 0$, (B.6) implies that $\bar{c}_H^* \leq \bar{c}_L^*$. So

$$\rho + \bar{\lambda} + \lambda_{HL} \left(1 - \frac{\bar{v}_L}{\bar{v}_H} \right) - \sigma \bar{c}_H^* \geq \rho + \bar{\lambda} + \lambda_{LH} \left(1 - \frac{\bar{v}_H}{\bar{v}_L} \right) - \sigma \bar{c}_L^*,$$

which, together with (B.7) and $\sigma > 1$, contradicts the fact that $\bar{r}_H^* > \bar{r}_L^*$. Therefore by contradiction $\bar{v}_H > \bar{v}_L$.¹ We rewrite (B.3) as

$$\rho + \bar{\lambda} + \lambda_{i,-i} \left(1 - \frac{\bar{v}_{-i}}{\bar{v}_i} \right) - \bar{r}_i^* = -\sigma (\bar{r}_i^* - \bar{c}_i^*). \quad (\text{B.8})$$

Because $\bar{v}_L < \bar{v}_H < 0$, $\frac{\bar{v}_L}{\bar{v}_H} > 1 > \frac{\bar{v}_H}{\bar{v}_L}$. In addition $\bar{r}_H^* > \bar{r}_L^*$, therefore

$$\rho + \bar{\lambda} + \lambda_{HL} \left(1 - \frac{\bar{v}_L}{\bar{v}_H} \right) - \bar{r}_H^* < \rho + \bar{\lambda} + \lambda_{LH} \left(1 - \frac{\bar{v}_H}{\bar{v}_L} \right) - \bar{r}_L^*.$$

This inequality, together with (B.3) and with the fact that $\sigma > 0$, implies $\bar{r}_H^* - \bar{c}_H^* > \bar{r}_L^* - \bar{c}_L^*$. $\square$

It is easy to see that Theorem 1, Theorem 2 and Theorem 3 (Part 1) extend easily to cases with general CRRA utility functions. While other results rely more crucially on log utility function.

¹(B.6) then implies that $\bar{c}_H^* > \bar{c}_L^*$.

C AK Growth Model

In this section we assume that $u(c) \equiv \log c$ and $F_i(k, l) = A_i k$ where $A_H > A_L > 0$. We further assume type-preserving wealth redistribution, and no population growth. These assumptions allow us to sharply characterize the balanced growth paths.

Theorem 1 does not tell us which case happens depending on the exogenous parameters of the model. Under log utility, the following proposition completely characterizes the equilibrium.²

Proposition C.1. *In a stationary BGP, one of the following three cases happens:*

Case 1: (Low m) If

$$\frac{M_L}{1 + \frac{A_H - A_L}{\lambda_{HL} + \lambda_{LH}}} > m > 0, \quad (\text{C.1})$$

then $\bar{r} = A_L$ and $X^* < 1 - m$.

Case 2: (Intermediate m) If

$$M_L > m > \frac{M_L}{1 + \frac{A_H - A_L}{\lambda_{HL} + \lambda_{LH}}}, \quad (\text{C.2})$$

then

$$\bar{r} = A_H - \frac{(1 - m)\lambda_{HL} - m\lambda_{LH}}{m} \in (A_L, A_H)$$

and $X^* = 1 - m$.

Case 3: (High m) If

$$1 > m > M_L, \quad (\text{C.3})$$

then $r^* = A_H$ and $X^* > 1 - m$

Proof. Because $u(c) = \log c$. From the HJB equation, (17), the first order condition in consumption implies that $\bar{c}_i^* = \rho + \bar{\lambda}$.

We look at three cases in Theorem 1 and derive the conditions that determine which case actually happens in equilibrium for each set of exogenous parameters.

Case 1: In this case $\bar{r} = A_L < A_H$. Therefore $(\bar{k}_H^*, \bar{b}_H^*) = \left(\frac{1}{1-m}, -\frac{m}{1-m}\right)$. From the bond market clearing condition (20),

$$-\frac{m}{1-m}X^* + \bar{b}_L^*(1 - X^*) = 0$$

Therefore, $\bar{b}_L^* = \frac{m}{1-m} \frac{X^*}{1-X^*}$ and $\bar{k}_L^* = 1 - \bar{b}_L^*$. Because $\bar{k}_L^* \geq 0$, $X^* \leq 1 - m$.

²One minor difference relative to Theorem 1 is that the growth rate of the economy is now totally endogenous, instead of being determined exogenously by g_x and n .

Equation (A.3) becomes:

$$0 = \frac{A_H - A_L}{1 - m} X^* (1 - X^*) + (1 - X^*) \lambda_{LH} - X^* \lambda_{HL}. \quad (\text{C.4})$$

Let $f(x)$ denote the right hand-side (with x standing for X^*). $f(x)$ is quadratic in x and the coefficient on the leading term, x^2 , is negative. In addition $f(0) > 0 > f(1)$. Therefore, there exists a unique $X^* \in (0, 1)$ that solves $f(X^*) = 0$ (the other root is negative).

In order for $X^* < 1 - m$, it is necessary and sufficient that $f(1 - m) < 0$, which is equivalent to (C.1).

Case 2: In this case $r^* \in (A_L, A_H)$. Therefore $(\bar{k}_H^*, \bar{b}_H^*) = \left(\frac{1}{1-m}, -\frac{m}{1-m}\right)$ and $(\bar{k}_L^*, \bar{b}_L^*) = (0, 1)$. The bond market clearing condition (20) implies that $X^* = 1 - m$. Together with equation (C.4), we obtain

$$0 = (A_H - \bar{r}) m - (1 - m) \lambda_{HL} + m \lambda_{LH}.$$

Or

$$A_H - \bar{r} = \frac{(1 - m) \lambda_{HL} - m \lambda_{LH}}{m} < A_H - A_L,$$

since $\bar{r} > A_L$. This inequality is equivalent to (C.2).

Case 3: In this case $\bar{r} = A_H$. Following the steps for Case 1, we arrive at (C.3). $\square$

The case of log utility also allows us to determine in closed forms the Pareto tail index of the stationary wealth distribution which is a solution to a quadratic equation similar to (28). Therefore, we can also characterize how the degree of financial friction affects the tail index, or equivalently top end wealth inequality.

Proposition C.2 (Financial Kutznet's Curve). *In stationary BGPs, top end wealth inequality varies in m differently depending on which case in Proposition C.1 the equilibrium belongs to:*

Case 1 (Low m): Top end wealth inequality is increasing in m .

Case 2 (Intermediate m): Top end wealth inequality is decreasing in m .

Case 3 (High m): Top end wealth inequality is independent of m .

Proof. By Theorem 3, the tail index is a function of $\bar{g}_L^*$ (and $\lambda_{HL}, \lambda_{LH}$).

Case 1: As argued above, to show that $\frac{d\eta_1}{dm} > 0$ in this case, we just need to show $\frac{\partial \bar{g}_L^*}{\partial m} > 0$.

Indeed, from the portfolio choices in Case 1 of Theorem 1 and by definition, the growth rate of the economy is given by

$$\bar{g}^* = X^* \left(\frac{A_H - mA_L}{1 - m} - \rho - \bar{\lambda} \right) + (1 - X^*) (A_L - \rho - \bar{\lambda}).$$

Therefore

$$\bar{g}_L^* = A_L - \rho - \bar{\lambda} - \bar{g}^* = -\frac{A_H - A_L}{1 - m} X^*.$$

Differentiating the last expression with respect to m , we get

$$\frac{d\bar{g}_L^*}{dm} = -\frac{A_H - A_L}{(1 - m)^2} X^* - \frac{A_H - A_L}{1 - m} \frac{dX^*}{dm}.$$

In addition, from the equation that determines X^* , (C.4), using the Implicit Function Theorem, we obtain

$$\frac{dX^*}{dm} = -\frac{\frac{A_H - A_L}{(1 - m)^2} X^* (1 - X^*)}{\frac{A_H - A_L}{1 - m} (1 - 2X^*) - (\lambda_{LH} + \lambda_{HL})} > 0,$$

where the inequality comes from the fact that X^* is the higher root of (C.4), and thus, $\frac{A_H - A_L}{1 - m} (1 - 2X^*) - (\lambda_{LH} + \lambda_{HL}) < 0$.

Therefore, $\frac{d\bar{g}_L^*}{dm} < 0$. By Theorem 4 (Part 1), $\frac{d\eta_1}{dm} > 0$.

Case 2: Similar to Case 1, to show that $\frac{d\eta_1}{dm} < 0$ in this case, we just need to show $\frac{d\bar{g}_L^*}{dm} > 0$.

From the portfolio choices in Case 2 of Theorem 1 and by definition, the growth rate of the economy is given by

$$\bar{g}^* = X^* \left(\frac{A_H - m\bar{r}}{1 - m} - \rho - \bar{\lambda} \right) + (1 - X^*) (\bar{r} - \rho - \bar{\lambda}).$$

Therefore

$$\begin{aligned} \bar{g}_L^* &= \bar{r} - \rho - \bar{\lambda} - \bar{g}^* = -X^* \frac{A_H - \bar{r}}{1 - m} \\ &= -\frac{(1 - m)\lambda_{HL} - m\lambda_{LH}}{m}. \end{aligned}$$

So

$$\frac{d\bar{g}_L^*}{dm} = \frac{1}{m^2} \lambda_{HL} > 0.$$

By Theorem 4 (Part 1), $\frac{d\eta_1}{dm} < 0$.

Case 3: In this case $\bar{g}_H^* = \bar{g}_L^* = 0$, therefore relative wealth distribution does not change over time. $\square$

Numerically, Moll (2012) finds a similar result in a production economy with a continuum of investment productivity types and we borrow the term “Financial Kutznet’s Curve” from his analysis. As he explains, in Case 1, top end inequality is increasing in

m because of the “leverage effect,” i.e., higher m allows the more productive agents to borrow more at an interest rate determined by the rate of returns to the less productive agents, this magnifies the differences in returns and increases wealth inequality. However, at higher m , i.e., in Case 2, top end wealth inequality is decreasing because of the “return equalization effect,” i.e. higher m increases the demand to borrow by the more productive agents which pushes up the interest rate earned by the less productive agents (which is strictly higher than their own rate of return). This increase in interest rate reduces the differences in returns and decreases wealth inequality. If m is even higher, one reaches Case 3 in which the interest rate earned by the less productive agents is pushed up to the rate of return of the more productive agents. There is essentially no wealth inequality (in a stationary BGP) in this case.

D Live-Fast-Die-Young Dynamics

Our model features the “live-fast-die-young” dynamics described in [Gabaix et al. \(2016\)](#), as our calibration features a relatively short duration of the high return state (10 years) and relatively high returns of type H to match the tail index. As shown by [Gabaix et al.](#), such dynamics generate fast transition for wealth distribution, including tail index and top wealth shares. In this online appendix, we argue further that a shorter duration of high return, i.e., higher λ_{HL} , generates a higher transition speed for wealth distribution. We fix other parameters as in the benchmark calibration, but vary the switching rate λ_{HL} to be $1/5$ and $1/2$ from $1/10$ in the benchmark. We simultaneously recalibrate λ_{LH} and initial wealth distribution to keep the tail index and top 1% wealth share to be the same as in the benchmark calibration. Then we consider the transition dynamics with the non-uniform corporate tax cut, as described in Section 6.1, in models with different λ_{HL} .³

Figure [D.1](#) reports the results. As shown, a higher λ_{HL} corresponds to a lower duration of high return states, and leads to a faster transition speed of both top wealth shares and tail index. Notice the top 1% shares and tail indices in the starting and ending periods agree across models by construction. As for the sensitivity with respect to the λ_{HL} parameter, the half life for the transition of top 1 % wealth share is around 40 years in the benchmark calibration and around 20 years in the model with $\lambda_{HL} = 1/2$.

³For fair a comparison, we fix the final corporate tax rate for type L to be 35%, and recalibrate the final corporate tax rate for type H to keep the final tail index the same as in the benchmark transition exercise. This is different from the exercise in [Gabaix et al. \(2016, Online Appendix I.4\)](#) in which the final steady-state tail index changes as they vary the rate of switching from high to low growth and the speed of convergence is partially affected by the change in tail index.

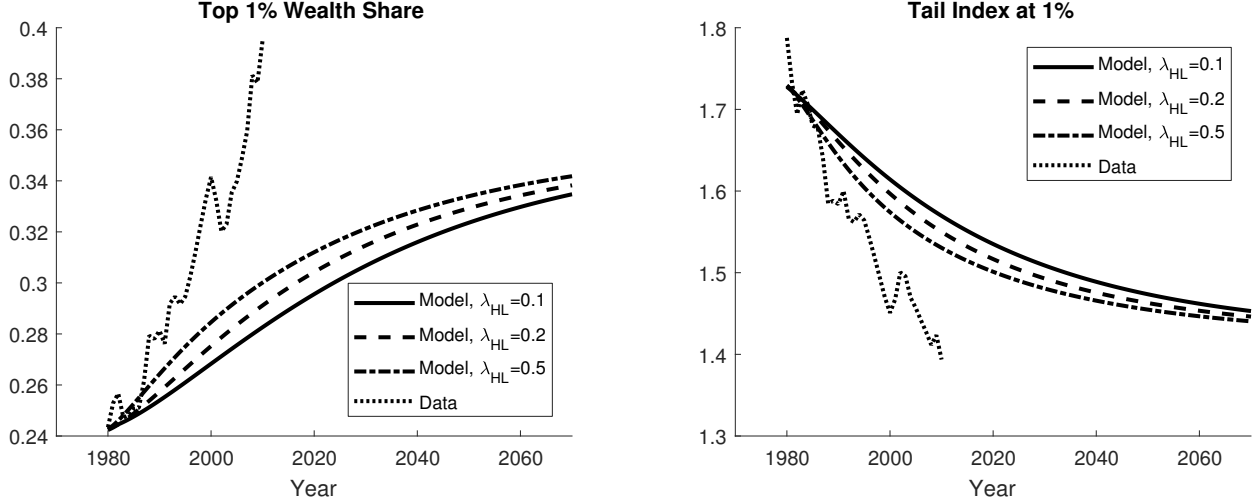


Figure D.1: Transition Paths of Top Wealth Inequality after Corporate Tax Cuts

E Welfare Analysis along Transition Paths

Since the value function of agents is homogeneous in wealth, the model generates very neat welfare analysis. Consider the value function V^h in Section 3, rewritten here

$$V^h(t, i, x, w) = v(t, i) + \frac{1}{\rho + \bar{\lambda}} \log(w + xq_t).$$

As shown, the value consists of two parts. The first part $v(t, i)$ only depends on investment productivity type i , and the second part $\frac{1}{\rho + \bar{\lambda}} \log(w + xq_t)$ depends on current wealth and labor productivity type x^4 . From the HJB equation for $v(t, i)$ in (8), we further see that $v(t, i)$ is fully determined by the paths of investment returns. Therefore, the welfare changes can be decomposed into changes in investment returns and changes in human wealth.

Notice with log utility, the consumption equivalent variation CEV solves the following:

$$\begin{aligned} \frac{\log(1 + CEV(i, x, w))}{(\rho + \bar{\lambda})} &= V_{new}^h(0, i, x, w) - V^h(0, i, x, w), \\ CEV(i, x, w) &\approx (\rho + \bar{\lambda})[V_{new}^h(0, i, x, w) - V^h(0, i, x, w)], \end{aligned}$$

where V_{new} is the value after a shock hits. We calculate welfare changes at the average level and for agents with different wealth and labor productivity levels at the initial

⁴Notice that w (and the distribution of w) is a state variable that does not change on impact, q_t is a forward looking variable that changes immediately when the shock hits.

stationary distribution. The gain from increased returns depends only on the investment productivity types. The loss from decline in human wealth is larger for agents with higher labor productivity x , but declines as wealth w goes large, as shown by $\frac{1}{\rho+\bar{\lambda}} \log(w + xq_t)$.

F Elasticity of Substitution

Table F.1: Steady State Statistics with Different Elasticities of Substitution

		KY	EY	Interest Rate	Tail Index	Top 1%
				Rate	Index	Wealth Share
Initial Steady State		3.00	0.650	3.00%	-2.00	24.3%
Nonuniform corporate tax cut	$\gamma = 1.5$	3.08	0.646	3.01%	-1.55	35.3%
	Cobb-Douglas	3.03	0.650	3.00%	-1.51	37.9%
	$\gamma = 0.7$	3.01	0.651	3.00%	-1.49	39.3%
	$\gamma = 1.25$	3.05	0.648	3.01%	-1.53	36.7%
Financial deregulation and corporate tax cut	$\gamma = 1.5$	3.21	0.640	3.02%	-1.40	42.6%
	Cobb-Douglas	3.00	0.650	3.00%	-1.31	56.8%
	$\gamma = 0.7$	2.92	0.651	2.99%	-1.27	66.5%
	$\gamma = 1.25$	3.09	0.647	3.01%	-1.35	49.4%

While the main calibration in Section 5 uses the estimate for the elasticity of substitution from [Piketty and Zucman \(2015\)](#), the literature is still far from reaching a consensus on its estimate. For example, [Karabarbounis and Neiman \(2014\)](#) report an average of 1.25, smaller than [Piketty and Zucman's](#) and [Oberfield and Raval \(2014\)](#) report a significantly lower estimate, 0.75. In this appendix, we assess the effects of corporate tax cut and financial deregulation reforms on aggregate variables and wealth inequality in models calibrated with different elasticities of substitution between capital and labor. Table F.1 reports statistics for the final steady states after two reforms described in Section 6 that generate the dynamics of aggregate statistics and top wealth inequality consistent with data under our benchmark calibration $\gamma = 1.5$.

First of all, the sufficient statistics formula described in Theorem 3 does not depend on the particular production function, i.e., the effects of production functions on wealth inequality are reflected in the aggregate statistics. Therefore, targeting the same aggregate statistics gives the same tail index, and with the same initial distribution function and earnings distribution, gives the same wealth distribution. This is extremely convenient for comparing reforms in models with different production function parameters as these

models agree on wealth distribution once aggregate statistics are set to match the same targets.

Now, after policy changes, wealth inequality move in the same direction in all calibrations. However, capital to output ratio and labor share move in opposite directions if the elasticity of substitution is greater than 1 and in the same direction if the elasticity of substitution is smaller than 1. Labor share remains constant if the elasticity of substitution is exactly equal to 1, i.e., Cobb-Douglas production functions.⁵ All in all, only models with the elasticity strictly greater than one can generate the joint dynamics of increasing wealth inequality, increasing capital to output ratio, and decreasing labor share observed in the U.S. data since the 1980s.

G Earnings Inequality and Wealth Inequality

In this online appendix, we examine the effects of earnings inequality on wealth inequality in the benchmark model in Section 3 and in an extension with stochastic life-cycles.

G.1 Benchmark Model

In this subsection we investigate the importance of earnings inequality for wealth inequality the benchmark model in Section 3. We consider two families of distributions for permanent earnings: log-normal and Pareto distributions.

First, we consider the log-normal distribution of x_t as in the baseline calibration. Table G.1 (columns 2-4) shows different statistics of the wealth distribution when we vary σ_x . As σ_x varies from 0.75 to 1.25, the top 1% income share, displayed in the last row in the table, increases from 5.7% to 12.7%. This variation in income share is in the range of observed wage income shares in the U.S. documented in [Piketty and Saez \(2003, Figure IX\)](#). Despite the significant variation in income inequality captured by the top 1% income share, the top end inequality does not vary significantly. In particular, both top 1% and top 0.1% wealth share varies by less than 1% from the baseline calibration (second column).

Second, we consider cases in which x_t follows a Pareto distribution with Pareto exponent ζ :

$$\Pr(x_t \geq x) \propto x^{-\zeta}.$$

From the top income shares estimates of [Piketty and Saez \(2003\)](#), we can back out the

⁵The intuition for this result is discussed at length in [Piketty \(2014\)](#). For example, for the case with homogenous returns, $EY = 1 - \alpha A^{\gamma/(\gamma-1)} (KY)^{(\gamma-1)/\gamma}$ which is increasing in KY if $\gamma < 1$ and decreasing in KY if $\gamma > 1$.

Table G.1: Top Wealth Shares with Different Earnings Distributions

	Data (1980)	Log normal, $\sigma_x =$			Pareto, $\zeta =$		
		0.75	0.93 (benchmark)	1.25	3.0	2.2	1.5
Top 1% income share	8.2%	5.7%	8.2%	12.7%	5.4%	8.2%	9.5%
Top 10% wealth share	67.1%	64.3%	64.7%	65.2%	64.0%	64.2%	64.7%
Top 5% wealth share	47.2%	47.2%	47.5%	47.8%	47.0%	47.2%	47.4%
Top 1% wealth share	24.1%	24.1%	24.3%	24.6%	23.8%	23.9%	24.2%
Top 0.1% wealth share	11.0%	11.0%	11.6%	12.4%	10.4%	10.8%	11.5%

estimates of ζ to be between 1.5 and 3 (lower exponent corresponds to higher earnings inequality). Table G.1 (column 4-7) shows different statistics of the wealth distribution when we vary ζ . We can see again that varying earnings inequality by changing the exponent of the productivity distribution does not change wealth inequality significantly.

A caveat of this exercise is that, for tractability, we only model *permanent* earnings inequality since x_t/G_t is constant over agents' lifetime. When changes in earnings inequality are associated with the temporary components of earnings, the effects on wealth inequality is likely to be more important. Indeed, [Kaymak and Poschke \(2016\)](#) show that rising earnings inequality accounts for more than two thirds of the increase in top end wealth inequality in the U.S. from the 1970s until recent years.⁶ In their model, wealth inequality arises from very large, and temporary earnings shocks a la [Castaneda et al. \(2003\)](#). Because these shocks are temporary, agents with high earnings shocks save a significant fraction of the earnings leading to high growth rates of their wealth and thus high wealth inequality. Consequently, earnings inequality and its changes are quantitatively important determinants of wealth inequality as shown by [Castaneda et al. \(2003\)](#) (in steady states) and [Kaymak and Poschke \(2016\)](#) (over transitional paths). In our model, earnings shocks are permanent, therefore, they do not affect saving rates. The high growth rates of wealth come instead from high investment returns. Consequently, changes in earnings inequality matter little for wealth inequality in our model. Following the same reasoning, the following subsection, Online Appendix G.2, presents a stochastic life-cycle model, in which earnings shocks are not permanent, and therefore changes in the earnings process also have stronger effects on wealth inequality.

⁶Empirically, [Saez and Zucman \(2016\)](#) also argue for changing earning inequality as one of the main causes of changing wealth inequality, the other being changing saving rates.

G.2 Model with Life-Cycles

For tractability, we incorporate life-cycles to the benchmark model using stochastic retirement and death shocks. We assume that agents are born with productivity $x_t > 0$ which grows at rate g_x . Over their lifetime, young agents are hit by Poisson shock of intensity λ_r upon which they retire. Retired agents do not receive labor income from human wealth and face a death shock of $\bar{\lambda}_r$. This leaves the fraction of young agents in the economy

$$M_Y = \frac{\bar{\lambda}_r + n}{\lambda_r + \bar{\lambda}_r + n},$$

where n is the population growth rate. To keep the aggregate labor supply the same as the benchmark, we assume that the mean labor productivity $M_x = 1/M_Y$.

Let Q_t denote the present discounted value of future labor income for each agent, i.e. human wealth (as if they never retire):

$$Q_t^h = \int_0^\infty \exp\left(-\int_0^{t'} r_{t+t'_1} dt'_1\right) e_{t+t'} x_{t+t'}^h dt',$$

which satisfies

$$\frac{dQ_t^h}{dt} = r_t Q_t^h - e_t x_t^h.$$

Since x_t^h grows at the rate g_x , we also have $Q_t^h = q_t x_t^h$, where

$$q_t = \int_0^\infty \exp\left(-\int_0^{t'} (r_{t+t'_1} - g_x) dt'_1\right) e_{t+t'} dt'.$$

The dynamics of q_t can be described by

$$\frac{dq_t}{dt} = (r_t - g_x) q_t - e_t.$$

Let $V^e(t, x, w)$ denote the value function of the employed agent. The HJB equation for V^e is

$$\begin{aligned} (\rho + \bar{\lambda}) V^e - \frac{\partial V^e}{\partial t} &= \max_{c,k,b} \log(c) + \frac{\partial V^e}{\partial x} g_x x_t + \frac{\partial V^e}{\partial w} (R_{i_t^h}(e_t)k + r_t b - c + e_t x_t) \\ &\quad + \lambda_{i_t^h, -i_t^h} \left(V^e(t, -i_t^h, x_t^h, w_t^h) - V^e(t, i_t^h, x_t^h, w_t^h) \right) \\ &\quad + \lambda_r \left(V^r(t, i_t^h, w_t^h) - V^e(t, i_t^h, x_t^h, w_t^h) \right) \end{aligned} \quad (\text{G.1})$$

where the maximization problem is subject to the constraints $w_t^h = k + b$ and $0 \leq k$ and

Table G.2: Top Wealth Share with Uninsurable Retirement Shock

	Baseline	$\lambda_r = 0$	$\lambda_r = \frac{1}{65}$			$\lambda_r = \frac{1}{25}$		
			$\sigma_x = 0.93$	0.75	1.25	$\sigma_x = 0.93$	0.75	1.25
Top 10%	64.7%	58.5%	41.0%	40.3%	42.3%	35.8%	33.1%	39.7%
Top 5%	47.5%	43.7%	29.1%	28.8%	29.9%	24.0%	22.1%	27.3%
Top 1%	24.3%	23.0%	13.4%	13.4%	13.6%	9.4%	8.8%	11.0%
Top 0.1%	11.6%	11.1%	5.6%	5.4%	5.8%	3.0%	2.8%	3.2%

$0 \leq mk + b + Q_t^h$, and where V^r denote the value function of the retired agent which satisfies

$$\begin{aligned}
(\rho + \bar{\lambda}_r) V^r - \frac{\partial V^r}{\partial t} = \max_{c,k,b} \log(c) + \frac{\partial V^r}{\partial w} (R_{i_t^h}(e_t)k + r_t b - c) \\
+ \lambda_{i_t^h, -i_t^h} \left(V^r(t, -i_t^h, w_t^h) - V^r(t, i_t^h, w_t^h) \right),
\end{aligned}$$

subject to $w_t^h = k + b$ and $0 \leq k$ and $0 \leq mk + b$.

Due to the uninsurable retirement risk, the solutions to young agents' decision problems are not linear in total wealth and hence it is not sufficient to track only the type-specific aggregate wealth to solve the equilibrium. Instead, one needs to track the joint distribution of age, productivity type and wealth and loses much of the analytical tractability in our benchmark model. To solve for the stationary BGP in this model, we extend the method described in [Achdou et al. \(2015\)](#) to allow for heterogeneous returns.

Table G.2 reports the top wealth share in the stationary equilibrium with retirement shock. The first column is the baseline economy without retirement shock copied from Table 3. The second column corresponds to an intermediate case where the arrival rate of the retirement shock is set to zero, but young agents cannot borrow, i.e., $w_t^h \geq 0$. Such constraint automatically arises when retirement shock hits with positive rate since retiring with zero or negative wealth generates utility of minus infinity. Therefore, the constraint on borrowing itself creates precautionary saving motive for young agents to self insure the investment productivity shock even in the absence of positive retirement shock. Comparing the first and second column, we see that though the borrowing constraint and the induced precautionary saving creates more saving at the bottom and lowers the top wealth shares, the effects are overall very small, especially for the very top. This is consistent with results in [Benhabib et al. \(2015\)](#) that with labor income and capital income risk, agents' saving function keeps to be asymptotically linear in wealth as wealth becomes large, and the growth rate of wealth at the top is not affected by labor income risk.

The third column of Table G.2 sets the arrival rate of retirement shock $\lambda_r = 1/65$.

To calibrate a realistic fraction of young agents in the economy, we set the death rate of retired agents $\bar{\lambda}_r = 0.1$. This corresponds to an average retirement age of 65 and average life span of 10 years after retirement. Adding the realistic retirement shock lowers the top wealth share significantly. This is because retired agents face a higher rate of death shock, which significantly lowers their saving rate. Consequently, young agents also discount future wealth more and save less. The lower aggregate wealth in the economy drives up interest rate and differences in returns but the general equilibrium effects are small. The thinner tail is largely driven by the differences in saving rates; again, the saving function is asymptotically linear in wealth and the growth rate of wealth at the top is not affected by precautionary saving motive due to the labor income risk. This is evident as we vary income risk by changing the standard deviation of labor productivity (as done in Subsection G.1). As shown in the fourth and fifth columns, the top wealth shares do not move much. Notice that since agents receive zero income after retirement, the differences in labor productivity when they are young are indeed close to permanent with respect to the whole life span.

The labor income risk matters more for top wealth shares as income shocks become more transitory. To illustrate this point, in the last three columns of Table G.2, we set the arrival rate of retirement shock at a much higher level $\lambda_r = 1/25$, and vary labor productivity of the young agents the same way as before. As shown, a higher rate of retirement shock further lowers top wealth shares since young agents save less as they expect to be more likely to enter the retirement stage when they discount future wealth more heavily. But top wealth shares vary even more as we change the standard deviation of labor productivity. This is because a higher rate of retirement also means the income differences of young agents are more transitory, and for precautionary saving motive the dispersion of transitory income shocks have larger effects on agents' saving rates upon receiving different labor income shocks. We expect if agents face labor income process with highly transitory shocks, the dispersion of labor income shocks is crucial in shaping the wealth distribution, a mechanism highlighted in Kaymak and Poschke (2016).

H Additional Empirical Analysis

One of the main objects studied in our paper is the tail index of the wealth distribution, assuming that the distribution follows the power law at the right tail. However, the existing literature on top end wealth inequality has never explicitly investigated at what point in the wealth distribution the right tail is well approximated by the power law (Figure 2 in Diamond and Saez, 2011 addresses the same question for income distribution). For

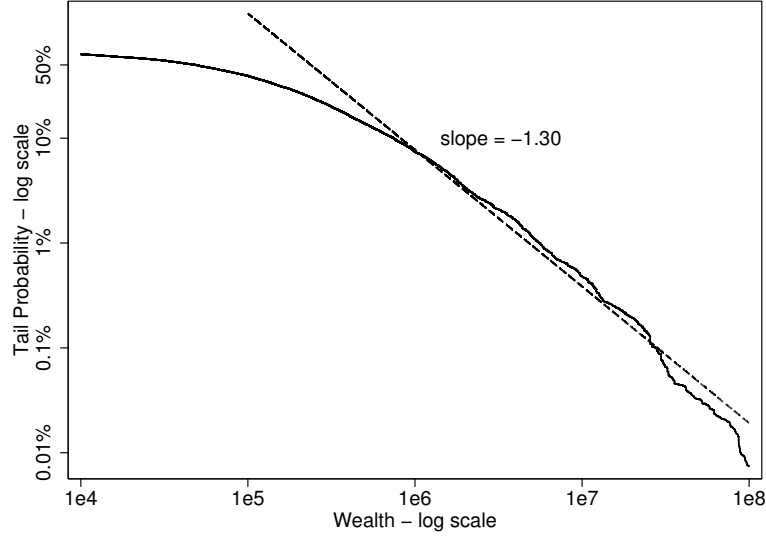


Figure H.1: The distribution of wealth from SCF in 2010.

example, in two prominent studies and surveys of Pareto power law for wealth distribution, [Gabaix \(2009\)](#) writes casually “It seems that the tail exponent of wealth is rather stable, perhaps around 1.5 [...]” and similarly, [Benhabib et al. \(2011\)](#) write “The top 1% of the richest households in the United States hold over 33% of wealth and the top end of the wealth distribution obeys a Pareto law, the standard statistical model for heavy upper tails.[...]” The authors, as well as many others, do not specify at what point in the wealth distribution, the power law is a good approximation. Therefore we did our own investigation using the SCF data in 2010. Figure [H.1](#) plot log tail probability (or equivalently log rank) as a function of log wealth. It shows that starting from around the top 10% in the wealth distribution ($\approx \$750,000$), the distribution is well-approximated by a power law with coefficient around 1.30:

$$\log \Pr(W \geq w) = \text{const} - 1.30 \cdot \log w.$$

The very top of the distribution is thinner than what is predicted by the power law. But [Vermeulen \(2017\)](#) shows that we can improve the fit by adding observations with extreme wealth from Forbes’ billionaire list.

In the remaining of this online appendix, we present additional empirical analyses on returns to wealth using PSID data.

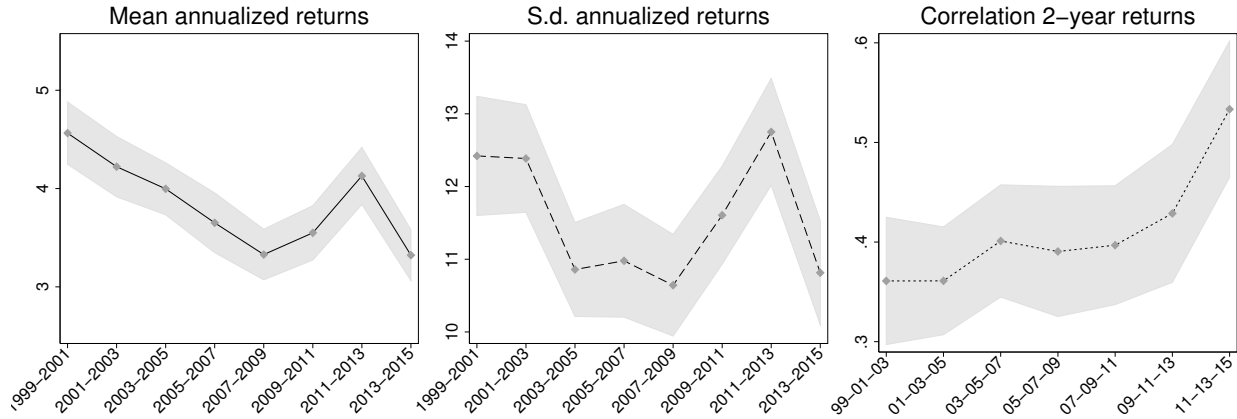


Figure H.2: Statistics of 2-year returns to wealth

Notes: Lines are mean estimates. Shaded areas are 2 standard error bands. Standard errors in parentheses are computed based on bootstrap procedures with 200 repetitions.

H.1 Correlation of returns to wealth in later waves of PSID

After 1999, PSID surveys households once every two years and contains wealth information every wave. In order to compute the 2-year returns to wealth, we need to observe capital income every year. We use a linear interpolation to impute capital income in the year not surveyed. For example, the return from 1999 to 2001 is defined as $(\text{capital income in 1999} + \text{capital income in 2000}) / (\text{core asset in 1999})$. Capital income in 2000 is surveyed in year 2001 but capital income in 1999 is not in the survey. Hence we impute capital income in 1999 = $(\text{capital income in 1998} + \text{capital income in 2000}) / 2$, of which capital income in 1998 is surveyed in year 1999.

Figure H.2 depicts the mean and standard deviation statistics of 2-year returns without capital gains, and the correlation of 2-year returns over time. As shown, there is substantial heterogeneity in returns: the standard deviation of annualized returns stays stably around 11% compared to the mean annualized returns of around 4% during this episode. Returns to wealth are highly persistent as shown in the third panel: the 2-year correlation stays significantly positive at around 0.4. Compared to results based on 5-year returns in Section 2, both standard deviations and correlations are higher due to shorter horizons in between. There is a weak pattern that the correlation of returns has increased in recent years, but we do not want to over interpret this result since the estimates of these correlations are with relatively large standard errors.

H.2 Persistence in returns indicated by wealth mobility

In this subsection, we show that the conditional wealth mobility is indicative of persistent returns to wealth. We exploit the panel feature of PSID and assess wealth mobility by inspecting the transition probability from one wealth class to another between survey waves.

We first divide households into 3 wealth classes based on their net wealth in a particular year t . Each wealth class consists of equal number of observations (weights adjusted). We call the bottom 1/3 households bottom class, middle 1/3 middle class, and top 1/3 top class. We further categorize households based on the returns to wealth that they received in the same year⁷. We call households which received returns lower than the median within each wave the low return group, and which received returns higher than the median the high return group. Then we compute the transition probability between wealth classes from year t to the next survey year $t + s$, conditional on being of a particular return group in year t .

Table H.1: 5-Year Transition Probability Between Wealth Classes, 1984–1999

$t + s \rightarrow$	Bottom	Middle	Top	Bottom	Middle	Top
$t \downarrow$	Low Return			High Return		
Bottom	0.52	0.41	0.08	0.37	0.49	0.14
Middle	0.18	0.59	0.23	0.11	0.57	0.32
Top	0.05	0.20	0.75	0.02	0.10	0.89

Tables H.1 reports the results for 5-year transition probability for samples from $t \in \{1984, 1989, 1994\}$ and $s = 5$. Each row corresponds to the wealth class in the starting year and each column corresponds to the wealth class in the ending year. For example, in Table H.1, 0.52 in the first row and first column means that the probability of being in bottom class in year $t + s$ conditional on being in the bottom class and low return group in year t is 52%. These statistics are constructed for each wave of survey and then averaged across waves.

We highlight in bold pairs of statistics to compare. First, the upward mobility, the probability of transiting from bottom class to top class, is higher if a household received a higher return to wealth in the starting year. This indicates that households which initially received higher returns were likely to persist in their higher returns and moved up the wealth class ladder. We would not expect to observe such patterns if returns to wealth

⁷Notice return here is defined as capital income in year t divided by wealth in year t , which is different from the return during year t to year $t + s$ defined in Section 2.

were i.i.d. draws. Similarly, the downward mobility - the probability of transiting from top class to bottom class - is lower if a household received a higher return to wealth in the starting year.

Table H.2: 5-Year Transition Probability Between Wealth Classes, 1984–1999

$t + s \rightarrow$	Bottom	Middle	Top	Bottom	Middle	Top
$ts \downarrow$	Staying Low			Low to High		
Bottom	0.41	0.48	0.10	0.26	0.62	0.12
Middle	0.12	0.62	0.26	0.08	0.63	0.29
Top	0.04	0.18	0.79	0.02	0.19	0.79
	High to Low			Staying High		
Bottom	0.28	0.55	0.17	0.25	0.53	0.22
Middle	0.11	0.52	0.37	0.06	0.61	0.33
Top	0.02	0.11	0.87	0.01	0.09	0.91

To strengthen the arguments that the persistence in wealth classes is indicative of the persistence in returns to wealth, we compute the same transition probability further conditional on return groups in both starting and ending survey years. Table H.2 reports the results. "Staying Low" corresponds to households who started within the low return group in year t and stayed in the low return group in year $t + s$. Definitions are similar for "Low to High", "High to Low", and "Staying High" groups. As shown, coming from the low return group, households were more likely to transit out of the bottom class if they switched to the high return group. Similarly, coming from a high return group, households were more likely to stay in the top class if they continued to receive high returns. Nevertheless, even conditioning on return groups in both years, households which received higher returns in the starting year were still more likely to transit out of the bottom class or stay in the top class. This finding indicates that the initial returns do persist for many years.

I Wealth Inequality in the Standard Neo-Classical Growth Model

The textbook Ramsey-Cass-Koopman neoclassical growth model does not feature wealth inequality because the agents in the model put equal weights on the utility of their offsprings and on their own. In the decentralized version of the model, this assumption implies that family assets are equally divided among family members at every instant. A simple way to obtain wealth inequality in the model is to assume that the agents put zero

weights on the utility of their offsprings. Under this setup, the offsprings are born with zero financial assets and start out with only human wealth, i.e., the present discounted value of future labor income.

Human wealth grows at the same rate as the growth rate of economy's income per capita, x , driven by the exogenous growth rate of labor augmenting technological change. Existing agents save their assets at the equilibrium interest rate r and consume at the rate c , so their total wealth (human plus financial wealth) grows at the rate $r - c - x$ relative to the total wealth of the new born. Therefore if $r - c - x > 0$, wealth inequality will grow without bound. In order to obtain bounded inequality, we can assume death shocks (arriving at the Poisson rate λ), and wealth of the dying agents are redistributed equally, or according to a thin tail redistribution function, to the new borns.

Under these assumptions, we show that the stationary wealth distribution has Pareto tail and the tail index is given by:

$$\theta = \frac{\lambda + n}{r - c - x}.$$

Lower tail index corresponds to higher top end wealth inequality. Therefore this formula confirms [Piketty \(2014\)](#)'s intuition: lower population growth rate, or higher interest rate, or lower the technological growth rate all correspond to higher top end wealth inequality.

The disadvantage of this formula is that the consumption rate c might not be observable. Using the equilibrium conditions to solve out for c , we show that

$$\theta = \left(1 + \frac{\lambda}{n}\right) \left(1 + \frac{EY}{KY} \frac{1}{r - x}\right),$$

where EY and KY are the labor share and capital-output ratio. This formula tells us that lower labor share or higher capital-output ratio correspond to higher top end wealth inequality. This result is also consistent with [Piketty \(2014\)](#)'s narratives.

In the next Subsection, we reproduce the standard neoclassical growth model from [Barro and Sala-i Martin \(2004\)](#) to set up the notations and make it explicit why there is no wealth inequality in such a model. In Subsection [I.2](#), we present a simple modification that leads to a fat-tail wealth distribution and derive the formulae for the tail index.

I.1 The Standard Neo-Classical Growth Model

Consider the standard neoclassical growth model (Ramsey-Cass-Koopman) as presented in [Barro and Sala-i Martin \(2004\)](#) with population growth and exogenous labor-augmenting

technological growth.

Time t is continuous and runs from 0 to ∞ . Population at time t is

$$L(t) = e^{nt}$$

where $n > 0$ is population growth rate. $C(t)$ is total consumption at time t and $c(t) \equiv \frac{C(t)}{L(t)}$ is consumption per capita.

Each household (agent) maximizes overall utility, U , as given by

$$U = \int_0^\infty u(c(t))e^{nt}e^{-\rho t}dt, \quad (\text{I.1})$$

where $\rho > 0$ is the discount rate and the instantaneous utility function:⁸

$$u(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma}.$$

Implicitly, (I.1) assumes that each agent puts equal weigh on her utility as well as on the utility of her offsprings.

Let A_t denote the total assets in the economy with earn the (equilibrium) rate of returns $r(t)$. In addition, $w(t)$ denotes the wage rate paid per unit of labor services. Therefore, the evolution of the total asset A_t is:

$$\frac{dA(t)}{dt} = r(t)A(t) + L(t)w(t) - C(t).$$

The assumption that each agent puts equal weigh on her utility as well as on the utility of her offsprings translates to total assets being divided equally across agents. Each agent then owns

$$a(t) = \frac{A(t)}{L(t)}.$$

Consequently, there is no wealth inequality in this economy. From the evolution of the total assets,

$$\frac{da(t)}{dt} = r(t)a(t) + w(t) - c(t) - na(t). \quad (\text{I.2})$$

At time t , competitive firms produce goods, pay wages for labor input, and make rental payments for capital inputs. Each firm has access to the constant returns to scale production technology:

⁸ $u(c) = \log c$ if $\sigma = 1$.

$$Y(t) = F(K(t), L(t)T(t)) = L(t)T(t)f\left(\frac{K(t)}{L(t)T(t)}\right)$$

where $T(t)$ is the level of technology which grows at the rate $x \geq 0$: $T(t) = e^{xt}$.

In a competitive equilibrium, firms rent capital from the households: $K(t) = A(t)$, and

$$r(t) = F_K(K(t), L(t)T(t)) - \delta K(t) = f'(\hat{k}(t)) - \delta \hat{k}(t),$$

where $\hat{k}(t) = \frac{K(t)}{L(t)} \frac{1}{T(t)} = \frac{a(t)}{T(t)}$. Similarly

$$w(t) = T(t)F_L(K(t), L(t)T(t)) = T(t) \left(f(\hat{k}(t)) - \hat{k}(t)f'(\hat{k}(t)) \right).$$

Plugging in these identities into (I.2), we obtain

$$\dot{\hat{k}}(t) = f(\hat{k}(t)) - \hat{c}(t) - (x + n + \delta)\hat{k}(t), \quad (\text{I.3})$$

where $\hat{c}(t) = \frac{c(t)}{T(t)}$.

From the households' optimization problem, we obtain the Euler equation:

$$\frac{\dot{c}(t)}{c(t)} = \frac{r(t) - \rho}{\sigma},$$

which implies

$$\frac{\dot{\hat{c}}(t)}{\hat{c}(t)} = \frac{\dot{c}(t)}{c(t)} - x = \frac{f'(\hat{k}(t)) - \delta - \rho - \sigma x}{\sigma}. \quad (\text{I.4})$$

Setting $\dot{\hat{k}}(t) = \dot{\hat{c}}(t) = 0$, (I.3) and (I.4) provide two equations that determine the steady state $(\hat{k}^*, \hat{c}^*)$ of the economy.

I.2 The Standard Neo-Classical Growth Model with Wealth Inequality

We consider the environment as in the standard model presented in the previous section with exogenous population growth and labor augmenting technological growth. However, in order to generate wealth inequality in the model, we assume that the agents do not put weight on the utility of their offsprings. To prevent wealth inequality to grow without bound, we assume that the agents are hit with death shocks, arriving at the Poisson rate λ . Upon her death, an agent's wealth is divided equally among the new borns.

The inter-temporal expected utility of an agent i born at time t is then

$$\tilde{U} = \int_0^\infty e^{-(\rho+\lambda)t'} u(c_{t+t'}^i) dt'. \quad (\text{I.5})$$

The asset (financial wealth) of agent i evolves according to

$$\dot{a}^i(t) = r(t)a^i(t) + w(t) - c^i(t).$$

Let h_t denote the human wealth of each agent, i.e., the present discounted value of future wages:

$$h(t) = \int_t^\infty e^{-\int_t^{t_1} r(s)ds} w(t_1) dt_1,$$

or in differential form:

$$\dot{h}(t) = r(t)h(t) - w(t).$$

Then the total wealth of agent i , $b^i(t) = a^i(t) + h(t)$ evolves according to:

$$\dot{b}^i(t) = r(t)b^i(t) - c^i(t). \quad (\text{I.6})$$

Agent i then maximizes the objective function (I.5) subject to constraint (I.6). It is easy to see that the maximization problem is homogenous. Therefore the optimal consumption is linear in total wealth

$$c_t^i = \hat{c}_t b_t^i,$$

where $\hat{c}(t)$ depends on the whole path of future interest rates. The HJB equation that describes how $\hat{c}_t$ depends on r_t can be found from a HJB equation similar to (B.1). The following derivations also apply under the assumption that $\hat{c}_t$ is exogenously given.

In a competitive equilibrium, we have

$$K(t) = \int_{\mathcal{I}(t)} a^i(t) di,$$

where $\mathcal{I}(t)$ denote the set of agents at time t and $K(t)$ denote the aggregate capital.

Aggregating (I.6) across agents, and let $B(t) = \int_{\mathcal{I}(t)} b^i(t) di$, we have

$$\dot{B}(t) = (r(t) - \hat{c}(t))B(t) + nL(t)h(t),$$

where the last term captures the human wealth of the new borns.

We also have $B(t) = \int_{\mathcal{I}(t)} a^i(t) di + L(t)h(t) = K(t) + L(t)h(t)$. Therefore,

$$\dot{B}(t) = \dot{K}(t) + nL(t)h(t) + L(t)(r(t)h(t) - w(t)).$$

Comparing the two expressions for $\dot{B}(t)$, we arrive at

$$\dot{K}(t) = (r(t) - \hat{c}(t))K_t + (w(t) - \hat{c}(t)h(t))L(t).$$

Consequently,

$$\dot{\hat{k}}(t) = \frac{d}{dt} \left(\frac{K(t)}{T(t)L(t)} \right) = (r(t) - \hat{c}(t))\hat{k}(t) + (\hat{w}(t) - \hat{c}(t)\hat{h}(t)) - (n + x)\hat{k}(t),$$

where $\hat{w}(t) = \frac{w(t)}{T(t)}$ and $\hat{h}(t) = \frac{h(t)}{T(t)}$.

In a balanced growth path:

$$(r^* - \hat{c}(r^*))\hat{k}^* + (\hat{w}^* - \hat{c}(r^*)\hat{h}^*) - (n + x)\hat{k}^* = 0 \quad (\text{I.7})$$

and

$$\hat{h}^* = \frac{\hat{w}^*}{r^* - x}.$$

This equation determines the balanced growth path.

Now, we turn our attention to the stationary wealth distribution in the balanced growth path. Let $\omega^i(t) = \frac{b^i(t)}{B(t)/L(t)}$. After algebra simplifications, we arrive at

$$\dot{\omega}(t) = (r^* - \hat{c}(r^*) - x)\omega(t),$$

conditional on the death shocks not hitting at t and $t + dt$. If an agent is hit by the death shock, the agent is replaced by new borns with financial wealth $\frac{K(t)}{L(t)} \frac{\lambda}{\lambda + n}$ and human wealth $h(t)$, which corresponds to normalized wealth

$$\frac{\frac{K(t)}{L(t)} \frac{\lambda}{\lambda + n} + h(t)}{B(t)/L(t)} = \frac{\hat{k}^* \frac{\lambda}{\lambda + n} + \hat{h}^*}{\hat{k}^* + \hat{h}^*}.$$

Therefore, the measure of agents with relative wealth exceeding ω , $\mathcal{M}_t(\omega(t) \geq \omega)$, satisfies:

$$\begin{aligned} & \mathcal{M}_{t+\Delta t}(\omega(t + \Delta t) \geq \omega) \\ &= (1 - \lambda\Delta t)\mathcal{M}_t(\omega(t) + (r^* - \hat{c}(r^*) - x)\omega(t)\Delta t \geq \omega) \\ &+ (\lambda + n)L(t)\Delta t \mathbf{1}_{\left\{ \frac{\hat{k}^* \frac{\lambda}{\lambda + n} + \hat{h}^*}{\hat{k}^* + \hat{h}^*} \geq \omega \right\}} + o(\Delta t). \end{aligned}$$

Subtracting $\mathcal{M}_t(\omega, t)$ from both sides and dividing them by Δt , then taking the limit $\Delta t \rightarrow 0$, we obtain the following PDE for $p(t, \omega) = \frac{\mathcal{M}_t(\omega(t) \geq \omega)}{L_t}$:

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial \omega} \omega(r^* - \hat{c}(r^*) - x) \Delta t - (\lambda + n)p(t, \omega) + (\lambda + n) \mathbf{1}_{\left\{ \frac{\hat{k}^* \frac{\lambda}{\lambda+n} + \hat{h}^*}{\hat{k}^* + \hat{h}^*} \geq \omega \right\}}.$$

In the stationary BGP $\frac{\partial p}{\partial t} \equiv 0$ and $p(t, \omega) \equiv p^*(\omega)$, so

$$\frac{\partial p^*(\omega)}{\partial \omega} \omega(r^* - \hat{c}(r^*) - x) = -(\lambda + n)p^*(\omega) + (\lambda + n) \mathbf{1}_{\left\{ \frac{\hat{k}^* \frac{\lambda}{\lambda+n} + \hat{h}^*}{\hat{k}^* + \hat{h}^*} \geq \omega \right\}}.$$

This leads to the Pareto distribution $p^*(\omega) = (\frac{\omega}{\omega^*})^{-\theta}$ with $\omega^* = \frac{\hat{k}^* \frac{\lambda}{\lambda+n} + \hat{h}^*}{\hat{k}^* + \hat{h}^*}$ and the tail index θ :

$$\theta = \frac{\lambda + n}{r^* - \hat{c}(r^*) - x}.$$

From (I.7), after simplification, we have

$$r^* - \hat{c}(r^*) - x = n \frac{\hat{k}^*}{\hat{h}^* + \hat{k}^*}.$$

Therefore

$$\begin{aligned} \theta &= \left(1 + \frac{\lambda}{n}\right) \left(1 + \frac{\hat{h}^*}{\hat{k}^*}\right) \\ &= \left(1 + \frac{\lambda}{n}\right) \left(1 + \frac{EY}{KY} \frac{1}{r^* - x}\right), \end{aligned}$$

where $EY = \frac{\hat{w}^*}{\hat{y}^*}$ and $KY = \frac{\hat{k}^*}{\hat{y}^*}$.

References

- Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2015). Heterogeneous agent models in continuous time. mimeo.
- Barro, R. and X. Sala-i Martin (2004). *Economic Growth* (2nd Edition ed.). MIT Press.
- 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.

- Benhabib, J., A. Bisin, and S. Zhu (2015). The wealth distribution in bewley economies with capital income risk. *Journal of Economic Theory* 159, Part A, 489 – 515.
- Castaneda, A., J. Diaz-Gimenez, and J.-V. Rios-Rull (2003). Accounting for the u.s. earnings and wealth inequality. *Journal of Political Economy* 111(4), pp. 818–857.
- Diamond, P. and E. Saez (2011, December). The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives* 25(4), 165–90.
- Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics* 1(1), 255–294.
- Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). The dynamics of inequality. *Econometrica* 84(6), 2071–2111.
- Karabarbounis, L. and B. Neiman (2014). The global decline of the labor share*. *The Quarterly Journal of Economics* 129(1), 61.
- Kaymak, B. and M. Poschke (2016). The evolution of wealth inequality over half a century: The role of taxes, transfers and technology. *Journal of Monetary Economics* 77, 1 – 25.
- Moll, B. (2012). Inequality and financial development: A power-law kuznets curve. Working paper, Princeton University.
- Oberfield, E. and D. Raval (2014). Micro data and macro technology. Princeton University Working Paper.
- Piketty, T. (2014). *Capital in the 21st century*. Harvard University Press.
- Piketty, T. and E. Saez (2003). Income inequality in the united states, 1913-1998. *The Quarterly Journal of Economics* 118(1), 1–41.
- Piketty, T. and G. Zucman (2015). Chapter 15 - wealth and inheritance in the long run. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume 2 of *Handbook of Income Distribution*, pp. 1303 – 1368. Elsevier.
- Saez, E. and G. Zucman (2016). Wealth inequality in the united states since 1913: Evidence from capitalized income tax data. *The Quarterly Journal of Economics*.
- Vermeulen, P. (2017). How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, n/a–n/a.