Wealth Inequality and Asset Prices*

Matthieu Gomez †

December, 2024

Abstract

Wealthy households disproportionately invest in equity, causing equity returns to generate large and persistent fluctuations in top wealth inequality. Motivated by this observation, I study the joint dynamics of asset prices and wealth inequality in a model where a subset of agents ("entrepreneurs") hold levered positions on the economy. In the model, as in the data, the wealth distribution is stochastic and it exhibits a Pareto tail, with a tail index that depends on the logarithmic average return of top households. The model features a feedback loop between asset prices and wealth inequality, which amplifies the effect of aggregate shocks on the economy. The model, calibrated to the U.S. data, can account for a substantial portion of the fluctuations in asset prices and top wealth shares over the 20th century.

*This article is a substantially revised version of the first chapter of my dissertation (job market paper). An earlier version was circulated with the title "Asset Prices and Wealth inequality". I am indebted to my thesis advisors Valentin Haddad, Atif Mian, Wei Xiong, and Motohiro Yogo for their guidance. I am also grateful to Jarda Borovička, Nicolae Gârleanu, Xavier Gabaix, Leonid Kogan, Hanno Lustig, Ben Moll, Stavros Panageas, Dimitris Papanikolaou, José Scheinkman, David Sraer, and Alexis Toda for helpful discussions and comments.

†Columbia University; mg3901@columbia.edu

1 Introduction

Recent empirical studies have documented important fluctuations in top wealth inequality over time.[1] Volatile stock market returns may account for these fluctuations, though this explanation has remained relatively underexplored. Conversely, a large theoretical literature in asset pricing examines the role of household heterogeneity in shaping asset prices but seldom considers its implication for the wealth distribution.

In this paper, I use newly available data on the wealth distribution to investigate the relationship between asset prices and wealth inequality. I first document that wealthy households disproportionately invest in equity, causing equity returns to generate large and persistent fluctuations in top wealth inequality. Motivated by this fact, I build an asset pricing model where agents have heterogeneous exposures to aggregate income shocks. The model generates a feedback loop between top wealth shares and asset prices: a higher-than-expected shock in economic growth increases inequality as wealthier households are more exposed to aggregate shocks. In equilibrium, this raises asset prices due to the greater demand for assets from wealthier households. These higher asset valuations feed back into inequality, thus continuing the cycle. I show that such a mechanism can account for a substantial portion of the observed fluctuations in asset prices and top wealth shares.

The paper proceeds in three steps. I first use recently available data on top wealth inequality to document that wealthy households are twice as exposed to equity returns as the rest of the population. Formally, in response to a realized stock return of 10%, the average wealth in the economy increases by 4.3% while the average wealth in the top 0.01% increases by 7.8%. As a result, equity returns generate fluctuations in top wealth shares: in response to a realized stock return of 10%, the share of wealth owned by the top 0.01% increases by 3.5% (= 7.8% − 4.3%). Using local project methods, I show that this increase is very persistent over time, with minimal reversion over the next ten years. Given that equity returns are very volatile and almost serially uncorrelated, this mechanism can generate large fluctuations in top wealth shares over time.

To interpret this evidence, I develop a general equilibrium economy model where agents have heterogeneous exposures to aggregate income shocks (e.g., shocks to aggregate productivity). In the model, a subset of the population ("entrepreneurs") hold concentrated stakes in their firms, while the remaining population ("households") freely trade bonds and equity. This setup effectively implies that entrepreneurs hold levered claims on the economy, amplifying their gains during economic upturns and their losses during downturns. In the model, as in the data, the wealth distribution is stochastic. Despite the presence of aggregate shocks, I show that the wealth distribution exhibits a Pareto tail, and I derive a simple characterization of its tail index based on the time-averaged drift and variance of the log wealth of top households.

Third, I explore the quantitative implications of the model by calibrating it to the U.S. data.

[1] See, for instance, Wolff (2002), Kopczuk and Saez (2004), and Saez and Zucman (2016).

I discipline the heterogeneity across agents using moments related to the wealth distribution. In particular, I use the elasticity of the average wealth in the top 0.01% with respect to equity returns to discipline the aggregate risk exposure of entrepreneurs relative to households and the tail index of the wealth distribution to discipline their saving rate (as the tail index is directly informative on the average growth rate of top households).

The model generates an "excess" volatility of stock market returns relative to the volatility of aggregate income shocks due to a feedback loop between asset prices and wealth inequality: when a positive shock hits, investors at the top of the wealth distribution gain more than the rest; that is, wealth inequality increases. As wealth is redistributed towards wealthier agents, the aggregate demand for assets increases; that is, asset valuations increase. To better understand the mechanism, I derive an exact decomposition for the volatility of asset valuations as a sum of changes in future risk-free rates and changes in future expected excess returns (à la Campbell-Shiller). Relative to the original decomposition by Campbell and Shiller (1988), this decomposition is exact and can be computed analytically at every point of the state space. After applying this new methodology, I show that the volatility of asset returns in the model is mainly driven by changes in future expected excess returns in good times and by changes in future risk-free rates in bad times.

Finally, I use the calibrated model to trace the full impulse response of top wealth shares to aggregate shocks. This analysis extends my empirical findings, which were limited to estimating the response of top wealth shares over a few years, as standard errors grow prohibitively large beyond this horizon due to the limited sample period. The calibrated model reveals that aggregate shocks induce exceptionally persistent movements in top wealth share: it takes approximately 40 years for the impact of an aggregate income shock on the wealth share of the top 0.01% to be divided by three. Economically, this comes from the fact that a given change in the wealth of households at the top only dissipates when new generations, unaffected by the shock, reach top percentiles—a process that unfolds gradually over decades.

The model generates sizable fluctuations in top wealth inequality over time, given the persistence of top wealth shares and the lack of correlation in equity returns over time. Quantitatively, I find that the calibrated model can account for about 40% of the actual standard deviation in top wealth shares observed in the data. While the model accurately matches the business cycle dynamics of top wealth inequality, it cannot fully capture the overall U-shape of inequality over the 20th century — that is, the large decline in the 40s and the steep increase starting from 1980. In sum, the model's core mechanism — the disproportional exposure of wealthy households to aggregate shocks — can explain a substantial portion, but not all, of the actual fluctuations in top wealth shares. Hence, this mechanism complements but does not replace other drivers of wealth inequality put forth in the literature, such as changes in return dispersion, taxes, or saving behaviors.[2]

[2] A non-exhaustive list of papers focused on the low-frequency fluctuations of top wealth inequality for the U.S.

Literature review. This paper is motivated by a growing literature documenting the dynamics of top wealth shares in the U.S. (Kopczuk and Saez, 2004, Saez and Zucman, 2016, Smith et al., 2023). In response to these findings, a number of macro papers have studied the role of changes in taxes (Kaymak and Poschke, 2016; Hubmer et al., 2021; Cao and Luo, 2017), changes in labor income (Kaymak et al., 2018), or changes in idiosyncratic shocks (Atkeson and Irie, 2022; Gomez, 2023) on top wealth inequality. In contrast to these papers, I focus on examining the effect of excess stock market returns (in the model, shocks in aggregate income) on top wealth inequality, both in the short run and in the longer run.

On the empirical side, this paper contributes to a large literature examining the heterogeneity in equity holdings across the distribution of households (Guiso et al., 1996; Carroll, 2000; Campbell, 2006; Wachter and Yogo, 2010; Roussanov, 2010; Bach et al., 2020; Kacperczyk et al., 2018). In particular, Parker and Vissing-Jørgensen (2010) document that the income of top percentiles became more exposed to aggregate shocks at the turn of the 20th century. Mankiw and Zeldes (1991) and Malloy et al. (2009) document that the consumption of rich stockholders is more exposed to stock market returns. In contemporaneous work, Kuhn et al. (2020) use Survey of Consumer Finances (SCF) data to measure how a rise in stock market returns affects the share of wealth held by the top 10%. In contrast, my empirical findings provide two novel insights. First, leveraging newly available data on top wealth inequality, I show that wealth exposure to stock market returns increases sharply further up the distribution. While Kuhn et al. (2020) find that a 10% higher-than-expected equity return raises the top 10% wealth share by 0.25%, I estimate that it raises the top 0.01% share by 3.5%—an order of magnitude larger.[3] Second, I employ local projections to estimate the full impulse response of top wealth shares to stock market returns, revealing that that the effect of stock market shocks on top wealth shares is highly persistent over time—a result that plays a central role in explaining the historical fluctuations of top wealth shares.

On the theoretical side, my paper contributes to the theoretical literature on wealth inequality. While the existing literature focuses on deterministic economies, I study a Markovian economy, where the wealth distribution *varies over time*. Using tools from large deviations theory, I show that the wealth distribution exhibits a Pareto tail, as in the deterministic case, and that its tail index can be characterized analytically. Hence, this paper shows that the characterizations of tail indices obtained in deterministic random growth models can be extended to more realistic, time-varying economies (see Champernowne, 1953 for a seminal paper, and Toda, 2014 for an extension where aggregate shocks average out over any non-infinitesimal time period). The fact that the thickness of the tail depends on the *time-averaged logarithmic* wealth growth of top households connects my paper to Kelly (1956), Blume et al. (1992) and Borovička (2020), who stress the importance

includes Kaymak and Poschke (2016), Benhabib et al. (2019), Cao and Luo (2017), Mian et al. (2020), Hubmer et al. (2021), Atkeson and Irie (2022), and Gomez and Gouin-Bonenfant (2024).

[3] These estimates also exceed those of Bach et al. (2020) for Sweden, who report that a 10% domestic equity return raises the wealth of households in the top 0.01% by 5.3%. By comparison, my estimates imply that a 10% equity return in the United States raises the wealth of top households by 9.8%.

of this quantity for long-run survival in infinite-horizon economies. Finally, my paper complements recent work by Luttmer (2012), Gabaix et al. (2016), and Cao and Luo (2017), who study the transition dynamics of the wealth distribution between two steady states. While these papers emphasize that transition dynamics are slow after a permanent change in the dynamics of individual wealth, I emphasize the flip side of this phenomenon: that one-time shocks in individual wealth have very persistent effects on the wealth distribution.

This paper also contributes to the large asset pricing literature with heterogeneous agents (Dumas, 1989; Guvenen 2009; Chan and Kogan, 2002; Basak and Cuoco, 1998; Gomes and Michaelides, 2008; Brunnermeier and Sannikov, 2014; He and Krishnamurthy, 2012; Gârleanu and Panageas, 2015). An open question in the literature is whether there is enough heterogeneity across households to account for the excess volatility of asset prices in equilibrium.[4] My paper advances on this question by using two key moments related to wealth inequality—the elasticity of top wealth shares to stock market returns and the tail index of the wealth distribution—to discipline the degree of heterogeneity across households. Another contribution of this paper is to develop an exact version of Campbell and Shiller (1988)'s decomposition of innovations in the price-dividend ratio in continuous-time economies. This decomposition, which can be computed analytically, is particularly useful for analyzing the excess volatility of returns in non-linear asset pricing models, such as models with heterogeneous agents.

More generally, this paper contributes to the growing literature on the effect of inequality on asset prices. The work of Gollier (2001) is an early example that examines theoretically the importance of the wealth distribution for asset prices. Barczyk and Kredler (2016) and Favilukis (2013) also study the role of changes in wage inequality on asset prices. More recently, Auclert and Rognlie (2017) and Straub (2019) study the effect of a secular rise in income inequality on interest rates. Toda and Walsh (2020) document that fluctuations in income inequality negatively predict future excess stock returns. Eisfeldt et al. (2023) discuss the joint relation between the wealth distribution and asset prices across markets with different expertise.

Roadmap. The rest of the paper is organized as follows. In Section 2, I document the dynamics of top wealth shares following equity returns. In Section 3, I build a perpetual-youth endowment economy model in which agents have heterogeneous exposures to aggregate shocks and I characterize the shape of the wealth distribution implied by the model. In Section 4, I calibrate the model using U.S. data; I show that the model can jointly match moments related to asset prices and wealth inequality. In Section 5, I study the impulse response function of top wealth shares to aggregate income shocks.

[4]For instance, Cochrane (2017) writes that "[the heterogeneous agents] model faces challenges and opportunities in the microdata just as the idiosyncratic risk model does. Do the 'high-beta rich' really lose so much in bad times? Can the model quantitatively account for return predictability? But that investigation has not really started."

2 Empirical results

In this section, I explore how stock market returns influence the dynamics of top wealth shares. Section 2.1 presents the data, Section 2.2 discusses the findings, and Section 2.3 examines the robustness of my results.

2.1 Data

I am interested in measuring the changes in the wealth distribution following stock market returns. Therefore, I need yearly estimates of the wealth distribution that cover several business cycles. For the baseline analysis, I use the latest version of the series of top wealth shares constructed from income tax returns by Saez and Zucman (2016) (2022 vintage), which spans from 1913 to 2020.[5] The dataset also includes a series for the average wealth in the economy. In robustness checks, I also use two alternative data series on top wealth shares found in the literature for smaller time samples: Smith et al. (2023), which spans from 1966 to 2016, and Kopczuk and Saez (2004), which spans from 1916 to 2000. It is important to note that all these series measure a *time-averaged* distribution of wealth in a given year rather than pinpointing wealth at a specific moment within the year.

I supplement these series of top wealth shares with the list of the wealthiest 400 Americans constructed by *Forbes* every September since 1982, which offers an unparalleled view on the right tail of the wealth distribution. The list is created by a dedicated staff of the magazine, based on a mix of public and private information.[6] To be consistent with the other data series, I focus on a given percentile group rather than on a given number of households (the two concepts differ in the presence of population growth). More precisely, I focus on the percentile that includes the entirety of households in the Forbes 400 list in 2017 (264 households in 1982) — it corresponds to approximately 3% of agents in the top 0.01%. I will refer to this top percentile as the top 400 in the rest of the paper.

I use the series of stock market returns and risk-free rates from Goyal et al. (2024) for asset pricing data. Stock returns correspond to the S&P 500 index returns from 1926 and returns from Robert Shiller's website beforehand. The risk-free rate corresponds to the Treasury-bill rate.[7]

[5] This series improves on the initial published series by updating the time sample and by incorporating several methodological innovations developed, among others, by Smith et al. (2023).

[6] *Forbes* reports that "we pored over hundreds of Securities Exchange Commission documents, court records, probate records, federal financial disclosures and Web and print stories. We took into account all assets: stakes in public and private companies, real estate, art, yachts, planes, ranches, vineyards, jewelry, car collections and more. We also factored in debt. Of course, we don't pretend to know what is listed on each billionaire's private balance sheet, although some candidates do provide paperwork to that effect."

[7] For the series constructed from tax return data, I construct yearly stock returns by cumulating monthly stock returns from January to December. For the series constructed from Forbes data, I construct yearly returns by cumulating monthly stock returns from October to September, consistently with the fact that the ranking tries to report the distribution of wealth in September of each year.

2.2 Findings

Response of the average wealth in top percentiles. I estimate the effect of realized stock market returns on the average wealth in top percentiles using local projection methods (Jordà, 2005). Formally, I regress the excess growth of the average wealth in a given top percentile p at different horizons on excess stock market returns:

$$\log\left(\frac{W_{p,t+h}}{W_{p,t-1}}\right) - (h+1)\log R_{f,t} = \alpha_{p,h} + \beta_{p,h}(\log R_{M,t} - \log R_{f,t}) + \epsilon_{p,t+h}, \tag{1}$$

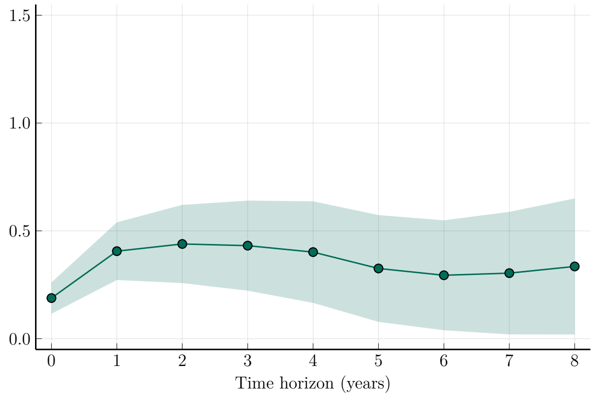
where $h \geq 0$ denotes the horizon, $W_{p,t}$ denotes the average wealth of households in the top percentile p in year t, $\log R_{M,t}$ denotes the log stock market return, and $\log R_{f,t}$ denotes the log risk-free rate. Note that both the growth of the average wealth in a top percentile (the dependent variable) and the stock market return (the independent variable) are adjusted by the risk-free rate, as a way to capture expected changes in these variables (e.g., expected inflation). I will discuss alternative specifications in Section 2.3. Following Herbst and Johannsen (2021), I estimate standard errors using heteroskedasticity-consistent (Huber-White) estimators.[8]

Figure 1 plots the estimates of $\beta_{p,h}$ for $0 \leq h \leq 8$ and $p \in \{100\%, 1\%, 0.1\%, 0.01\%, \text{Top } 400\}$. There are three important observations. First, the estimates increase monotonically across top percentiles. Second, within each top percentile, the elasticities initially increase with the horizon. One reason is that $W_{p,t}$ represents the *time-averaged* wealth in a given percentile during the year; as a result, the effect of the cumulative stock market return in year t is only fully incorporated by year $t+1$ (i.e., at $h = 1$ rather than at $h = 0$). Another potential reason is that a large share of wealth in top percentiles is held in privately held assets, whose valuations tend to react sluggishly to changes in the stock market.[9] Third, the effect of stock market returns tends to be very persistent, with little mean reversion over time (note, however, that these estimates become less precise as the horizon grows). As we will see in the model below, this reflects that top percentiles mean-revert very slowly after shocks.

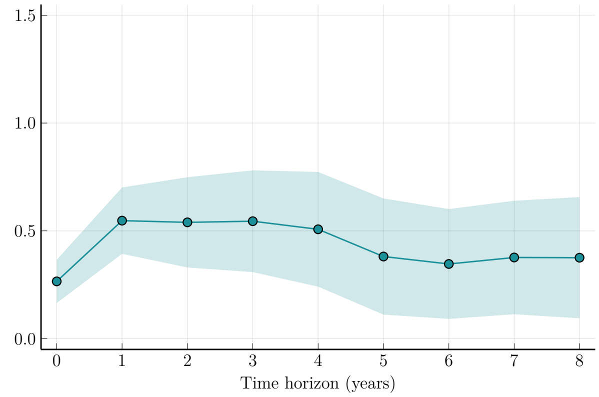
Table 1 reports the estimates of $\beta_{p,h}$ for $h = 3$, corresponding to the horizon at which the impulse responses peak for top percentiles. The estimates increase with top percentiles, from $\beta = 0.43$ for the average household to $\beta = 0.78$ for households in the top 0.01% and $\beta = 0.94$ for households in the top 400. In short, these estimates suggest that the wealth of households in the right tail of the wealth distribution tends to be twice as exposed to stock market returns relative to the average household in the economy.

[8]While Jordà (2005) recommends using Newey-West standard errors, Herbst and Johannsen (2021) stress that they can be downward biased in finite samples and recommend using heteroskedasticity-consistent standard errors. In line with their results, I find that robust standard errors give me wider standard errors than Newey-West, so I report the former ones to be conservative.

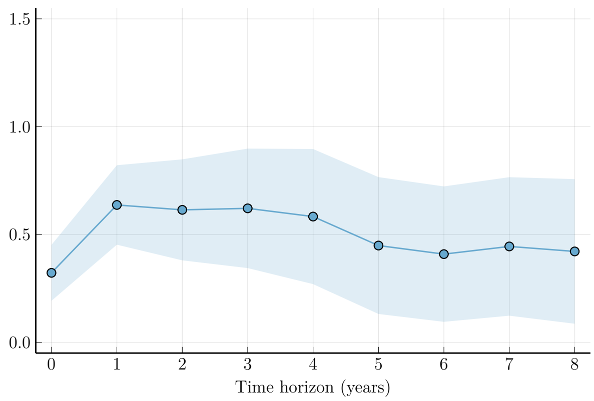
[9]This is particularly true for estimates from *Forbes*, who often use the valuation implied by the last financing round. Relatedly, I find a similar pattern for measures of wealth constructed from estate tax returns (Kopczuk and Saez, 2004), where an external appraiser does the valuation of non-tradable assets.



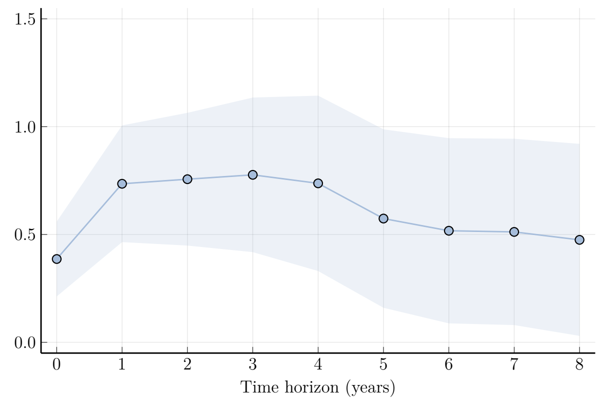
(a) Top 100%



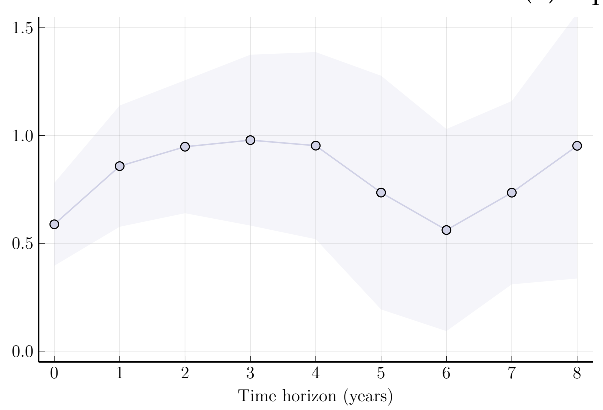
(b) Top 1%



(c) Top 0.1%



(d) Top 0.01%



(e) Top 400

Figure 1: Response of the average wealth in top percentiles to excess stock returns

Notes: The figure reports the estimates for $\beta_{p,h}$ estimated via the regression (1) for $0 \leq h \leq 8$ as well as their 5%–95% confidence intervals using heteroskedasticity consistent standard errors. Each figure corresponds to a different top percentile. Figure 1a corresponds to the average wealth of U.S. households ($p = 100\%$). Figures 1b-1d correspond to the top 1%, 0.1%, 0.01% using data from Saez and Zucman (2016) (2022 vintage). Figure 1e corresponds to Forbes 400.

Response of top percentile wealth shares. The previous results show that stock market returns have higher effects on the average wealth in top percentiles than on the average wealth in the economy. Mechanically, this means that stock market returns tend to increase top percentile wealth

Table 1: Wealth exposure to stock returns across top percentiles

	Top 100%	Top 1%	Top 0.1%	Top 0.01%	Top 400
	(1)	(2)	(3)	(4)	(5)
Panel A: Average wealth					
Excess returns	0.43***	0.54***	0.62***	0.78***	0.98***
	(0.11)	(0.12)	(0.14)	(0.18)	(0.20)
Adjusted R^2	0.16	0.20	0.19	0.18	0.31
Time sample	1914-2016	1914-2016	1914-2016	1914-2016	1984-2014
N	103	103	103	103	31
Panel B: Wealth share					
Excess returns		0.11**	0.19**	0.35**	0.59***
		(0.05)	(0.09)	(0.14)	(0.20)
Adjusted R^2		0.05	0.04	0.06	0.22
Time sample		1914-2016	1914-2016	1914-2016	1984-2014
N		103	103	103	31

Notes: Panel A reports the results of regressing the four-year growth of the average wealth in a given percentile group on excess stock returns; that is, equation (1) with $h = 3$. Panel B reports the same regression using the four-year growth of the top wealth share as the left-hand-side variable; that is, equation (2) with $h = 3$.. Each column corresponds to a different top percentile. Column (1) corresponds to the average U.S. household ($p = 100\%$). Columns (2)–(4) correspond to increasing top percentiles in the wealth distribution using data from Saez and Zucman (2016) (2022 vintage). Column (5) corresponds to Forbes 400. Estimation is done via OLS. Standard errors are in parentheses and are estimated using heteroskedasticity consistent standard errors. *,** ,*** indicate significance at the 10%, 5%, 1% levels, respectively.

shares. To see this formally, I estimate regressions of the form

$$\log\left(\frac{S_{p,t+h}}{S_{p,t-1}}\right) = a_{p,h} + b_{p,h}(\log R_{M,t} - \log R_{f,t}) + e_{p,t+h}, \tag{2}$$

where $S_{p,t} \equiv pW_{p,t}/W_{100\%,t}$ denotes the share of aggregate wealth owned by individuals in the top percentile p (i.e., the top percentile wealth share). Note that (2) can be obtained by taking the difference of (1) between p and $p = 100\%$: intuitively, the exposure of the share of wealth owned by a top percentile is the difference between the exposure of the average wealth in the top percentile and the average wealth in the population; that is, $b_{p,h} = \beta_{p,h} - \beta_{100\%,h}$. Still, running this specification allows me to test whether this difference is statistically significant.

Panel B of Table 1 reports the estimates for $b_{p,h}$ at the four-year horizon. Consistently with the discussion above, the estimate 0.35 for the top 0.01% corresponds precisely to the difference between the wealth exposure of households in the top 0.01% and the average household in the economy; that is, $0.35 = 0.78 - 0.43$. Note that the difference is significant at the 1% level. Finally, similarly to Figure 1, Online Appendix Figure A5 reports the impulse response of top wealth shares obtained by plotting the estimated $b_{p,h}$ from (2) for $0 \leq h \leq 8$ for different top percentiles.

2.3 Robustness checks

I now explore the robustness of my findings across three key aspects: empirical specifications, alternative data sources, and shifts in the composition of households in top percentiles. I briefly summarize the results below, relegating the reader to Online Appendix B for details.

Alternative specifications. For my baseline results, I estimated local projections using simple univariate regressions (i.e., excess stock returns as the only regressors). One reason is that these univariate regressions allow for a straightforward mapping between the response of the average wealth in a top percentile (1) and the response of the top percentile wealth share (2).

In the spirit of local projections, however, I now augment the specifications with predetermined controls; that is, variables known at time $t-1$ that capture the information available at that time. These variables help isolate the effect of *unexpected* stock market returns on the wealth distribution. Consistently with the usual intuition for omitted variable biases, I focus on variables that either correlate with the treatment (excess returns) or the outcome (growth of average wealth in top percentile / top percentile wealth share). In Online Appendix B.1, I show that augmenting the baseline specification with these controls does not significantly change the response of top wealth shares to stock market returns (Table OA1). Intuitively, this comes from the fact that fluctuations in excess stock returns are hard to predict (i.e., they are only weakly correlated with variables known at time $t-1$).

Another potential concern is that, when the treatment exhibits serial correlation, local projections capture both the direct effect of a higher-than-average treatment *and* its indirect effect through higher future treatments on average. To isolate the direct effect, Alloza et al. (2020) suggests augmenting local projections with controls for future realized treatments. In Online Appendix B.1, I show that I obtain similar results when doing so (Table OA2), indicating that the response of top wealth shares obtained in my baseline specification is entirely driven by the direct effect of higher stock returns at $h = 0$, rather than by indirect effects through higher (or lower) average returns at $h \geq 1$. This reflects that the serial correlation in excess stock returns is close to zero empirically.[10]

Alternative data sources. There is substantial uncertainty about the historical dynamics of top wealth shares. In Online Appendix B.2, I show that my results remain similar when using the two main alternative data series for top wealth shares available from the literature: the series from Kopczuk and Saez (2004), constructed from estate tax returns, and the series from Smith et al. (2023). The conclusion is that, while these series disagree on the low-frequency fluctuations in top wealth shares, they tend to imply similar responses of top wealth shares to stock market returns.

[10] More generally, note that I will rely on these empirical results to discipline a model (see Section 4). Hence, some degree of misspecification in these regressions is acceptable, provided I consistently apply the same specification in both the model and the data.

As a further robustness check, I also show in Online Appendix B.2 that my estimates for the elasticity of top percentile wealth to stock market returns are consistent with the share of wealth invested in equity in different part of the wealth distribution, as reported in the Survey of Consumer Finances.

Accounting for composition effects. Top percentiles do not include the same individuals over time. As a result, changes in the average wealth in a top percentile can be driven by the wealth changes of individuals *initially* in the top percentile (an "intensive" term) or by changes in the composition of individuals in the top percentile (an "extensive" term). Do these composition effects matter for my estimates?

To answer this question, I decompose the growth of the average wealth in the top 400 into these two terms using the same methodology as Gomez (2023). As shown in Online Appendix B.3 (Figure OA3), I find the response of the average wealth in the top 400 to stock market returns is almost entirely driven by the intensive term rather than by the extensive term. Said differently, high stock market returns increase the average wealth in the top 400 because they increase the average wealth of agents who were *initially* in the top 400, not because they increase the arrival of new fortunes in the top 400.[11]

3 A model of wealth inequality with aggregate shocks

Motivated by the reduced-form evidence presented in the previous section, I now build an asset pricing model in which certain agents ("entrepreneurs") are required to hold a large share of their wealth in equity. Because the relative proportion of entrepreneurs increases in the right tail of the wealth distribution, higher equity returns increase top wealth shares, as in the data. Section 3.1 presents the model, Section 3.2 solves for the Markovian equilibrium, Section 3.3 characterizes the wealth distribution implied by the model, and Section 3.4 discusses potential extensions.

3.1 Setup

The model is a continuous time, pure-exchange economy with two types of agents: "households," who can freely trade firms, and "entrepreneurs," who must remain disproportionately exposed to the firms they are born with.

Demography. The demographic structure follows the perpetual youth model of Blanchard (1985), wherein agents face a constant hazard rate of death δ and the population grows at rate η. Consequently, over a short interval $\mathrm{d}t$, a proportion $\delta\,\mathrm{d}t$ of the population dies, while a proportion

[11]This empirical fact is consistent with Proposition 4, which shows that composition changes become second-order as the time-horizon tends to zero.

$(\delta + \eta)\,\mathrm{d}t$ is born. In the model (as in the data), these demographic forces play an essential role in making the wealth distribution stationary.

A proportion π_E of agents are born as "entrepreneurs" while the rest $\pi_H = 1 - \pi_E$ are born as "households". I denote $\mathbb{I}_{Ht}$ the set of households, $\mathbb{I}_{Et}$ the set of entrepreneurs, and $\mathbb{I}_t \equiv \mathbb{I}_{Ht} \cup \mathbb{I}_{Et}$ the set of all agents in the economy at time t.

Endowment. Aggregate income per capita Y_t follows a geometric random walk; that is,

$$\frac{\mathrm{d}Y_t}{Y_t} = g\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t, \tag{3}$$

where $(Z_t)_{t\in\mathbb{R}}$ is a standard Brownian motion that represents aggregate shocks, g represents the growth rate of the economy per capita, and σ represents the volatility of aggregate income.

Each agent is born with a tree that delivers a stochastic flow of income. Formally, each tree i produces an income flow $Y_{it} = s_{it} Y_t$, where s_{it} evolves as

$$\frac{\mathrm{d}s_{it}}{s_{it}} = -\phi\,\mathrm{d}t + \nu\,\mathrm{d}B_{it},$$

where $(B_{it})_{t\in\mathbb{R}}$ is a standard Brownian motion that represents shocks specific to the tree i, ϕ represents the rate of depreciation of the tree, and ν represents its idiosyncratic volatility. For the income of all trees in existence to sum up to aggregate income, the initial value of s_{it} for trees at birth must average to $(\eta + \phi)/(\eta + \delta)$.[12]

Finally, I assume that the wealth of agents who die is redistributed to newborn agents; that is, newborns are endowed with new trees and old trees from deceased agents. I assume that the distribution of this initial endowment among newborns is independent of their types and that it follows a log-normal distribution with variance ν_0^2.

Markets. Agents in the economy can trade risk-free claims in zero net supply and claims to trees. Denote r_t the risk-free rate and p_t the market value of a tree relative to its income.[13] We guess that the process p_t evolves according to

$$\frac{\mathrm{d}p_t}{p_t} = \mu_{pt}\,\mathrm{d}t + \sigma_{pt}\,\mathrm{d}Z_t, \tag{4}$$

[12] Indeed, in this case, one can integrate the income flow of all trees in existence to obtain

$$\int_{s=-\infty}^{t} (\eta + \delta)|\mathbb{I}_t| e^{-\eta(t-s)} \left(\frac{\eta + \phi}{\eta + \delta} e^{-\phi(t-s)} Y_t \right) \mathrm{d}s = |\mathbb{I}_t| Y_t.$$

Here, and in the rest of the paper, $|\mathbb{X}|$ denotes the mass of a set $\mathbb{X}$.

[13] It is identical across trees as they all have the same law of motions for income.

where μ_{pt} and σ_{pt} will be determined in equilibrium. The instantaneous return of holding tree i between t and $t + \mathrm{d}t$ is the sum of it income yield and the growth in its market value:[14]

$$\begin{aligned}\frac{\mathrm{d}R_{it}}{R_{it}} &= \frac{1}{p_t}\,\mathrm{d}t + \frac{\mathrm{d}\left(Y_{it}p_t\right)}{Y_{it}p_t} \\ &= \underbrace{\left(\frac{1}{p_t} + g - \phi + \mu_{pt} + \sigma\sigma_{pt}\right)}_{\equiv \mu_{Rt}}\mathrm{d}t + \underbrace{\left(\sigma + \sigma_{pt}\right)}_{\equiv \sigma_{Rt}}\mathrm{d}Z_t + \nu\,\mathrm{d}B_{it},\end{aligned} \tag{5}$$

where the second line uses Ito's lemma.

Households. Households have Duffie and Epstein (1992) preferences, which correspond to the continuous-time version of the recursive preferences of Epstein and Zin (1989). More precisely, the welfare of a household i with consumption process $(C_{it})_{t\in\mathbb{R}}$ is defined recursively by

$$\begin{aligned} V_{it} &= \mathrm{E}_t\left[\int_t^\infty f(C_{iu}, V_{iu})du\right], \\ \text{with } f(C,V) &= \rho\frac{1-\gamma}{1-1/\psi}V\left(\frac{C^{1-1/\psi}}{((1-\gamma)V)^{\frac{1-1/\psi}{1-\gamma}}} - 1\right). \end{aligned}$$

These preferences are characterized by three parameters: the subjective discount rate (SDR) ρ, the elasticity of intertemporal substitution (EIS) ψ, and the coefficient of relative risk aversion (RRA) γ.[15]

Households can freely sell their initial tree and use the proceeds to invest in a diversified portfolio of trees. Formally, household $i \in \mathbb{I}_{Ht}$ chooses a share of wealth invested in a diversified portfolio of trees, α_{it}, and a consumption rate $c_{it} = C_{it}/W_{it}$ to maximize their welfare. The Hamilton-Jacobi-Bellman (HJB) equation corresponding to this problem is

$$\begin{aligned} 0 &= \max_{\alpha_{it}, c_{it}}\left\{f\left(c_{it}W_{it}, V_{it}\right)\mathrm{d}t + \mathrm{E}_t[\mathrm{d}V_{it}]\right\} \\ \text{with } \frac{\mathrm{d}W_{it}}{W_{it}} &= \left(r_t + \alpha_{it}(\mu_{Rt} - r_t) - c_{it}\right)\mathrm{d}t + \alpha_{it}\sigma_{Rt}\,\mathrm{d}Z_t. \end{aligned} \tag{6}$$

Given homothetic preferences and linear budget constraints, we know that all households will choose the same share of wealth invested in equity and consumption rate (irrespective of their wealth), which we denote by α_{Ht} and c_{Ht}, respectively.

[14] Here, R_{it}/R_{is} denotes the *cumulative* return of owning the tree i up from s to t.

[15] As shown in Gârleanu and Panageas (2015), the SDR ρ should be interpreted as the sum of a subjective impatience rate $\hat{\rho}$ and of the hazard rate of death δ.

Entrepreneurs. In contrast with households, entrepreneurs are required to hold an exogenous share of wealth α_{Et} in the tree they are born with:

$$\alpha_{Et} = \min\left(\alpha_E, \frac{\int_{i\in\mathbb{I}_t} W_{it}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i}\right), \tag{7}$$

where $\alpha_E > 1$. The upper bound on the risk exposure α_{Et} ensures that entrepreneurs are not required to own more trees than there exist in the economy. This constraint will rarely bind in equilibrium, but it is necessary to solve the model globally.[16] For simplicity, I take this equity constraint as exogenous, and I remain agnostic about its origin. As in Di Tella (2017), this constraint could be motivated by a moral hazard or asymmetric information problem. Alternatively, the over-exposure of entrepreneurs could represent optimism in their projects (Moskowitz and Vissing-Jørgensen, 2002), a preference for idiosyncratic volatility (Roussanov, 2010), or a higher risk tolerance (Gârleanu and Panageas, 2015).

For simplicity, I assume that entrepreneurs have Epstein-Zin utility with an EIS of one. This value fits with the estimates for Vissing-Jørgensen (2002) for the EIS of individuals at the top of the wealth distribution.[17] With these assumptions, the wealth of an entrepreneur $i \in \mathbb{I}_{Et}$ evolves as

$$\frac{\mathrm{d}W_{it}}{W_{it}} = (r_t + \alpha_{Et}(\mu_{Rt} - r_t) - \rho_E)\,\mathrm{d}t + \alpha_{Et}\sigma_{Rt}\,\mathrm{d}Z_t + \alpha_{Et}\nu\,\mathrm{d}B_{it}. \tag{8}$$

The law of motion for wealth directly depends on the entrepreneurs' fixed equity share α_{Et} and their fixed consumption rate $c_{Et} = \rho$, which will allow for a transparent calibration of these parameters based on the observed wealth dynamics of agents at the top of the wealth distribution. Finally, note that the risk aversion of entrepreneurs does not matter for the equilibrium, as it neither affects their consumption rate (which is pinned down by ρ_E) nor their share of wealth invested in equity (which is pinned down by α_E).

Finally, note that the wealth of entrepreneurs in (8) is exposed to the idiosyncratic risk of the tree they are born with. While this does not affect the aggregate demand for goods and assets in equilibrium (as entrepreneurs have a fixed equity share and consumption rate), it is important to generate a realistic wealth distribution.

Equilibrium. An equilibrium for the model is defined as a set of price processes $(r_t)_{t\in\mathbb{R}}$, $(p_t)_{t\in\mathbb{R}}$ and decision processes for the households $(c_{Ht})_{t\in\mathbb{R}}$, $(\alpha_{Ht})_{t\in\mathbb{R}}$ such that

1. Given the price processes, the decision processes solve the household problem (6).

[16] More precisely, this constraint will bind less than 0.01% of the time in the calibrated model.

[17] In contrast, I allow for the EIS of households to differ from one, which is consistent with micro- evidence for the average household (Vissing-Jørgensen, 2002; Best et al., 2020).

2. The market for goods and risky assets clear; that is

$$\int_{i\in\mathbb{I}_{Et}} \rho_E W_{it}\,\mathrm{d}i + \int_{i\in\mathbb{I}_{Ht}} c_{Ht} W_{it}\,\mathrm{d}i = Y_t|\mathbb{I}_t|, \tag{9}$$

$$\int_{i\in\mathbb{I}_{Et}} \alpha_{Et} W_{it}\,\mathrm{d}i + \int_{i\in\mathbb{I}_{Ht}} \alpha_{Ht} W_{it}\,\mathrm{d}i = p_t Y_t|\mathbb{I}_t|. \tag{10}$$

By Walras's law, the market for risk-free claims clears automatically.

3.2 Solving the model

I now outline the main steps in deriving the solution in this section (see Appendix C.1 for more details).

Household optimal policy. We guess that the value function of households takes the form

$$V_{it} = \frac{(\chi_t W_{it})^{1-\gamma}}{1-\gamma}, \tag{11}$$

where the process χ_t, which captures the investment opportunities faced by the households, follows a diffusion process

$$\frac{\mathrm{d}\chi_t}{\chi_t} = \mu_{\chi t}\,\mathrm{d}t + \sigma_{\chi t}\,\mathrm{d}Z_t,$$

where $\mu_{\chi t}$ and $\sigma_{\chi t}$ will be determined in equilibrium. Plugging (11) into the household's HJB (6) and applying Ito's lemma gives

$$\begin{aligned} 0 = \max_{c_{Ht},\alpha_{Ht}} \Bigg\{ & \frac{\rho}{1-1/\psi}\left(\left(\frac{c_{it}}{\chi_t}\right)^{1-1/\psi} - 1\right) \\ & + r_t + \alpha_{it}(\mu_{Rt} - r_t) - c_{Ht} + \mu_{\chi t} - \frac{\gamma}{2}\left(\alpha_{Ht}^2\sigma_{Rt}^2 + \sigma_{\chi t}^2 - 2\frac{1-\gamma}{\gamma}\alpha_{Ht}\sigma_{Rt}\sigma_{\chi t}\right)\Bigg\}. \end{aligned} \tag{12}$$

The first-order conditions of this problem give

$$c_{Ht} = \rho^\psi \chi_t^{1-\psi}, \tag{13}$$

$$\alpha_{Ht} = \frac{1}{\gamma}\frac{\mu_{Rt} - r_t}{\sigma_{Rt}^2} + \frac{1-\gamma}{\gamma}\frac{\sigma_{\chi t}}{\sigma_{Rt}}. \tag{14}$$

Markov equilibrium. Households and entrepreneurs' policy functions are linear in wealth. As a result, the distribution of wealth within each type does not matter for aggregate demand; only the distribution of wealth between types does. Accordingly, I look for a Markovian equilibrium where the (endogenous) state variable is the share of aggregate wealth owned by entrepreneurs: $x_t = \int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i / \left(\int_{i\in\mathbb{I}_t} W_{it}\,\mathrm{d}i\right)$.

Using this notation, the market clearing equations (9) and (10) can be rewritten as:

$$x_t \rho_E + (1 - x_t) c_{Ht} = \frac{1}{p_t}, \tag{15}$$

$$x_t \alpha_{Et} + (1 - x_t) \alpha_{Ht} = 1. \tag{16}$$

The first equation says that the wealth-weighted average consumption rate equals the income yield of the tree, while the second equation says that the wealth-weighted average equity share equals one.

We have five unknown functions of x: $r_t = r(x_t)$, $p_t = p(x_t)$, $\chi_{Ht} = \chi_H(x_t)$, $\alpha_{Ht} = \alpha_H(x_t)$, and $c_{Ht} = c_H(x_t)$. The market clearing equations (15) and (16) and the optimization conditions for households (12), (13), and (14) constitute a system of five equations. To solve for the equilibrium, it remains to solve for the law of motion of the endogenous state variable x_t using Ito's lemma:

Proposition 1. *The law of motion of x_t is given by*

$$\mathrm{d}x_t = \mu_{xt}\,\mathrm{d}t + \sigma_{xt}\,\mathrm{d}Z_t, \textit{ where}$$

$$\mu_{xt} \equiv x_t(1-x_t)\left((\alpha_{Et} - \alpha_{Ht})(\mu_{Rt} - r_t) + c_{Ht} - c_{Et} - (\alpha_{Et} - \alpha_{Ht})\sigma_{Rt}^2 + (\eta + \delta + \phi)\left(\frac{\pi_E}{x_t} - \frac{\pi_H}{1-x_t}\right)\right)$$

$$\sigma_{xt} \equiv x_t(1-x_t)(\alpha_{Et} - \alpha_{Ht})\sigma_{Rt}.$$

The volatility of x_t corresponds to the difference in risk exposure between entrepreneurs and households. The drift of x_t is the sum of four terms: the difference in portfolio returns between entrepreneurs and households, the difference in their consumption rates, an Ito's term that accounts for the difference in their risk exposures, and a demography term related to the overlapping generation setting (i.e., due to population growth and death). Due to the demography term, we have $\mu_{xt}(0) > 0$ and $\mu_{xt}(1) < 0$. Together with $\sigma_{xt}(0) = \sigma_{xt}(1) = 0$, this ensures that the boundaries $x_t = 0$ and $x_t = 1$ are not absorbing states.

In the rest of the paper, I assume this Markov equilibrium exists, is unique, and that that the endogenous equilibrium quantities $r(\cdot), p(\cdot), \chi_H(\cdot), \alpha(\cdot)$ and $c_H(\cdot)$ are all twice differentiable with respect to x. Given the law of motion from Proposition 1, this implies that the process $(x_t)_{t\in\mathbb{R}}$ is recurrent with a unique stationary density (Karlin and Taylor, 1981 Chapter 15).[18]

3.3 The cross-sectional distribution of wealth

I now study the cross-sectional distribution of wealth implied by the model. Since the economy grows over time, I focus on individual wealth *normalized* by the average wealth in the economy: $w_{it} \equiv W_{it}/(p_t Y_t)$. Ito's lemma provides the following expression for the law of motion of w_{it} for

[18] See Proposition 3.2 in Borovička (2020) as well as its Online Appendix for an exploration of the existence and smoothness of policy function in heterogeneous agent models, as well as a detailed analysis of probability measure in heterogeneous agent models.

household i of type $j \in \{E, H\}$:[19]

$$\begin{aligned}
\frac{\mathrm{d}w_{it}}{w_{it}} &= \mu_{wjt}\,\mathrm{d}t + \sigma_{wjt}\,\mathrm{d}Z_t + \nu_{wjt}\,\mathrm{d}B_{it}, \text{ where} \\
\mu_{wjt} &\equiv r_t + \alpha_{jt}(\mu_{Rt} - r_t) - c_{jt} - g - \mu_{pt} - \sigma\sigma_{pt} - (\alpha_{jt} - 1)\sigma_{Rt}^2 \\
\sigma_{wjt} &\equiv (\alpha_{jt} - 1)\sigma_{Rt} \\
\nu_{wjt} &\equiv 1_{j=E}\alpha_{Et}\nu.
\end{aligned} \tag{17}$$

Cumulative distribution function. I first analyze the cumulative distribution function of wealth. The law of motion (17) implies that log wealth follows a random walk, with type-specific and time-varying drift and volatility. Hence, within each type and cohort, the distribution of wealth is log-normal. As a result, the overall distribution of wealth across types and cohorts is a mixture of log-normal distributions, which leads to the following proposition.

Proposition 2. *The cumulative distribution of wealth at time t is given by:*[20]

$$\mathbb{P}_t\left(w_{it} \le w | i \in \mathbb{I}_t\right) = \sum_{j \in \{E,H\}} \pi_j \int_{-\infty}^{t} (\eta + \delta) e^{-(\eta+\delta)(t-s)} \Phi\left(\frac{\log w - \mu_{j,s\to t}}{\nu_{j,s\to t}}\right) \mathrm{d}s, \tag{18}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal distribution, and $\mu_{j,s\to t}$ and $\nu_{j,s\to t}^2$ denote the cross-sectional mean and variance of log wealth of individuals of type $j \in \{E, H\}$ born at time $s \le t$

$$\begin{aligned}
\mu_{j,s\to t} &\equiv \log\left(\frac{\eta + \delta + \phi}{\eta + \delta}\right) - \frac{1}{2}\nu_{j,s\to t}^2 + \int_s^t \left(\mu_{wju} - \frac{1}{2}\sigma_{wju}^2\right) \mathrm{d}u + \int_s^t \sigma_{wju}\,\mathrm{d}Z_u \\
\nu_{j,s\to t}^2 &\equiv \nu_0^2 + \int_s^t \nu_{wju}^2\,\mathrm{d}u.
\end{aligned}$$

This proposition expresses the cross-sectional distribution of (normalized) wealth as a mixture of log-normal distributions, where the mixture weights $\pi_j(\eta + \delta)e^{-(\eta+\delta)(t-s)}$ correspond to the relative mass of individuals of type j in the cohort born at time s. The cross-sectional mean $\mu_{j,s\to t}$ and variance $\nu_{j,s\to t}^2$ of the log-normal distributions vary across cohorts and types, reflecting the heterogeneity in their ages, in the economic conditions since their birth, and in their policy functions. In Online Appendix C.2, I derive a similar analytical expression for the total normalized wealth above a threshold (or, equivalently, for the *share* of aggregate wealth owned by individuals above that threshold), and, more generally, for any moment of normalized wealth above a threshold.

Right tail of the wealth distribution. While the distribution of wealth depends non trivially on the history of preceding aggregate shocks, we can obtain a simple characterization of its right tail.

[19] This can be obtained by combining the law of motion of individual wealth (6) and (8) with the laws of motion of Y_t (3) and p_t (4).

[20] Here, and in the rest of the paper, $\mathbb{P}_t$ (resp. $\mathbb{E}_t$) denotes the probability (resp. expectation) with respect to the cross-sectional distribution of normalized wealth at time t.

Definition 1. The distribution of a random variable Z has a right Pareto tail if there exists $\zeta \in (0,\infty)$ such that

$$\log \mathbb{P}(Z \geq z) \sim -\zeta \log z \quad \text{as } z \to \infty. \tag{19}$$

ζ is called the *tail index* of the distribution.[21]

Intuitively, we say that a wealth distribution has a right Pareto tail when the log rank of individuals (the left-hand-side) becomes asymptotically proportional to their log wealth (the right-hand-side) as wealth converges to infinity.[22] A lower coefficient of proportionality ζ corresponds to a larger increase in wealth for a given decrease in rank, that is, to a higher level of inequality. The following proposition, proved using tools from large deviations theory, states that the wealth distribution implied by the model has a right Pareto tail, and derives a simple formula for its tail index.

Proposition 3 (Tail index)**.** *Let ζ_j be the positive root of the equation:*[23]

$$\zeta_j \mathrm{E}\left[\mu_{wjt} - \frac{1}{2}\sigma_{wjt}^2\right] + \frac{1}{2}\zeta_j(\zeta_j - 1)\mathrm{E}\left[\nu_{wjt}^2\right] - (\eta + \delta) = 0 \quad \textit{for } j \in \{H, E\}, \tag{20}$$

where E *denotes the expectation with respect to the stationary density of x. At any time t,*

1. *The distribution of wealth within type $j \in \{E, H\}$ has a right Pareto tail with tail index ζ_j if $\zeta_j < \infty$.*
2. *The distribution of wealth has a right Pareto tail with tail index* $\min(\zeta_E, \zeta_H)$*.*

The first part of the proposition states that ζ_j corresponds to the tail index for the wealth distribution within type $j \in \{E, H\}$. Since the overall wealth distribution is a mixture of these two distributions, it inherits the minimum of these two tail indices, resulting in a tail index of $\zeta = \min(\zeta_E, \zeta_H)$. Moreover, the fact that the share of wealth owned by entrepreneurs is a stationary process within $(0, 1)$ implies that $\min(\zeta_E, \zeta_H) > 1$; that is, the right tail of the wealth distribution is thinner than Zipf's law. In the remainder of this section, I assume that the right tail for entrepreneurs is "thicker" than the right tail of households; that is $\zeta_E < \zeta_H$ (this will also hold in the calibrated model).[24] As shown in the Proof of Proposition 3, this implies that the relative proportion of entrepreneurs converges to one in the right tail of the wealth distribution.

[21]Here, and in the rest of the paper $f(z) \sim g(z)$ as $z \to \infty$ for two functions $f(\cdot)$ and $g(\cdot)$ means $\frac{f(z)}{g(z)} \to 1$ as $z \to \infty$.

[22] This definition, which characterizes the asymptotic limit of the log CDF, often appears in the literature (e.g., Nakagawa, 2007 or Beare and Toda, 2022). Note that it is weaker than the statement $\mathbb{P}(Z \geq z) \sim Cz^{-\zeta}$ as $z \to \infty$, which would hold if the economy was deterministic (see, for instance, Reed, 2001).

[23]When $j = E$, the positive root always exists as $\mathrm{E}[\nu_{wjt}^2] > 0$. When $j = H$, ζ_H should be understood as the limit of the positive root as idiosyncratic volatility tends to zero; that is, $\zeta_H = (\eta + \delta)/\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]$ if $\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right] > 0$, and $+\infty$ otherwise.

[24]A sufficient condition is that entrepreneurs grow faster than households in average; that is, $\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right] \leq \mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma_{wEt}^2\right]$.

To better understand the analytical characterization of the tail index given in (20), it is useful to discuss it within the context of the existing literature. It is well known that in a static economy where individual wealth follows a geometric diffusion with drift μ, idiosyncratic volatility ν, and death rate $\eta + \delta$, the stationary wealth distribution has a tail index given by the positive root of[25]

$$\zeta\mu + \frac{1}{2}\zeta(\zeta - 1)\nu^2 - (\eta + \delta) = 0. \tag{21}$$

Proposition 3 extends this fundamental result in two ways: to an economy where the dynamics of individual wealth *varies over time* and *is exposed to aggregate shocks*. First, to account for time-varying wealth dynamics, the geometric drift and variance of wealth μ and ν^2 must be replaced by their time-averaged counterparts; that is, $\mathrm{E}\left[\mu_{wjt}\right]$ and $\mathrm{E}\left[\nu^2_{wjt}\right]$. Second, to account for the presence of aggregate shocks, the geometric drift must be adjusted by an Ito term, $-\frac{1}{2}\mathrm{E}\left[\sigma^2_{wjt}\right]$, which captures the negative effect of aggregate shocks on the average logarithmic growth of individuals at the top.

An alternative interpretation of (20) can be obtained by dividing each term by ζ_j:

$$\mathrm{E}\left[\mu_{wjt} - \frac{1}{2}\sigma^2_{wjt}\right] + \frac{1}{2}(\zeta_j - 1)\mathrm{E}\left[\nu^2_{wjt}\right] - \frac{1}{\zeta_j}(\eta + \delta) = 0. \tag{22}$$

As in Gomez (2023), the left-hand side can be interpreted as the logarithmic growth rate of top wealth shares in the right tail of the distribution. Indeed, the first term, $\mathrm{E}\left[\mu_{wjt} - \frac{1}{2}\sigma^2_{wjt}\right]$, corresponds to the time-averaged logarithmic growth of the wealth of agents in the top (an intensive margin) while the two other terms capture the effect of composition changes due to idiosyncratic returns and demographic forces (an extensive margin). Equation (22) says that, for top wealth shares to neither grow or shrink on average over time, their logarithmic growth must average to zero, which pins down the tail index ζ_j.

One surprising implication of this proposition is that the tail index does not vary with aggregate shocks or with the state of the economy. To understand the intuition, recall that the fraction of entrepreneurs tends to one in the right tail of the wealth distribution. As a result, when an aggregate shock occurs, all agents in the right tail move by the same relative amount, which implies that the tail index (a measure of inequality within the wealthy) remains unchanged.[26] The next paragraph discusses this phenomenon in more detail.

Exposure of top wealth shares to aggregate shocks. I now analyze the response of top wealth shares to aggregate shocks. As in the previous section, I denote $S_{p,t}$ the share of aggregate wealth

[25] See, for instance, Champernowne (1953) and Reed (2001).

[26] The fact that the tail index does not depend on the state of the economy (i.e., that aggregate shocks do not affect the tail index, even after some time) is more subtle. It hinges critically on the Markovian structure of the economy, which allows the application of the law of large numbers to the past economic conditions faced by individuals in the right tail of the distribution (see the proof of the proposition in Appendix A for more detail).

owned by individuals in the top percentile p.

Proposition 4. *The instantaneous response of top wealth shares to aggregate shocks can be expressed as*

$$\frac{\partial \log S_{p,t}}{\partial Z_t} = \sum_{j \in \{E,H\}} F_{p,t}(j) \sigma_{wjt}, \tag{23}$$

where $F_{p,t}(j)$ denotes the fraction of wealth in the top percentile p at time t owned by individuals of type $j \in \{E, H\}$.[27]

This proposition expresses the exposure of a top percentile wealth share to aggregate shocks as the wealth-weighted average wealth exposure of individuals in the top percentile. In other words, while the composition of households in the top percentile changes over time, these composition changes do not affect the instantaneous exposure of top wealth shares.[28] In Section 5, we will generalize this proposition and show that composition changes do affect the response of top wealth shares over non-infinitesimal horizons (see Proposition 7).

This proposition implies that the exposure of top wealth shares increases across the wealth distribution and, thus, that the model can generate the reduced-form evidence documented above. To see why, note that, at the bottom of the wealth distribution (i.e., as $p \to 1$), top wealth shares converge to one, and thus the risk exposure of top wealth shares converges to zero. In contrast, in the right tail of the wealth distribution (i.e., as $p \to 0$), the fraction of entrepreneurs converges to one, and thus the risk exposure of top wealth shares converges to $\sigma_{wEt} = (\alpha_E - 1)\sigma_{Rt} > 0$.[29]

We can also use this proposition to characterize the effect of an aggregate shock on the ratio between two consecutive top wealth shares, which is often used as a measure of inequality in the tail.[30] For any given $p \in (0, 1)$, consider the effect of an aggregate shock on the ratio between the wealth share in the top $0.1p$ and in the top p:

$$\frac{\partial \log \left(S_{0.1p,t}/S_{p,t}\right)}{\partial Z_t} = \left(\sigma_{wEt} - \sigma_{wHt}\right) \left(F_{0.1p,t}(E) - F_{p,t}(E)\right). \tag{24}$$

As long as the fraction of entrepreneurs in the top $0.1p$ is higher than in the top p, aggregate shocks increase the share of wealth owned by the top $0.1p$ more than the top p. Note that, as

[27] An analytical expression for $F_{p,t}(j)$ is derived in Equation 44 in the proof of the proposition (Appendix A). Note that the left-hand side in (23) can also be seen as the instantaneous volatility of $\log S_{p,t}$; that is, $\mathrm{d}\log S_{p,t} - \mathrm{E}_t\left[\mathrm{d}\log S_{p,t}\right] = \left(\partial \log S_{p,t}/\partial Z_t\right)\mathrm{d}Z_t$.

[28] This is consistent with the reduced-form evidence discussed in Online Appendix B.3.

[29] More formally, the proof of Proposition 4 in Appendix A shows that

$$\lim_{p \to 1} \frac{\partial \log S_{p,t}}{\partial Z_t} = 0; \quad \lim_{p \to 0} \frac{\partial \log S_{p,t}}{\partial Z_t} = \sigma_{wEt}.$$

[30] For instance, Jones and Kim (2016) propose to use $\zeta_{p,t} \equiv 1 \Big/ \left(1 + \log_{10}\left(\frac{S_{0.1p,t}}{S_{p,t}}\right)\right)$ as an empirical proxy for the tail index of a distribution. When the distribution is exactly Pareto, this quantity equals the tail index. However, when the distribution only has a right Pareto tail, the equality only holds in the limit $p \to 0$.

$p \to 0$, the fraction of entrepreneurs converges to one, so the right-hand side of (24) converges to zero, consistently with the fact that the tail index of the wealth distribution (as a limiting concept) does not respond to aggregate shocks.

In conclusion, we have obtained simple formulas for the tail index of the wealth distribution (Proposition 3) and the exposure of top wealth shares (Proposition 4). In particular, we have shown that (i) the tail index of the wealth distribution depends on the average growth rate of entrepreneurs, while (ii) the limiting exposure of top wealth shares to aggregate shocks depends on the wealth exposure of entrepreneurs. These results will play a key role in calibrating the model, as they suggest that one can calibrate the consumption rate and the wealth exposure of entrepreneurs by targeting these two moments.

3.4 Extensions

For the sake of parsimony, the environment is highly stylized. I now briefly discuss three extensions of the baseline model that would make it more realistic: (i) distinguishing between labor and capital income, (ii) adding hand-to-mouth households, and (iii) making the heterogeneity in initial endowment arbitrary. I show that these extensions would not affect asset prices nor the two key moments of the wealth distribution discussed above: its tail index and the elasticity of top wealth shares to aggregate shocks.

Distinction between labor and capital income. For the sake of simplicity, agents in the model are endowed with only one kind of tree. In reality, agents earn both labor and capital income. The distinction between these two does not matter for the individual optimization problem, as markets are dynamically complete in our model (see, for instance, Gârleanu and Panageas, 2015). However, the distinction matters when mapping the model to the data: wealth, in the model, corresponds to the capitalized value of all future income promised to an individual (i.e., "total wealth"), while observed wealth, in the data, only corresponds to the capitalized value of future capital income (i.e., "financial wealth").[31] In Online Appendix C.3, however, I show that the tail index of the wealth distribution and the elasticity of top wealth shares to stock market returns are unchanged whether analyzing the distribution of "financial wealth" or "total wealth". This justifies my choice of abstracting from the distinction between these two concepts in the baseline model.

Hand-to-mouth households. I now turn to the presence of hand-to-mouth households. In the model, households can freely trade in financial markets. In reality, a lot of households face financial frictions. To account for this fact, the model could be extended to assume that a third type of agent consumes the income they are endowed with every period. The key point is that these agents would not matter for asset prices as they do not trade assets. Furthermore, they would not

[31] See, for instance, Catherine et al. (2020) and Greenwald et al. (2022b).

affect the elasticity of top wealth shares to stock returns or the tail index of the wealth distribution as they would not appear in top percentiles.

Heterogeneity among newborns. Finally, while I have assumed that the initial distribution of wealth among newborns is log-normally distributed, this parametric assumption (as well as the value of its standard deviation) does not affect asset prices as agents have homothetic preferences. Furthermore, as long as the initial distribution is thin-tailed, its shape does not affect the right tail of the wealth distribution. As a result, its does not affect the elasticity of top wealth shares to stock market returns nor the tail index of the wealth distribution.[32]

4 Quantitative analysis

I now turn to the quantitative implications of the model. Section 4.1 presents the calibration, Section 4.2 discusses the equilibrium, and Section 4.3 discusses the excess volatility of asset returns implied by the model.

4.1 Parameters

The model has thirteen parameters that I calibrate to match moments related to the U.S. economy.

Demography and endowment. I start with the five parameters related to demography (η, δ) and to the endowment process (g, σ, ϕ). The population growth rate η is chosen to match the annual population growth in the U.S. since 1913; that is, $\eta = 1.5\%$. The death rate δ is chosen to match the annual death rate of households in the top 0.5% estimated by Kopczuk and Saez (2004); that is, $\delta = 2.5\%$. This value is roughly consistent with the 2.2% annual death rate in the top 400 for the 1983-2017 period measured in Gomez (2023).

The drift g and volatility σ of the endowment process are chosen to match, respectively, the average and standard deviation of the growth of time-averaged annual consumption per capita; that is, $g = 2\%$ and $\sigma = 4\%$. The depreciation rate of trees is chosen to match the 2.5pp difference between the growth rate of dividends in the economy and the dividend growth of existing firms in the economy (respectively, $g + \eta$ and $g - \phi$ in the model), which gives $\phi = 1\%$.[33]

Wealth dynamics of entrepreneurs. I now turn to the four parameters related to the wealth dynamics of entrepreneurs ($\alpha_E, \nu, \rho_E, \pi_E$). The share of wealth invested in equity by entrepreneurs,

[32] The proof follows from writing the (normalized) wealth of an agent at time t as the product of their (normalized) wealth at birth and the cumulative growth of their (normalized) wealth since birth. Because these two variables are independent, the tail index of their products is the minimum of the tail indices of each variable (see, for instance, Gabaix, 2009).

[33] More precisely, Gârleanu et al. (2015) document a 2pp difference between the growth rate of dividends in the economy and the dividend growth of existing firms in the S&P 500. I adjust this number to account for the fact that acquisition accounts for 0.5pp of the growth of assets of firms in Compustat.

α_E, is chosen to match the regressions of the growth of top wealth shares on equity returns estimated in Section 2. To interpret these regressions, remember that agents only trade all-equity firms (or trees) in the model. In reality, firms issue a mix of debt and equity, and therefore, levered equity corresponds to a levered claim on the underlying firms. Following Modigliani-Miller logic, the instantaneous return on this levered equity (i.e., the "stock market return") is:

$$\frac{\mathrm{d}R_{Mt}}{R_{Mt}} = (r_t + \lambda(\mu_{Rt} - r_t))\,\mathrm{d}t + \lambda\sigma_{Rt}\,\mathrm{d}Z_t, \tag{25}$$

where λ denotes the market leverage of the corporate sector (i.e., the ratio between the market value of all liabilities and the market value of equity).[34] As a result, regressing the instantaneous growth of aggregate wealth on stock market returns in the model estimates $1/\lambda$, while regressing the growth of the average wealth of entrepreneurs on stock market returns estimates α_E/λ. Together with the estimates reported in Table 1, this implies $\lambda = 2.3$ and $\alpha_E = 2$.[35]

The idiosyncratic volatility of trees, ν is chosen to match the 20% annual cross-sectional dispersion of the wealth growth for agents at the top of the wealth distribution ($\sqrt{\mathrm{E}\left[\alpha_{Et}^2\nu^2\right]}$ in the model), as measured in Gomez (2023); that is, $\nu = 10\%$. This value is consistent with similar reduced-form evidence from Sweden Bach et al. (2020) in Sweden, as well as existing calibrations by Angeletos (2007) and Benhabib et al. (2011).

The entrepreneur consumption rate, ρ_E, is chosen to match the tail index of the wealth distribution. More precisely, I use the expression for the tail index ζ given in Proposition 3, together with the calibrated values for (δ, η, ν), to back out the average logarithmic growth of entrepreneurs relative to the economy $\mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma_{wEt}^2\right]$. Klass et al. (2006) and Vermeulen (2018) measure a power law exponent for the wealth distribution of $\zeta = 1.5$. Plugging this number into (22) implies an estimate for the average logarithmic growth of entrepreneurs relative to the economy of $\mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma_{wEt}^2\right] \approx 1.7\%$.

In a second step, I use this estimate to infer the consumption rate of entrepreneurs. More precisely, the average logarithmic growth of entrepreneurs relative to the economy can be written as the difference between the average logarithmic growth of entrepreneurs and the logarithmic growth rate of the economy:[36]

$$\mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma_{wEt}^2\right] = \underbrace{\mathrm{E}\left[r_t + \alpha_{Et}(\mu_{Rt} - r_t) - \frac{1}{2}\alpha_{Et}^2\sigma_{Rt}^2\right]}_{\text{Average logarithmic return of entrepreneurs}} - \rho_E - \underbrace{\left(g - \frac{1}{2}\sigma^2\right)}_{\text{Logarithmic growth rate of economy}}. \tag{26}$$

Given the values of (g, σ), the logarithmic growth rate of the economy is 1.9%. To estimate en-

[34] See Barro (2006) for a similar approach.

[35] More precisely, using the exposure for the top 0.01% gives $\alpha_E = 1.8$ while the exposure for the top 400 gives $\alpha_E = 2.3$. I use the average between the two to calibrate α_E.

[36] Note that I used the fact that the timed-average drift of log asset prices is zero; that is, $\mathrm{E}\left[\mu_{pt} - \frac{1}{2}\sigma_{pt}^2\right] = 0$.

trepreneurs' average logarithm return, I use moments on asset prices over the sample 1913−2020 (i.e., the time sample for which we have data on top wealth shares). As reported in Table 3, the average logarithmic (real) risk-free rate is $\mathrm{E}\left[r_t\right] = 0.3\%$, the average logarithmic stock market return is $\mathrm{E}\left[r_t + \lambda(\mu_{Rt} - r_t) - \frac{1}{2}\lambda^2\sigma_{Rt}^2\right] = 6.4\%$, and its standard deviation is $\sqrt{\mathrm{E}\left[\lambda^2\sigma_{Rt}\right]} = 19.3\%$. Combining these estimates gives an average logarithm return for entrepreneurs of 5.8%. Plugging these estimates into (26) implies a consumption rate of entrepreneurs $\rho_E = 2.2\%$.[37]

I then pick the population share of entrepreneurs to match the proportion of households that report having more than two-thirds of their wealth invested in public or private equity (Online Appendix Table A4); that is, $\pi_E = 9\%$. Note that this parameter is a bit difficult to pin down since, in reality, there exists a continuum between households and entrepreneurs. Fortunately, the sensitivity analysis reported in Online Appendix Table A8 shows that the model's implication for asset prices is not particularly sensitive to the value for π_E. In any case, note that our parameter value is roughly consistent with the 7.5% proportion of entrepreneurs reported in Cagetti and De Nardi (2006), based on the proportion of U.S. households who are self-employed and who own a business for which they have an active management role.

Household preferences. I calibrate the remaining three parameters related to households' preferences (their SDR ρ, their EIS ψ, and their RRA γ) to match four asset price moments jointly: the average and standard deviation of the risk-free rate and of stock market returns from 1913 to 2020, which are reported in Table 3. Formally, denote $\theta \equiv (\rho, \pi_E, \gamma)$ the vector of parameters and $m(\theta)$ the vector of moments implied by these parameters after simulating the model; that is, the average and standard deviation of the risk-free rate and of stock market returns. I pick the vector of parameters $\hat{\theta}$ which minimizes the distance $(\hat{m} - m(\theta))' (\hat{m} - m(\theta))$, where $\hat{m}$ denotes the four moments in the data. For the sake of realism, I only search for an RRA γ and an inverse EIS $1/\psi$ below 20, as well as a SDR ρ below 10%.

Table 2 reports the set of parameters that minimize $(\hat{m} - m(\theta))' (\hat{m} - m(\theta))$. I estimate a relatively high SDR ($\rho = 10\%$), a high RRA ($\gamma = 10.3$), and a low EIS ($\psi = 0.05$). Note that such a low EIS is consistent with evidence from the microdata for the average household (Vissing-Jørgensen, 2002; Best et al., 2020). It is also consistent with existing calibrations of asset pricing models with heterogeneous agents (Guvenen, 2009 ; Gârleanu and Panageas, 2015).

Distribution of initial wealth. Finally, I pick the standard deviation of the distribution of log wealth for newborns, ν_0, to match the level of the share of wealth owned by the top 1%, which averages 33% between 1913 and 2020. This gives me $\nu_0 = 1.6$. As highlighted in Section 3, this parameter solely affects the level of top wealth shares — it does not affect asset prices or other inequality moments such as the tail index of the wealth distribution or the limiting exposure of

[37] In Online Appendix D.1, I show that I obtain similar results if I directly estimate the consumption rate of top entrepreneurs using individual data from Forbes 400 instead.

top wealth shares — which is why I calibrate it after all other parameters have been set.[38]

Table 2: Parameters

Description	Symbol	Value	Target
Demography and endowment			
Population growth rate	η	1.5%	Growth rate number of U.S. households
Death hazard rate	δ	2.5%	Death rate at the top
Endowment growth rate	g	2%	Per capita growth rate of consumption
Endowment volatility	σ	4%	SD of time-averaged consumption
Tree depreciation rate	ϕ	1%	Growth rate public firms
STD initial endowment	ν_0	1.6	Average wealth share top 1%
Entrepreneurs' dynamics			
Entrepreneur equity share	α_E	2	Regression growth top 0.01% wealth on stock returns
Tree idiosyncratic volatility	ν	10%	Dispersion wealth growth at the top
Entrepreneur SDR	ρ_E	2.2%	Tail index of wealth distribution
Entrepreneur pop. share	π_E	9%	Pct. agents with more than half of wealth in equity
Households' preferences			
Household SDR	ρ	10%	Asset price moments
Household EIS	ψ	0.05	Asset price moments
Household RRA	γ	10.3	Asset price moments

Notes: This table summarizes the calibration discussed in Section 4.1. Each parameter is given at the annual frequency.

Table 3: Targeted moments

Moments	Data	Model
Average interest rate	0.3%	3.4%
Standard deviation interest rate	0.9%	0.7%
Average stock market return	6.4%	6.3%
Standard deviation stock market return	19.3%	18.5%

Notes: The table reports the moments in the data (measured over the 1913-2020 period) and in the calibrated model. The interest rate is the nominal interest rate deflated by the inflation rate. Each moment is given at the annual frequency.

4.2 Studying the equilibrium

I now examine how well the calibrated model matches asset prices and top wealth inequality. I also analyze the impulse response of important economic quantities in the calibrated model.

[38]As discussed above, for the sake of simplicity, the model does not differentiate between capital and labor income, so the notion of wealth in the model encompasses both financial wealth and human capital. While the distinction does not matter for two inequality moments I focus on, the tail index of the distribution and the wealth exposure of top percentiles (Section 3.4), it does matter for the average level of top shares as labor income is more equally distributed than capital income. Calibrating ν_0 to match the average level of the top 1% income share, which averages 16% on a pre-tax and pre-transfer basis (see Piketty et al., 2018), would imply a lower value for $\nu_0 = 0.8$. In any case, the value for ν_0 does not affect equilibrium asset prices or the impulse response of top wealth shares to aggregate shocks, which is the focus of the rest of the paper.

Matching asset prices. Table 3 reports the asset price moments implied by the calibrated model. The calibrated model matches very well the average of stock market returns (6.4% in the data versus 6.3% in the model) as well as their standard deviation (19.3% in the data versus 18.5% in the model). Note, however, that the calibrated model tends to overestimate the level of the interest rate (0.3% in the data versus 3.4% in the data), even though it matches well its low standard deviation (0.9% in the data versus 0.7% in the model). Note that measuring asset price moments starting from 1871, as in Gârleanu and Panageas (2015), would give an average interest rate of 2.8%, closer to the one implied by the model.

The fact that the calibrated model implies an interest rate higher than the data is due to some tension in the model between matching the high standard deviation of returns and matching a low interest rate. To understand this, observe the asset price volatility rises with the heterogeneity in consumption rates between households and entrepreneurs.[39] Given that the consumption rate of entrepreneurs is pinned down by the tail index of the wealth distribution, this implies that, in the model, the standard deviation of returns increases with the consumption rate of households. However, a high household consumption rate results in a high average consumption rate across the economy, necessitating a high interest rate to clear the goods market.[40] Consistently with this discussion, Online Appendix Table A8, which reports the sensitivity of asset price moments to the calibrated parameters, shows that household preferences that increase the standard deviation of returns also increase the average interest rate.

Second, the calibrated model captures well the effect of excess stock market returns on top wealth inequality. More precisely, Online Appendix Figure OA4 shows that local projections of the average wealth in top percentiles on excess stock returns in the model and in the data are very similar, for all top percentiles $p \in \{100\%, 1\%, 0.1\%, 0.01\%, \text{Top } 400\}$ and horizons $0 \leq h \leq 8$ (Online Appendix Figure A4 presents the same exercise for top wealth shares). This good fit partly reflects that λ (resp. α_E) was chosen to match the response of the average wealth in the economy (resp. in the top 0.01%) to excess stock returns. What is nontrivial is that the model matches very well (i) the gradual increase in the wealth exposure to stock market returns across the wealth distribution (in the model, this is driven by the gradual increase of the proportion of entrepreneurs in the right tail), as well as (ii) the slow rate of decay of these local projections with the horizon. I will discuss the impulse response of top wealth shares to aggregate shocks in more detail in Section 5.

Impulse response functions. I now examine the impulse response of asset prices and expected returns to aggregate shocks. For any quantity that depends smoothly on the state variable $g_t = g(x_t)$, I denote its Infinitesimal Impulse Response Function (IIRF) as the effect of an infinitesimal

[39] Indeed, differentiating the market clearing condition for the goods market (15) give $\partial_x \log p = p\,(c_{Ht} - \rho_E - (1-x)\partial_x c_{Ht})$.

[40] Again, this can be seen through the market clearing condition for the goods market (15).

aggregate shock on its expected value at horizon h starting from some initial state x:[41]

$$\text{IIRF}_g(x_t, h) \equiv \frac{\partial \text{E}\left[g(x_{t+h})|x_t = x\right]}{\partial Z_t} = \text{E}\left[\frac{\partial x_{t+h}}{\partial x_t}\partial_x g(x_{t+h})|x_t = x\right]\sigma_x(x), \tag{27}$$

where $\partial x_{t+h}/\partial x_t$ denotes the stochastic derivative of the process $(x_t)_{t\in\mathbb{R}}$ at time $t+h$ with respect to its value at time t. This process equals one at time t and then evolves with the law of motion:

$$\left(\mathrm{d}\frac{\partial x_{t+h}}{\partial x_t}\right) \Big/ \left(\frac{\partial x_{t+h}}{\partial x_t}\right) = \partial_x \mu_x(x_{t+h})\,\mathrm{d}h + \partial_x \sigma_x(x_{t+h})\,\mathrm{d}Z_{t+h}. \tag{28}$$

One key advantage of working in continuous-time is that this impulse response function can be computed analytically, even though it depends non-linearly with the horizon h and the state variable x.[42] Figure 2 plots $\text{E}\left[\text{IIRF}_g(x,h)\right]$, the average IIRF across the state space, as a function of the horizon h for several important quantities in the model: the price-to-income ratio p_t, the wealth-to-consumption ratio of households $1/c_{Ht}$, the risk-free rate r_t, and expected log stock market returns. These plots summarize the key mechanism at the heart of the model: in response to an aggregate shock, the share of wealth owned by entrepreneurs increases (as they own levered positions in risky assets), which increases asset prices and decreases expected returns in equilibrium (as they have a higher demand for assets). As a complement to these impulse response functions, I also plot in Online Appendix Figure A6 the same quantities as a function of the state variable x.

These impulse response functions reveal that aggregate shocks generate persistent effects on equilibrium prices. As shown in Online Appendix D.3, all infinitesimal impulse response functions decay at the same exponential rate, which is the "spectral gap" of the infinitesimal generator associated with the process $(x_t)_{t\in\mathbb{R}}$. This decay rate is 0.06 in the calibrated model, which means that it takes more than a decade for the effect of an aggregate shock on asset valuations to decay by half ($\log 2/0.06 \approx 12$ years). This decay rate results from the combination of two weak mean-reverting forces for x_t: first there is a mechanical force due to population renewal (death and population growth), second there is an economic force due to the equilibrium increase in asset valuations following an increase in x_t, which decreases the growth rate of entrepreneurs relative to households.[43]

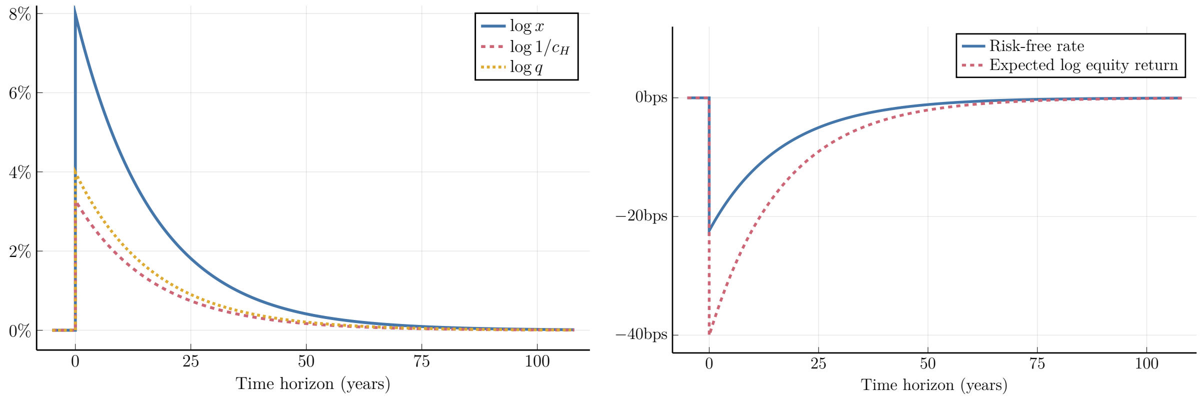
[41] See Borovička et al. (2014) and Alvarez and Lippi (2022) for related definitions.

[42] The extended Feynman-Kac formula stated in Lemma A3 (Online Appendix E.3) implies that IIRF_g can be obtained by solving the linear PDE

$$\partial_h \text{IIRF}_g(x,h) = \partial_x \mu_x(x)\text{IIRF}_g(x,h) + \Big(\mu_x(x) + \sigma_x(x)\partial_x\sigma_x(x)\Big)\partial_x \text{IIRF}_g(x,h) + \frac{1}{2}\sigma_x^2(x)\partial_{xx}\text{IIRF}_g(x,h),$$

with initial condition $\text{IIRF}_g(x,0) = \partial_x g(x)$.

[43] Since the decay rate due to demographic forces is $\delta + \eta = 0.04$, the remainder can be interpreted as the effect due to the equilibrium decline in equity returns.



(a) Wealth-to-consumption ratios

(b) Expected log returns

Figure 2: Impulse response functions of economic quantities

Notes: This figure plots the average infinitesimal impulse response function of different quantities; that is, $h \to \mathrm{E}[\mathrm{IIRF}_g(x, h)]$ for different functions of the state variable $g(\cdot)$. The expected log equity return corresponds to the expected log return of unlevered equity; that is, $r + \mu_R - \frac{1}{2}\sigma_R^2$. These graphs can be interpreted as the first-order response to a one standard deviation annual shock in aggregate income.

4.3 Model-implied dynamics of asset prices

Feedback loop. The calibrated model successfully replicates the high volatility of stock market returns despite the low volatility of aggregate income (Table 3). This arises from a feedback loop between asset prices and wealth inequality: an increase in asset prices amplifies wealth inequality, as entrepreneurs hold levered equity positions, while greater wealth inequality, in turn, drives up asset prices due to entrepreneur' higher demand for assets. We can formalize this feedback loop through two equations. The first equation relates the volatility of the state variable to the volatility of asset returns:[44]

$$\sigma_x = x(\alpha_E - 1)\sigma_R, \tag{29}$$

which says that as long as entrepreneurs own levered positions in equity ($\alpha_{Et} > 1$), the volatility of the state variable increases with the volatility of equity returns. On the other hand, the definition of returns (5) gives:

$$\sigma_R = \sigma + \sigma_p = \sigma + \partial_x \log p \times \sigma_x, \tag{30}$$

which says that, as long as entrepreneurs have a higher demand for assets (i.e., $\partial_x \log p > 0$), asset returns' volatility increases with the state variable's volatility. Combining these two equations

[44] This equation can be obtained by combining the volatility of x from Proposition 1 $\sigma_{xt} = x_t(1 - x_t)(\alpha_{Et} - \alpha_{Ht})\sigma_{Rt}$ with the market clearing condition for equity (16) $x_t\alpha_{Et} + (1 - x_t)\alpha_{Ht} = 1$. This equation reflects the fact that when an aggregate shock $\mathrm{d}Z_t$ hits the economy, the average wealth of entrepreneurs increases by $\alpha_E\sigma_R\,\mathrm{d}Z_t$ while the average wealth in the economy increases by $\sigma_R\,\mathrm{d}Z_t$. As a result, the share of wealth owned by entrepreneurs increases by $(\alpha_E - 1)\sigma_R\,\mathrm{d}Z_t$.

allows us to solve for the volatility of stock market returns, σ_R:[45]

$$\underbrace{\sigma_R}_{\substack{\text{Return}\\ \text{volatility}}} = \underbrace{\frac{1}{1-(\alpha_E-1)x\partial_x \log p}}_{\text{Multiplier} \geq 1} \times \underbrace{\sigma.}_{\substack{\text{Income}\\ \text{volatility}}} \tag{31}$$

The volatility of equity returns, σ_R, is the product between the volatility of aggregate income, σ, and a multiplier. Intuitively, this multiplier increases with the relative risk exposure of entrepreneurs $\alpha_E - 1$ and with the elasticity of asset valuations to the share of wealth owned by entrepreneurs $x\partial_x \log p$.[46] In the calibrated model, we have $\alpha_E = 2.0$, $\mathrm{E}[x] \approx 0.24$, and $\mathrm{E}[\partial_x \log p] \approx 2.09$, which gives a multiplier around 2.0. In other words, the calibrated model generates a volatility of equity returns σ_R that is twice as high as the volatility of aggregate income σ (and, therefore, a volatility of stock market returns $\lambda\sigma_R$ that is four times as high as the volatility of aggregate income). Hence, the model can generate the excess volatility of stock market returns.

An exact decomposition for the volatility of asset valuations. The endogenous response of asset valuations to aggregate shocks plays a key role in generating volatile asset returns (remember that (30) gives $\sigma_R = \sigma + \sigma_p$). In the spirit of Campbell and Shiller (1988), I now relate this endogenous response of asset valuations to changes in future risk-free rates and excess equity returns.

Proposition 5. *The volatility of asset valuations can be decomposed into two terms, corresponding to the present value of changes in future risk-free rates and in future (log) excess returns, respectively.*

$$\begin{aligned}\sigma_p(x) = & \underbrace{-\mathrm{E}\left[\int_0^\infty e^{-\int_0^t \frac{1}{p(x_s)}\,\mathrm{d}s} \frac{\partial x_t}{\partial x_0}\partial_x r(x_t)\,\mathrm{d}t \middle| x_0 = x\right]\sigma_x(x)}_{\text{Risk-free rate channel}} \\ & \underbrace{-\mathrm{E}\left[\int_0^\infty e^{-\int_0^t \frac{1}{p(x_s)}\,\mathrm{d}s} \frac{\partial x_t}{\partial x_0}\partial_x \left(\mu_R - \frac{1}{2}\sigma_R^2 - r\right)(x_t)\,\mathrm{d}t \middle| x_0 = x\right]\sigma_x(x).}_{\text{Excess return channel}}\end{aligned} \tag{32}$$

where $\partial x_t / \partial x_0$ denotes the stochastic derivative of the process $(x_t)_{t\in\mathbb{R}}$.

This equation says that the response of asset valuations to aggregate shocks, σ_p, can be written as the present value of changes in future expected returns. This term can be decomposed into the contribution of change in future risk-free rates ("risk-free rate channel") and changes in future excess returns ("excess return channel"). Relative to the log-linearization introduced by Campbell and Shiller (1988), this equation has three key advantages.[47]

[45] One way to understand this equation is that, following an aggregate income shock, the share of wealth owned by entrepreneurs increases via (29), which, in turn, increases valuations via (30), which then increases the share of wealth owned by entrepreneurs even more via (29)... Summing all of these rounds gives σ_R as the sum of a geometric series $\sigma_R = \sum_{k=0}^{\infty} \left(x(\partial_x \log p)(\alpha_E - 1)\right)^k \sigma$, which is equivalent to (31).

[46] If either term is null, this multiplier is simply equal to one.

[47] This decomposition, which is new to my knowledge, holds in any asset pricing model in which the growth rate

First, this decomposition is exact, which is helpful as Campbell and Shiller (1988)'s log linearization can have large errors in nonlinear models (e.g., Pohl et al., 2018). Second, each term can be computed numerically using a version of the Feynman-Kac formula.[48] Third, the decomposition can be computed at each point of the state space x, so it can be used to examine the relative effect of fluctuations in risk-free and expected equity returns in different parts of the state space.

In terms of magnitude, I find that the risk-free rate channel and the excess return channel account for, respectively, 53% and 47% of the volatility of asset valuations. The impulse responses for log expected returns plotted in Figure 2b shows that the risk free-rate and the expected excess return contribute equally to the response in expected log returns to aggregate shocks. As shown in Online Appendix Figure A7, this average value masks a large heterogeneity across the state space: in particular, news about interest rates become a relatively larger source of asset price fluctuations as x, the share of wealth owned by entrepreneurs, approaches zero.[49]

Finally, one can easily extend the result of Proposition 5 to obtain a similar decomposition for the response of asset valuations over any horizon $h \geq 0$.[50] Figure 3 plots the resulting decomposition of the average infinitesimal impulse response function, $h \to \mathrm{E}\left[\mathrm{IIRF}_{\log p}(x,h)\right]$, as a sum of the effect of changes in future risk-free rates and future expected excess returns.

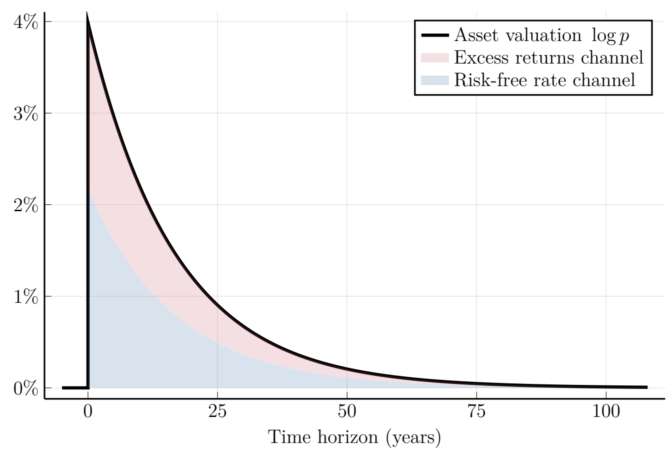


Figure 3: Decomposing the impulse response of asset valuations to aggregate shocks

Notes: This figure plots the average impulse response function for asset valuations, $h \to \mathrm{E}\left[\mathrm{IRRF}_{\log p}(x,h)\right]$, as well as its decomposition into a "risk-free rate channel" and an "excess return channel", similarly to (32). The graph can be interpreted as the first-order response to a one standard deviation annual shock in aggregate income.

---

of cashflows and expected returns are smooth functions of some Markovian process. This essentially includes all "textbook" asset pricing models (e.g., Campbell and Cochrane, 1999, Bansal and Yaron, 2004, Wachter, 2013, He and Krishnamurthy, 2013, Gârleanu and Panageas, 2015...). See the proof of Proposition 5 in Appendix A for details.

[48]See the proof of the proposition in Appendix A for details.

[49]This comes from the fact that the gradient of the interest rate with respect to the state variable x increases as it approaches zero (see Online Appendix Figure A6.

[50]Formally, the response of asset valuations to an aggregate shock at horizon h can be written as

$$\mathrm{IIRF}_{\log p}(x,h) = \mathrm{E}\left[\frac{\partial x_h}{\partial x_0}\partial_x \log p(x_t)|x_0 = x\right]\sigma_x(x) = -\mathrm{E}\left[\int_h^\infty e^{-\int_h^t \frac{1}{p(x_s)}\,\mathrm{d}s}\frac{\partial x_t}{\partial x_0}\partial_x\left(\mu_R - \frac{1}{2}\sigma_R^2\right)(x_t)\,\mathrm{d}t\Big|x_0 = x\right]\sigma_x(x).$$

Comparison with the literature. Having described the implications of my model for asset prices, I now briefly emphasize two differences between my model and the existing asset pricing literature. First, the key financial friction in my model is that a subset of agents ("entrepreneurs") must maintain a constant share of their wealth invested in their firms. This contrasts with the standard financial friction used in the asset pricing literature, where entrepreneurs are collectively required to hold a fixed fraction of the corporate sector in the aggregate (e.g., Basak and Cuoco, 1998, Guvenen, 2009, Brunnermeier and Sannikov, 2014, He and Krishnamurthy, 2013. . .). This alternative approach implies that the share of wealth invested in equity by entrepreneurs is counter-cyclical.[51] In Online Appendix D.4, I discuss the difference between the two approaches for asset prices and show that there is no evidence for such cyclicality in the micro-data, which justifies my approach.

The second difference with the existing literature is that my model is built and calibrated to match the wealth distribution. This allows me to discipline the degree of agent heterogeneity using observable moments, instead of treating it as a free parameter. To illustrate the importance of this step, I show in Online Appendix D.4 that changing moments related to the wealth distribution would have large implications on asset prices. Relatedly, I show that existing asset pricing models tend to imply a wealth distribution that moves too much or is too unequal relative to the data. This suggests that they tend to overestimate the degree of heterogeneity between agents.

5 Impulse response function of top wealth shares

I now use the calibrated model to analyze the full impulse response of top wealth shares to aggregate shocks. This analysis complements the empirical approach in Section 2, which focused on estimating the short-term responses of top wealth shares (less than ten years) due to the rapid decline in the precision of local projections over longer horizons.

5.1 Impulse response of surviving individuals

I first examine the effect of an aggregate shock on the average normalized wealth of "surviving" entrepreneurs. Here, and in the rest of the section, the term "surviving" entrepreneurs refers to the subset of entrepreneurs who remain alive following the realization of the aggregate shock up to the horizon of interest. Formally, the response of the average normalized wealth of surviving entrepreneurs at horizon h starting from an economy in state x is defined as:[52]

$$\epsilon(x_t, h) \equiv \frac{\partial}{\partial Z_t} \mathrm{E}_t \left[\log \mathbb{E}_{t+h} \left[w_{i,t+h} | i \in \mathbb{I}_{E,t} \cap \mathbb{I}_{E,t+h} \right] \right]. \tag{33}$$

[51] To see why, note that, after a sequence of negative shocks, entrepreneurs are relatively poorer, which means that they must invest a larger fraction of their wealth in equity to hold a given fraction of the corporate sector. Within the asset pricing literature, Di Tella (2017) uses a similar constraint as this paper; however, that paper allows entrepreneurs to undo their exposure to aggregate risk, which limits the comparison between the two papers.

[52] The fact that it depends on the state of the economy purely through the value of the state variable at that time is proved in Proposition 6 below.

The following proposition gives an analytical characterization of this impulse response function.

Proposition 6. *The effect of an aggregate shock on the average wealth of surviving entrepreneurs at horizon h and starting from an economy in state x is given by*

$$\epsilon(x,h) = \sigma_{wE}(x) + \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\partial_x\left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)(x_t)\,\mathrm{d}t\Big|x_0 = x\right]\sigma_x(x). \tag{34}$$

This proposition expresses the impulse response of surviving entrepreneurs as the sum of two terms. The first term corresponds to the "instantaneous" effect of the aggregate shock on the normalized wealth of entrepreneurs (i.e., at $h = 0$). The second term corresponds to the effect of the aggregate shock on the (logarithmic) growth rate of surviving entrepreneurs going forward (i.e., for $h > 0$), due to the endogenous rise in asset valuations. The second term can be computed numerically using a version of the Feynman-Kac formula.[53]

To fix ideas, consider what would happen if asset valuations did not react to aggregate shocks. In this case, an aggregate income shock $\sigma\,\mathrm{d}Z_t$ would permanently increase the normalized wealth of surviving entrepreneurs by $(\alpha_E - 1)\sigma\,\mathrm{d}Z_t$; that is, the impulse response of surviving entrepreneurs would simply be given by $\epsilon(x,h) = (\alpha_E - 1)\sigma$.

Figure 4 plots the average impulse response of surviving entrepreneurs $\mathrm{E}[\epsilon(x,h)]$ together with this baseline income response $(\alpha_E - 1)\sigma$. The difference between the two captures the effect of the endogenous response in asset valuations following an aggregate shock. Initially (at $h = 0$), this difference is positive: due to the rise in asset valuations after an aggregate shock, entrepreneurs' wealth reacts twice as much as their income on impact; that is, $\sigma_{wE}(x) = (\alpha_E - 1)\sigma_R > (\alpha_E - 1)\sigma$. In the longer term, however, the difference turns negative; that is, the normalized wealth of entrepreneurs "underreacts" to the initial income shock. This comes from the fact that the rise in asset valuations reduces the growth rate of entrepreneurs going forward since it makes them earn lower returns and pushes them to consume more (see Online Appendix D.5 for more details). Overall, this figure illustrates that the endogenous response of asset valuations amplifies the initial response of entrepreneurs' wealth while dampening their long-term response.

5.2 Impulse response of top wealth shares

I now turn to the impulse response of top wealth shares. The next proposition characterizes the effect of an aggregate shock on top wealth shares at any horizon $h \geq 0$. It can be seen as a generalization of Proposition 4, which characterized the *instantaneous* response of top wealth shares to aggregate shocks (i.e., the case $h = 0$).

Proposition 7. *The effect of an aggregate shock at time t on the share of wealth owned by a top percentile*

[53] See the proof of the proposition in Appendix A for details.

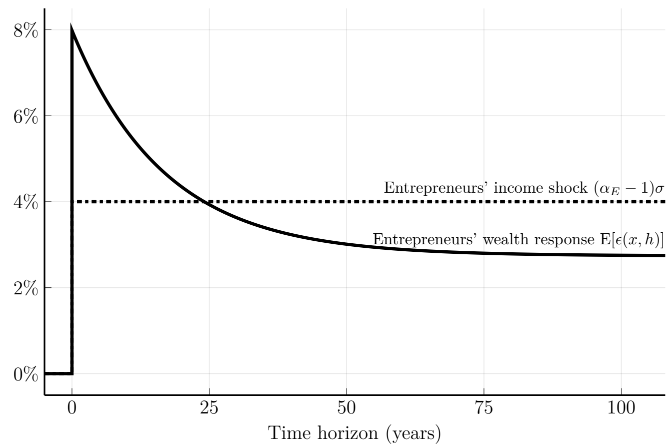


Figure 4: Average impulse response of the wealth of surviving entrepreneurs

Notes: This figure plots the average impulse response of the (log) average (normalized) wealth of surviving entrepreneurs (i.e., $h \to \mathrm{E}\left[\epsilon(x,h)\right]$), together with the entrepreneurs' income shock $(\alpha_E - 1)\sigma$, which can be interpreted as the impulse response if asset valuations were fixed over time. Hence, the difference between the two lines reflects the effect of endogenous changes in asset valuations. The graph can be interpreted as the first-order response to a one standard deviation annual shock in aggregate income.

p at time $t+h$ is given by

$$\frac{\partial \mathrm{E}_t\left[\log S_{p,t+h}\right]}{\partial Z_t} = \sum_{j\in\{E,H\}} \mathrm{E}_t\left[\sigma_{wj}(x_t)F_{p,t+h}(j,t)\right] + \sum_{j\in\{E,H\}} \mathrm{E}_t\left[\int_t^{t+h} \frac{\partial x_s}{\partial x_t}\partial_x\left(\mu_{wj} - \frac{1}{2}\sigma_{wj}^2\right)(x_s)F_{p,t+h}(j,s)\,\mathrm{d}s\right]\sigma_x(x_t),$$

where $F_{p,t+h}(j,s)$ denotes the fraction of wealth in the top percentile p at time $t+h$ owned by individuals of type $j \in \{E,H\}$ born before time s.[54]

To understand the intuition behind this expression, it is helpful to compare it to the impulse response of surviving entrepreneurs $\epsilon(x,h)$ discussed in Proposition 6. The key difference is that the effect of aggregate shocks on the wealth of surviving individuals is now mediated by weights $F_{p,t+h}(j,s)$. These weights account for the fact that the effect of an aggregate shock on the wealth dynamics at time s impacts the top wealth share at $t+h$ only through the fraction of individuals in the top percentile who were alive at that time. Another difference is the summation sign across agent types: top percentiles include both entrepreneurs and households, so one needs to consider the effect of aggregate shocks on both types of agents.

I now use this proposition to describe the impulse response of top wealth shares in two important limiting cases.[55] On the one hand, fixing the top percentile p, the response of the top wealth share converges to zero in the limit $h \to \infty$. Intuitively, as the horizon h grows, surviving individuals in the top percentile cede their places to younger individuals, born long after the shock. On

[54] An analytical expression for $F_{p,t+h}(j,s)$ is derived in Equation 46 in the proof of the proposition (Appendix A). The proof also details the two (accurate) approximations used to obtain the result.

[55] The proof of Proposition 7 in Appendix A contains a heuristic derivation for these two limits.

the other hand, fixing the horizon h, the response of the top wealth share converges to the impulse response of surviving entrepreneurs $\epsilon(x,h)$ in the limit $p \to 0$. This limit comes from the fact that, at any horizon, the fraction of individuals in the top percentile born before the shock converges to one in the right tail of the wealth distribution, so the impulse response of the top percentile converges to the impulse response of surviving entrepreneurs. Overall, while the first limit says that all top wealth shares ultimately mean-revert as the horizon grows, the second limit says that the speed of mean-reversion becomes arbitrarily low in the right tail of the wealth distribution. Intuitively, wealth in top percentiles is older and thus retains the effect of past aggregate shocks for longer.

To visualize these forces, Figure 5 plots the impulse response of the share of wealth owned by a top percentile $p \in \{1\%, 0.1\%, 0.01\%, 0.001\%\}$ to an aggregate shock up to a horizon of 100 years using simulated data. On impact (i.e., $h = 0$), higher top percentiles respond more to aggregate shocks, consistently with the data. In the model, this comes from the fact that the relative proportion of entrepreneurs increases in the distribution's right tail (see the discussion of Proposition 4). As the horizon grows, top wealth shares mean-revert towards zero. As seen above, this mean-reversion results from the combination of two distinct forces. First, the rise in asset valuations after an aggregate shock means that surviving entrepreneurs earn lower returns and consume relatively more going forward. Second, younger generations, who are not directly impacted by the shock, slowly replace older generations in top percentiles. The higher the top percentile, the longer it takes for existing agents to be replaced by newborns and, therefore, the longer it takes for the effect of the aggregate shock to dissipate fully. Quantitatively, it takes 30 years for the effect of the initial shock to be divided by three for the top 1%. In comparison, this reduction takes 60 years for the top 0.01%.

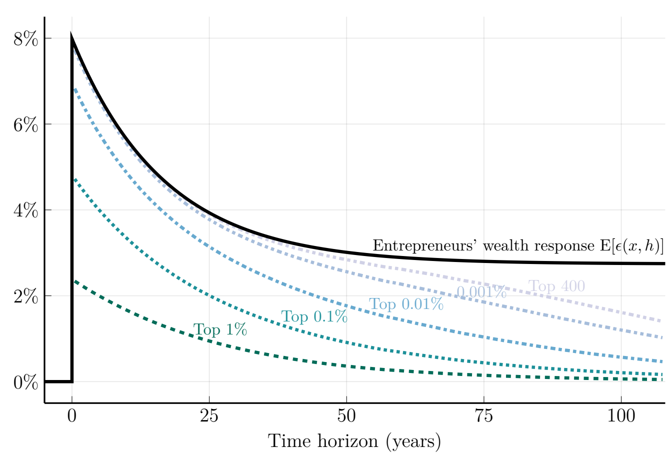


Figure 5: Impulse response of top wealth shares

Notes: This figure plots the infinitesimal impulse response for the average (normalized) wealth of surviving entrepreneurs; that is, $h \to E[\epsilon(x,h)]$. The figure also plots the impulse responses of the (log) share of wealth owned by each top percentile to aggregate income shocks, estimated using local projections on a very long sample of simulated data. The graph can be interpreted as the first-order response to a one standard deviation annual shock in aggregate income.

5.3 Standard deviation of top wealth shares

So far, I have focused on the dynamics of top wealth shares in response to aggregate income shocks. I now use this analysis to study the overall standard deviation of top wealth shares generated by the model, as aggregate income shocks cumulate over time. The two concepts are intimately linked as the standard deviation of top wealth shares roughly corresponds to the area below its impulse response function.[56]

The first line of Table 4 reports the standard deviation of (log) top wealth shares in the calibrated model using simulated data. The key observation is that this standard deviation increases monotonically in the right tail. Glancing at Figure 5 reveals that this increase in the standard deviation of top wealth shares at the top is due to the combination of two forces (i) aggregate shocks have larger effects on higher top percentiles on impact (due to the higher fraction of entrepreneurs in top percentiles) (ii) higher top percentiles take a longer time to mean-revert (due to the longer time it takes for newborns to reach top percentiles).

Table 4: Standard deviation of log top wealth shares

	Top 1%	Top 0.1%	Top 0.01%	Top 0.001%
Model (long sample)	0.09	0.18	0.28	0.35
Model (short sample)	0.07	0.07	0.07	0.07
Data	0.20	0.33	0.46	

For the sake of parsimony, I have focused on a model in which the *only* reason top wealth shares fluctuate over time is that wealthier agents have a higher wealth exposure to aggregate shock. In reality, there may be additional sources of fluctuations in top wealth shares. One interesting question is: How much can fluctuations in top wealth shares be explained by the model? To facilitate the comparison between the model and the data, I compare the standard deviation of top wealth shares in the data with the averaged standard deviation of top wealth shares in the calibrated model, obtained by averaging the estimated standard deviations across simulated samples with the same length as the data (105 years).[57] The results, reported in Table 4, reveal that the calibrated model can account for approximately 40% of the standard deviation of top wealth

[56] Formally, the Clark-Ocone formula allows us to write the logarithm of the share of wealth owned by a top percentile p as some time t as the infinite moving average of past aggregate shocks

$$\log S_{p,t} = \mathrm{E}[\log S_{p,t}] + \int_{-\infty}^{t} \mathrm{E}_s \left[\frac{\partial \log S_{p,t}}{\partial Z_s}\right] \mathrm{d}Z_s.$$

Ito's isometry implies

$$\mathrm{Var}[\log S_{p,t}] = \mathrm{E}\left[\int_{-\infty}^{t} \mathrm{E}_s \left[\frac{\partial \log S_{p,t}}{\partial Z_s}\right]^2 \mathrm{d}s\right] = \mathrm{E}\left[\int_{h=0}^{\infty} \mathrm{E}_0 \left[\frac{\partial \log S_{p,h}}{\partial Z_0}\right]^2 \mathrm{d}h.\right]$$

[57] This is because the naive estimate for the standard deviation of a persistent process suffers from a downward bias in finite sample. This is why I do not report the standard deviation of the Top 400 wealth share in the data, as the time sample is too low to be informative (35 years).

shares in the data (or, equivalently, a quarter of their variances). Hence, the heterogeneous exposure of agents to aggregate income shocks can explain a sizable fraction, but not all, of the actual fluctuations in top wealth shares observed in the data.

A complementary way to assess "how much" of the fluctuations in top wealth inequality can be explained by the model is to compare the *realized* dynamics of top wealth shares between the data and the model, after feeding the model with the realization for excess returns across the 20th century. This exercise, done in Online Appendix D.6, reveals that the model can explain the persistent decline in wealth inequality during the Great Depression, its rise immediately after WW2, and its rise during the dot-com bubble. However, the model has a hard time reproducing the overall U-shape of top wealth shares; that is, it cannot fully explain the secular decline of wealth inequality in the 1940s nor its secular rise in the 1980s. This result suggests that, to fully account for the dynamics of top wealth shares over the 20th century, one would need to augment the model with additional sources of fluctuations in the economy, such as changes in taxes or regulations (Hubmer et al., 2021, Cao and Luo, 2017), changes in idiosyncratic returns (Benhabib et al., 2019, Atkeson and Irie, 2022, Gomez, 2023), or changes in labor income inequality (Rosen, 1981, Gabaix and Landier, 2008, Terviö, 2008).

6 Conclusion

This paper studies the interplay between wealth inequality and asset prices. First, I document that the wealth of households at the top of the wealth distribution is twice as exposed to stock market returns as the wealth of the average household. Since stock market returns are volatile and uncorrelated, and top wealth shares mean-revert very slowly, this heterogeneous exposure generates long-lived fluctuations in top wealth inequality.

Motivated by this fact, I develop a model where different households have different exposures to aggregate shocks. While the wealth distribution implied by the model is stochastic, I show that it still exhibits a right Pareto tail, and that its tail index can be characterized analytically. Hence, this paper extends the characterization of tail indices obtained in deterministic random growth models (starting with Champernowne, 1953) to more realistic, stochastic economies.

Third, I calibrate the model to U.S. moments related to asset prices and wealth inequality. The model features a feedback loop between wealth inequality and asset prices following an aggregate shock. The model generates particularly large fluctuations in the right tail of the wealth distribution, as higher percentiles are more exposed to aggregate risk and take longer to mean revert.

For simplicity, I only consider shocks to aggregate income in the model. However, the interplay I describe between asset prices and wealth inequality would also appear with shocks that redistribute aggregate income between labor and capital (Greenwald et al., 2022a, Moll et al., 2022), shocks between young and old households (Gârleanu et al., 2012), or monetary policy shocks (Silva, 2016, Kekre and Lenel, 2022). Moreover, this interplay could also affect real quantities through changes in corporate investment policies or labor supply. Exploring these effects requires moving away from an endowment economy, which I leave for future research.

Data Availability Statement

The data and code underlying this research is available on Zenodo at https://doi.org/10.5281/zenodo.14248965.

References

Alloza, Mario, Jesus Gonzalo, and Carlos Sanz, "Dynamic effects of persistent shocks," 2020.

Alvarez, Fernando and Francesco Lippi, "The Analytic Theory of a Monetary Shock," *Econometrica*, 2022, *90* (4), 1655–1680.

Angeletos, George-Marios, "Uninsured idiosyncratic investment risk and aggregate saving," *Review of Economic dynamics*, 2007, *10* (1), 1–30.

Atkeson, Andrew G and Magnus Irie, "Rapid Dynamics of Top Wealth Shares and Self-Made Fortunes: What Is the Role of Family Firms?," *American Economic Review: Insights*, 2022, *4* (4), 409–24.

Auclert, Adrien and Matthew Rognlie, "Aggregate demand and the top 1 percent," *American Economic Review*, 2017, *107* (5), 588–92.

Bach, Laurent, Laurent E Calvet, and Paolo Sodini, "Rich pickings? Risk, return, and skill in household wealth," *American Economic Review*, 2020, *110* (9), 2703–2747.

Bansal, Ravi and Amir Yaron, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 2004, *59* (4), 1481–1509.

Barczyk, Daniel and Matthias Kredler, "Inequality and asset prices," *Working Paper*, 2016.

Barro, Robert J, "Rare Disasters and Asset Markets in the Twentieth Century," *The Quarterly Journal of Economics*, 2006, pp. 823–866.

Basak, Suleyman and Domenico Cuoco, "An equilibrium model with restricted stock market participation," *The Review of Financial Studies*, 1998, *11* (2), 309–341.

Beare, Brendan K and Alexis Akira Toda, "Determination of Pareto exponents in economic models driven by Markov multiplicative processes," *Econometrica*, 2022, *90* (4), 1811–1833.

Benhabib, Jess, Alberto Bisin, and Mi Luo, "Wealth distribution and social mobility in the US: A quantitative approach," *American Economic Review*, 2019, *109* (5), 1623–47.

__ , __ , and Shenghao Zhu, "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents," *Econometrica*, 2011, *79* (1), 123–157.

Best, Michael Carlos, James S Cloyne, Ethan Ilzetzki, and Henrik J Kleven, "Estimating the elasticity of intertemporal substitution using mortgage notches," *The Review of Economic Studies*, 2020, *87* (2), 656–690.

Blanchard, Olivier J, "Debt, Deficits, and Finite Horizons," *Journal of Political Economy*, 1985, *93* (2), 223–247.

Blume, Lawrence, David Easley et al., "Evolution and market behavior," *Journal of Economic theory*, 1992, *58*, 9–9.

Borovička, Jaroslav, "Survival and long-run dynamics with heterogeneous beliefs under recursive preferences," *Journal of Political Economy*, 2020, *128* (1), 206–251.

__, Lars Peter Hansen, and José A Scheinkman, "Shock elasticities and impulse responses," *Mathematics and Financial Economics*, 2014, *8* (4), 333–354.

Brunnermeier, Markus K and Yuliy Sannikov, "A Macroeconomic Model with a Financial Sector," *The American Economic Review*, 2014, *104* (2), 379–421.

Cagetti, Marco and Mariacristina De Nardi, "Entrepreneurship, frictions, and wealth," *Journal of political Economy*, 2006, *114* (5), 835–870.

Campbell, John Y, "Household finance," *The Journal of Finance*, 2006, *61* (4), 1553–1604.

__ and John H Cochrane, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 1999, *107* (2), 205–251.

__ and Robert J Shiller, "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, 1988, *1* (3), 195–228.

Cao, Dan and Wenlan Luo, "Persistent heterogeneous returns and top end wealth inequality," *Review of Economic Dynamics*, 2017, *26*, 301–326.

Carroll, Christopher D, "Portfolios of the Rich," in "Household Portfolios: Theory and Evidence," MIT Press Cambridge, 2000.

Catherine, Sylvain, Max Miller, and Natasha Sarin, "Social Security and Trends in Wealth Inequality," *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*, 2020.

Champernowne, David G, "A model of income distribution," *The Economic Journal*, 1953, *63* (250), 318–351.

Chan, Yeung Lewis and Leonid Kogan, "Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices," *Journal of Political Economy*, 2002, *110* (6), 1255–1285.

Cochrane, John H, "Macro-Finance," *Review of Finance*, 03 2017, *21* (3), 945–985.

Duffie, Darrell and Larry G Epstein, "Stochastic Differential Utility," *Econometrica*, 1992, *60* (2), 353–394.

Dumas, Bernard, "Two-Person Dynamic Equilibrium in the Capital Market," *Review of Financial Studies*, 1989, *2* (2), 157–188.

Eisfeldt, Andrea L., Hanno Lustig, and Lei Zhang, "Complex Asset Markets," *The Journal of Finance*, 2023, *78* (5), 2519–2562.

Epstein, Larry G and Stanley E Zin, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 1989, pp. 937–969.

Favilukis, Jack, "Inequality, stock market participation, and the equity premium," *Journal of Financial Economics*, 2013, *107* (3), 740–759.

Gabaix, Xavier, "Power Laws in Economics and Finance," *Annual Review of Economics*, 2009, *1* (1), 255–294.

_ and Augustin Landier, "Why has CEO Pay Increased So Much?*," *The Quarterly Journal of Economics*, 02 2008, *123* (1), 49–100.

_ , Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "The dynamics of inequality," *Econometrica*, 2016, *84* (6), 2071–2111.

Gârleanu, Nicolae and Stavros Panageas, "Young, Old, Conservative, and Bold: The Implications of Heterogeneity and Finite Lives for Asset Pricing," *Journal of Political Economy*, 2015, *123* (3), 670–685.

_ , Leonid Kogan, and Stavros Panageas, "Displacement Risk and Asset Returns," *Journal of Financial Economics*, 2012, *105* (3), 491–510.

_ , Stavros Panageas, Dimitris Papanikolaou, and Jianfeng Yu, "Drifting apart: The pricing of assets when the benefits of growth are not shared equally," 2015. Working Paper.

Gollier, Christian, "Wealth inequality and asset pricing," *The Review of Economic Studies*, 2001, *68* (1), 181–203.

Gomes, Francisco and Alexander Michaelides, "Asset Pricing with Limited Risk Sharing and Heterogeneous Agents," *Review of Financial Studies*, 2008, *21* (1), 415–448.

Gomez, Matthieu, "Decomposing the growth of top wealth shares," *Econometrica*, 2023, *91* (3), 979–1024.

_ and Émilien Gouin-Bonenfant, "Wealth inequality in a low rate environment," *Econometrica*, 2024, *92* (1), 201–246.

Goyal, Amit, Ivo Welch, and Athanasse Zafirov, "A comprehensive 2022 look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 2024, *37* (11), 3490–3557.

Greenwald, Daniel L, Martin Lettau, and Sydney C Ludvigson, "How the Wealth was Won: Factor Shares as Market Fundamentals," 2022. Working Paper.

_ , Matteo Leombroni, Hanno Lustig, and Stijn Van Nieuwerburgh, "Financial and total wealth inequality with declining interest rates," 2022. Working Paper.

Guiso, Luigi, Tullio Jappelli, and Daniele Terlizzese, "Income risk, Borrowing Constraints, and Portfolio Choice," *The American Economic Review*, 1996, pp. 158–172.

Guvenen, Fatih, "A Parsimonious Macroeconomic Model for Asset Pricing," *Econometrica*, 2009, *77* (6), 1711–1750.

He, Zhigu and Arvind Krishnamurthy, "A model of capital and crises," *The Review of Economic Studies*, 2012, *79* (2), 735–777.

He, Zhiguo and Arvind Krishnamurthy, "Intermediary asset pricing," *American Economic Review*, 2013, *103* (2), 732–70.

Herbst, Edward and Benjamin K Johannsen, "Bias in local projections," 2021. FEDS Working Paper No. 2020-010R1.

Hubmer, Joachim, Per Krusell, and Anthony A Smith Jr, "Sources of US wealth inequality: Past, present, and future," *NBER Macroeconomics Annual*, 2021, *35* (1), 391–455.

Jones, Charles I and Jihee Kim, "A Schumpeterian Model of Top Income Inequality," 2016. Working Paper.

Jordà, Òscar, "Estimation and inference of impulse responses by local projections," *American economic review*, 2005, *95* (1), 161–182.

Kacperczyk, Marcin, Jaromir Nosal, and Luminita Stevens, "Investor sophistication and capital income inequality," *Journal of Monetary Economics*, 2018.

Karlin, Samuel and Howard E Taylor, *A second course in stochastic processes*, Elsevier, 1981.

Kaymak, Barış and Markus Poschke, "The evolution of wealth inequality over half a century: The role of taxes, transfers and technology," *Journal of Monetary Economics*, 2016, *77*, 1–25.

Kaymak, Barıs, David Leung, Markus Poschke et al., "Accounting for the determinants of wealth concentration in the US," in "Meeting Papers" 2018.

Kekre, Rohan and Moritz Lenel, "Monetary policy, redistribution, and risk premia," *Econometrica*, 2022, *90* (5), 2249–2282.

Kelly, John L, "A new interpretation of information rate," *IRE Transactions on Information Theory*, 1956, pp. 185–189.

Klass, Oren S, Ofer Biham, Moshe Levy, Ofer Malcai, and Sorin Solomon, "The Forbes 400 and the Pareto Wealth Distribution," *Economics Letters*, 2006, *90* (2), 290–295.

Kopczuk, Wojciech and Emmanuel Saez, "Top wealth shares in the United States, 1916–2000: Evidence from estate tax returns," *National Tax Journal*, 2004, *57* (2), 445–487.

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins, "Income and wealth inequality in America, 1949–2016," *Journal of Political Economy*, 2020, *128* (9), 3469–3519.

Luttmer, Erzo GJ, "Slow convergence in economies with firm heterogeneity," *Federal Reserve Bank of Minneapolis working paper*, 2012, (696).

Malloy, Christopher J, Tobias J Moskowitz, and Annette Vissing-Jørgensen, "Long-run stockholder consumption risk and asset returns," *The Journal of Finance*, 2009, *64* (6), 2427–2479.

Mankiw, N Gregory and Stephen P Zeldes, "The Consumption of Stockholders and Nonstockholders," *Journal of Financial Economics*, 1991, *29* (1), 97–112.

Mian, Atif R, Ludwig Straub, and Amir Sufi, "The saving glut of the rich," 2020. Working paper.

Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo, "Uneven growth: automation's impact on income and wealth inequality," *Econometrica*, 2022, *90* (6), 2645–2683.

Moskowitz, Tobias J and Annette Vissing-Jørgensen, "The returns to entrepreneurial investment: A private equity premium puzzle?," *American Economic Review*, 2002, *92* (4), 745–778.

Nakagawa, Kenji, "Application of Tauberian theorem to the exponential decay of the tail probability of a random variable," *IEEE Transactions on Information Theory*, 2007, *53* (9), 3239–3249.

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

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman, "Distributional national accounts: methods and estimates for the United States," *The Quarterly Journal of Economics*, 2018, *133* (2), 553–609.

Pohl, Walter, Karl Schmedders, and Ole Wilms, "Higher order effects in asset pricing models with long-run risks," *The Journal of Finance*, 2018, *73* (3), 1061–1111.

Reed, William J, "The Pareto, Zipf and other power laws," *Economics letters*, 2001, *74* (1), 15–19.

Rosen, Sherwin, "The economics of superstars," *The American Economic Review*, 1981, *71* (5), 845–858.

Roussanov, Nikolai, "Diversification and its Discontents: Idiosyncratic and Entrepreneurial Risk in the Quest for Social Status," *Journal of Finance*, 2010, *65* (5), 1755–1788.

Saez, Emmanuel and Gabriel Zucman, "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," *The Quarterly Journal of Economics*, 2016, *131* (2), 519–578. Updated series (February 2022 vintage) can be downloaded from `https://gabriel-zucman.eu/usdina/`.

Silva, Dejanir H, "The risk channel of unconventional monetary policy," 2016. Working Paper.

Smith, Matthew, Owen Zidar, and Eric Zwick, "Top Wealth in America: New Estimates under Heterogeneous Returns," *The Quarterly Journal of Economics*, 2023, *138* (1), 515–573.

Straub, Ludwig, "Consumption, savings, and the distribution of permanent income," 2019. Working Paper.

Tella, Sebastian Di, "Uncertainty shocks and balance sheet recessions," *Journal of Political Economy*, 2017, *125* (6), 2038–2081.

Terviö, Marko, "The difference that CEOs make: An assignment model approach," *American Economic Review*, 2008, *98* (3), 642–668.

Toda, Alexis Akira, "Incomplete market dynamics and cross-sectional distributions," *Journal of Economic Theory*, 2014, *154*, 310–348.

— **and Kieran James Walsh**, "The equity premium and the one percent," *The Review of Financial Studies*, 2020, *33* (8), 3583–3623.

Vermeulen, Philip, "How fat is the top tail of the wealth distribution?," *Review of Income and Wealth*, 2018, *64* (2), 357–387.

Vissing-Jørgensen, Annette, "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 2002, *110* (4), 825–853.

Wachter, Jessica A, "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *The Journal of Finance*, 2013, *68* (3), 987–1035.

— **and Motohiro Yogo**, "Why do Household Portfolio Shares Rise in Wealth?," *Review of Financial Studies*, 2010, *23* (11), 3929–3965.

Wolff, Edward N, *Top heavy: The increasing inequality of wealth in America and what can be done about it*, New Press, 2002.

Appendix

A Proofs

Proof of Proposition 1. By definition, we have $x_t = \left(\int_{i\in\mathbb{I}_{E,t}} W_{it}\,\mathrm{d}i\right) / \left(\int_{i\in\mathbb{I}_{E,t}} W_{it}\,\mathrm{d}i + \int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i\right)$. Applying Ito's lemma gives:

$$\mathrm{d}x_t = x_t(1-x_t)\left(\frac{\mathrm{d}\left(\int_{i\in\mathbb{I}_{E,t}} W_{it}\,\mathrm{d}i\right)}{\int_{i\in\mathbb{I}_{E,t}} W_{it}\,\mathrm{d}i} - \frac{\mathrm{d}\left(\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i\right)}{\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i} - \left(\frac{\mathrm{d}\left[\int_{i\in\mathbb{I}_{E,t}} W_{it}\,\mathrm{d}i\right]}{\int_{i\in\mathbb{I}_{E,t}} W_{it}\,\mathrm{d}i} - \frac{\mathrm{d}\left[\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i\right]}{\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i}\right)\frac{\mathrm{d}\left[\int_{i\in\mathbb{I}_{it}} W_{it}\,\mathrm{d}i\right]}{\int_{i\in\mathbb{I}_{it}} W_{it}\,\mathrm{d}i}\right). \tag{35}$$

Now, the instantaneous change in the aggregate wealth within each group $j \in \{E, H\}$ is given by:

$$\mathrm{d}\left(\int_{i\in\mathbb{I}_{jt}} \mathrm{d}W_{it}\,\mathrm{d}i\right) = \underbrace{\int_{i\in\mathbb{I}_{jt}} \mathrm{d}W_{it}\,\mathrm{d}i}_{\text{contribution of surviving agents}} \underbrace{+\,|\mathbb{I}_{jt}|(\eta+\phi+\delta)p_tY_t\,\mathrm{d}t}_{\text{contribution of newborns}} \underbrace{-\,\delta\left(\int_{i\in\mathbb{I}_{jt}} W_{it}\,\mathrm{d}i\right)\mathrm{d}t.}_{\text{contribution of deceased}}.$$

Dividing by $\int_{i\in\mathbb{I}_{jt}} \mathrm{d}W_{it}\,\mathrm{d}i$ and rearranging gives

$$\frac{\mathrm{d}\left(\int_{i\in\mathbb{I}_{jt}} \mathrm{d}W_{it}\,\mathrm{d}i\right)}{\int_{i\in\mathbb{I}_{jt}} W_{it}\,\mathrm{d}i} = \frac{\int_{i\in\mathbb{I}_{jt}} \mathrm{d}W_{it}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{jt}} W_{it}\,\mathrm{d}i} + (\eta+\delta+\phi)\frac{p_tY_t}{\left(\int_{i\in\mathbb{I}_{jt}} W_{it}\,\mathrm{d}i\right)/|\mathbb{I}_{jt}|}\,\mathrm{d}t - \delta\,\mathrm{d}t.$$

Plugging this into the law of motion of x_t (35) gives

$$\mathrm{d}x_t = x_t(1-x_t)\left(\frac{\int_{i\in\mathbb{I}_{E,t}} \mathrm{d}W_{it}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{E,t}} W_{it}\,\mathrm{d}i} - \frac{\int_{i\in\mathbb{I}_{Ht}} \mathrm{d}W_{it}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i} + (\eta+\delta+\phi)\left(\frac{\pi_E}{x_t} - \frac{\pi_H}{1-x_t}\right)\mathrm{d}t - \left(\frac{\int_{i\in\mathbb{I}_{E,t}} \mathrm{d}W_{it}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{E,t}} W_{it}\,\mathrm{d}i} - \frac{\int_{i\in\mathbb{I}_{Ht}} \mathrm{d}W_{it}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i}\right)\frac{\mathrm{d}(p_tY_t)}{p_tY_t}\,\mathrm{d}t\right).$$

Combining this expression with the law of motion of W_{it} for households (6) and entrepreneurs (8) gives the result. □

Proof of Proposition 2. Equation (17) implies the following law of motion for the logarithm of normalized wealth for individual i in group $j \in \{E, H\}$:

$$\mathrm{d}\log w_{it} = \left(\mu_{wjt} - \frac{1}{2}\sigma_{wjt}^2 - \frac{1}{2}\nu_{wjt}^2\right)\mathrm{d}t + \sigma_{wjt}\,\mathrm{d}Z_t + \nu_{wjt}\,\mathrm{d}B_{it}. \tag{36}$$

Integrating over time, this implies that, for an individual born at time $s \le t$,

$$\log w_{it} = \log w_{is} + \int_s^t \left(\mu_{wju} - \frac{1}{2}\sigma_{wju}^2 - \frac{1}{2}\nu_{wju}^2\right)\mathrm{d}u + \int_s^t \sigma_{wju}\,\mathrm{d}Z_u + \int_s^t \nu_{wju}\,\mathrm{d}B_{iu}. \tag{37}$$

We assumed in Section 3.1 that the logarithm of normalized wealth at birth was normally distributed with mean $\log\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right) - \frac{1}{2}\nu_0^2$ and variance ν_0^2 (Section 3.1). Together with (37), this implies that the cross-sectional distribution of log normalized wealth is normal with mean $\mu_{j,s\to t} = \log\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right) - \frac{1}{2}\nu_0^2 + \int_s^t \left(\mu_{wju} - \frac{1}{2}\sigma_{wju}^2 - \frac{1}{2}\nu_{wju}^2\right) \mathrm{d}u + \int_s^t \sigma_{wju}\, \mathrm{d}Z_u$ and variance $\nu_{j,s\to t}^2 = \nu_0^2 + \int_s^t \nu_{wju}^2\, \mathrm{d}u$. Finally, denoting a_{it} the age of an individual $i \in \mathbb{I}_{it}$ at time t, we get:

$$\begin{aligned}
\mathbb{P}\left(w_{it} \le w | i \in \mathbb{I}_{jt}\right) &= \int_0^\infty \mathbb{P}\left(a_{it} = a | i \in \mathbb{I}_{jt}\right) \mathbb{P}\left(w_{it} \le w | a_{it} = a, i \in \mathbb{I}_{jt}\right) \mathrm{d}a \\
&= \int_{-\infty}^t \mathbb{P}\left(a_{it} = t-s | i \in \mathbb{I}_{jt}\right) \mathbb{P}\left(w_{it} \le w | a_{it} = t-s, i \in \mathbb{I}_{jt}\right) \mathrm{d}s \qquad (s \equiv t-a) \\
&= \int_{-\infty}^t (\eta+\delta) e^{-(\eta+\delta)(t-s)} \Phi\left(\frac{\log w - \mu_{j,s\to t}}{\nu_{j,s\to t}}\right) \mathrm{d}s.
\end{aligned}$$

□

Proof of Proposition 3. Similarly to many results in large deviations theory, the proof is structured in three steps. In the first step, I use the existence of cross-sectional moments of individual wealth of order lower than ζ_j to prove that the limit *superior* of $\log \mathbb{P}_t(w_{it} \ge w | i \in \mathbb{I}_{jt}) / \log w$ is lower than $-\zeta_j$ for $j \in \{E, H\}$. In the second step, we use the law of large numbers to prove that the limit *inferior* of $\log \mathbb{P}_t(w_{it} \ge w | i \in \mathbb{I}_{jt}) / \log w$ is higher than $-\zeta_j$ for $j \in \{E, H\}$. In the third step, we combine the two preceding to show that the limit of $\log \mathbb{P}_t(w_{it} \ge w | i \in \mathbb{I}_{jt}) / \log w$ is exactly $-\zeta_j$ for $j \in \{E, H\}$, which implies that the limit of $\log \mathbb{P}_t(w_{it} \ge w | i \in \mathbb{I}_t) / \log w$ is exactly $-\min(\zeta_H, \zeta_E)$. As the distribution of wealth is itself random, all of these statements should be understood as holding at any point in time $t \in \mathbb{R}$ almost surely (i.e., with probability one).
Step 1. In this step, we prove

$$\limsup_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \ge w | i \in \mathbb{I}_{jt})}{\log w} \le -\zeta_j \text{ for } j \in \{E, H\}. \tag{38}$$

To show this, we first prove that $m_{jt}(\xi) \equiv \mathbb{E}_t\left[w_{it}^\xi | i \in \mathbb{I}_{jt}\right]$, the ξ-th cross sectional moment of wealth within group j at time t, is finite for $0 \le \xi < \zeta_j$. Applying Ito's lemma to the definition of $m_{jt}(\xi)$ gives:

$$\mathrm{d}m_{jt}(\xi) = \mathrm{d}\left(\frac{1}{|\mathbb{I}_{jt}|}\int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\, \mathrm{d}i\right) = m_{jt}(\xi)\left(\frac{\mathrm{d}\left(\int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\, \mathrm{d}i\right)}{\int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\, \mathrm{d}i} - \frac{\mathrm{d}\,|\mathbb{I}_{jt}|}{|\mathbb{I}_{jt}|}\right). \tag{39}$$

Now, the instantaneous change in $\int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\, \mathrm{d}i$ is given by

$$\mathrm{d}\left(\int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\, \mathrm{d}i\right) = \underbrace{\int_{i\in\mathbb{I}_{jt}} \mathrm{d}w_{it}^\xi\, \mathrm{d}i}_{\text{contribution of surviving agents}} \underbrace{+ (\eta+\delta)|\mathbb{I}_{jt}| e^{\xi \log\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right) + \frac{1}{2}\xi(\xi-1)\nu_0^2}\, \mathrm{d}t}_{\text{contribution of newborns}} - \underbrace{\delta\left(\int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\, \mathrm{d}i\right)\mathrm{d}t.}_{\text{contribution of deceased}}$$

Dividing by $\int_{i\in\mathbb{I}_{jt}} w_{it}^{\xi}\,\mathrm{d}i$ and rearranging gives

$$\frac{\mathrm{d}\left(\int_{i\in\mathbb{I}_{jt}} w_{it}^{\xi}\,\mathrm{d}i\right)}{\int_{i\in\mathbb{I}_{jt}} w_{it}^{\xi}\,\mathrm{d}i} = \frac{\int_{i\in\mathbb{I}_{jt}} \mathrm{d}w_{it}^{\xi}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{jt}} w_{it}^{\xi}\,\mathrm{d}i} + (\eta+\delta)\frac{e^{\xi\log\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right)+\frac{1}{2}\xi(\xi-1)\nu_0^2}}{m_{jt}(\xi)}\,\mathrm{d}t - \delta\,\mathrm{d}t.$$

Plugging this into (39) gives

$$\mathrm{d}m_{jt}(\xi) = \frac{\int_{i\in\mathbb{I}_{jt}} \mathrm{d}w_{it}^{\xi}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{jt}} w_{it}^{\xi}\,\mathrm{d}i} m_{jt}(\xi) + (\eta+\delta)\left(e^{\xi\log\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right)+\frac{1}{2}\xi(\xi-1)\nu_0^2} - m_{jt}(\xi)\right)\mathrm{d}t.$$

Now, applying Ito's lemma to (17) gives the law of motion of w_{it}^{ξ} for $i \in \mathbb{I}_{jt}$:

$$\frac{\mathrm{d}w_{it}^{\xi}}{w_{it}^{\xi}} = \left(\xi\mu_{wjt} + \frac{1}{2}\xi(\xi-1)\left(\sigma_{wjt}^2 + \nu_{wjt}^2\right)\right)\mathrm{d}t + \xi\sigma_{wjt}\,\mathrm{d}Z_t.$$

Combining the two previous equations gives

$$\begin{aligned}\mathrm{d}m_{jt}(\xi) = &\left(\xi\mu_{wjt} + \frac{1}{2}\xi(\xi-1)\left(\sigma_{wjt}^2+\nu_{wjt}^2\right) - (\eta+\delta)\right) m_{jt}(\xi)\,\mathrm{d}t + \xi\sigma_{wjt} m_{jt}(\xi)\,\mathrm{d}Z_t \\ &+ (\eta+\delta)\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right)^{\xi} e^{\frac{1}{2}\xi(\xi-1)\nu_0^2}\,\mathrm{d}t.\end{aligned} \tag{40}$$

Given our assumptions on the smoothness of the policy functions, we can apply Lemma A1 (stated and proved in Online Appendix E.1) which implies that the stochastic process $m_{jt}(\xi)$ remains finite a.s. as long as

$$\mathrm{E}\left[\xi\left(\mu_{wjt} - \frac{1}{2}\sigma_{wjt}^2\right) + \frac{1}{2}\xi(\xi-1)\nu_{wjt}^2 - (\eta+\delta)\right] < 0.$$

Combined with the definition of ζ_j (20), this implies that $m_{jt}(\xi)$ remains finite a.s. for $\xi \in (0, \zeta_j)$.[58]

We now use this result to derive an upper bound on the limit superior of the tail probability. For any $\xi \in (0, \zeta_j)$, Markov inequality implies

$$\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{jt}) = \mathbb{P}_t(w_{it}^{\xi} \geq w^{\xi} | i \in \mathbb{I}_{jt}) \leq \frac{m_{jt}(\xi)}{w^{\xi}}.$$

Taking logarithms, dividing by $\log w$, and passing to the limit $w \to \infty$ gives

$$\limsup_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{jt})}{\log w} \leq -\xi.$$

As this inequality holds for any $0 < \xi < \zeta_j$, it also holds in the limit $\xi \to \zeta_j$, which gives (38).

[58] In this case, we can also write $m_{jt}(\xi)$ in an integral form:

$$m_{jt}(\xi) = \left(\int_{-\infty}^{t} (\eta+\delta) e^{\int_s^t \left(\left(\xi(\mu_{wju} - \frac{1}{2}\sigma_{wju}^2) + \frac{1}{2}\xi(\xi-1)\nu_{wju}^2 - (\eta+\delta)\right)\mathrm{d}u + \xi\sigma_{wju}\,\mathrm{d}Z_u\right)}\,\mathrm{d}s\right)\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right)^{\xi} e^{\frac{1}{2}\xi(\xi-1)\nu_0^2}.$$

Step 2. In this step, we prove

$$\liminf_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{jt})}{\log w} \geq -\zeta_j \text{ for } j \in \{E, H\}. \tag{41}$$

To do so, we start by rewriting the probability of wealth being higher than a certain threshold from Proposition 2. Denoting $\overline{\Phi}(\cdot) \equiv 1 - \Phi(\cdot)$ the counter-cumulative distribution function of a standard normal variable, we have

$$\mathbb{P}_t(w_{it} \geq w | j \in \mathbb{I}_{jt}) = \int_0^\infty (\eta + \delta) e^{-(\eta+\delta)a} \overline{\Phi}\left(\frac{\log w - \mu_{j,t-a\to t}}{\nu_{j,t-a\to t}}\right) \mathrm{d}a. \tag{42}$$

In words, the mass of individuals above wealth w can be written as the integral of the mass of individuals above wealth w and with age a, evaluated over all ages $a \in (0, \infty)$.

We first consider the case $j = H$ (households). When $\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right] \leq 0$, $\zeta_H = \infty$ and (41) is trivial. Otherwise, we are in the case $\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right] > 0$. The law of large numbers implies that $\frac{\mu_{H,t-a\to t}}{a} \to \mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]$ as $a \to \infty$. As a result, for any $\epsilon > 0$, there exists a_0 such that $\mu_{H,t-a\to t} \geq \frac{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]}{1+\epsilon} a$ for $a \geq a_0$. In words, there exists an age a_0 so that all cohorts older than a_0 have a lifetime average log wealth growth higher than $\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]/(1+\epsilon)$. This implies that, for any $\log w \geq \frac{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]}{1+\epsilon} a_0$, we have $\mu_{H,t-a\to t} \geq \log w$ for $a \geq \frac{1+\epsilon}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]} \log w$. Combining this inequality with (42) implies that, for any $\log w \geq \frac{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]}{1+\epsilon} a_0$,

$$\begin{aligned}
\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{Ht}) &\geq \int_{\frac{1+\epsilon}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]} \log w}^\infty (\eta + \delta) e^{-(\eta+\delta)a} \overline{\Phi}\left(\frac{\log w - \mu_{H,t-a\to t}}{\nu_{H,t-a\to t}}\right) \mathrm{d}a \\
&\geq \int_{\frac{1+\epsilon}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]} \log w}^\infty (\eta + \delta) e^{-(\eta+\delta)a} \frac{1}{2} \mathrm{d}a \\
&= \frac{1}{2} e^{-(\eta+\delta)\frac{1+\epsilon}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]} \log w}.
\end{aligned}$$

Taking logarithms, dividing by $\log w$, and passing to the limit $w \to \infty$ gives

$$\liminf_{w\to\infty} \frac{\log \mathbb{P}\left(w_{it} \geq w | i \in \mathbb{I}_{Ht}\right)}{\log w} \geq -(1+\epsilon) \frac{\eta + \delta}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]}.$$

Since this inequality holds for any $\epsilon > 0$, we can pass to the limit $\epsilon \to 0$ to obtain (41).

$$\liminf_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{Ht})}{\log w} \geq -\frac{\eta + \delta}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]} = -\zeta_H.$$

We now consider the case $j = E$ (entrepreneurs) — while the intuition remains the same, the mathematical derivation is more intricate due to the presence of idiosyncratic volatility. The law of large numbers implies that $\frac{\mu_{E,t-a\to t}}{a} \to \mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma_{wEt}^2 - \frac{1}{2}\nu_{wEt}^2\right]$ and $\frac{\nu_{E,t-a\to t}^2}{a} \to \mathrm{E}\left[\nu_{wEt}^2\right]$ as $a \to \infty$. Together with the asymptotic behavior of $\overline{\Phi}$, this implies that, for any $\alpha > 0$,

$$\lim_{w\to\infty} \frac{\log\left(e^{-(\eta+\delta)\alpha\log w}\overline{\Phi}\left(\frac{\log w-\mu_{E,t-\alpha\log w\to t}}{\nu_{E,t-\alpha\log w\to t}}\right)\right)}{\log w} = -I(\alpha), \text{ where}$$

$$I(\alpha) \equiv (\eta+\delta)\alpha + \frac{1}{2}\frac{\left(1-\alpha\mathrm{E}\left[\mu_{wEt}-\frac{1}{2}\sigma_{wEt}^2-\frac{1}{2}\nu_{wEt}^2\right]\right)^2}{\alpha\mathrm{E}\left[\nu_{wEt}^2\right]},$$

where the convergence is locally uniform in α. Hence, for any $\alpha > 0$ and any $\epsilon > 0$, there exists $\gamma > 0$ and $\overline{w} \geq 0$ such that, for any $w \geq \overline{w}$ and $v \in (\alpha-\gamma, \alpha+\gamma)$,

$$e^{-(\eta+\delta)v\log w}\overline{\Phi}\left(\frac{\log w-\mu_{E,t-v\log w\to t}}{\nu_{E,t-v\log w\to t}}\right) \geq e^{-(1+\epsilon)I(\alpha)\log w}.$$

Without loss of generality, we can assume $\gamma \leq \alpha$. Combining the previous inequality with (42) gives, for any $w \geq \underline{w}$,

$$\begin{aligned}
\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{E,t}) &= \int_0^\infty (\eta+\delta)e^{-(\eta+\delta)a}\overline{\Phi}\left(\frac{\log w-\mu_{E,t-a\to t}}{\nu_{E,t-a\to t}}\right)\mathrm{d}a \\
&= \int_0^\infty (\eta+\delta)e^{-(\eta+\delta)v\log w}\overline{\Phi}\left(\frac{\log w-\mu_{E,t-v\log w\to t}}{\nu_{E,t-v\log w\to t}}\right)(\log w)\,\mathrm{d}v \qquad (v \equiv a/\log w) \\
&\geq \int_{\alpha-\gamma}^{\alpha+\gamma} (\eta+\delta)e^{-(\eta+\delta)v\log w}\overline{\Phi}\left(\frac{\log w-\mu_{E,t-v\log w\to t}}{\nu_{E,t-v\log w\to t}}\right)(\log w)dv \\
&\geq 2\gamma(\eta+\delta)e^{-(1+\epsilon)I(\alpha)\log w}(\log w).
\end{aligned}$$

Taking logarithms, dividing by $\log w$, and passing to the limit $w \to \infty$ gives

$$\liminf_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{E,t})}{\log w} \geq -(1+\epsilon)I(\alpha).$$

Since $\alpha > 0$ and $\epsilon > 0$ were chosen arbitrarily, we get

$$\liminf_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{E,t})}{\log w} \geq -\inf_{\alpha\in(0,\infty)} I(\alpha). \tag{43}$$

The minimum of the function $\alpha \to I(\alpha)$ on $(0,\infty)$ is attained for

$$\alpha = \frac{1}{\sqrt{\mathrm{E}\left[\mu_{wEt}-\frac{1}{2}\sigma_{wEt}^2-\frac{1}{2}\nu_{wEt}^2\right]^2 + 2(\eta+\delta)\mathrm{E}\left[\nu_{wEt}^2\right]}}$$

and the corresponding value at this point is ζ_E. Hence, (43) implies

$$\liminf_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{E,t})}{\log w} \geq -\zeta_E.$$

Step 3. Combining the results proved in the two previous steps gives that

$$\lim_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{jt})}{\log w} = -\zeta_j \text{ for } j \in \{E, H\},$$

which constitutes the first part of the proposition. Now, we have

$$\mathbb{P}_t \left(w_{it} \geq w | i \in \mathbb{I}_t\right) = \sum_{j\in\{E,H\}} \pi_j \mathbb{P}_t \left(w_{it} \geq w | i \in \mathbb{I}_{jt}\right).$$

Combining the two previous equations gives

$$\lim_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_t)}{\log w} = -\min(\zeta_E, \zeta_H).$$

To conclude, note that $\mathrm{E}\left[\nu_{wEt}^2\right] > 0$ implies that $\zeta_E < \infty$, and, therefore, $\zeta = \min(\zeta_E, \zeta_H) < \infty$. The fact that x has a stationary distribution implies that the cross-sectional moment of order one is finite for both types, which implies $\min(\zeta_E, \zeta_H) > 1$.

Finally, note that assuming $\zeta_E < \zeta_H$ implies

$$\lim_{w\to\infty} \log\left(\frac{\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{Ht})}{\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_t)}\right) = -\infty,$$

which implies that the relative fraction of households (resp. entrepreneurs) converges to zero (resp. one) in the right tail. □

Proof of Proposition 4. The proof is in two steps. In the first step, we express the instantaneous response of top wealth shares in terms of the instantaneous response of the cumulative distribution function of wealth. In the second step, we use the expression for the cumulative distribution function from Proposition 2 to obtain its instantaneous response to aggregate shocks.

Step 1. Consider a given top percentile $p \in (0, 1)$. Denote $q_{p,t}$ the top quantile, which solves the equation

$$\mathbb{P}_t \left(w_{it} \geq q_{p,t} | i \in \mathbb{I}_t\right) = p.$$

The top percentile wealth share corresponds to the sum of normalized wealth above this quantile threshold

$$S_{p,t} = \int_{q_{p,t}}^{\infty} w \, \mathrm{d}\mathbb{P}_t \left(w_{it} \leq w | i \in \mathbb{I}_t\right).$$

Integrating by parts gives

$$S_{p,t} = p q_{p,t} + \int_{q_{p,t}}^{\infty} \mathbb{P}_t \left(w_{it} \geq w | i \in \mathbb{I}_t\right) \mathrm{d}w.$$

Differentiating with respect to an aggregate shock gives

$$\begin{aligned}
\frac{\partial S_{p,t}}{\partial Z_t} &= p\frac{\partial q_{p,t}}{\partial Z_t} - \mathbb{P}_{t+h}(w_{it} \geq q_{p,t} | i \in \mathbb{I}_t)\frac{\partial q_{p,t}}{\partial Z_t} + \int_{q_{p,t}}^{\infty} \frac{\partial \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_t)}{\partial Z_t} \mathrm{d}w \\
&= \int_{q_{p,t}}^{\infty} \frac{\partial \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_t)}{\partial Z_t} \mathrm{d}w.
\end{aligned}$$

This equation expresses the instantaneous response of a percentile wealth share to an aggregate shock as the integral of the instantaneous response of the cumulative distribution function of wealth above the quantile (conveniently, the response of the top quantile to the aggregate shock does not appear in the expression).

Step 2. Combining the previous equation with the expression for the CDF given by Proposition 2 gives

$$\frac{\partial S_{p,t}}{\partial Z_t} = \int_{w=q_{p,t}}^{\infty} \left(\sum_{j \in \{E,H\}} \pi_j \int_{s=-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} \phi\left(\frac{\log w - \mu_{j,s\to t}}{\nu_{j,s\to t}}\right) \frac{1}{\nu_{j,s\to t}} \left(\frac{\partial \mu_{j,s\to t}}{\partial Z_t}\right) \mathrm{d}s \right) \mathrm{d}w,$$

where $\phi(z) \equiv \partial_Z \Phi(z)$ denotes the density of a standard normal variable. Exchanging the order of integrations gives

$$\begin{aligned}
\frac{\partial S_{p,t}}{\partial Z_t} &= \sum_{j \in \{E,H\}} \pi_j \left(\int_{s=-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} \left(\int_{w=q_{p,t}}^{\infty} \frac{1}{\nu_{j,s\to t}} \phi\left(\frac{\log w - \mu_{j,s\to t}}{\nu_{j,s\to t}}\right) \mathrm{d}w \right) \mathrm{d}s \right) \sigma_{wjt} \\
&= \sum_{j \in \{E,H\}} \pi_j \left(\int_{s=-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} \left(\int_{x=\log q_{p,t}}^{\infty} \frac{e^x}{\nu_{j,s\to t}} \phi\left(\frac{x - \mu_{j,s\to t}}{\nu_{j,s\to t}}\right) \mathrm{d}x \right) \mathrm{d}s \right) \sigma_{wjt} \\
&= \sum_{j \in \{E,H\}} \pi_j \left(\int_{s=-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} e^{\mu_{j,s\to t} + \frac{1}{2}\nu_{j,s\to t}^2} \overline{\Phi}\left(\frac{\log q_{p,t} - \mu_{j,s\to t}}{\nu_{j,s\to t}} - \nu_{j,s\to t}\right) \mathrm{d}s \right) \sigma_{wjt},
\end{aligned}$$

where the last equation uses Lemma A2 (stated and proved in Online Appendix E.2) with $\xi = 1$. Next, Proposition A1 (stated and proved in Online Appendix C.2) gives, in the special case $\xi = 1$,

$$\mathbb{E}_t \left[w_{it} 1_{w_{it} \geq q_{p,t}} | i \in \mathbb{I}_{jt} \right] = \int_{-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} e^{\mu_{j,s\to t} + \frac{1}{2}\nu_{j,s\to t}^2} \overline{\Phi}\left(\frac{\log q_{p,t} - \mu_{j,s\to t}}{\nu_{j,s\to t}} - \nu_{j,s\to t}\right) \mathrm{d}s.$$

Combining the two expressions implies

$$\frac{1}{S_t} \frac{\partial S_{p,t}}{\partial Z_t} = \sum_{j \in \{E,H\}} F_{p,t}(j) \sigma_{wjt},$$

where $F_{p,t}(j)$ denotes the fraction of wealth in the top percentile p at time t owned by individuals of type j; that is,

$$\begin{aligned}
F_{p,t}(j) &\equiv \frac{\mathbb{E}_t \left[w_{it} 1_{w_{it} \geq q_{p,t}} | i \in \mathbb{I}_{jt} \right]}{\sum_{j \in \{E,H\}} \pi_j \mathbb{E}_t \left[w_{it} 1_{w_{it} \geq q_{p,t}} | i \in \mathbb{I}_{jt} \right]} \\
&= \frac{\pi_j \int_{-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} e^{\mu_{j,s\to t} + \frac{1}{2}\nu_{j,s\to t}^2} \overline{\Phi}\left(\frac{\log q_{p,t} - \mu_{j,s\to t}}{\nu_{j,s\to t}} - \nu_{j,s\to t}\right) \mathrm{d}s}{\sum_{j \in \{E,H\}} \pi_j \int_{-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} e^{\mu_{j,s\to t} + \frac{1}{2}\nu_{j,s\to t}^2} \overline{\Phi}\left(\frac{\log q_{p,t} - \mu_{j,s\to t}}{\nu_{j,s\to t}} - \nu_{j,s\to t}\right) \mathrm{d}s}.
\end{aligned} \tag{44}$$

We can use this expression to characterize the instantaneous response of top wealth shares to aggregate shocks in the limit $p \to 0$. As shown in the proof of Proposition 3, the assumption that $\zeta_E \leq \zeta_H$ implies that the fraction of entrepreneurs tends to one in the right tail of the distribution; that is,

$$\lim_{w \to \infty} \frac{\pi_E \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{E,t})}{\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_t)} = 1.$$

As a result, we get

$$\lim_{p\to 0} \frac{\pi_E \int_{q_{p,t}}^{\infty} \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{E,t})\, \mathrm{d}w}{\int_{q_{p,t}}^{\infty} \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_t)\, \mathrm{d}w} = 1.$$

Integrating by parts both the numerator and the denominator implies

$$\lim_{p\to 0} \frac{\pi_E \int_{q_{p,t}}^{\infty} w\, \mathrm{d}\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{E,t})\, \mathrm{d}w}{\int_{q_{p,t}}^{\infty} w\, \mathrm{d}\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_t)\, \mathrm{d}w} = 1,$$

which implies $\lim_{p\to 0} F_{p,t}(E) = 1$ (and therefore $\lim_{p\to 0} F_{t,p}(H) = 0$). Hence, we get

$$\lim_{p\to 0} \frac{\partial \log S_{p,t}}{\partial Z_t} = \sigma_{wEt}.$$

□

Proof of Proposition 5. I first prove a more general statement that might be helpful in other contexts. I then apply this result to the specific model discussed in the paper.

Step 1. Consider an asset with income flow $\mathrm{E}_t[\mathrm{d}\log D_t] = g_D(x_t)$ and required return $\mathrm{E}_t[\mathrm{d}\log R_t] = g_R(x_t)$, where $g_D(\cdot)$ and $g_R(\cdot)$ are both smooth functions of a Markov process (x_t). Denote $p(x_t)$ the ratio of the asset value to its income. We can express the derivative p with respect x as the weighted sum of the derivative of g_D and g_R with respect x:

$$\partial_x \log p(x) = \mathrm{E}\left[\int_0^{\infty} e^{-\int_0^t \frac{1}{p(x_s)}\, \mathrm{d}s} \frac{\partial x_t}{\partial x_0} \partial_x \left(g_D(x_t) - g_R(x_t)\right) \mathrm{d}t \Big| x_0 = x\right]. \tag{45}$$

To show this, start with the market pricing equation

$$\mathrm{E}_t[\mathrm{d}\log R_t] = g_R(x_t)\, \mathrm{d}t = \frac{1}{p(x_t)}\, \mathrm{d}t + g_D(x_t)\, \mathrm{d}t + \mathrm{E}_t\left[\mathrm{d}\log p(x_t)\right].$$

Differentiating with respect to x_0 gives

$$\frac{\partial x_t}{\partial x_0} \partial_x g_R(x_t)\, \mathrm{d}t = -\frac{\partial x_t}{\partial x_0} \frac{1}{p(x_t)} \partial_x \log p(x_t)\, \mathrm{d}t + \frac{\partial x_t}{\partial x_0} \partial_x g_D(x_t)\, \mathrm{d}t + \mathrm{E}_t\left[\mathrm{d}\left(\frac{\partial x_t}{\partial x_0} \partial_x \log p(x_t)\right)\right].$$

Rearranging,

$$\mathrm{E}_t\left[\mathrm{d}\left(e^{-\int_0^t \frac{1}{p(x_s)}\, \mathrm{d}s} \frac{\partial x_t}{\partial x_0} \partial_x \log p(x_t)\right)\right] = e^{-\int_0^t \frac{1}{p(x_s)}\, \mathrm{d}s} \frac{\partial x_t}{\partial x_0} \partial_x \left(g_R(x_t) - g_D(x_t)\right).$$

Integrating forward gives (45), which concludes the proof of this statement.

Step 2. In our particular model, $g_D(x)$ is constant and $g_R(x) = \mu_R(x) - \frac{1}{2}\sigma_R^2(x)$, so (45) implies

$$\partial_x \log p(x) = -\mathrm{E}\left[\int_0^{\infty} e^{-\int_0^t \frac{1}{p(x_s)}\, \mathrm{d}s} \frac{\partial x_t}{\partial x_0} \partial_x \left(\mu_R(x_t) - \frac{1}{2}\sigma_R^2(x_t)\right) \mathrm{d}t \Big| x_0 = x\right].$$

Adding and subtracting by $\partial_x r(x)$ in the right-hand-side and multiplying both sides by $\sigma_x(x)$ gives the result (32).

I now briefly discuss how to compute this decomposition using numerical methods. The extended Feynman-Kac formula stated in Lemma A3 (Online Appendix E.3) implies that, for any f, the function $u(x) \equiv \mathrm{E}\left[\int_0^\infty e^{-\int_0^h \frac{1}{p_s}\,\mathrm{d}s} \frac{\partial x_h}{\partial x_0} f(x_h)\,\mathrm{d}h \middle| x_0 = x\right]$ solves the linear ODE

$$0 = f(x) + \left(\partial_x \mu(x) - \frac{1}{p(x)}\right) u(x) + \Big(\mu_x(x) + \sigma_x(x)\partial_x\sigma_x(x)\Big)\partial_x u(x) + \frac{1}{2}\sigma_x^2(x)\partial_{xx}u(x).$$

Applying this result with $f(x) = r(x)$ and $f(x) = \mu_R(x) - \frac{1}{2}\sigma_R^2(x)$ allows me to compute the risk-free rate channel and the excess return channel, respectively. $\square$

Proof of Proposition 6. We can express the normalized wealth of an entrepreneur at time $t+h$ in terms of their wealth at time t:

$$w_{i,t+h} = w_{i,t} e^{\int_t^{t+h}\left(\left(\mu_{wEs} - \frac{1}{2}\sigma_{wEs}^2 - \frac{1}{2}\nu_{wjs}^2\right)\mathrm{d}s + \sigma_{wEs}\,\mathrm{d}Z_s + \nu_{wEs}\,\mathrm{d}B_{is}\right)}.$$

Averaging across all individuals in $\mathbb{I}_{E,t} \cap \mathbb{I}_{E,t+h}$:

$$\mathbb{E}_{t+h}\left[w_{i,t+h} | i \in \mathbb{I}_{E,t} \cap \mathbb{I}_{E,t+h}\right] = \mathbb{E}_t\left[w_{it} | i \in \mathbb{I}_{E,t} \cap \mathbb{I}_{E,t+h}\right] e^{\int_t^{t+h}\left(\left(\mu_{wEs} - \frac{1}{2}\sigma_{wEs}^2\right)\mathrm{d}s + \sigma_{wEs}\,\mathrm{d}Z_s\right)}.$$

Taking logarithms and the expectations at time t

$$\mathrm{E}_t\left[\log \mathbb{E}_{t+h}\left[w_{i,t+h} | i \in \mathbb{I}_{E,t} \cap \mathbb{I}_{E,t+h}\right]\right] = \log \mathbb{E}_t\left[w_{it} | i \in \mathbb{I}_{E,t} \cap \mathbb{I}_{E,t+h}\right] + \mathrm{E}_t\left[\int_t^{t+h}\left(\mu_{wEs} - \frac{1}{2}\sigma_{wEs}^2\right)\mathrm{d}s\right].$$

Differentiating with respect to an aggregate shock at time t:

$$\frac{\partial}{\partial Z_t}\mathrm{E}_t\left[\log \mathbb{E}_{t+h}\left[w_{i,t+h} | i \in \mathbb{I}_{E,t} \cap \mathbb{I}_{E,t+h}\right]\right] = \sigma_{wEt}(x_t) + \partial_{x=x_t}\mathrm{E}\left[\int_0^h \left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)\mathrm{d}s | x_0 = x\right]\sigma_x(x_t),$$

which proves the result. Note that the right-hand side only depends on the horizon h and the value of x at t, which justifies the notation $\epsilon(x,h)$.

I now briefly discuss how to compute this quantity using numerical methods. The extended Feynman-Kac formula stated in Lemma A3 (Online Appendix E.3) implies that the function $u(x,h) \equiv \mathrm{E}\left[\int_0^h \frac{\partial x_h}{\partial x_0}\left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)(x_h)\,\mathrm{d}h\right|$ solves the linear PDE

$$\partial_h u(x,h) = \left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)(x) + \partial_x\mu(x)u(x,h) + \Big(\mu_x(x) + \sigma_x(x)\partial_x\sigma_x(x)\Big)\partial_x u(x,h) + \frac{1}{2}\sigma_x^2(x)\partial_{xx}u(x,h),$$

with initial condition $u(x,0) = 0$. After solving this PDE, one can obtain $\epsilon(x,h) = \sigma_{wE}(x) + u(x,h)\sigma_x(x)$. $\square$

Proof of Proposition 7. We can follow the same steps as in Step 1 of the proof of Proposition 4 to obtain the response of the top wealth share to an aggregate shock at horizon h in terms of the response of the CDF of the wealth distribution:

$$\frac{\partial S_{p,t+h}}{\partial Z_t} = \int_{q_{p,t+h}}^{\infty} \frac{\partial \mathbb{P}_{t+h}(w_{it+h} \geq w | i \in \mathbb{I}_{t+h})}{\partial Z_t} \, \mathrm{d}w.$$

Combining with the expression for the CDF of wealth given in Proposition 2 gives

$$\frac{\partial S_{p,t+h}}{\partial Z_t} = \int_{w=q_{p,t+h}}^{\infty} \left(\sum_{j \in \{E,H\}} \pi_j \int_{s=-\infty}^{t+h} (\eta + \delta) e^{-(\eta+\delta)(t+h-s)} \phi \left(\frac{\log w - \mu_{j,s\to t+h}}{\nu_{j,s\to t+h}} \right) \right.$$
$$\left. \times \frac{1}{\nu_{j,s\to t+h}} \left(\frac{\partial \mu_{j,s\to t+h}}{\partial Z_t} - \frac{\log w - \mu_{j,s\to t+h}}{\nu_{j,s\to t+h}} \frac{\partial \nu_{j,s\to t+h}}{\partial Z_t} \right) \mathrm{d}s \right) \mathrm{d}w,$$

where $\phi(z) \equiv \partial_Z \Phi(z)$ denotes the density of a standard normal variable. Next, the expressions for $\mu_{j,s\to t+h}$ and $\nu_{j,s\to t+h}$ given in Proposition 2 imply

$$\frac{\partial \mu_{j,s\to t+h}}{\partial Z_t} = \sigma_{wj}(x_t) 1_{s\leq t} + \sigma_x(x_t) \int_t^{t+h} \frac{\partial x_u}{\partial x_t} \left(\partial_x \left(\mu_{wj} - \frac{1}{2}\sigma_{wj}^2 \right) (x_u) \, \mathrm{d}u + \partial_x \sigma_{wj}(x_u) \, \mathrm{d}Z_u \right) 1_{s\leq u}$$

and

$$\frac{\partial \nu_{j,s\to t+h}}{\partial Z_t} = 1_{j=E} \sigma_x(x_t) \frac{\nu^2}{\nu_{j,s\to t+h}} \int_t^{t+h} \alpha_E(x_u) \frac{\partial x_u}{\partial x_t} \partial_x \alpha_E(x_u) \, \mathrm{d}u.$$

The fact that $\nu_{j,s\to t+h}$ potentially reacts to aggregate shocks is an artifact of the model: to make sure that entrepreneurs never own collectively more than 100% of equity, I assumed that, whenever x_t is higher than $1/\alpha_E$, the required equity share of entrepreneurs decreases to $\alpha_E(x) = 1/x$ (see Equation 7). Because this situation happens in less than 0.01% of the time in the calibrated model, and because taking this into account would complicate the formulas further, I make the approximation $\partial \nu_{j,s\to t+h}/\partial Z_t \approx 0$ in the rest of the derivation.

As a result,

$$\frac{\partial S_{p,t+h}}{\partial Z_t} \approx \sum_{j\in\{E,H\}} \pi_j \int_{w=q_{p,t+h}}^{\infty} \int_{s=-\infty}^{t+h} (\eta+\delta) e^{-(\eta+\delta)(t+h-s)} \frac{1}{\nu_{j,s\to t+h}} \phi \left(\frac{\log w - \mu_{j,s\to t+h}}{\nu_{j,s\to t+h}} \right)$$
$$\times \left(\sigma_{wj}(x_t) 1_{s\leq t} + \sigma_x(x_t) \int_{u=t}^{t+h} \frac{\partial x_u}{\partial x_t} \left(\partial_x \left(\mu_{wj} - \frac{1}{2}\sigma_{wj}^2 \right)(x_u) \, \mathrm{d}u + \partial_x \sigma_{wj}(x_u) \, \mathrm{d}Z_u \right) 1_{s\leq u} \right) \mathrm{d}s \, \mathrm{d}w.$$

Exchanging the order of integrations gives

$$\frac{\partial S_{p,t+h}}{\partial Z_t} \approx \sum_{j\in\{E,H\}} \sigma_{wj}(x_t) W_{p,t+h}(j,t)$$
$$+ \sum_{j\in\{E,H\}} \sigma_x(x_t) \int_t^{t+h} \frac{\partial x_s}{\partial x_t} \left(\partial_x \left(\mu_{wj} - \frac{1}{2}\sigma_{wj}^2 \right)(x_s) \, \mathrm{d}s + \partial_x \sigma_{wj}(x_s) \, \mathrm{d}Z_s \right) W_{p,t+h}(j,s),$$

where $W_{p,t+h}(j,s)$ denotes the total normalized wealth at time $t+h$ for individuals of type j in the top percentile p born before time $s \leq t+h$:

$$W_{p,t+h}(j,s) \equiv \pi_j \int_{-\infty}^{s} (\eta+\delta)e^{-(\eta+\delta)u} \int_{q_{t+h,p}}^{\infty} \frac{1}{\nu_{j,u\to t+h}} \phi\left(\frac{\log w - \mu_{j,u\to t+h}}{\nu_{j,u\to t+h}}\right) \mathrm{d}w\, \mathrm{d}u$$
$$= \pi_j \int_{-\infty}^{s} (\eta+\delta)e^{-(\eta+\delta)u} e^{\mu_{j,u\to t+h}+\frac{1}{2}\nu^2_{j,u\to t+h}} \overline{\Phi}\left(\frac{\log q_{p,t+h} - \mu_{j,u\to t+h}}{\nu_{j,u\to t+h}} - \nu_{j,u\to t+h}\right) \mathrm{d}u,$$

where the second line uses Lemma A2 (state and proved in Online Appendix E.2) with $\xi = 1$. Dividing by $S_{p,t+h} = W_{p,t+h}(E, t+h) + W_{p,t+h}(H, t+h)$ gives

$$\frac{\partial \log S_{p,t+h}}{\partial Z_t} \approx \sum_{j\in\{E,H\}} \sigma_{wj}(x_t) F_{p,t+h}(j,t)$$
$$+ \sum_{j\in\{E,H\}} \sigma_x(x_t) \int_t^{t+h} \frac{\partial x_s}{\partial x_t} \left(\partial_x \left(\mu_{wj} - \frac{1}{2}\sigma^2_{wj}\right)(x_s)\, \mathrm{d}s + \partial_x \sigma_{wj}(x_s)\, \mathrm{d}Z_s\right) F_{p,t+h}(j,s),$$

where $F_{p,t+h}(j,s)$ denotes the *fraction* of wealth in the top percentile p at time $t+h$ owned by individuals of type j born before time $s \leq t+h$:

$$F_{p,t+h}(j,s) \equiv \frac{W_{p,t+h}(j,s)}{W_{p,t+h}(E,t+h) + W_{p,t+h}(H,t+h)}$$
$$= \frac{\pi_j \int_{-\infty}^{s} (\eta+\delta)e^{-(\eta+\delta)u} e^{\mu_{j,u\to t+h}+\frac{1}{2}\nu^2_{j,u\to t+h}} \overline{\Phi}\left(\frac{\log q_{p,t+h} - \mu_{j,u\to t+h}}{\nu_{j,u\to t+h}} - \nu_{j,u\to t+h}\right) \mathrm{d}u}{\sum_{j\in\{E,H\}} \pi_j \int_{-\infty}^{t+h} (\eta+\delta)e^{-(\eta+\delta)u} e^{\mu_{j,u\to t+h}+\frac{1}{2}\nu^2_{j,u\to t+h}} \overline{\Phi}\left(\frac{\log q_{p,t+h} - \mu_{j,u\to t+h}}{\nu_{j,u\to t+h}} - \nu_{j,u\to t+h}\right) \mathrm{d}u}. \quad (46)$$

Taking the expectation at time t gives

$$\frac{\partial \mathrm{E}_t[\log S_{p,t+h}]}{\partial Z_t} \approx \mathrm{E}_t\left[\sum_{j\in\{E,H\}} \sigma_{wj}(x_t) F_{p,t+h}(j,t)\right]$$
$$+ \mathrm{E}_t\left[\sum_{j\in\{E,H\}} \int_t^{t+h} \frac{\partial x_s}{\partial x_t} \left(\partial_x \left(\mu_{wj} - \frac{1}{2}\sigma^2_{wj}\right)(x_s)\, \mathrm{d}s + \partial_x \sigma_{wj}(x_s)\, \mathrm{d}Z_s\right) F_{p,t+h}(j,s)\right] \sigma_x(x_t).$$

To simplify this further and make the resulting formula more intuitive, the second (and last) approximation I make is $\mathrm{cov}(\partial_x \sigma_{wjt}(x_s)\, \mathrm{d}Z_u, F_{p,t+h}(j,u)) \approx 0$; that is, I neglect the second-order term that accounts for the fact that aggregate shocks between t and $t+h$ impact both the relative wealth exposure of entrepreneurs and the fraction of wealth in the top percentile at time $t+h$ owned by agents of type j born before u. This term is small as long as $\partial_x \sigma_{wjt}(x_s)$ and/or $\partial F_{p,t+h}(j,u)/\partial Z_u$ are small.[59] This approximation allows me to abstract from the term in $\partial_x \sigma_{wj}(x_s)\, \mathrm{d}Z_s$ in the previous equation, which proves the main formula in the

[59] The fact that the actual impulse responses of top wealth shares plotted in Figure 5 are close to $\epsilon(x,h)$ for top percentiles is visual evidence that this second-order term is small.

proposition:

$$\frac{\partial \mathrm{E}_t[\log S_{p,t+h}]}{\partial Z_t} \approx \mathrm{E}_t\left[\sum_{j\in\{E,H\}} \sigma_{wj}(x_t) F_{p,t+h}(j,t)\right] + \mathrm{E}_t\left[\sum_{j\in\{E,H\}} \int_t^{t+h} \frac{\partial x_s}{\partial x_t} \partial_x \left(\mu_{wj} - \frac{1}{2}\sigma_{wj}^2\right)(x_s) F_{p,t+h}(j,s)\, \mathrm{d}s\right] \sigma_x(x_t). \quad (47)$$

I now use this proposition to describe the impulse response of top wealth shares as $h \to \infty$ or $p \to 0$. First, consider the case where p is fixed and $h \to \infty$. The key observation is that, given two times $t < s$ and a top percentile p, the fraction of individuals in the top percentile p at time $t+h$ who are born before time s tends to zero as the horizon h tends to infinity; that is, $\lim_{h\to\infty} F_{p,t+h}(j,s) = 0$ a.s. This result comes from the fact that this fraction is bounded above by the relative mass of individuals born before s, $e^{-(\delta+\eta)(t+h-s)}$, which itself tends to zero. Combining this with (47) gives $\lim_{h\to\infty} \partial \mathrm{E}_t[\log S_{p,t+h}]/\partial Z_t = 0$.

Second, consider the case where h is fixed and $p \to 0$. The key observation is that, given a time t and a horizon h, the fraction of individuals in the top percentile p at time $t+h$ who are born after time t tends to zero as the top percentile p tends to zero. This comes from the fact that the cross-sectional average growth of the cohorts born between t and $t+h$ is bounded above by $\sup_{u\in(t,t+h)} |\mu_{ju}| h + \sup_{u\in(t,t+h)} |\sigma_{ju}| \sup_{u\in(t,t+h)} |Z_{t+h} - Z_u|$, and so the wealth distribution of these agents at time $t+h$ has thin tails. Because the overall distribution of entrepreneurs has a thick tail, this implies $\lim_{p\to 0} F_{p,t+h}(j,t) = 1_{j=E}$ a.s.. Combining with (47) gives

$$\lim_{p\to 0} \frac{\partial \mathrm{E}_t[\log S_{p,t+h}]}{\partial Z_t} = \epsilon(x,h),$$

which concludes the proof.

□

Online Appendix for "Wealth Inequality and Asset Prices"

MATTHIEU GOMEZ

B Appendix for Section 2

B.1 Alternative specifications

Controlling for predetermined variables. To assess the robustness of the empirical findings discussed in Section 2.2, I now augment the baseline specifications with a set of variables known at time $t-1$; that is, I estimate the models:

$$\begin{aligned}\log\left(\frac{W_{p,t+h}}{W_{p,t-1}}\right)-(h+1)\log R_{f,t}=\alpha_{p,h}+\beta_{p,h}\left(\log R_{M,t}-\log R_{f,t}\right)+\sum_{c\in\mathcal{C}}\gamma_c Z_{c,t-1}+\epsilon_{p,t+h},\\ \log\left(\frac{S_{p,t+h}}{S_{p,t-1}}\right)=a_{p,h}+b_{p,h}(\log R_{M,t}-\log R_{f,t})+\sum_{c\in\mathcal{C}}\gamma_c Z_{c,t-1}+e_{p,t+h},\end{aligned}\tag{A1}$$

where $\{Z_{c,t}\}_{c\in\mathcal{C}}$ denotes a set of additional controls. The special case $\mathcal{C}=\varnothing$ (no controls) corresponds to the baseline specifications discussed in the main text. The advantage of adding predetermined variables is that they help to capture any information known at time $t-1$, making it easier to interpret the estimates from local projections as the effect of "unexpected" stock market returns on the dependent variable.

By analogy with the omitted variable bias, it is particularly important to add variables correlated with expected excess stock returns that may independently affect the expected growth of the dependent variable. For the first set of variables, I add the dividend-price ratio, a known predictor of excess returns, and a five-year rolling average of past excess returns. For the second set of variables, I add two lags of the dependent variables (i.e., $\log W_{p,t-1}$ and $\log W_{p,t-2}$ for the average wealth in the top percentile and $\log S_{p,t-1}$ and $\log S_{p,t-2}$ for the top percentile wealth share). Note that adding *two* lags means that I implicitly control for the dependent variable's lagged growth, which helps capture low-frequency changes in wealth inequality. I also add the cross-sectional variance in stock market returns, following works by Atkeson and Irie (2022) and Gomez (2023), who stress the role of this quantity in determining the low-frequency fluctuations in top wealth shares.[1]

Table A1 reports the results. Overall, I find estimates for $\beta_{p,3}$ and $b_{p,3}$ that are highly similar to those obtained in the baseline specification (Table 1). The intuition is that most of the variations in excess stock market returns come from variations in *unexpected* excess stock market returns, which are uncorrelated with variables known at time $t-1$.[2]

[1] More precisely, I control by the synthetic "between" term constructed in Section 5 of Gomez (2023); that is, $(\zeta_{t-1}-1)/2\nu_{t-1}^2$, where ζ_{t-1} is a local measure of the Pareto exponent in year $t-1$, $\zeta_t=1/(1-\log_{10}(S_{0.1\%,t}/S_{0.01\%,t}))$ and ν_{t-1}^2 is the proportional to the cross-sectional variance of stock market returns for public firms in year $t-1$.

[2] One outlier is the Forbes 400, where I observe lower estimates. This deviation appears to be linked to a spike in cross-sectional variance just before the tech bubble burst, a period marked by lower excess returns. Hence, this deviation appears to be a byproduct of the limited time frame of the data, as Appendix B.3 shows that the reaction of

Table A1: Exposure to stock market returns after controlling for predetermined variables

	Top 100%	Top 1%	Top 0.1%	Top 0.01%	Top 400
	(1)	(2)	(3)	(4)	(5)
Panel A: Average wealth					
Excess returns	0.38***	0.50***	0.60***	0.79***	0.81***
	(0.09)	(0.11)	(0.13)	(0.19)	(0.26)
Lagged average wealth	0.04	−0.17	−0.11	−0.00	−0.97
	(0.25)	(0.22)	(0.22)	(0.24)	(0.72)
Two-year lagged average wealth	−0.06	0.17	0.13	0.01	0.82
	(0.26)	(0.23)	(0.22)	(0.25)	(0.61)
Lagged cross-sectional variance of returns	−2.20***	−1.24	−0.59	0.10	−6.06
	(0.78)	(0.88)	(1.03)	(1.34)	(3.81)
Lagged dividend-price ratio	−0.11**	−0.08	−0.08	−0.13	−0.13
	(0.05)	(0.06)	(0.08)	(0.12)	(0.24)
Five-year average lagged excess returns	−0.84***	−0.77**	−0.66	−0.66	0.22
	(0.31)	(0.36)	(0.42)	(0.50)	(1.26)
Adjusted R^2	0.24	0.25	0.23	0.21	0.35
N	102	102	102	102	30
Panel B: Wealth share					
Excess returns		0.12**	0.22**	0.40***	0.37*
		(0.05)	(0.09)	(0.14)	(0.18)
Lagged wealth share		−0.04	−0.07	−0.09	−0.60
		(0.29)	(0.27)	(0.26)	(0.39)
Two-year lagged wealth share		−0.12	−0.15	−0.19	0.25
		(0.28)	(0.26)	(0.24)	(0.35)
Lagged cross-sectional variance of returns		1.30***	2.52***	3.79***	−4.61***
		(0.38)	(0.70)	(1.10)	(1.36)
Lagged dividend-price ratio		−0.02	−0.07**	−0.19***	−0.04
		(0.02)	(0.03)	(0.06)	(0.12)
Five-year average lagged excess returns		−0.04	0.03	0.02	0.16
		(0.11)	(0.20)	(0.30)	(0.72)
Adjusted R^2		0.26	0.29	0.33	0.50
N		102	102	102	30

Notes: This table reports the coefficients obtained in the regression of the growth of the average wealth in top percentiles on excess stock returns controlling for a set of predetermined variables; that is, equation (A1) with $h = 3$. Estimation is done via OLS. Standard errors are in parentheses and are estimated using heteroskedasticity consistent standard errors. *,** ,*** indicate significance at the 10%, 5%, 1% levels, respectively.

Controlling for future excess returns. As seen in Figure 1 in the main text, the exposure of the wealth in top percentiles to stock market returns tends to build over time (especially for the top 400). As discussed in the main text, my interpretation is that the data on top wealth inequality is not very precise (average of data sources over a year rather than at a point in time) and that private assets may take time to respond to changes in valuation.

An alternative reason might be that excess stock market returns themselves are correlated: in this case, impulse response functions estimated by local projections measure both the effect of the current shock in the treatment *and* the effect of the current shock on future treatments. As discussed in Alloza et al. (2020),

Forbes 400 is driven by changes in the average wealth in the top rather than by the arrival of new fortunes at the top (which is what the cross-sectional variance is supposed to control for).

one way to isolate the first effect is to add future treatments as controls in the local projection specification; that is,

$$\log\left(\frac{W_{p,t+h}}{W_{p,t-1}}\right) - (h+1)\log R_{f,t} = \alpha_h + \sum_{i=0}^{h} \beta_{h,i}\left(\log R_{M,t+i} - \log R_{f,t+i}\right) + \epsilon_{p,t,t+h},$$
$$\log\left(\frac{S_{p,t+h}}{S_{p,t-1}}\right) = a_{p,h} + \sum_{i=0}^{h} \beta_{h,i}\left(\log R_{M,t+i} - \log R_{f,t+i}\right) + e_{p,t+h}. \tag{A2}$$

Table A2 reports the result for $h = 3$. One can see that the coefficients on contemporaneous excess stock returns are similar to the ones obtained in the baseline specifications (Table 1). The reason is that excess stock returns are not significantly correlated over time, so there is little difference in controlling for future excess stock returns.

Table A2: Exposure to stock market returns after controlling for future excess returns

	Top 100%	Top 1%	Top 0.1%	Top 0.01%	Top 400
	(1)	(2)	(3)	(4)	(5)
Panel A: Average wealth					
Excess returns	0.48***	0.62***	0.70***	0.86***	1.04***
	(0.09)	(0.09)	(0.10)	(0.14)	(0.14)
Excess returns (1y lead)	0.39***	0.50***	0.60***	0.77***	0.87***
	(0.10)	(0.09)	(0.09)	(0.12)	(0.12)
Excess returns (2y lead)	0.40***	0.57***	0.66***	0.75***	0.60***
	(0.09)	(0.09)	(0.10)	(0.14)	(0.10)
Excess returns (3y lead)	0.12	0.19**	0.25**	0.29**	0.36**
	(0.08)	(0.09)	(0.11)	(0.14)	(0.16)
Adjusted R^2	0.44	0.62	0.61	0.54	0.73
N	103	103	103	103	31
Panel B: Wealth share					
Excess returns		0.13***	0.22***	0.38***	0.61***
		(0.04)	(0.08)	(0.12)	(0.20)
Excess returns (1y lead)		0.11***	0.21***	0.38***	0.46***
		(0.04)	(0.07)	(0.11)	(0.16)
Excess returns (2y lead)		0.17***	0.26***	0.34***	0.30**
		(0.04)	(0.08)	(0.12)	(0.12)
Excess returns (3y lead)		0.07	0.13	0.18	0.13
		(0.04)	(0.08)	(0.13)	(0.13)
Adjusted R^2		0.21	0.17	0.18	0.39
N		103	103	103	31

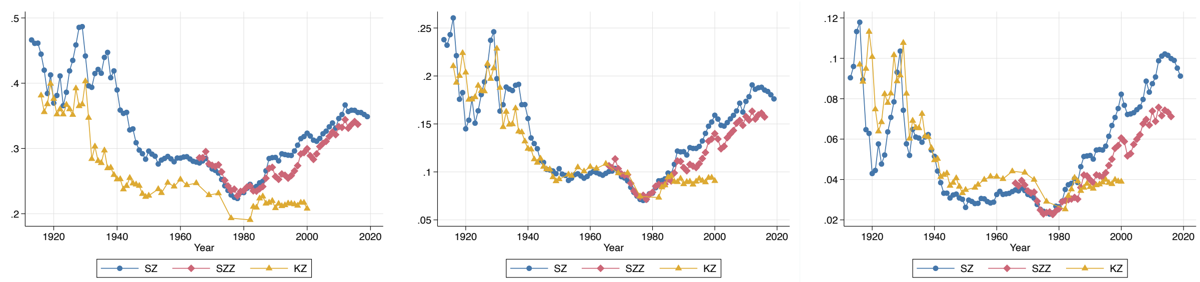
Notes: This table reports the coefficients obtained in the regression of the four-year growth of the average wealth in top percentiles on excess stock returns controlling for future excess stock returns; that is, equation (A2) with $h = 3$. Estimation is done via OLS. Standard errors are in parentheses and are estimated using heteroskedasticity consistent standard errors. *,** ,*** indicate significance at the 10%, 5%, 1% levels, respectively.

Finally, note that while some amount of misspecification in my local projection specifications is inevitable (excess stock market returns are very close but not identical to unexpected stock market returns), these estimates are still informative as long as I use the same specification when comparing the model to the data (which I do in Appendix Figures A4 and A5).

B.2 Alternative data sources

Alternative series on top wealth shares. Due to data limitations, there is substantial uncertainty about the historical evolution of top wealth shares. In my baseline results, I focus on the updated series of top wealth shares constructed by Saez and Zucman (2016) (2022 vintage), which improves on the series released at the time of publication by incorporating several methodological improvements described in Saez and Zucman (2020) and Saez and Zucman (2022).

Two major alternative series for top wealth shares are available in the literature. The first alternative is the series constructed from income tax returns by Smith et al. (2023) from 1966 to 2016. While this series initially disagreed with Saez and Zucman (2016), subsequent updates in both series largely reconciled these differences, leaving only some discrepancies for the top 0.01%. The second alternative is the series constructed from estate tax returns by Kopczuk and Saez (2004) from 1916 to 2000. This series estimates the living's wealth distribution from the deceased's wealth distribution using the mortality multiplier technique, which amounts to weighting each estate tax return by the inverse probability of death (depending on age and gender). One concern with this methodology is that the evolution of death rates at the top may have diverged from the rest of the population. I refer the reader to Saez and Zucman (2020), Smith et al. (2023), and Saez and Zucman (2022) for a more thorough discussion of the difference between these three series. Figure A1 compares the evolution of top wealth shares across the tree series.



(a) Top 1% (b) Top 0.1% (c) Top 0.01%

Figure A1: Alternative series for top wealth shares

Notes: The figure plots three alternative series for top wealth shares. The label "SZ" denotes the series from Saez and Zucman (2016) (2022 vintage), used for my baseline results. The label "SZZ" denotes the series from Smith et al. (2023) while the label "KZ" denotes the series from Kopczuk and Saez (2004); ; both correspond to the versions available at their respective publication dates.

To assess how much my results depend on the data source for top wealth shares, I re-estimate the baseline specification discussed in the main text (1) with $h = 3$ using these two alternative series. For the sake of comparison, it is important to hold the time sample constant across these exercises; hence, I replace the dependent variable (the four-year growth of the average wealth in each top percentile) with the one from Saez and Zucman (2016) in years in which they are missing (i.e., pre-1966 Smith et al. (2023) and post-2000 for Kopczuk and Saez (2004)).

Panel A of Table A3 reports the results using the series from Smith et al. (2023) while Panel B reports the results using the series from Kopczuk and Saez (2004). Overall, I find that the three main data series on top wealth inequality give similar results for the response of top percentiles to stock market returns. In particular, the elasticity for the top 0.01% is 0.80 using the data from Smith et al. (2023) and 0.72 using the data from Kopczuk and Saez (2004), which are close to the baseline estimate 0.78 obtained using the data from Saez and Zucman (2016) (Table 1). To economize on space, I do not report the corresponding estimates obtained for the growth of top wealth shares: as discussed in Section 2.2, they can simply be obtained by subtracting the elasticity of the average wealth in each percentile by the elasticity of the average wealth in the economy (which is the same across data sources). These results suggest that while the three series of top wealth shares have different implications for the low-frequency fluctuations of top wealth shares, they largely agree on the response of top wealth shares to excess stock market returns.

Table A3: Exposure to stock market returns using alternative series for top wealth shares

	Top 100%	Top 1%	Top 0.1%	Top 0.01%
	(1)	(2)	(3)	(4)
Panel A: Wealth data from Kopczuk and Saez (2004)				
Excess returns	0.43***	0.60***	0.67***	0.72***
	(0.11)	(0.14)	(0.15)	(0.17)
Adjusted R^2	0.16	0.22	0.24	0.20
Time sample	1914-2016	1914-2016	1914-2016	1914-2016
N	103	103	103	103
Panel B: Wealth data from Smith et al. (2023)				
Excess returns	0.43***	0.55***	0.64***	0.80***
	(0.11)	(0.12)	(0.14)	(0.18)
Adjusted R^2	0.16	0.20	0.19	0.19
Time sample	1914-2016	1914-2016	1914-2016	1914-2016
N	103	103	103	103

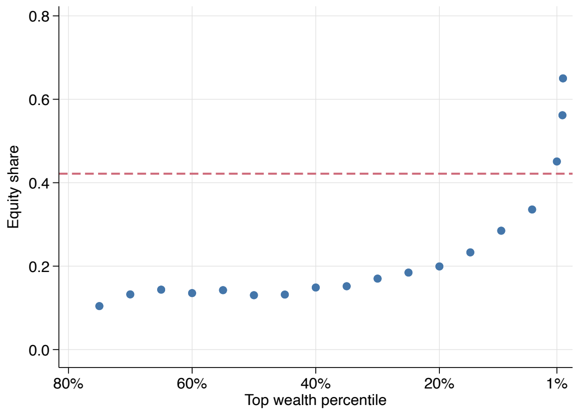
Notes: This table reports the coefficients obtained in the regression of the four-year growth of the average wealth in top percentiles on excess stock returns; that is, equation (1) with $h = 3$ using data from Smith et al. (2023) (Panel A) and from Kopczuk and Saez (2004) (Panel B). Estimation is done via OLS. Standard errors are in parentheses and are estimated using heteroskedasticity consistent standard errors. *,** ,*** indicate significance at the 10%, 5%, 1% levels, respectively.

Evidence from portfolio holdings. One additional data source on the wealth distribution is the Survey of Consumer Finances (SCF). I now show that the estimates for the equity exposure of different top percentiles, as reported in Table 1, line up with the share of wealth invested in equity in different top percentiles, as reported in the SCF.[3]

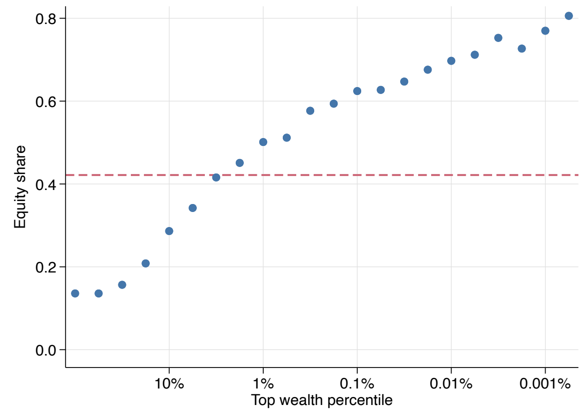
Figure A2a plots the average equity share within percentile bins across the wealth distribution, where the equity share is defined as the total investment in equity over financial wealth, as reported in the SCF. The equity share is essentially flat at 0.2 over most of the wealth distribution, but increases sharply within the top 1%. Figure A2b plots the equity share with respect to the *log* top percentiles, showing that the equity share is approximately linear in the log percentile at the top of the distribution.

The first row of Table A4 returns the wealth-weighted average equity share in all top percentiles. The key observation is that these equity shares line up almost perfectly with the response of the average wealth in top percentiles to stock market returns, as reported in Table 1: for instance, the average equity share in the top 0.01% is 0.75 (last column of Table A4), which is consistent with the fact that a 1% excess stock market return increases the average wealth in the top 0.01% by 0.78% on average (last column of Table 1). Table A4 also decomposes the equity share into several subcomponents, revealing that the increase in the

[3]The survey is a repeated cross-section of about 4,000 households per survey year, including a high-wealth sample. The survey is conducted every three years, from 1989 to 2019. The respondents provide information on their financial wealth, including their public and private equity investments.



(a) Linearly-spaced top percentiles



(b) Loglinearly-spaced top percentiles

Figure A2: Equity share across the wealth distribution

Notes: Figure A2a plots the average equity share within 20 linearly spaced percentile bins in the wealth distribution. Figure A2b plots the average equity share within 20 logarithmically spaced percentile bins in the wealth distribution. The horizontal line represents the wealth-weighted average equity share in the economy. The equity share is defined as the sum of private and public equity divided by net worth: (`equity` + `bus`) / `net worth`. Data from SCF 1989-2019.

average equity share across the wealth distribution is mainly driven by an increase in the share of wealth invested in private equity, consistently with the model discussed in Section 3.

The second to last row of the table reports the fraction of entrepreneurs in each top percentile, where an entrepreneur is defined as a household investing more than two-thirds of their wealth in equity. The last row reports the fraction of income in each percentile that takes the form of labor income.

Table A4: Average equity share in top percentiles

	Top percentiles			
	Top 100%	Top 1%	Top 0.1%	Top 0.01%
Equity share	0.42	0.61	0.68	0.75
Public equity	0.21	0.23	0.21	0.19
Directly held	0.12	0.16	0.17	0.16
Indirectly held	0.09	0.06	0.04	0.03
Private equity	0.21	0.38	0.47	0.56
Actively managed	0.18	0.33	0.40	0.47
Non actively managed	0.02	0.05	0.07	0.09
Proportion of entrepreneurs	0.09	0.41	0.57	0.69
Labor income / Total income	0.68	0.34	0.23	0.14

Notes: The equity share is defined as the sum of private and public equity divided by net worth: (`equity + bus`) / `networth`. "Entrepreneurs" are defined as households investing more than half of their wealth in equity. The share of labor income in total income is defined as `wageinc/income`. Data from SCF 1989-2019.

B.3 Accounting for composition changes

One can always decompose the growth of the average wealth in a top percentile into two terms: an intensive term that captures the wealth growth of households initially in the top percentile (whether or not they remain in the top percentile by the end of the period) and an extensive term that captures the effect of composition changes on the average wealth in the top percentile (due to idiosyncratic shocks and demographic forces).

Following Gomez (2023), I decompose the growth of the average wealth in the top 400 into these two terms. More precisely, I construct the intensive term as

$$\text{Intensive term}_t \equiv \log\left(\frac{\sum_{i\in\mathcal{P}_{t-1}\setminus\mathcal{D}_t} W_{i,t}}{\sum_{i\in\mathcal{P}_{t-1}\setminus\mathcal{D}_t} W_{i,t-1}}\right),$$

where $\mathcal{P}_{t-1}\setminus\mathcal{D}_t$ denotes the set of individuals in the top 400 at time $t-1$ who do not die between $t-1$ and t. I then obtain the extensive term as a residual; that is, as the difference between the logarithmic growth of top wealth shares and the intensive term:[4]

$$\text{Extensive term}_t \equiv \log\left(\frac{W_{p,t}}{W_{p,t-1}}\right) - \text{Intensive term}_t.$$

I then estimate the baseline specification (1) after replacing the dependent variable with each of these two terms; that is,

$$\begin{aligned}\sum_{t\le s\le t+h} \text{Intensive term}_s - (h+1)\log R_{f,t} &= \alpha_{p,h}^{\text{intensive}} + \beta_{p,h}^{\text{intensive}}(\log R_{M,t} - \log R_{f,t}) + \epsilon_{p,t+h}^{\text{intensive}},\\ \sum_{t\le s\le t+h} \text{Extensive term}_s &= \alpha_{p,h}^{\text{extensive}} + \beta_{p,h}^{\text{extensive}}(\log R_{M,t} - \log R_{f,t}) + \epsilon_{p,t+h}^{\text{extensive}}.\end{aligned} \tag{A3}$$

Note that because these regressions are univariate, the "intensive" and "extensive" coefficients exactly sum up to the coefficients obtained for the total growth in the average wealth in Forbes 400; that is, $\beta_{p,h}^{\text{intensive}} + \beta_{p,h}^{\text{extensive}} = \beta_{p,h}$.

Figure A3 plots the resulting estimates for $\beta_{p,h}^{\text{intensive}}$ and $\beta_{p,h}^{\text{extensive}}$ for $0 \le h \le 8$ as well as their 95% confidence intervals. I find that almost all of the response in the growth of the average wealth in the top 400 is due to change in the wealth of agents initially in the top ("intensive" term), rather than changes in composition effects ("extensive" term). Note that this is consistent with the model discussed in Section 3, in which the larger response of the average wealth in top percentiles to stock market returns is driven by the larger wealth exposure of individuals in the top percentile.

C Appendix for Section 3

C.1 Solving the model

I now describe how I solve the model step by step by rewriting the equilibrium as an ODE on χ.

Step 1. I first express p and its derivatives in terms of χ and its derivatives. Market clearing for goods (15) gives

$$p(x) = \frac{1}{x\rho_E + (1-x)\rho^{\psi}\chi(x)^{1-\psi}}.$$

[4] Gomez (2023) further decomposes this extensive term into a between and demography terms, which correspond, respectively, to the effect of idiosyncratic shocks and demographic forces.

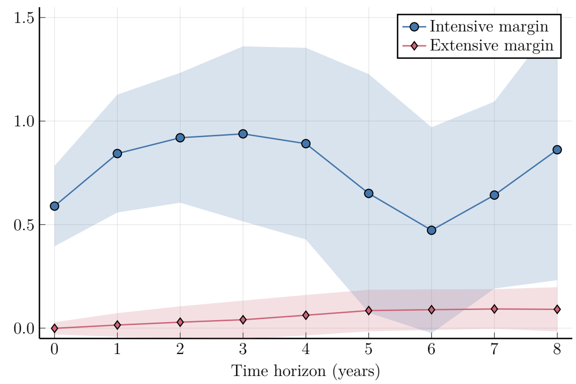


Figure A3: Decomposing the response of the average wealth in the top 400

Notes: The figure reports the estimates for $\beta^{\text{intensive}}_{p,h}$ and $\beta^{\text{extensive}}_{p,h}$ from the regression model (A3) for $0 \leq h \leq 8$ as well as their 5%–95% confidence intervals using heteroskedasticity consistent standard errors. At each horizon $0 \leq h \leq 8$, the intensive and extensive estimates sum up exactly to the coefficients plotted in the last panel of Figure 1; that is, $\beta^{\text{intensive}}_{p,h} + \beta^{\text{extensive}}_{p,h} = \beta_{p,h}$.

Differentiating with respect to x gives

$$\partial_x p(x) = -p(x)^2 \left(\rho_E - \rho^\psi \chi(x)^{1-\psi} + (1-x)\rho^\psi(1-\psi)\chi(x)^{-\psi}\partial_x\chi(x)\right).$$

Differentiating a second time gives

$$\partial_{xx} p(x) = p(x)^2 \Bigg(2p(x)\left(\rho_E - \rho^\psi \chi(x)^{1-\psi} + (1-x)\rho^\psi(1-\psi)\chi(x)^{-\psi}\partial_x\chi(x)\right)^2 + 2\rho^\psi(1-\psi)\chi(x)^{-\psi}\partial_x\chi(x) \\ +(1-x)\rho^\psi(1-\psi)p(x)\left(\chi(x)^{-\psi}\partial_{xx}\chi(x) - \psi\chi(x)^{-\psi-1}(\partial_x\chi(x))^2\right)\Bigg).$$

Step 2. I then express the volatility of the state variable in terms of the p and its derivatives. Combining Ito's lemma with Proposition 1 gives

$$\sigma_x(x) = x(\alpha_E(x)-1)\left(\sigma + \frac{\partial_x p(x)}{p(x)}\sigma_x(x)\right)$$
$$\implies \sigma_x(x) = \frac{x(\alpha_E(x)-1)\sigma}{1 - x(\alpha_E(x)-1)\frac{\partial_x p(x)}{p(x)}}.$$

where $\alpha_E(x) = \min\left(\alpha_E, \frac{1}{x}\right)$. This equation reflects the feedback loop discussed in Section 4.3. This equation allows me to pin down the volatility of χ and p using Ito's lemma:

$$\sigma_\chi(x) = \frac{\partial_x\chi(x)}{\chi(x)}\sigma_x(x)$$
$$\sigma_p(x) = \frac{\partial_x p(x)}{p(x)}\sigma_x(x).$$

Moreover, market clearing for the risky asset (16) gives an expression for $(\mu_R - r)(x)$:

$$1 = x_t \alpha_E(x) + (1-x)\left(\frac{1}{\gamma}\frac{(\mu_R - r)(x)}{\sigma_R(x)^2} + \frac{1-\gamma}{\gamma}\frac{\sigma_\chi(x)}{\sigma_R(x)}\right)$$
$$\implies (\mu_R - r)(x) = \frac{1 - x\alpha_E(x)}{1-x}\gamma\sigma_R(x)^2 + (\gamma - 1)\sigma_\chi(x)\sigma_R(x), \tag{A4}$$

where $\sigma_R(x) = \sigma + \sigma_p(x)$.

Step 3. I then obtain the drift of the state variable $\mu_x(x)$ from Proposition 1, which allows me to obtain the drift of χ and p using Ito's lemma:

$$\mu_\chi(x) = \frac{\partial_x \chi(x)}{\chi(x)}\mu_x(x) + \frac{1}{2}\frac{\partial_{xx}\chi(x)}{\chi(x)}\sigma_x(x)^2,$$
$$\mu_p(x) = \frac{\partial_x p(x)}{p(x)}\mu_x(x) + \frac{1}{2}\frac{\partial_{xx}p(x)}{p(x)}\sigma_x(x)^2.$$

Finally, subtracting the expression for $(\mu_R - r)(x)$ in (A4) from the definition of $\mu_R(x)$ (5) gives an expression for $r(x)$.

Step 4. Plugging these quantities into the household's HJB equation (5) gives the ODE for the function χ. I solve the ODE using an accelerated finite difference method. Formally, I solve for $\boldsymbol{\chi} = [\chi_1, \ldots, \chi_N]$, a vector of length N corresponding to the value of the function χ on a discretized grid between 0 and 1.

Denote $F(\boldsymbol{\chi})$ the finite difference scheme corresponding to a model, where the solution satisfies $F(\boldsymbol{\chi}) = 0$. I solve for $\boldsymbol{\chi}$ using an iteration method. I start from an initial guess $\boldsymbol{\chi}_0 = [1, \ldots, 1]$ and then iterates using the equation:

$$0 = F(\boldsymbol{\chi}_{i+1}) - \frac{\boldsymbol{\chi}_{i+1} - \boldsymbol{\chi}_i}{\Delta}. \tag{A5}$$

Each update requires solving a non-linear equation (it corresponds to a fully implicit Euler method). Economically, each update can be thought of as solving for the value function today given the value function in Δ time. I solve this non-linear equation using a Newton-Raphson method. The Newton-Raphson method converges if the initial guess is close to the solution. Since $\boldsymbol{\chi}_i$ converges towards $\boldsymbol{\chi}_{i+1}$ as Δ tends to zero, one can always choose Δ low enough so that the inner steps converge. Therefore, I adjust Δ as follows. If the inner iteration does not converge, I decrease Δ. If the inner iteration converges, I increase Δ. After a few successful implicit time steps, Δ is large, making the algorithm like Newton-Raphson. In particular, the convergence is quadratic around the solution. I stop the iteration when $F(\boldsymbol{\chi}_i)$ is small enough.

This method corresponds to a method used in the fluid dynamics literature called the Pseudo-Transient Continuation method. The algorithm with only one inner iteration and Δ constant corresponds to Achdou et al. (2022) (it corresponds to a semi-implicit Euler method). Allowing multiple inner iterations and adjusting Δ dynamically are important to ensure convergence of this non-linear PDE. This solution method is helpful for solving other asset pricing models globally. I uploaded it as an online package for Julia `https://github.com/matthieugomez/EconPDEs.jl`

C.2 Cross-sectional moments of wealth above a threshold

The following proposition characterizes the $\xi-$th cross-sectional moment of wealth above some threshold.

Proposition A1. *For any wealth threshold $q \in \mathbb{R}$ and exponent $\xi \in [0, \zeta_j)$, we have*

$$\mathbb{E}_t\left[w_{it}^{\xi} 1_{w_{it} \geq q} | i \in \mathbb{I}_t\right] = \sum_{j \in \{E,H\}} \pi_j \int_{-\infty}^{t} (\eta + \delta) e^{-(\eta+\delta)(t-s)} e^{\xi \mu_{j,s\to t} + \frac{1}{2}\xi^2 \nu_{j,s\to t}^2} \overline{\Phi}\left(\frac{\log q - \mu_{j,s\to t}}{\nu_{j,s\to t}} - \xi \nu_{j,s\to t}\right) \mathrm{d}s.$$

where $\mu_{j,s\to t}$ and $\nu_{j,s\to t}$ are defined in Proposition 2 and $\overline{\Phi}(\cdot) = 1 - \overline{\Phi}$ denotes the counter-CDF of a standard normal variable.

The special case $\xi = 0$ characterizes the mass of individuals above a certain threshold (i.e., the cumulative distribution function given in Proposition 2). In contrast, the special case $\xi = 1$ characterizes the total wealth owned by individuals above a certain threshold (i.e., the top wealth share). Another interesting special case is the limit $q \to -\infty$, in which case this proposition characterizes the cross-sectional moment of order ξ of the wealth distribution (or, equivalently, the Laplace transform of the distribution of log wealth).

Proof. The law of iterated expectations gives

$$\mathbb{E}_t\left[w_{it}^{\xi} 1_{w_{it} \geq q} | i \in \mathbb{I}_t\right] = \sum_{j \in \{E,H\}} \pi_j \mathbb{E}_t\left[w_{it}^{\xi} 1_{w_{it} \geq q} | i \in \mathbb{I}_{jt}\right].$$

In turn, the ξ-th moment within each type can be expressed as

$$\mathbb{E}_t\left[w_{it}^{\xi} 1_{w_{it} \geq q} | i \in \mathbb{I}_{jt}\right] = \int_{-\infty}^{t} (\eta + \delta) e^{-(\eta+\delta)(t-s)} \mathbb{E}_t\left[w_{it}^{\xi} 1_{w_{it} \geq q} | a_{it} = t - s, i \in \mathbb{I}_{jt}\right] \mathrm{d}s,$$

where a_{it} denotes the age of individual i at time t. We know from the proof of Proposition 2 that, within the cohort born at time $s \leq t$, log wealth is normally distributed with mean $\mu_{j,s\to t}$ and standard deviation $\nu_{j,s\to t}$. Hence, applying Lemma A2 gives

$$\mathbb{E}_t\left[w_{it}^{\xi} 1_{w_{it} \geq q} | a_{it} = t - s, i \in \mathbb{I}_{jt}\right] = e^{\xi \mu_{j,s\to t} + \frac{1}{2}\xi^2 \nu_{j,s\to t}^2} \overline{\Phi}\left(\frac{\log q - \mu_{j,s\to t}}{\nu_{j,s\to t}} - \xi \nu_{j,s\to t}\right).$$

Combining the previous three equations gives the result. □

C.3 Distinguishing between labor and capital income

All income in the model is produced by trees. As a result, the concept of wealth in the model (the capitalized value of all income produced by trees) encompasses both financial wealth (the capitalized value of capital income) and human capital (the capitalized value of labor income). In the data, however, we only observe financial wealth (as human capital is not traded). I now argue that the distinction between the two does not matter for two key moments used in Section 4 to calibrate the model: the elasticity of top percentiles to stock market returns and the tail index of the wealth distribution. I examine this point in two contexts: firstly, about the actual wealth distribution observed in the U.S., and secondly, within the framework of the model.

In the data. Let $A_{p,t}$ represent the average financial wealth and $H_{p,t}$ the average human capital in the top percentile p at time t. Denote $\omega_p \equiv \mathrm{E}[H_{p,t}/(A_{p,t} + H_{p,t})]$ the average ratio of human capital to total wealth in the top percentile p. The growth of total wealth between two periods can be written as a weighted average of the growth of financial assets and the growth of human capital:

$$\log\left(\frac{A_{t+1}+H_{t+1}}{A_t+H_t}\right) \approx \omega_p \log\left(\frac{A_{p,t+1}}{A_{p,t}}\right) + (1-\omega_p)\log\left(\frac{H_{p,t+1}}{H_{p,t}}\right). \tag{A6}$$

Projecting this approximation on stock returns, as in Equation (1) in the main text, gives

$$\begin{aligned} \beta_{A+H,p} &\approx (1-\omega_p)\beta_{A,p} + \omega_p \beta_{H,p} \\ \implies \beta_{A+H,p} - \beta_{A,p} &\approx \omega_p(\beta_{H,p} - \beta_{A,p}). \end{aligned}$$

This equation says that the difference between the exposure of total wealth, $\beta_{A+H,p}$, and the exposure of financial wealth, $\beta_{A,p}$ (i.e., the bias in inferring the exposure of total wealth from the exposure of financial wealth) is the product of (i) the share of human capital in total wealth ω_p (ii) the difference between the exposure of human capital and financial wealth $\beta_{H,p} - \beta_{A,p}$.

This equation suggests that the difference between the exposure of total wealth and the financial wealth $\beta_{A+H,p} - \beta_{A,p}$ is likely to be small for agents at the top of the wealth distribution (e.g., $p = 0.01\%$) since $\lim_{p\to 0}\omega_p = 0$. For instance, the IRS reports that labor income represents 13.2% of total income for the top 400 tax returns in the U.S. on average from 1992 to 2014 (see Appendix Table A5). Assuming the same capitalization rate for human capital and financial assets, this suggests that human capital represents one-tenth of total wealth for agents at the top of the wealth distribution; that is $\omega_p \approx 13.2\%$.[5]

This equation also suggests that the difference between the exposure of total wealth and the financial wealth $\beta_{A+H,p} - \beta_{A,p}$ is likely to be small for the average household in the economy ($p = 100\%$) as $\beta_{H,p} \approx \beta_{A,p}$. Indeed, at the aggregate level, labor and capital income are co-integrated, which implies that their permanent response to aggregate shocks must be equal. In conclusion, this discussion suggests that the risk exposure of the *financial* wealth of agents in percentile p is a good approximation for the exposure of their *total* wealth for p close to zero or p close to one.

Similarly, the distinction between "total wealth" and "financial wealth" does not matter for the tail index of the wealth distribution either, as, empirically, most of the wealth in the top takes the form of observable financial wealth (e.g., firm ownership) rather than unobserved human capital.

In the model. I now present a simple way to incorporate the distinction between labor and capital income in the model. I derive a condition under which the wealth of agents in the right tail of the wealth distribution takes the form of financial wealth rather than human capital (as in the data). Under this condition, the distinction between human capital and financial wealth does not matter for the elasticity of top percentiles to stock market returns and the tail index of the wealth distribution, as in the data.

Formally, I assume that a portion χ of initial wealth endowed to an individual at birth takes the form of human capital (i.e., trees giving labor income). I assume that this income grows at rate $\delta - \phi$ relative to the economy and disappears when the individual dies, so that this income, on average, depreciates at rate ϕ relative to the economy. Hence, capital and labor income grow at the same rate on average. As a result, all trees have the same market value-to-income ratio. All the equations in the model are unchanged, as the consumption and portfolio decisions only depend on total wealth. In particular, this distinction between

[5] Similarly, Appendix Table A4 reports that the share of labor income in the top 0.01% is 14% using data from the Survey of Consumer Finances.

labor and capital income does not affect asset prices.

What does the distribution of financial wealth look like in this model? As shown in Proposition 3, the distribution of "total" wealth is Pareto with tail index $\min(\zeta_H, \zeta_E)$. In contrast, one can show that human capital is distributed with a Pareto tail with tail index $(\delta+\eta)/(\delta-\phi)$ if $\delta > \phi$, or $+\infty$ $\delta < \phi$.[6] As a result, financial wealth, which is the difference between total wealth and human capital, inherits the Pareto tail of total wealth as long as $\min(\zeta_H, \zeta_E) \leq (\delta+\eta)/(\delta-\phi)$. This condition is satisfied whenever the growth rate of the type of agents making it to the right tail of the wealth distribution is higher than the growth rate of their labor income. When this condition holds (which is the case in the calibrated version of the model), the ratio of human capital to total wealth tends to zero in the right tail of the wealth distribution. So, as in the data, the distinction between total wealth and financial wealth does not matter for the tail index of the wealth distribution or for the elasticity of top wealth shares to stock market returns.

D Appendix for Section 4

D.1 Additional evidence on the consumption rate at the top

In the main text, I pick the consumption rate of entrepreneurs ρ_E to match the tail index of the wealth distribution ζ. As discussed in Section 4.1, this calibration can be decomposed into two steps. In the first step, I use the formula for the tail index given obtained in Proposition 3, together with the values of $(\zeta, \nu, \delta, \eta)$, to infer a value for the average growth rate of households in top percentiles relative to the economy. In the second step, I use the fact that this latter quantity equals the average log return of entrepreneurs minus their consumption rates minus the average log growth rate of the average wealth in the economy. Estimating separately the log return of entrepreneurs and the growth rate of the economy allows me to obtain an implied value for the consumption rate of entrepreneurs (as a residual).

I now discuss an alternative calibration that focuses on estimating the consumption rate of entrepreneurs in Forbes 400. The key advantage of focusing on Forbes 400 is that, due to its panel dimension, I can *directly* measure the growth rate of households initially in the top 400 (instead of backing it out from the tail index of the wealth distribution ζ). I can then use data on asset returns (as well as some data on wages received and taxes paid by the Top 400) to obtain an implied value for the consumption of entrepreneurs as a fraction of their financial wealth. The disadvantage of this method is that the consumption rate of top entrepreneurs from 1982 to 2017 may not be representative of their average consumption rate over the 20th century (indeed, that time period coincides with a steep increase in top wealth inequality).

The advantage of this method, relative to the one in the main text, is that it is more direct, as one can directly measure the average wealth growth of top households relative to the economy using panel data. Its disadvantage, however, is that the data from Forbes 400 only covers a very particular time, where top wealth shares increase dramatically (which potentially reflects a steep decrease in the average consumption rate of top entrepreneurs over that period).

I now formalize this alternative methodology. I start from the following "model-free" budget constraint for the financial wealth of households in the top percentile between year t and year $t+1$

$$\overline{W}_{t+1} = \overline{R}_{t+1}\left(\overline{W}_t + \overline{Y}_{t+1} - \overline{T}_{t+1} - \overline{C}_{t+1}\right),$$

[6] To see why, note that human capital at time t of an agent with age a_{it} is $\chi e^{(\delta-\phi)a_{it}} w_{i,t-a_{it}}$. Because age is exponentially distributed with rate parameter $\eta+\delta$, $e^{(\delta-\phi)a_{it}}$ is Pareto distributed with tail index $(\eta+\delta)/(\delta-\phi)$.

where $\overline{W}_t$ (resp. $\overline{W}_{t+1}$) denotes the average financial wealth of these households at the end of year t (resp. at the end of year $t+1$), $\overline{R}_{t+1}$ denotes their wealth-weighted average portfolio returns in year $t+1$, $\overline{Y}_{t+1}$ their average labor income, $\overline{T}_{t+1}$ their average taxes and $\overline{C}_{t+1}$ their average consumption. Taking logs and rearranging:

$$\begin{aligned}\log\left(\frac{\overline{W}_{t+1}}{\overline{W}_t}\right) &= \log \overline{R}_{t+1} + \log\left(1 + \frac{\overline{Y}_{t+1} - \overline{T}_{t+1} - \overline{C}_{t+1}}{\overline{W}_t}\right) \\ &\approx \log \overline{R}_{t+1} + \frac{\overline{Y}_{t+1}}{\overline{W}_t} - \frac{\overline{T}_{t+1}}{\overline{W}_t} - \frac{\overline{C}_{t+1}}{\overline{W}_t}. \end{aligned} \tag{A7}$$

where the second line is valid at the first-order in $(\overline{Y}_{t+1} - \overline{T}_{t+1} - \overline{C}_{t+1})/\overline{W}_t$ (which is indeed small at the annual frequency). I then estimate the log average return $\log \overline{R}_{t+1}$ as

$$\log \overline{R}_{t+1} \equiv \log R_{f,t+1} + \beta_{Top400}\left(\log R_{M,t+1} - \log R_{f,t+1}\right),$$

where $\beta_{Top400} = 0.98$ was estimated via (1). I estimate the average labor income and taxes paid using IRS tabulations on the top 400 income tax returns (ranked by Adjustable Gross Income) from 1992 to 2014.[7] I then obtain the consumption rate $\overline{C}_{t+1}/\overline{W}_t$ as a residual.

Table A5 reports the results. The average annual wealth growth of households in Forbes 400 over the 1982-2017 time period is 5.1% (in real terms); their average log annual return is 7.6% (in real terms); their average labor income represents 0.7% of their financial wealth; while their average taxes represent −1.2% of their financial wealth. Using Equation (A7), this implies that their annual consumption represents approximately 2% of their financial wealth. Interestingly, this value ends up being very similar to the value of ρ_E required to match the tail index of the wealth distribution in Section 4.1, which gave $\rho_E = 2.2\%$.

Table A5: Estimating the annual consumption rate in the Top 400

Average wealth growth	= Portfolio return	+ Labor income	− Taxes	− Consumption
5.1	7.6	0.7	–1.2	–2.0

Notes: The table reports the average, over the 1982-2017 period, of each term given in Equation (A7) (all in percentage term). More precisely, the first column corresponds to $\log\left(\overline{W}_{t+1}/\overline{W}_t\right)$, the second column corresponds to $\log \overline{R}_{t+1}$, the third column to $\overline{Y}_{t+1}/\overline{W}_t$, the fourth column to $-\overline{T}_{t+1}/\overline{W}_t$, and the last column to the residual $-\overline{C}_{t+1}/\overline{W}_t$.

D.2 Equilibrium quantities in the calibrated model

Table A6 reports the average level of top wealth shares in the data and in the calibrated model. One can see that the model matches the average level of top wealth shares across top percentiles very well. Note, however, that the model slightly overestimates the share of wealth owned by the top 400 (remember that the top 400 corresponds to the top 0.03% of the top 0.01%, in the model as in the data), which reflects that the tail index of the wealth distribution is slightly smaller in the model (1.42) relative to the data (1.5).

Figure A4 compares the estimates from local projections of the average wealth in a top percentile onto the excess stock market returns, contrasting model simulations with empirical data. The "data" estimates correspond exactly to the ones plotted in Figure 1 in the main text. To facilitate the comparison between the

[7] See https://www.irs.gov/pub/irs-soi/14intop400.pdf.

Table A6: Average level of top wealth shares in the model versus the data

	Top 1%	Top 0.1%	Top 0.01%	Top 400
Data	0.328	0.139	0.055	0.009
Model	0.318	0.135	0.062	0.022

Notes: The table reports the average level of top wealth shares in the data (using Saez and Zucman (2016) series from 1913 to 2020 and Forbes magazine from 1982 to 2017) and in the model (using simulated data).

model and the data, the "model" estimates are obtained by running the specification (1) on simulated data from the calibrated model, using the same number of years as in the data and averaging across simulations.[8] Moreover, consistently with the construction of top wealth shares in the data, I construct the average wealth in a top percentile p in a given year t as the average between the wealth at the end of year $t-1$ and the wealth at the end of year t. As a result, in the simulations, as in the data, the effect of the stock market return peaks at the one-year horizon.

As a complement to the impulse response functions plotted in the main text (Figure 2), Figure A6 reports important economic quantities as a function of the state variable x. More precisely, Figure A6a plots the drift and volatility of the state variable, x, and its associated stationary density represented as a shaded area. Figure A6b plots the price-to-income ratio $p(x)$ and the wealth-to-consumption ratio $c_H(x)$ of households. Finally, Figure A6c plots the risk-free rate and the expected log stock market return (i.e., the return on levered equity).

Figure A7 plots the decomposition for $\sigma_p(x)$ in terms of a "risk-free rate channel" and an "excess return channel" defined in Proposition 5 as a function of x.

D.3 Decay rate of infinitesimal impulse response functions

As discussed in Section 4.2, the law of motion of the stochastic derivative of the process $(x_t)_{t\in\mathbb{R}}$ is

$$\left(\mathrm{d}\frac{\partial x_{t+h}}{\partial x_t}\right) \Big/ \left(\frac{\partial x_{t+h}}{\partial x_t}\right) = \partial_x \mu_x(x_{t+h})\,\mathrm{d}h + \partial_x \sigma_x(x_{t+h})\,\mathrm{d}Z_{t+h}.$$

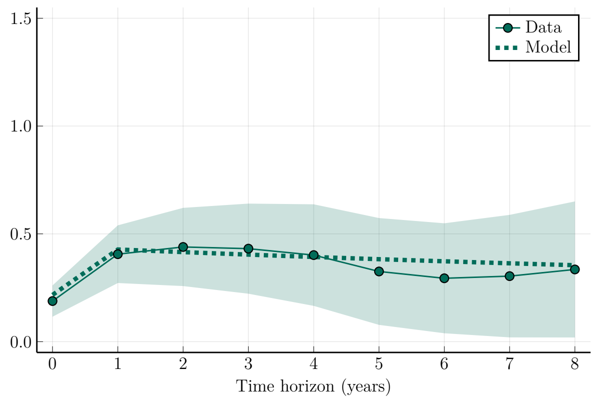
Using the terminology of Hansen and Scheinkman (2009), the stochastic derivative of the process $(x_t)_{t\in\mathbb{R}}$ is a multiplicative functional, and, as such, it can be written as the product of three terms:

$$\frac{\partial x_t}{\partial x_0} = e^{-\kappa t}\frac{\phi(x_0)}{\phi(x_t)}\hat{M}_t, \tag{A8}$$

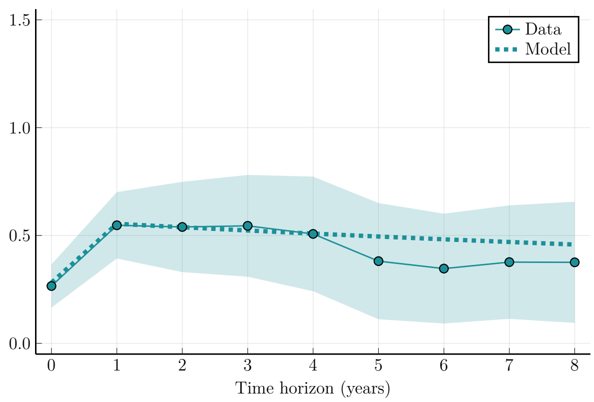
where $\hat{M}_t$ is a local martingale and $(\phi, -\kappa)$ denote, respectively, the principal eigenvector and eigenvalue of the operator:

$$\mathbb{T} : g \to \lim_{h\to 0}\frac{1}{h}\left(\mathrm{E}\left[\frac{\partial x_h}{\partial x_0} g(x_h)|x_0 = x\right] - g(x)\right). \tag{A9}$$

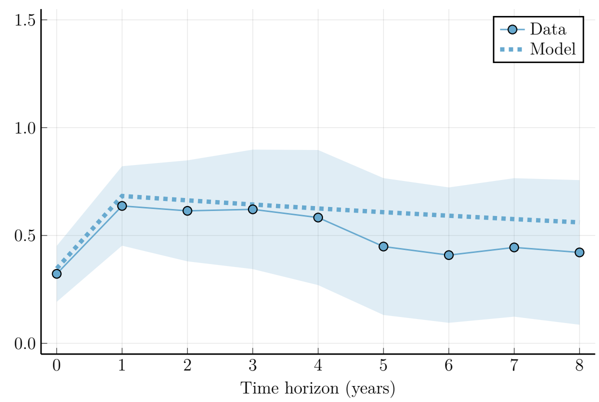
[8] This is to adjust for the finite sample bias of local projections, which is stressed by Herbst and Johannsen (2021). That being said, I obtain very similar estimates after running local projections on a very long sample, which indicates that this bias is small in the calibrated model.



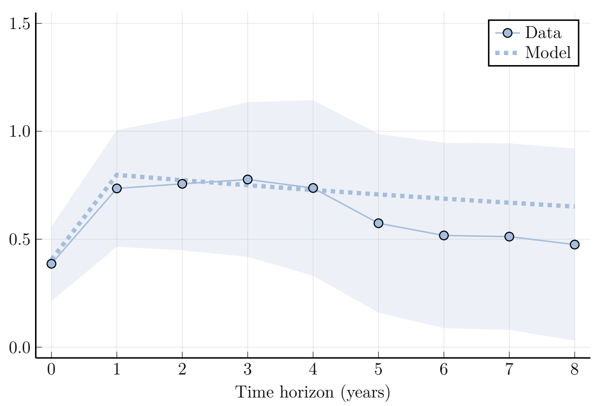
(a) Top 100%



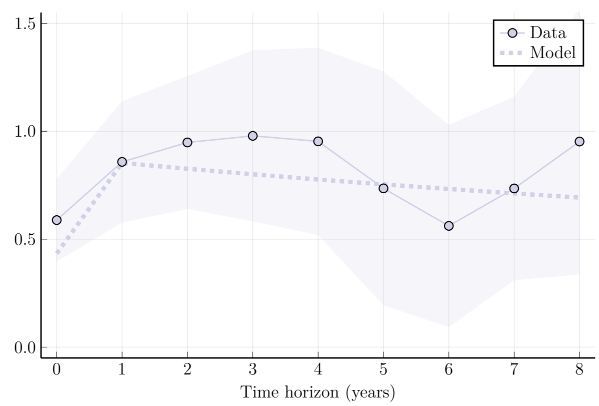
(b) Top 1%



(c) Top 0.1%



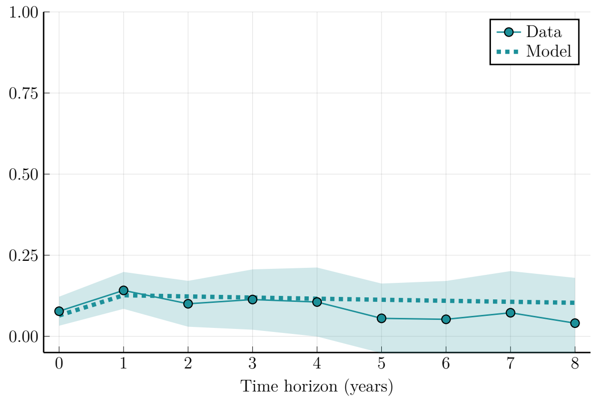
(d) Top 0.01%



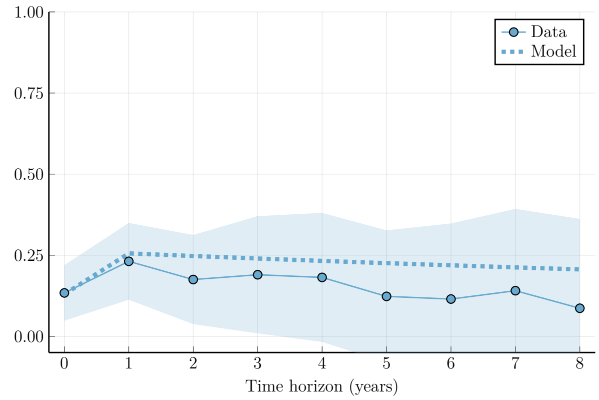
(e) Top 400

Figure A4: Response of wealth in top percentiles to excess stock returns in the model versus the data

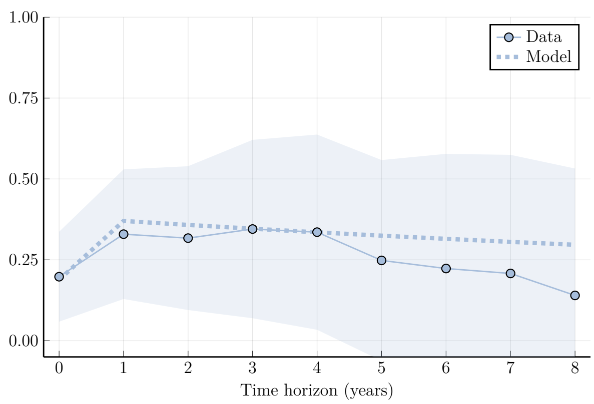
Notes: The figure reports the estimates for $\beta_{p,h}$ estimated via the regression (1) for $0 \leq h \leq 8$, as well as their 5%–95% confidence intervals using heteroskedasticity consistent standard errors. Each figure corresponds to a different top percentile. Figure A4a corresponds to all U.S. households ($p = 100\%$). Figures A4b-A4d correspond to the top 1%, 0.1%, 0.01% using data from Saez and Zucman (2016). Figure A4e corresponds to Forbes 400. The dotted lines represent the estimates obtained after running the same regressions on simulated data from the calibrated model. More precisely, I run the regression on subsamples with the same number of years as in the data ($T = 105$), and I report the average of these estimates across subsamples.



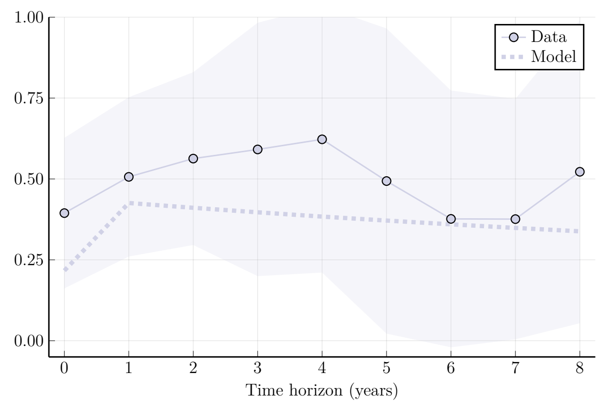
(a) Top 1%



(b) Top 0.1%



(c) Top 0.01%



(d) Top 400

Figure A5: Response of top wealth shares to excess stock returns in the model versus the data

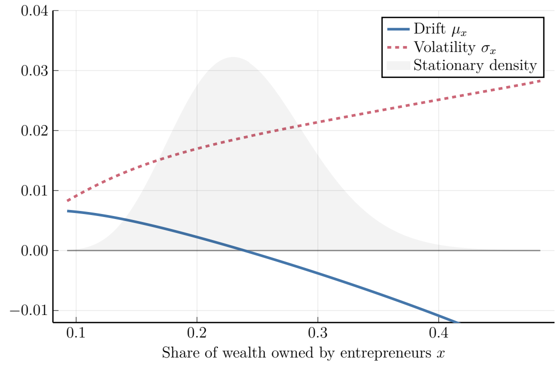
Notes: The figure reports the estimates for $b_{p,h}$ estimated via the regression (2) for $0 \leq h \leq 8$ as well as their 5%–95% confidence intervals using heteroskedasticity consistent standard errors. Each figure corresponds to a different top percentile. Figures A5a-A5c correspond to the top 1%, 0.1%, 0.01% using data from Saez and Zucman (2016). Figure A5d corresponds to Forbes 400. The dotted lines represent the estimates obtained after running the same regressions on simulated data from the calibrated model. More precisely, I run the regression on subsamples with the same number of years as in the data ($T = 105$), and I report the average of these estimates across subsamples.

Plugging the decomposition (A8) into the definition for the IIRF (Section 4.2) gives:

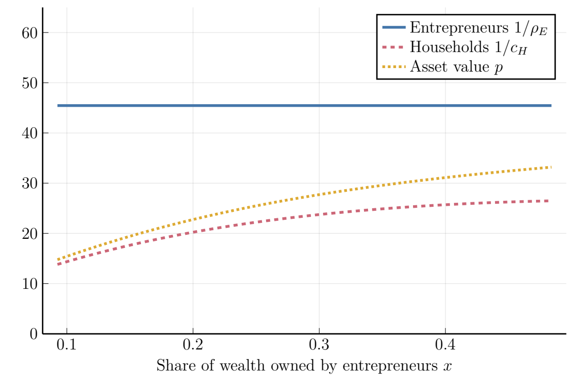
$$\text{IIRF}_g(x,h) = \mathrm{E}\left[\frac{\partial x_h}{\partial x_0}\partial_x g(x_h)\Big|x_0 = x\right] = e^{-\kappa h}\phi(x_0)\mathrm{E}\left[\hat{M}_h\frac{\partial_x g(x_h)}{\phi(x_h)}\Big|x_0 = x\right].$$

This equation implies that $\text{IIRF}_g(x,h)$ decays with the horizon h at an exponential rate κ. Note that this decay rate does not depend on the initial state x nor on the function $g(\cdot)$.[9] In the calibrated model, computing numerically κ as the (opposite) of the principal eigenvalue of the operator $\mathbb{T}$ defined in (A8) yields $\kappa \approx 0.06$.

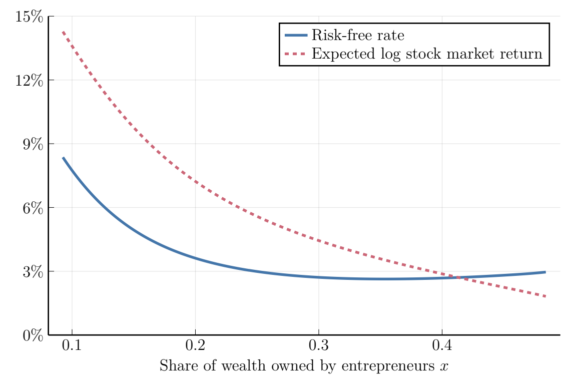
[9]One could show that κ also corresponds to the second largest eigenvalue of the infinitesimal generator associated with the process $(x_t)_{t\in\mathbb{R}}$.



(a) Drift and volatility of the state variable



(b) Wealth-to-consumption ratios



(c) Expected asset returns

Figure A6: Economic quantities across the state space

Notes: This figure plots equilibrium quantities in the model in terms of the share of aggregate wealth owned by entrepreneurs, x. Expected log stock market return corresponds to the expected stock market return of levered equity; that is, $r + \lambda(\mu_R - r) - \frac{1}{2}\lambda^2\sigma_R^2$ (see Section 4.1). The bounds of the x-axis correspond to the 0.01% and the 99.99% quantiles of the state variable.

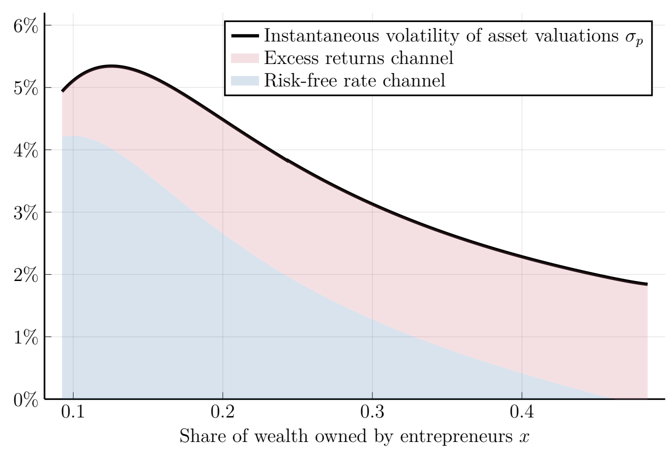


Figure A7: Decomposing the impulse response of asset valuations to aggregate shocks

Notes: The figure plots the decomposition of the instantaneous volatility of asset valuations, $\sigma_p(x)$, as well as its decomposition into a "risk-free rate channel" and an "expected excess return channel" (Proposition 5). The bounds of the x-axis correspond to the 0.01% and the 99.99% quantiles of the state variable.

D.4 Comparison with the asset pricing literature

I now emphasize two differences between my model and the existing asset pricing literature with heterogeneous agents.

Financial friction. The key financial friction in my model is that the equity share of entrepreneurs must remain constant over time. This friction differs from the financial acceleration literature, which typically assumes that entrepreneurs must collectively hold a constant fraction of the corporate sector. This alternative approach implies that the equity share of entrepreneurs is counter-cyclical (i.e., that entrepreneurs must hold a larger fraction of their wealth in equity in bad times when they account for a smaller fraction of overall wealth).

The difference between the two approaches matters for asset prices; in particular, the approach taken by the financial acceleration literature tends to imply more significant fluctuations in the equity premium. One striking instance is that the equity premium converges to infinity as the share of wealth owned by entrepreneurs converges to zero in this latter approach, as entrepreneurs must accept to hold an arbitrarily large share of wealth invested in equity (much higher than one).[10] In contrast, in my model, as the share of wealth owned by entrepreneurs converges to zero, the equity premium converges to the one that would obtain in an economy without entrepreneurs; that is, $\gamma\sigma^2$.

Which financial friction is more consistent with the micro-data? To answer this question, I test whether the wealth exposure of top households to stock market returns depends on previous stock market returns. Formally, I run the same regression as in Section 1, except that I now allow the wealth exposure of top households to vary with past stock market returns:

$$\begin{aligned}\log\left(\frac{W_{p,t+3}}{W_{p,t-1}}\right) - (h+1)\log R_{f,t} = \alpha_p &+ \beta_p(\log R_{M,t} - \log R_{f,t}) \\ &+ \delta_p(\log R_{M,t-1} - \log R_{f,t-1})(\log R_{M,t} - \log R_{f,t}) \\ &+ \gamma_p(\log R_{M,t-1} - \log R_{f,t-1}) \\ &+ \epsilon_{p,t+3},\end{aligned} \tag{A10}$$

A positive (resp. negative) coefficient δ_p for the interaction term between current returns and past returns would reveal a pro-cyclical (resp. counter-cyclical) exposure of the average wealth in top percentiles to stock market returns. Table A7 reports the results. I find none of the estimates for $\delta_{p,h}$ are statistically different from zero at the 10% level; if anything, the coefficients are positive, suggesting that top households' wealth exposure is pro-cyclical rather than counter-cyclical. Overall, while the standard errors of the estimates are substantial, the micro-data seem to reject the assumption of counter-cyclical leverage for top households.[11]

Role of wealth inequality moments. The second difference between my approach and the existing literature is that I build and calibrate the model to match the level and dynamics of the wealth distribution. Ensuring that the model matches moments related to the wealth distributions disciplines the degree

[10] In turn, market clearing for goods then typically requires the interest rate to converge to minus infinity to compensate for the rise in equity premia (see He and Krishnamurthy, 2013 for an example).Brunnermeier and Sannikov, 2014 avoids this by allowing non-entrepreneurs to own equity directly at the cost of lower capital productivity.

[11] Relatedly, Adrian and Shin (2010) also present some evidence that the leverage of financial intermediaries is procyclical.

Table A7: Testing the cyclicality of the wealth exposure of top percentiles

	Top 100%	Top 1%	Top 0.1%	Top 0.01%
	(1)	(2)	(3)	(4)
Excess returns	0.42***	0.55***	0.64***	0.79***
	(0.12)	(0.14)	(0.16)	(0.21)
Excess returns × Excess returns (lagged)	−0.09	0.11	0.32	0.28
	(0.54)	(0.66)	(0.78)	(0.94)
Excess returns (lagged)	0.13	0.16	0.17	0.25
	(0.10)	(0.12)	(0.15)	(0.18)
Constant	0.04**	0.03	0.03	0.04
	(0.02)	(0.02)	(0.02)	(0.03)
Adjusted R^2	0.16	0.21	0.19	0.19
Time sample	1914-2016	1914-2016	1914-2016	1914-2016
N	103	103	103	103

Notes: This table reports the coefficients obtained in the regression of the growth of the average wealth in top percentiles on excess stock returns and their interaction with lagged excess stock market returns; that is, equation (A10). Estimation is done via OLS. Standard errors are in parentheses and are estimated using heteroskedasticity consistent standard errors. *,** ,*** indicate significance at the 10%, 5%, 1% levels, respectively.

of heterogeneity between agents, which, in turn, disciplines the extent to which asset valuations respond to aggregate shocks. Intuitively, the lower the level or the volatility of top wealth inequality, the lower the degree of heterogeneity across agents, and, as a result, the lower the volatility of asset returns in equilibrium. I now quantify the impact of wealth inequality moments on asset price moments.

As discussed in Section 4.1, the consumption rate of entrepreneurs ρ_E is calibrated to match the tail index ζ, and its derivative is given by $\partial \rho_E / \partial \zeta = \alpha_E^2 \nu^2 / 2 - (\delta + \eta) / \zeta^2 \approx 0.038$. Similarly, the equity share of entrepreneurs α_E is determined by the average wealth exposure of top percentiles to stock market returns, with a derivative given by $\partial \alpha_E / \partial \beta_{0.01\%} = \lambda = 2.3$. Next, Table A8 reports the sensitivity of asset price moments to model parameters. Combining these two numbers allows us to compute the derivative of asset price moments with respect to wealth inequality moments.

To take a specific example, suppose that the tail index of the wealth distribution was equal to 1.7 instead of 1.5 (i.e., a lower level of wealth inequality). Then, our calibration exercise would imply that the consumption rate of entrepreneurs ρ_E is higher by $0.038 \times (1.7 - 1.5) \approx 0.8\%$, leading to a drop in the standard deviation of asset returns in equilibrium by $-6.77 \times 0.8\% \approx -5.4\%$ (Table A8). Similarly, suppose that the stock market exposure for the average wealth in top percentiles was equal to 0.7 instead of 0.85 (i.e., a lower level of wealth inequality volatility). Then, our calibration exercise would imply a decrease in the equity share of entrepreneurs α_E by $2.3 \times (0.7 - 0.85) \approx -0.35$, leading to a drop in the standard deviation of asset returns in equilibrium by $-0.15 \times 0.35 \approx -5\%$ (Table A8). Overall, these two examples show that the values of wealth inequality moments play a critical role in shaping the dynamics of asset prices in the model.[12]

Existing asset pricing models with heterogeneous agents typically do not try to match the cross-sectional distribution of wealth across households. Still, for the sake of comparison, Table A9 reports the ratio between the wealth exposure of entrepreneurs and the average wealth exposure in the economy in several

[12] For transparency, this exercise does not take into account that, after changes in wealth inequality moments, the re-estimated values for ρ, γ, and ψ could change. When these values are allowed to change, one can still show that targeting a lower (resp. higher) level or volatility of wealth inequality generates a worse (resp. better) fit for the model.

Table A8: Sensitivity analysis

	Entrepreneurs' parameters			Households' parameters		
	ρ_E	α_E	π_E	ρ	γ	ψ
Average risk-free rate	1.00	0.02	0.05	0.15	−0.004	0.169
Standard deviation risk-free rate	−0.42	0.01	−0.01	0.08	−0.002	0.014
Average stock market return	1.09	−0.01	−0.16	0.03	0.001	0.046
Standard deviation stock market return	−6.77	0.15	0.00	0.92	−0.018	0.024

Notes: The table reports the derivative of each moment (rows) with respect to each parameter (columns) approximated using finite differences. More precisely, the table reports $\left(m_i(1.1 \times \theta_j) - m_i(0.9 \times \theta_j)\right) / (0.2 \times \theta_j)$ for each moment m_i and parameter θ_j.

leading asset pricing models. Relative to the reduced-form evidence, existing models typically overestimate the wealth exposure of top households relative to the average. An exception is Guvenen, 2009, who focuses on the heterogeneity between stockholders and non-stockholders in the U.S. data and therefore generates top households who are only slightly more levered than the average household. My model and Gârleanu and Panageas (2015) are the only one that produce a non-degenerate stationary distribution within agent types as they both use an overlapping generation setup. However, the latter model generates a wealth distribution with a tail index of 1.17, much lower than the data (that is, it generates too much wealth inequality). In conclusion, existing asset pricing models tend to imply a wealth distribution that moves too much and/or is too unequal relative to the data, which suggests that they imply too much heterogeneity in asset demands relative to the data.

Table A9: Comparing wealth inequality moments with existing asset pricing models

Model	Wealth exposure top households / average	Tail index
Data (& baseline model)	2.00	1.5
Representative agent model	1.00	X
Guvenen (2009)	1.15	X
Brunnermeier and Sannikov (2014)	2.96	X
He and Krishnamurthy (2013)	2.45	X
Gârleanu and Panageas (2015)	2.65	1.17
Di Tella (2017)	2.83	X

Notes: The table reports the average ratio between the wealth exposure of top households and the wealth exposure of the average household in the economy. It also reports the tail index of the wealth distribution when the implied wealth distribution by the model admits a right Pareto tail.

D.5 Decomposing the response of the average wealth of surviving entrepreneurs

I now combine the impulse response function for the wealth of surviving entrepreneurs $\epsilon(x,h)$ (Proposition 6) with the fact that $\sigma_{wE} = (\alpha_E - 1)(\sigma + \sigma_p)$ to re-express $\epsilon(x,h)$ as a sum of two terms capturing the effects of (i) change in asset income and (ii) changes in asset valuation:

$$\epsilon(x,h) = \underbrace{(\alpha_E - 1)\sigma}_{\substack{\text{Change} \\ \text{in asset income}}} + \underbrace{(\alpha_E - 1)\sigma_p(x) + \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0} \partial_x \left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)(x_t)\,\mathrm{d}t \middle| x_0 = x\right] \sigma_x(x)}_{\substack{\text{Changes} \\ \text{in asset valuation}}}. \tag{A11}$$

The following proposition further decomposes the second term in this decomposition: the effect due to changes in asset valuations.

Proposition A2. *The effect of aggregate shocks on the average wealth of surviving entrepreneurs can be written as follows:*

$$\begin{aligned}\epsilon(x,h) \approx \quad & \underbrace{(\alpha_E - 1)\sigma}_{\text{Change in asset income}} \quad + \underbrace{\mathrm{E}\left[(\alpha_E - 1)\frac{\partial x_h}{\partial x_0}\partial_x \log p(x_h)\Big| x_0 = x\right]\sigma_x(x)}_{\text{Change in asset valuation: Revaluation effect } > 0} \\ & \underbrace{+ \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\partial_x\left(r + \alpha_E\left(\frac{1}{p} - r_t\right) - \frac{1}{2}\alpha_E(\alpha_E - 1)\sigma_R^2\right)(x_t)\,\mathrm{d}t\Big| x_0 = x\right]\sigma_x(x).}_{\text{Change in asset valuation: Accumulation effect } < 0}\end{aligned}$$

This proposition decomposes the effect of asset valuations changes on entrepreneurs' normalized wealth into two terms. The "revaluation effect" corresponds to its (positive) effect on the market value of assets owned by entrepreneurs, while the "accumulation effect" corresponds to its (negative) effect on the amount of assets they accumulate. The fact that the accumulation term is negative reflects that higher asset valuations mean that entrepreneurs receive less income per unit of wealth. Similarly to Proposition 7, this proposition is obtained under the approximation that α_{Et} remains constant (or, equivalently, $x_t < 1/\alpha_E$), which holds more than 99.99% of the time in the calibrated model.

Proof of Proposition A2. Using the expression for μ_{wE} and σ_{wE} obtained in Section 3.3, we have:

$$\begin{aligned}&\mu_{wE} - \frac{1}{2}\sigma_{wE}^2 \\ &= r + \alpha_E\left(\frac{1}{p} + g - \phi - \frac{1}{2}\sigma^2 + \mu_p + \frac{1}{2}\sigma_p^2 - r\right) - \frac{1}{2}\alpha_E(\alpha_E - 1)\sigma_R^2 - \rho_E - \left(g - \frac{1}{2}\sigma^2 + \mu_P - \frac{1}{2}\sigma_p^2\right).\end{aligned}$$

Plugging this formula into the expression for $\epsilon(x,h)$ obtained in Proposition 6, we get:

$$\begin{aligned}\epsilon(x,h) &= (\alpha_E - 1)(\sigma + \sigma_p(x)) \\ &+ \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_t}\partial_x\left(r + \alpha_E\left(\frac{1}{p} - \phi - r\right) - \frac{1}{2}\alpha_E(\alpha_E - 1)\sigma_R^2\right)(x_t)\,\mathrm{d}t\Big| x_0 = x\right]\sigma_x(x) \\ &+ \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_t}\partial_x\left((\alpha_E - 1)\left(g - \frac{1}{2}\sigma^2 + \mu_p - \frac{1}{2}\sigma_p^2\right)\right)(x_t)\,\mathrm{d}t\Big| x_0 = x\right]\sigma_x(x).\end{aligned} \tag{A12}$$

Moreover, integrating forward $\mathrm{E}_t\left[\mathrm{d}\log p(x_t)\right] = \mu_p(x_t) - \frac{1}{2}\sigma_p(x_t)^2$ gives

$$\log p(x) = \mathrm{E}\left[\int_0^h\left(\mu_p - \frac{1}{2}\sigma_p^2\right)(x_t)\,\mathrm{d}t\Big| x_0 = x\right] + \mathrm{E}\left[\log p(x_h)|x_0 = x\right].$$

Differentiating with respect to x and multiplying by $\sigma_x(x)$ gives

$$\sigma_p(x) = \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\left(\mu_p - \frac{1}{2}\sigma_p^2\right)(x_t)\,\mathrm{d}t|x_0 = x\right]\sigma_x(x) + \mathrm{E}\left[\frac{\partial x_h}{\partial x_0}\partial_x \log p(x_h)|x_0 = x\right]\sigma_x(x).$$

Plugging this expression into (A12) gives

$$\begin{aligned}\epsilon(x,h) &= (\alpha_E(x)-1)\sigma + (\alpha_E(x)-1)\mathrm{E}\left[\frac{\partial x_h}{\partial x_0}\partial_x \log p(x_h)|x_0 = x\right]\sigma_x(x) \\ &+ \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\partial_x\left(r+\alpha_E\left(\frac{1}{p}-\phi-r\right)-\frac{1}{2}\alpha_E(\alpha_E-1)\sigma_R^2\right)(x_t)\,\mathrm{d}t|x_0=x\right]\sigma_x(x) \\ &+ \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\partial_x\left((\alpha_E-1)\left(g-\frac{1}{2}\sigma^2\right)+(\alpha_E-\alpha_E(x_0))\left(\mu_p-\frac{1}{2}\sigma_p^2\right)\right)(x_t)\,\mathrm{d}t|x_0=x\right]\sigma_x(x).\end{aligned}$$

The third term becomes zero under the approximation that the wealth exposure of entrepreneurs is constant; that is $\alpha_E(x_t) \approx \alpha_E(x_0)$. □

Note that the accumulation effect is zero at $h = 0$ while the revaluation effect is zero as $h \to \infty$ (as valuation changes are purely transitory). Hence, the proposition implies the following expressions for the short-run and long-run response of entrepreneurs' wealth to an aggregate shock:

$$\begin{aligned}\epsilon(x,0) &= \underbrace{(\alpha_E-1)\sigma}_{\substack{\text{Change}\\ \text{in asset income}}} + \underbrace{(\alpha_E-1)(\partial_x \log p)\sigma_x(x)}_{\text{Change in asset valuation: Revaluation effect} > 0} \\ \epsilon(x,\infty) &\approx \underbrace{(\alpha_E-1)\sigma}_{\substack{\text{Change}\\ \text{in asset income}}} + \underbrace{\mathrm{E}\left[\int_0^\infty \frac{\partial x_t}{\partial x_0}\partial_x\left(r+\alpha_E\left(\frac{1}{p}-r_t\right)-\frac{1}{2}\alpha_E(\alpha_E-1)\sigma_R^2\right)(x_t)\,\mathrm{d}t\Big|x_0=x\right]\sigma_x(x).}_{\text{Change in asset valuation: Accumulation effect} < 0}\end{aligned}$$

In other words, the endogenous response of asset valuations to aggregate shocks amplifies entrepreneurs' wealth in the short run (due to the positive revaluation effect) but a dampening effect in the long run (due to the negative accumulation effect).

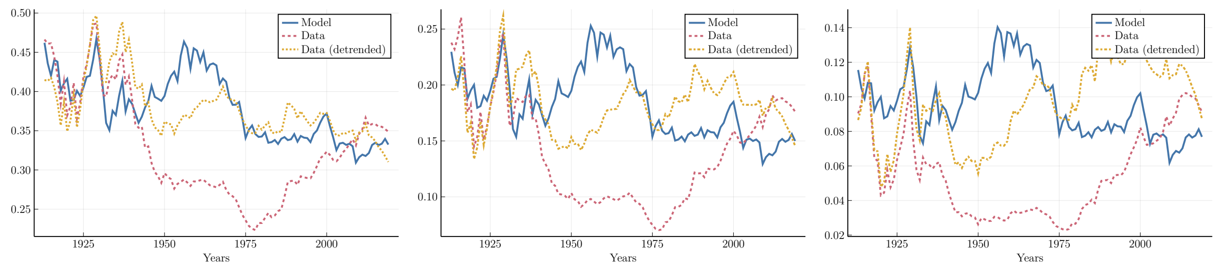
D.6 Historical dynamics of top wealth shares in the model and the data

In this paper, I have built a model that sheds light on the dynamics of asset prices and wealth inequality in response to aggregate shocks. One interesting question is: how much of the actual fluctuations in top wealth shares over the last hundred years can this mechanism explain? To answer this question, I feed the sequence of aggregate shocks that generates the realized sequence of excess returns between 1913 and 2020 into the calibrated model.

Figure A8 compares the time series of top wealth shares implied by the model with the actual realization of top wealth shares in the data (using data from Saez and Zucman, 2016), for $p \in \{1\%, 0.1\%, 0.01\%\}$. The model captures well business cycle fluctuations in top wealth inequality. However, it misses the overall U-shape of top wealth shares over the 20th century (in particular, the steep decline in the 40s and the steep increase starting in the 80s). To adjust for these low-frequency fluctuations in top wealth inequality, I also compare the series of top wealth shares implied by the calibrated model to a "detrended" version of realized top wealth shares, where "detrended" means that the series is adjusted for the structural break in the growth of top wealth shares.[13] One can see that the two series remain very close over the time sample.

[13]More precisely, I use a sup-Wald statistics to test for a structural break in the yearly growth of top wealth shares at an unknown break date for each top percentile $p \in \{1\%, 0.1\%, 0.01\%\}$. The test systematically suggests a structural break in 1979 at the 10% level for each top percentile. An alternative way to "detrend" realized top wealth shares would be to use the Hodrick-Prescott filter. However, this filter isolates business cycle dynamics, which is not adapted to our

To quantify the distance between the two series, I compute the relative root mean squared error between



(a) Top 1% (b) Top 0.1% (c) Top 0.01%

Figure A8: Historical dynamics of top wealth shares in the model versus the data

Notes: The figure reports the time series of top wealth shares implied by the model after feeding it with the time series of aggregate shocks generating the same realization of equity excess returns as in the data. The figure also reports the time series of realized top wealth shares from Saez and Zucman (2016) and a detrended version after adjusting for the structural break in the logarithmic growth of top wealth shares in 1979.

the two series, defined as

$$||(S_t)_{0\leq t\leq T} - (\hat{S}_t)_{0\leq t\leq T}|| \equiv \sqrt{\frac{\sum_{t=0}^{T} \left(\hat{S}_t - S_t\right)^2}{\sum_{t=0}^{T} S_t^2}}. \tag{A13}$$

where $\hat{S}_t$ denotes the model-implied series and S_t denotes the actual series of top wealth shares. As reported in Table A10, the typical difference between the model-implied top 1% wealth share and the actual (detrended) top wealth share, relative to the typical level of the latter, is approximately 11%. This difference increases to 25% for the top 0.01%.

Table A10: Distance of top wealth shares between the model and the data

	Top 1%	Top 0.1%	Top 0.01%
Model versus Data	0.25	0.49	0.90
Model versus Data (detrended)	0.11	0.22	0.34

Notes: The table reports the distance between the model-implied top wealth shares and the actual top wealth shares. The distance between the two series is computed as the relative root mean squared error between the two series, as defined in (A13).

It is important to note that top wealth shares are measured with errors (see Appendix B.2 for a detailed discussion). Hence, it is hard to know how much of the remaining discrepancy between the model and the data reflects model misspecification versus measurement error in top wealth shares. While measurement error in top wealth shares is a concern for this type of exercise, note that it does not necessarily bias the estimates obtained from local projections since top wealth shares appear on the left-hand side of the regressions.

purpose as aggregate shocks generate long-lived fluctuations in top wealth shares (see Figure 5).

E Additional lemmas

E.1 Stability of linear functionals

This section states and proves a lemma for linear functionals, which is used in the proof of Proposition 2.

Lemma A1. *Let $x_t \in \mathbb{R}$ be a one-dimensional diffusion process with a unique invariant probability measure. Denote* P *(resp.* E*) the probability measure (resp. expectation) with respect to the invariant probability measure of x. Consider the process*

$$\mathrm{d}M_t = \Big(\mu(x_t)M_t + b(x_t)\Big)\,\mathrm{d}t + \sigma(x_t)M_t\,\mathrm{d}Z_t, \tag{A14}$$

where $\mathrm{P}(b(x) \geq 0) = 1$, $\mathrm{P}(b(x) > 0) > 0$, and μ and σ are integrable with respect to the invariant probability measure. Then, we have:

(i) If $\mathrm{E}[\mu(x) - \frac{1}{2}\sigma(x)^2] < 0$, M_t does not converge to infinity a.s.

(ii) If $\mathrm{E}[\mu(x) - \frac{1}{2}\sigma(x)^2] > 0$, M_t converges to infinity a.s.

Proof. While this result is well known in the discrete-time case, I could not find a similar proof in the continuous-time case. I do the proof in two steps: I first bound the continuous-time process M_t by a discrete-time process, as in Maruyama and Tanaka (1957). I then apply the discrete-time results from Vervaat (1979) to characterize the limit of this discrete-time process.
Step 1. For $\tau > 0$, we have the following recurrence equation:

$$M_{t+\tau} = e^{\int_t^{t+\tau}\left(\mu(x_u) - \frac{1}{2}\sigma(x_u)^2\right)\mathrm{d}u + \int_t^{t+\tau}\sigma(x_u)\,\mathrm{d}Z_u} M_t + \int_t^{t+\tau} e^{\int_s^{t+\tau}\left(\mu(x_u) - \frac{1}{2}\sigma(x_u)^2\right)\mathrm{d}u + \int_s^{t+\tau}\sigma(x_u)\,\mathrm{d}Z_u} b(x_s)\,\mathrm{d}s.$$

Denote I the set of values that x_t can take. Take $a < b$, both in I. Define the sequence of stopping times $S_0 = 0$ and

$$\begin{aligned} T_n &\equiv \inf\{t > S_n; x_t = a\}, \\ S_{n+1} &\equiv \inf\{t > T_n; x_t = b\}. \end{aligned}$$

Define, for any $n \geq 0$,

$$\begin{aligned} X_n &\equiv M_{T_n}, \\ A_n &\equiv e^{\int_{T_n}^{T_{n+1}}(\mu(x_u) - \frac{1}{2}\sigma(x_u)^2)\,\mathrm{d}u + \int_{T_n}^{T_{n+1}}\sigma(x_u)\,\mathrm{d}Z_u}, \\ B_n &\equiv \int_{T_n}^{T_{n+1}} e^{\int_s^{T_{n+1}}(\mu(x_u) - \frac{1}{2}\sigma(x_u)^2)\,\mathrm{d}u + \int_s^{T_{n+1}}\sigma(x_u)\,\mathrm{d}Z_u} b(x_s)\,\mathrm{d}s. \end{aligned}$$

Note that X_n bounds the continuous time process M_t: for $t \in (T_n, T_{n+1}]$, $X_n \leq M_t \leq X_{n+1}$. In particular, M_t converges to infinity a.s. if and only if X_n converges to infinity a.s.
Step 2. The sequence X_n satisfies the following recurrence relation:

$$X_{n+1} = A_n X_n + B_n.$$

Moreover, by the strong Markov property, A_n and B_n i.i.d.[14] Moreover, A_1 is positive a.s., B_1 is non neg-

[14]This comes from the definition of A_n and B_n using the sequence of stopping times $(T_n)_{n\geq 0}$: whatever transpires between successive passages through a forms a sequence of independent events.

ative a.s. with $\mathrm{P}(B_1 > 0) > 0$ and $\mathrm{E}[\log B_1] < \infty$. From Theorem 1.6 of Vervaat (1979), X_n converges in distribution if $\mathrm{E}[\log A_1] < 0$ (i.e., if $\mathrm{E}\left[\int_{T_1}^{T_2}\left(\mu(x_u) - \frac{1}{2}\sigma(x_u)^2\right)\mathrm{d}u\right] > 0$). In contrast, X_n converges a.s. to infinity if $\mathrm{E}[\log A_1] > 0$.[15] Finally, from Theorem 4.2 Maruyama and Tanaka (1957), the recurrence of the process (x_t) implies that, for any integrable function f, $\mathrm{E}\left[\int_{T_1}^{T_2} f(x_u)\,\mathrm{d}u\right] \geq 0$ if and only if $\mathrm{E}\left[f(x)\right] \geq 0$, which concludes the proof. □

E.2 Expectation of a normal random variables above a threshold

The section states and proves a well-known lemma related to the expectation of the exponential of a normal variable above a threshold. This lemma is used in the proof of Proposition 4, Proposition 7, and Proposition 8.

Lemma A2. *Consider a normal random variable Z with mean μ and standard deviation ν. For any threshold $z \in \mathbb{R}$ and exponent $\xi \in \mathbb{R}$, we have*

$$\mathbb{E}\left[e^{\xi Z} 1_{Z \geq z}\right] = e^{\xi\mu + \frac{1}{2}\xi^2\nu^2}\overline{\Phi}\left(\frac{z-\mu}{\nu} - \xi\nu\right),$$

where $\overline{\Phi}(\cdot) \equiv 1 - \Phi$ denotes the counter-cumulative distribution function of a standard normal variable.

Proof.

$$\begin{aligned}
\mathbb{E}\left[e^{\xi Z} 1_{Z \geq z}\right] &= \int_z^\infty e^{\xi x}\frac{1}{\sqrt{2\pi\nu^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\nu^2}}\,\mathrm{d}x \\
&= \int_{(z-\mu)/\nu}^\infty e^{\xi(\mu+\nu y)}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\,\mathrm{d}y \\
&= e^{\xi\mu + \frac{1}{2}\xi^2\nu^2}\int_{(z-\mu)/\nu}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\xi\nu)^2}\,\mathrm{d}y \\
&= e^{\mu\xi + \frac{1}{2}\xi^2\nu^2}\int_{(z-\mu)/\nu-\xi\nu}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}\,\mathrm{d}s \\
&= e^{\mu\xi + \frac{1}{2}\xi^2\nu^2}\overline{\Phi}\left(\frac{z-\mu}{\nu} - \xi\nu\right),
\end{aligned}$$

where the second equality is obtained using the change of variable $y = (x-\mu)/\nu$ while the fourth equality is obtained using the change of variable $s = y - \xi\nu$. □

E.3 Feynman-Kac formula

This section states and proves an extension of the Feyman-Kac formula. This formula is used to compute analytically the IIRFs defined in the main text.

Lemma A3. *Consider a diffusion process*

$$\mathrm{d}x_t = \mu_x(x_t)\,\mathrm{d}t + \sigma_x(x_t)\,\mathrm{d}Z_t.$$

[15] Indeed, in this case, the law of large number implies $\sum_1^n \log A_k$ converges to infinity a.s. Since the function B_n is positive a.s., X_n is bounded below by $X_0 \prod_1^n A_k$, which implies that X_n also converges to infinity a.s.

Given a set of smooth functions f, ψ, μ_m, and σ_m, the function

$$u(x,h) \equiv \mathrm{E}\left[\int_0^h e^{\int_0^t \mu_m(x_s)\,\mathrm{d}s+\int_0^t \sigma_m(x_s)\,\mathrm{d}Z_s} f(x_t)\,\mathrm{d}t + e^{\int_0^h \mu_m(x_s)\,\mathrm{d}s+\int_0^h \sigma_m(x_s)\,\mathrm{d}Z_s}\psi(x_h)\Big|x_0 = x\right] \tag{A15}$$

can be computed numerically as the solution of the linear PDE

$$\partial_h u(x,h) = f(x) + \left(\mu_m(x) + \frac{1}{2}\sigma_m^2(x)\right)u(x,h) + \Big(\mu_x(x) + \sigma_x(x)\sigma_m(x)\Big)\partial_x u(x,h) + \frac{1}{2}\sigma_x^2(x)\partial_{xx}u(x,h)$$

with initial boundary condition $u(x,0) = \psi(x)$.

The traditional Feynman-Kac formula corresponds to the special case $\sigma_m(x) = 0$. This generalization is related to Hansen and Scheinkman (2009), who study similar "tilted" infinitesimal generators.

Proof. Equation (A15) implies the following recurrence relation for $0 < \tau < h$

$$u(x,h) = \mathrm{E}\left[\int_0^\tau e^{\int_0^t \mu_m(x_s)\,\mathrm{d}s+\int_0^t \sigma_m(x_s)\,\mathrm{d}Z_s} f(x_t)\,\mathrm{d}t + e^{\int_0^\tau \mu_m(x_s)\,\mathrm{d}s+\int_0^\tau m(x_s)\,\mathrm{d}Z_s}u(x_\tau, h-\tau)\Big|x_0 = x\right].$$

Subtracting by $u(x,h)$ on each side, dividing by τ, and passing to the limit $\tau \to 0$ gives:

$$0 = f(x) + \lim_{\tau\to 0}\frac{1}{\tau}\mathrm{E}\left[e^{\int_0^\tau \mu_m(x_s)\,\mathrm{d}s+\int_0^\tau \sigma_m(x_s)\,\mathrm{d}Z_s}u(x_\tau, h-\tau) - u(x,h)\Big|x_0 = x\right].$$

An application of Ito's lemma gives

$$0 = f(x) + \Big(\mu_m(x) + \frac{1}{2}\sigma_m^2(x)\Big)u(x) + \Big(\mu_x(x) + \sigma_x(x)\sigma_m(x)\Big)\partial_x u(x) + \frac{1}{2}\sigma_x^2(x)\partial_{xx}u(x) - \partial_h u(x,h).$$

The initial boundary condition is obtained by taking $h = 0$ in (A15). □

I now briefly discuss how I solve this linear PDE using finite difference methods. Consider an homogeneous discretized grid for x; that is $\boldsymbol{x} \equiv (i\Delta x)_{0\le i\le N-1}$, with $(N-1)\Delta x = 1$. Define $\mathbb{T}$ the $N \times N$ matrix that corresponds to the discretized version of the operator defined in (A9):

$$\begin{aligned}\mathbb{T} : u \to{} & \lim_{h\to 0}\frac{1}{h}\left(\mathrm{E}\left[e^{\int_0^h \mu_m(x_s)\,\mathrm{d}s+\int_0^h \sigma_m(x_s)\,\mathrm{d}Z_s}u(x_h)|x_0 = x\right] - u(x)\right)\\ ={} & \left(\mu_m(x) + \frac{1}{2}\sigma_m(x)^2\right)u + \Big(\mu_x(x) + \sigma_x(x)\sigma_m(x)\Big)\partial_x u + \frac{1}{2}\sigma_x^2(x)\partial_{xx}u.\end{aligned}$$

More precisely, for any vector $\boldsymbol{u} = (u_i)_{1\le i\le N}$, the vector $\mathbb{T}\boldsymbol{u}$ is a vector with i^{th} component:

$$\begin{aligned}(\mathbb{T}\boldsymbol{u})_i = {} & \Big(\mu_m(x_i) + \frac{1}{2}\sigma_m^2(x_i)\Big)u_i\\ & + \Big(\mu_x(x_i) + \sigma_x(x_i)\sigma_m(x_i)\Big)\left(1_{\mu_x(x_i)\ge 0}\frac{u_{i+1}-u_i}{\Delta x} + 1_{\mu_x(x_i)\le 0}\frac{u_i - u_{i-1}}{\Delta x}\right)\\ & + \frac{1}{2}\sigma_x^2(x_i)\left(\frac{u_{i+1}+u_{i-1}-2u_i}{\Delta x}\right),\end{aligned}$$

for $0 \le i \le N-1$. Consider a discretized time grid $(j\Delta h)_{0\le j\le T}$. Consistently with the initial condition $u(x,0) = g(x)$, I set $\boldsymbol{u}^0$, the discretized version of $u(\cdot,0)$ on the grid $\boldsymbol{x}$, to $\boldsymbol{u}^0 = (g(x_i))_{0\le i\le N-1}$. I then proceed by recurrence: given $\boldsymbol{u}^j$, the discretized version of $u(\cdot, j\Delta h)$ on the grid $\boldsymbol{x}$, I obtain $\boldsymbol{u}^{j+1}$, the discretized version of $u(\cdot, (j+1)\Delta h)$ on the grid $\boldsymbol{x}$, by solving the linear system

$$\frac{\boldsymbol{u}^{j+1} - \boldsymbol{u}^j}{\Delta h} = \boldsymbol{f} + \mathbb{T}\boldsymbol{u}^{j+1},$$

where $\boldsymbol{f} = (f(x_i))_{0 \leq i \leq N-1}$ denotes the discretized version of the function $f(\cdot)$ on the grid $\boldsymbol{x}$.

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "Income and wealth distribution in macroeconomics: A continuous-time approach," *The Review of Economic Studies*, 2022, *89* (1), 45–86.

Adrian, Tobias and Hyun Song Shin, "Liquidity and leverage," *Journal of financial intermediation*, 2010, *19* (3), 418–437.

Alloza, Mario, Jesus Gonzalo, and Carlos Sanz, "Dynamic effects of persistent shocks," 2020.

Atkeson, Andrew G and Magnus Irie, "Rapid Dynamics of Top Wealth Shares and Self-Made Fortunes: What Is the Role of Family Firms?," *American Economic Review: Insights*, 2022, *4* (4), 409–24.

Brunnermeier, Markus K and Yuliy Sannikov, "A Macroeconomic Model with a Financial Sector," *The American Economic Review*, 2014, *104* (2), 379–421.

Gârleanu, Nicolae and Stavros Panageas, "Young, Old, Conservative, and Bold: The Implications of Heterogeneity and Finite Lives for Asset Pricing," *Journal of Political Economy*, 2015, *123* (3), 670–685.

Gomez, Matthieu, "Decomposing the growth of top wealth shares," *Econometrica*, 2023, *91* (3), 979–1024.

Guvenen, Fatih, "A Parsimonious Macroeconomic Model for Asset Pricing," *Econometrica*, 2009, *77* (6), 1711–1750.

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

He, Zhiguo and Arvind Krishnamurthy, "Intermediary asset pricing," *American Economic Review*, 2013, *103* (2), 732–70.

Herbst, Edward and Benjamin K Johannsen, "Bias in local projections," 2021. FEDS Working Paper No. 2020-010R1.

Kopczuk, Wojciech and Emmanuel Saez, "Top wealth shares in the United States, 1916–2000: Evidence from estate tax returns," *National Tax Journal*, 2004, *57* (2), 445–487.

Maruyama, Gisirô and Hiroshi Tanaka, "Some properties of one-dimensional diffusion processes," *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics*, 1957, *11* (2), 117–141.

Saez, Emmanuel and Gabriel Zucman, "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," *The Quarterly Journal of Economics*, 2016, *131* (2), 519–578. Updated series (February 2022 vintage) can be downloaded from https://gabriel-zucman.eu/usdina/.

__ **and** __ , "Trends in US Income and Wealth Inequality: Revising After the Revisionists," Technical Report, National Bureau of Economic Research 2020.

__ **and** __ , "Top wealth in america: A reexamination," Technical Report, National Bureau of Economic Research 2022.

Smith, Matthew, Owen Zidar, and Eric Zwick, "Top Wealth in America: New Estimates under Heterogeneous Returns," *The Quarterly Journal of Economics*, 2023, *138* (1), 515–573.

Tella, Sebastian Di, "Uncertainty shocks and balance sheet recessions," *Journal of Political Economy*, 2017, *125* (6), 2038–2081.

Vervaat, Wim, "On a stochastic difference equation and a representation of non–negative infinitely divisible random variables," *Advances in Applied Probability*, 1979, *11* (4), 750–783.