Wealth Inequality and Asset Prices*

MATTHIEU GOMEZ †

First version: October 2016
This version: January 2024
[latest version]

Abstract

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 equilibrium model of the wealth distribution where agents have heterogeneous exposures to aggregate risk. While the wealth distribution is stochastic in the model, I show that it exhibits a Pareto tail, with a (time-invariant) index that depends on the average logarithmic return of top households. The model features a two-way feedback between asset prices and wealth inequality, which amplifies the response of top wealth inequality to aggregate income shocks in the short-run while dampening it in the medium-run. Aggregate shocks generate particularly large fluctuations in the right tail of the wealth distribution, as higher percentiles are more exposed to aggregate risk and take a longer time to mean revert.

*This is a substantially revised version of the first chapter of my dissertation (job market paper). 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, both at the business cycle and at lower frequencies.[1] What causes these fluctuations in wealth inequality? And what is the effect of these fluctuations in inequality on the aggregate economy?

In this paper, I study empirically and theoretically the interplay between asset prices and wealth inequality. I document that wealthy households disproportionately invest in equity, causing equity returns to generate large and persistent fluctuations in top wealth inequality. I then build a model with heterogeneous agents and aggregate income shocks that can match these impulse response functions. The model features a two-way feedback between asset prices and wealth inequality, which amplifies the response of top wealth inequality to aggregate income shocks in the short-run while dampening it in the medium-run.

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 relative to the rest of the population. 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 projection methods, I show that these effects appear to be very persistent over time.

Motivated by this empirical evidence, I then build an endowment economy model in which agents have heterogeneous exposures to aggregate income shocks. In the model, a subset of the population ("entrepreneurs") must hold concentrated positions in their firms, while the rest ("households") can freely trade equity. As entrepreneurs are more exposed to aggregate risk, aggregate shocks generate fluctuations in the distribution of wealth, as in the data.

I then characterize analytically the shape and the dynamics of the wealth distribution in the model. While the wealth distribution is stochastic in the model (as in the data), I show that it still exhibits a Pareto tail with a *time-invariant* tail index. The intuition is that wealthy individuals tend to have similar portfolios, and so, aggregate shocks do not affect the tail index of the wealth distribution (a measure of relative inequality within the rich). Moreover, I obtain a simple analytical formula for this tail index in terms of the average logarithmic return of top households relative to the growth of the economy.

In the last part of the paper, I explore this mechanism quantitatively. I calibrate the model using moments related to top wealth inequality and asset prices. In particular, I use the elasticity of the average wealth in the top 0.01% to discipline the aggregate risk exposure of entrepreneurs and the tail index of the wealth distribution to discipline their saving rate. I show that the calibrated model matches well the level and the dynamics of top wealth shares in response to excess stock returns.

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

The model generates a feedback loop between asset prices and top wealth inequality. After a positive aggregate income shock, wealthier investors gain more than the rest, i.e. wealth inequality increases. As wealth is redistributed towards wealthy agents, the aggregate demand for assets increases, which, in turn, increases asset valuations. As a result, the model can generate an "excess" volatility of stock market returns in response to aggregate income shocks. 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, can be computed analytically, and varies with the aggregate state of the economy. After applying this new methodology, I show that the volatility of asset returns in the model is mostly driven by changes in future expected excess returns in good times and by changes in future risk-free rates in bad times.

I then use the calibrated model to trace out the effect of aggregate shocks on top wealth inequality over the full horizon. This analysis complements my empirical results, which were solely focused on their short-term responses (as standard errors from local projections become excessively large after ten years). A key finding is that top wealth shares take a very long time to mean revert: more precisely, it takes approximately 40 years for the effect of an aggregate shock on the wealth share of the top 0.01% to be divided by three. Intuitively, this reflects the fact that the effect of aggregate shocks only fully dissipates when new generations reach top percentiles, a process which inherently takes time.

Due to this high persistence, the model generates sizable fluctuations in top wealth inequality over time. Quantitatively, I find that the calibrated model can account for about 40% of the actual standard deviation of top wealth shares observed in the data. In other words, the core mechanism of the model (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. More precisely, while the model matches well the business cycle dynamics of top wealth inequality, it cannot fully capture its low-frequency fluctuations (in particular its overall U-shape over the 20th century). This leaves room for further exploration of contributing factors discussed in the literature, such as changes in saving rates, taxes, or idiosyncratic shocks.[2]

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), changes in labor income (Kaymak et al., 2018), or changes in idiosyncratic shocks (Atkeson and Irie, 2022; Gomez, 2023) on top wealth inequality. Relative to these papers, I focus on examining the effect of excess stock market returns

[2] A non-exhaustive list of papers focused on the low-frequency fluctuations of top wealth inequality for the U.S. includes Kaymak and Poschke (2016), Benhabib et al. (2019), Mian et al. (2020), Hubmer et al. (2021), and Atkeson and Irie (2022).

(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) measures the elasticity of the top 10% wealth share to stock market returns using data from the Survey of Consumer Finances (SCF). Relative to this paper, I focus on the right tail of the wealth distribution (top 1% to top 400), which leads me to find estimates that are an order of magnitude larger.[3] I also focus on longer horizons (up to eight years), which reveals that the effect of stock market returns is very persistent over time, a fact that will play a central role in the calibrated model.

On the theoretical side, my paper contributes to random growth models of inequality. While standard models focus on deterministic economies, where the wealth distribution is time-invariant, I study a Markovian economy, where the wealth distribution is *stochastic*. Despite the presence of aggregate shocks, I obtain a simple characterization for the tail index of the wealth distribution, which depends on the *time-averaged logarithmic* wealth growth of top households. This result connects my paper to Kelly (1956), Blume et al. (1992), or Borovička (2020), who stress the importance of this quantity for long-run survival in infinite-horizon economies. Finally, my finding that aggregate shocks generate long-lived changes in top wealth shares complements recent work by Luttmer (2012) and Gabaix et al. (2016), that stress the slow transition dynamics of wealth distributions between two steady states.

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: is there 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 about the wealth distribution, 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 ver-

[3]More precisely, while Kuhn et al. (2020) show that a 10% equity return increases the wealth share of the top 10% by 0.25%, I show that it increases the wealth share of the top 0.01% by 3.5%. My estimates are also larger than the ones obtained by Bach et al. (2020) for Sweden, who estimate that a 1% (domestic) equity return increases the wealth of households in the top 0.01% by 0.53% — in comparison, I find that a 1% equity return increases the wealth of top U.S. households by 0.98%.

[4]Reflecting this sentiment, Cochrane (2017) writes that "[the heterogeneous agents] model faces challenges and opportunities in the micro data 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."

sion 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 to analyze 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 and I study the dynamics of asset prices and wealth inequality in response to aggregate shocks.

2 Empirical evidence

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

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

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

For asset pricing data, I use the series of stock market returns and risk-free rates from Welch and Goyal (2008) (2022 vintage). 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]

2.2 Empirical results

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

[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.

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 the fact that top percentiles mean-revert very slowly after shocks.

Table 1 reports the estimates of $\beta_{p,h}$ for $h = 3$, which corresponds to the horizon at which the impulse responses tend to 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.

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

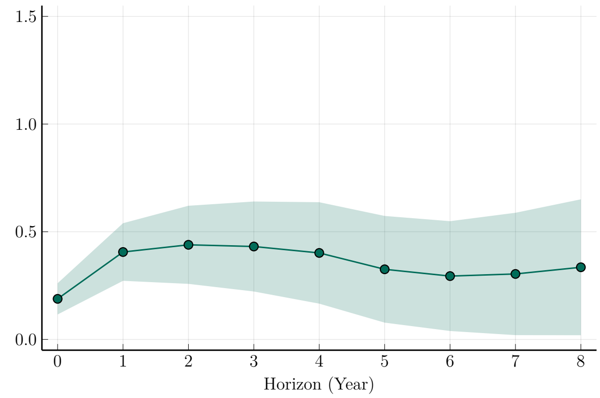
$$\log\left(\frac{S_{p,t+h}}{S_{p,t-1}}\right) = a_{p,h} + b_{p,h}(\log R_{M,t} - \log R_{f,t}) + e_{p,t+h}, \tag{1'}$$

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 (1′) 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 separately is useful to test whether this difference is statistically significant.

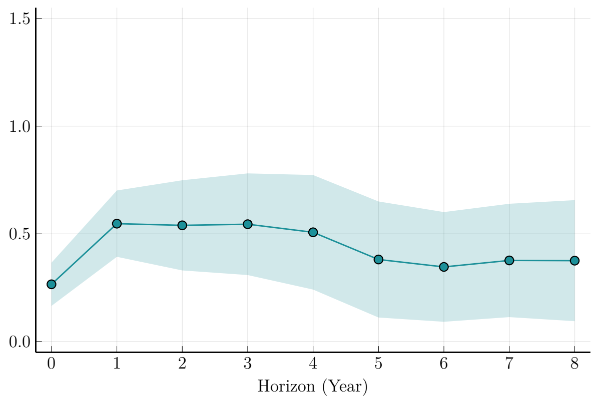
Panel B of Table 1 reports the estimates for $b_{p,h}$ at the four-year horizon. Consistently with the

[8]While Jordà (2005) recommends using Newey-West standard errors, Herbst and Johannsen (2021) stress that they can be downward biased in finite sample 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.

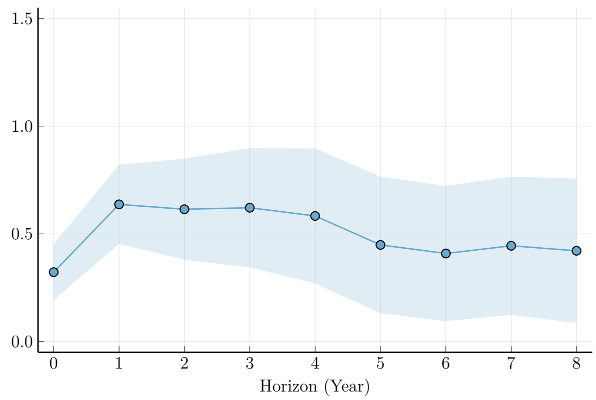
[9]This is particularly true for estimates from *Forbes*, who often use the valuation implied by the last financing round. Relatedly, I find a similar pattern for measures of wealth constructed from estate tax returns (Kopczuk and Saez, 2004), where the valuation of non-tradable assets is done by an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



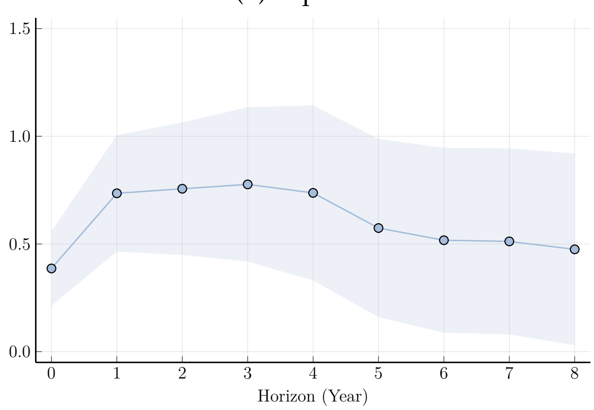
(a) Top 100%



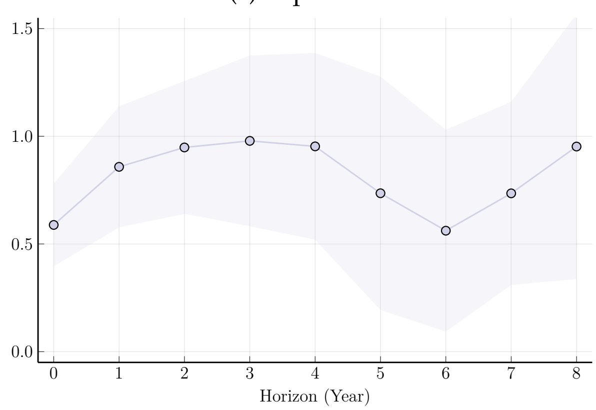
(b) Top 1%



(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.

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 (1′) 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.

discussion above, the estimate 0.35 for the top 0.01% corresponds exactly 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, Appendix Figure D6 reports the impulse response of top wealth shares obtained by plotting the estimated $b_{p,h}$ from (1′) 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 Appendix B for more details.

Alternative specifications. For my baseline results, I focused on estimating local projections using simple univariate regressions (i.e., using 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 (1′).

In the spirit of local projections, however, I now augment the specifications with pre-determined controls; that is, variables known at time $t - 1$ that capture the information available at that time. These variables help in isolating 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 Appendix B.1, I show that augmenting the baseline specification with these controls do not significantly change the response of top wealth shares to stock market returns (Table B1). Intuitively, this comes from the fact that fluctuations in excess stock returns are very 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 is serially correlated, local projections measure both the direct effect of a higher-than-average treatment *and* its indirect effect via higher future treatments on average. To isolate the direct effect, Alloza et al. (2020) suggest augmenting local projections with controls for future realized treatments. In Appendix B.1, I show that I obtain similar results when doing so (Table B2). This means that the response of top wealth shares obtained in my baseline specification is driven by the direct effect of higher stock returns at $h = 0$, rather than their indirect effects through higher (or lower) average returns at $h \geq 1$. This reflects the fact 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 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.

As a further robustness check, I also show in Appendix B.2 that my estimates for the elasticity of top percentile wealth to stock market returns is 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 Appendix B.3 (Figure B3), 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

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

agents who were *initially* in the top 400, not because they increase the arrival of new fortunes in the top 400. This is consistent with the model discussed below, which will generate these impulse responses through a higher equity exposure of agents in top percentiles.

3 A parsimonious model of heterogeneous risk exposures

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, while 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. Demographics follows the perpetual youth model of Blanchard (1985): agents face a constant hazard rate of death δ and population size grows at rate η. This implies that during a short period of time $\mathrm{d}t$, a proportion $\delta\,\mathrm{d}t$ of the population dies while a proportion $(\delta+\eta)\,\mathrm{d}t$ 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 of agents are born as "entrepreneurs" while the rest 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,

$$\frac{\mathrm{d}Y_t}{Y_t} = g\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t, \tag{2}$$

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 $Y_{it} = \vartheta_{it} Y_t$, where ϑ_{it} evolves as

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

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 ϑ_{it} for trees at birth must average to $(\eta+\phi)/(\eta+\delta)$.[11]

Finally, I assume that the wealth of agents who die is redistributed to newborn agents; that is, newborns are endowed with new trees as well as old trees from deceased agents..[12] 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 as well as 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

$$\frac{\mathrm{d}p_t}{p_t} = \mu_{pt}\,\mathrm{d}t + \sigma_{pt}\,\mathrm{d}Z_t, \tag{3}$$

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]

$$\begin{aligned}\frac{\mathrm{d}R_{it}}{R_{it}} &= \frac{1}{p_t}\,\mathrm{d}t + \frac{\mathrm{d}\left(Y_{it}p_t\right)}{Y_{it}p_t}\\ &= \underbrace{\left(\frac{1}{p_t} + g - \phi + \mu_{pt} + \sigma\sigma_{pt}\right)}_{\equiv\mu_{Rt}}\mathrm{d}t + \underbrace{\left(\sigma+\sigma_{pt}\right)}_{\equiv\sigma_{Rt}}\mathrm{d}Z_t + \nu\,\mathrm{d}B_{it},\end{aligned} \tag{4}$$

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,

[11] Indeed, in this case, one can integrate the income flow of all trees in existence with respect to their birth dates gives

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

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

[12] Note that Blanchard (1985) and Gârleanu and Panageas (2015) assume instead that the wealth of deceased agents is redistributed to all existing agents in proportion to their wealth (for instance, because existing agents participate in an annuity market). However, this assumption is increases the return of existing fortunes by δ every period, which leads to counterfactual implications on wealth inequality.

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

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

the welfare of a household i with consumption process $\{C_{it}\}$ is defined recursively by

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

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

$$\begin{aligned} 0 &= \max_{\alpha_{it}, c_{it}}\left\{f\left(c_{it}W_{it}, V_{it}\right)\mathrm{d}t + \mathrm{E}_t[\mathrm{d}V_{it}]\right\} \\ \text{with } \frac{\mathrm{d}W_{it}}{W_{it}} &= \left(r_t + \alpha_{it}(\mu_{Rt} - r_t) - c_{it}\right)\mathrm{d}t + \alpha_{it}\sigma_{Rt}\,\mathrm{d}Z_t. \end{aligned} \tag{5}$$

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.

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

$$\alpha_{Et} = \min\left(\alpha_E, \frac{\int_{i\in\mathbb{I}_t} W_{it}\,\mathrm{d}i}{\int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i}\right). \tag{6}$$

The upper bound on the risk exposure α_{Et} ensures that entrepreneurs are not required to own more trees than there exists in the economy. This constraint will almost never bind in equilibrium, but it is necessary to solve the model globally. 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).[16]

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

[16] A previous version of this paper modeled "entrepreneurs" as individuals with a higher risk tolerance, rather than individuals with a constraint on portfolio holdings — this alternative modelling choices leads to similar quantitative

For simplicity, I assume that entrepreneurs have Epstein-Zin utility with an EIS of one. This number corresponds to the estimates for Vissing-Jørgensen (2002) for the elasticity stockholders at the top of the wealth distribution. This choice also simplifies the calibration of the model because it implies that the consumption rate of entrepreneurs, denoted c_{Et}, is constant over time: $c_{Et} = \rho_E$, where ρ_E denotes their SDR.[17,18]

With these assumptions, the wealth of an entrepreneur $i \in \mathbb{I}_{Et}$ evolves as

$$\frac{\mathrm{d}W_{it}}{W_{it}} = (r_t + \alpha_{Et}(\mu_{Rt} - r_t) - \rho_E)\,\mathrm{d}t + \alpha_{Et}\sigma_{Rt}\,\mathrm{d}Z_t + \alpha_{Et}\nu\,\mathrm{d}B_{it}. \tag{7}$$

Note that this law of motion for wealth is a direct function of the entrepreneurs' fixed equity share α_{Et} and of their fixed consumption rate $c_{Et} = \rho$. This will make it easier to calibrate the model 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) nor their share of wealth invested in equity (which is pinned down by α_E).

Finally, note that the wealth of entrepreneurs in (7) is exposed to the idiosyncratic risk of the tree they are born with. This does not affect the aggregate demand for goods and assets in equilibrium (as entrepreneurs have a fixed equity share and consumption rate); however, 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 (5).

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

$$\int_{i\in\mathbb{I}_{Et}} \rho_E W_{it}\,\mathrm{d}i + \int_{i\in\mathbb{I}_{Ht}} c_{Ht} W_{it}\,\mathrm{d}i = Y_t|\mathbb{I}_t|, \tag{8}$$

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

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

results.

[17] Note that I allow the EIS of households to differ from one. As discussed in the Section 4.1, an EIS lower than one for households will make it easier for the model to match the high volatility of returns. It is consistent with micro-evidence for the average household (Vissing-Jørgensen, 2002; Best et al., 2020).

[18] There is no need to specify the risk aversion for entrepreneurs as its value does not affect their optimal policies: their consumption rate is pinned down by ρ_E while the share of wealth they invest in equity is pinned down by α_E.

3.2 Solving the model

I now outline the main steps in deriving the solution in this section (see Appendix C.1 for a detailed derivation).

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

$$V_{it} = \frac{(\chi_t W_{it})^{1-\gamma}}{1-\gamma}, \tag{10}$$

where the process χ_t, that captures the investment opportunities the 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 (10) into the household's HJB (5) and applying Ito's lemma gives

$$\begin{aligned} 0 = \max_{c_{Ht},\alpha_{Ht}} \Bigg\{ & \frac{\rho}{1-1/\psi}\left(\left(\frac{c_{it}}{\chi_t}\right)^{1-1/\psi} - 1\right) \\ & + r_t + \alpha_{it}(\mu_{Rt} - r_t) - c_{Ht} + \mu_{\chi t} - \frac{\gamma}{2}\left(\alpha_{Ht}^2\sigma_{Rt}^2 + \sigma_{\chi t}^2 - 2\frac{1-\gamma}{\gamma}\alpha_{Ht}\sigma_{Rt}\sigma_{\chi t}\right)\Bigg\}. \end{aligned} \tag{11}$$

The first-order conditions of this problem give

$$c_{Ht} = \rho^{\psi}\chi_t^{1-\psi}, \tag{12}$$

$$\alpha_{Ht} = \frac{1}{\gamma}\frac{\mu_{Rt} - r_t}{\sigma_{Rt}^2} + \frac{1-\gamma}{\gamma}\frac{\sigma_{\chi t}}{\sigma_{Rt}}. \tag{13}$$

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 (8) and (9) can be rewritten as:

$$x_t\rho_E + (1-x_t)c_{Ht} = \frac{1}{p_t}, \tag{8'}$$

$$x_t\alpha_{Et} + (1-x_t)\alpha_{Ht} = 1. \tag{9'}$$

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 (8′) and (9′) and the optimization conditions for households (11), (12), and (13) 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*

$$\mathrm{d}x_t = \mu_{xt}\,\mathrm{d}t + \sigma_{xt}\,\mathrm{d}Z_t, \textit{ where}$$
$$\mu_{xt} \equiv x_t(1-x_t)\left((\alpha_{Et}-\alpha_{Ht})(\mu_{Rt}-r_t) + c_{Ht} - c_{Et} - (\alpha_{Et}-\alpha_{Ht})\sigma_{Rt}^2 + (\eta+\delta+\phi)\left(\frac{\pi}{x_t} - \frac{1-\pi}{1-x_t}\right)\right)$$
$$\sigma_{xt} \equiv x_t(1-x_t)(\alpha_{Et}-\alpha_{Ht})\sigma_{Rt}.$$

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; that is, that the process $(x_t)_{t\in\mathbb{R}}$ has a stationary distribution (Karlin and Taylor, 1981).[19]

3.3 Characterizing the wealth distribution

I now study the cross-sectional distribution of wealth implied by the model. Because the economy grows over time, I focus on individual wealth *normalized* by the average wealth in the economy: $w_{it} \equiv W_{it}/(p_t Y_t)$. Using Ito's lemma, the law of motion of w_{it} for household i of type $j \in \{E, H\}$ is given by[20]

$$\begin{aligned}
\frac{\mathrm{d}w_{it}}{w_{it}} &= \mu_{wjt}\,\mathrm{d}t + \sigma_{wjt}\,\mathrm{d}Z_t + \nu_{wjt}\,\mathrm{d}B_{it}, \text{ where} \\
\mu_{wjt} &\equiv r_t + \alpha_{jt}(\mu_{Rt}-r_t) - c_{jt} - g - \mu_{pt} - \sigma\sigma_{pt} - (\alpha_{jt}-1)\sigma_{Rt}^2 \\
\sigma_{wjt} &\equiv (\alpha_{jt}-1)\sigma_{Rt} \\
\nu_{wjt} &\equiv 1_{j=E}\alpha_{Et}\nu.
\end{aligned} \tag{14}$$

This equation implies that, within each type, the cross-sectional distribution of wealth growth is log-normal. Given our assumption that wealth is log-normally distributed at birth, this implies that, within each type and cohort, the distribution of wealth is log-normal. This observation gives

[19] This stationary distribution can be solved numerically by solving the Kolmogorov Forward equation associated with the law of motion of the process.

[20] This can be obtained by combining the law of motion of individual wealth (5) and (7) with the law of motion of Y_t (2) and p_t (3).

rise to the subsequent formula.

Proposition 2. *The cumulative distribution of wealth within type $j \in \{E, H\}$ at time t is:*[21]

$$\mathbb{P}_t\left(w_{it} \leq w | i \in \mathbb{I}_{jt}\right) = \int_{-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} \Phi\left(\frac{\log w - \mu_{j,s\to t}}{\nu_{j,s\to t}}\right) \mathrm{d}s,$$

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

$$\mu_{j,s\to t} \equiv \log\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right) - \frac{1}{2}\nu^2_{j,s\to t} + \int_s^t \left(\mu_{wju} - \frac{1}{2}\sigma^2_{wju}\right) \mathrm{d}u + \int_s^t \sigma_{wju}\, \mathrm{d}Z_u$$
$$\nu^2_{j,s\to t} \equiv \nu_0^2 + \int_s^t \nu^2_{wju}\, \mathrm{d}u$$

This proposition expresses the distribution of (log normalized) wealth as a mixture of normal distributions corresponding to different cohorts. The mixture weights $(\eta+\delta)e^{-(\eta+\delta)(t-s)}$ correspond to the relative fraction of individuals at time t born at time s. The mean $\mu_{j,s\to t}$ and variance $\nu^2_{j,s\to t}$ of these normal distributions vary across cohorts and over time, reflecting the heterogeneity in cohort ages, the economic conditions at their birth, and the history of aggregate shocks they have experienced. While this Proposition focuses on the cumulative distribution of wealth, I obtain in Appendix C.3 a similar expression for the average normalized wealth above a threshold (or, equivalently, for the *share* of aggregate wealth owned by individuals above that threshold).

Note that Proposition 2 can be used to characterize the overall distribution of wealth across types, since it is a simple mixture of the distribution within types:

$$\mathbb{P}_t\left(w_{it} \leq w | i \in \mathbb{I}_{t}\right) = \pi \mathbb{P}_t\left(w_{it} \leq w | i \in \mathbb{I}_{Et}\right) + (1-\pi)\mathbb{P}_t\left(w_{it} \leq w | i \in \mathbb{I}_{Ht}\right). \tag{15}$$

Proposition 2 shows that the distribution of wealth depends non trivially of the history of preceding aggregate shocks. I now show that one can obtain a much simpler characterization for the wealth distribution by focusing on its right tail.

Definition 1. We say that the distribution of a random variable Y has a Pareto tail if there exists $\zeta \in (0, \infty)$ such that

$$\log \mathbb{P}(Y \geq y) \sim -\zeta \log y \text{ as } y \to \infty.$$

We call ζ the *tail index* of the distribution.[22]

Intuitively, this definition says that a distribution has a Pareto tail when its complementary cumulative distribution function is "close enough" to a Pareto distribution; that is, $\mathbb{P}(Y \geq y) \propto$

[21] 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.

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

$Cy^{-\zeta}$ as $y \to \infty$. A lower ζ corresponds to a lower decay rate; that is, a thicker tail. The next proposition gives a simple characterization for the tail index of the wealth distribution.

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

$$\zeta_j \mathrm{E}\left[\mu_{wjt} - \frac{1}{2}\sigma_{wjt}^2\right] + \frac{1}{2}\zeta_j(\zeta_j - 1)\mathrm{E}\left[\nu_{wjt}^2\right] = \eta + \delta \text{ for } j \in \{H, E\}, \tag{16}$$

where E *denotes the expectation with respect to the stationary density of x. Then*

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

The first part of the proposition says that ζ_j corresponds to the tail index for the wealth distribution within type $j \in \{E, H\}$. Because the overall wealth distribution is a mixture between these two distributions, it inherits the minimum between these two tail indices, which gives that its tail index is $\zeta = \min(\zeta_E, \zeta_H)$. In the rest 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 inequality 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. Finally, the fact that the share of wealth owned by entrepreneurs is a stationary process in $(0, 1)$ implies that $\min(\zeta_E, \zeta_H) > 1$; that is, the right tail of the wealth distribution is thinner than Zipf's law.

One surprising result of Proposition 3 is that the tail index does not vary over time or with the history of aggregate shocks. To understand why, remember that the fraction of entrepreneurs tends to one in the right tail of the wealth distribution. As a result, when an aggregate shock hits, all agents in the right tail move by the same amount (in relative term), which implies that the tail index (a measure of inequality within the rich) remains the same. Put differently, while top wealth shares do respond to aggregate shock, they tend to respond similarly as the top percentile tends to zero (or as wealth w tends to infinity). I discuss the impulse response of top wealth shares in more details in Section 4.4.

To understand the analytical characterization of the tail index given in (16), it is useful to discuss it in the context of the existing literature studying *stationary* distributions. It is well known that, in a static economy in which 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

[23] In the case of $j = H$, ζ_H should be understood as the limit of the positive root as $\mathrm{E}\left[\nu_{wHt}^2\right] \to 0$; 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, $\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right] \leq \mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma_{wEt}^2\right]$.

given by the positive root of[25]

$$\zeta\mu + \frac{1}{2}\zeta(\zeta - 1)\nu^2 = \eta + \delta. \tag{17}$$

Proposition 3 extends this fundamental result along two dimensions; that is, to an economy in which the dynamics of individual wealth *varies over time* and *is exposed to aggregate shocks*. To account for the first dimension (i.e., time-varying dynamics), μ and ν^2 must be replaced, respectively, by the time-averaged cross-sectional drift and variance of wealth growth; that is, $\mathrm{E}\left[\mu_{wjt}\right]$ and $\mathrm{E}\left[\nu_{wjt}^2\right]$. To account for the second dimension (i.e., aggregate shocks), the geometric drift must be adjusted by an Ito's term, $-\frac{1}{2}\mathrm{E}\left[\sigma_{wjt}^2\right]$, that captures the negative effect of aggregate shocks on the time-averaged logarithmic growth of individuals at the top.

I conclude this section by giving an alternative interpretation of Proposition 3. Dividing (16) by ζ_j gives:

$$\mathrm{E}\left[\mu_{wjt} - \frac{1}{2}\sigma_{wjt}^2\right] + \frac{1}{2}(\zeta_j - 1)\mathrm{E}\left[\nu_{wjt}^2\right] - \frac{1}{\zeta_j}(\eta + \delta) = 0. \tag{18}$$

This equation can be seen as a balance equation for top wealth shares. Indeed, the left-hand side corresponds to the time-averaged logarithmic growth of top wealth shares: the first term, $\mathrm{E}\left[\mu_{wjt} - \frac{1}{2}\sigma_{wjt}^2\right]$, corresponds to the time-averaged logarithmic growth of the wealth of agents initially in the top while the second and third terms account for the effect of composition changes in top percentiles due to idiosyncratic volatility and demographic forces, respectively (see Gomez, 2023 for a proof in the case of a stationary economy). For top wealth shares to neither grow or shrink on average over time, their logarithmic growth must average to zero, which gives (18).

3.4 Discussion

To focus on the intuition, I have described a parsimonious model of heterogeneous exposure to aggregate risk. I now briefly discuss three extensions of the baseline model that would make it more realistic: (i) distinction between labor and capital income, (ii) the presence of hand-to-mouth households, and (iii) arbitrary heterogeneity in initial endowment. I show that these extensions would not affect asset prices nor two key moments of the wealth distribution: its tail index and the elasticity of top wealth shares to stock market returns. This justifies my approach of focusing on these two moments when calibrating the model in the next section.

Capital and labor income. In the baseline model, agents only earn one type of income. In reality, agents earn both labor and capital income. This distinction potentially 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").[26] In Appendix

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

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

C.4, however, I show that the tail index of the wealth distribution as well as the elasticity of top wealth shares to stock market returns remain the same whether one considers the distribution of "financial wealth" or the distribution of "total wealth". This justifies my approach of abstracting away from the difference between the 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 agents simply consume 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 affect the elasticity of top wealth shares to stock returns or the tail index of the wealth distribution as they do 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 fact that entrepreneurs dominate in the right tail of the wealth distribution, and, 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.[27]

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, Section 4.3 studies the volatility of asset prices in the model, while Section 4.4 studies the impulse response function of entrepreneurs wealth and top wealth shares to aggregate shocks.

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 growth of the number households in the U.S. since 1913; that is, $\eta = 1.5\%$. The death rate δ is

[27] To better understand the latter claim, note that the normalized wealth of an agent at time t can be written as the product of normalized wealth at birth and the growth of normalized wealth since birth, where the two variables are independently distributed. As is well known in the literature, the tail index of the product of two independent random variables is equal to the minimum of the tail indices for each variables. As a result, the distribution of initial wealth for newborn agents does not affect the tail index of the overall wealth distribution as long as it has a thin tail.

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 $\phi = 1\%$.[28]

Wealth dynamics of entrepreneurs. I now turn to the four parameters related to the wealth dynamics of entrepreneurs $(\alpha_E, \nu, \rho_E, \pi)$. The share of wealth invested in equity by entrepreneurs, 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 in the model, agents only trade all-equity firms (or trees). In reality, firms issue a mix of debt and equity, and therefore levered equity correspond 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:

$$\frac{\mathrm{d}R_{Mt}}{R_{Mt}} = (r_t + \lambda(\mu_{Rt} - r_t))\,\mathrm{d}t + \lambda\sigma_{Rt}\,\mathrm{d}Z_t, \tag{19}$$

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).[29] 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 $\alpha_E = 2$.[30] Note that the estimate for α_E, while high, tends to be a bit lower than existing asset pricing models with heterogeneous stockholders, that do not use the dynamics of top wealth inequality to discipline this parameter. For instance, Gârleanu and Panageas (2015) calibration implies that households at the top of the wealth distribution have an average exposure to aggregate shocks $\alpha_E \approx 2.5$, while Di Tella (2017) calibration implies that the average exposure of financial intermediaries to aggregate shocks is $\alpha_E \approx 2.8$.

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

[28]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.

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

[30]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 too to calibrate α_E.

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 (18) 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:[31]

$$\mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma_{wEt}^2\right] = \underbrace{\mathrm{E}\left[r_t + \alpha_{Et}(\mu_{Rt} - r_t) - \frac{1}{2}\alpha_{Et}^2\sigma_{Rt}^2\right]}_{\substack{\text{Average logarithmic return}\\ \text{of entrepreneurs}}} - \rho_E - \underbrace{\left(g - \frac{1}{2}\sigma^2\right)}_{\substack{\text{Logarithmic growth rate}\\ \text{of economy}}}. \tag{20}$$

Given the values of (g, σ), the logarithmic growth rate of the economy is 1.9%. To obtain an estimate for the average logarithm return of entrepreneurs, 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 $\mathrm{E}[r_t] = 0.3\%$, 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 (20) implies a consumption rate of entrepreneurs $\rho_E = 2.2\%$.[32]

I then pick the population share of entrepreneurs to match the proportion of households that report that more than half of their wealth invested in equity in the Survey of Consumer Finances (Table B4); that is, $\pi = 9\%$. Note that this parameter is, comparatively, a bit difficult to pin down since, in reality, there exists a continuum between households and entrepreneurs. Fortunately, the sensitivity analysis reported in Appendix Table D7 shows that the implication of the model for asset prices is not particularly sensitive to the value for π. 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.

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

[32] While this method is a bit indirect, I show in Appendix D.1 that I obtain similar results if I estimate the consumption rate of top entrepreneurs using individual data from Forbes 400 instead.

Household preferences. I calibrate the remaining three parameters related to households' preferences (their SDR ρ, their EIS ψ, and their RRA γ) to jointly match four asset price moments: 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 $\theta \equiv (\rho, \pi, \gamma)$ 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 micro data 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).

Initial distribution. Finally, I pick the standard deviation of the initial distribution of logarithmic 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 shape of the wealth distribution, which is why I calibrate it after all other parameters have been set.

Table 2: Parameters

Description	Symbol	Value	Target
Demography and endowment			
Population growth rate	η	1.5%	Growth rate number U.S. hhs
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	9%	Pct. hhs with more than half of wealth in equity
Household 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%	4.0%
Standard deviation interest rate	0.9%	0.5%
Average equity return	6.4%	6.9%
Standard deviation equity return	19.3%	15.6%

Notes: The table reports the moments in the data (measured over the 1913-2020 period) and in the calibrated model. The interest rate is constructed as the nominal interest rate minus 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.

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 6.9% in the model) as well as their standard deviation (19.3% in the data versus 15.6% in the model). Note, however, that the calibrated model tends to overestimate the level of the interest rate (0.3% in the data versus 4.0% in the data), even though it matches well its low standard deviation (0.9% in the data versus 0.5% 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%, which is 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 why, note that the volatility of asset prices increases with the heterogeneity in consumption rates between households.[33] 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 consumption rate for households leads to a high average consumption rate in the economy, and, therefore, to a high interest rate to clear the goods market.[34] Consistently with this discussion, Appendix Table D7 reports the sensitivity of asset price moments to the calibrated parameters and 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, Appendix Figure D5 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 $0 \leq h \leq 8$ (Appendix

[33] Indeed, differentiating the market clearing condition for the goods market (8′) give $\partial_x \log p = p\,(c_{Ht} - \rho_E - (1-x)\partial_x c_{Ht})$.

[34] Again, this can be seen through the market clearing condition for the goods market (8′).

Figure D6 presents the same exercise for top wealth shares). This good fit partly reflects the fact 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 non trivial 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 4.4.

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 aggregate shock on its expected value at horizon h starting from the initial state x:[35]

$$\begin{aligned} \mathrm{IIRF}_g(x_t, h) &\equiv \frac{\partial \mathrm{E}\left[g(x_{t+h})|x_t = x\right]}{\partial Z_t} \\ &= \mathrm{E}\left[\frac{\partial x_h}{\partial x_0}\partial_x g(x_h)|x_0 = x\right]\sigma_x(x), \end{aligned} \tag{21}$$

where $\partial x_{t+h}/\partial x_t$ denotes the stochastic derivative of the process $(x_t)_{t\in\mathbb{R}}$ at horizon h with respect to its value at time t. This process equals one at $h = 0$ and then evolves with law of motion:

$$\left(\mathrm{d}\frac{\partial x_{t+h}}{\partial x_t}\right) \Big/ \left(\frac{\partial x_{t+h}}{\partial x_t}\right) = \partial_x \mu_x(x_{t+h})\,\mathrm{d}h + \partial_x\sigma_x(x_{t+h})\,\mathrm{d}Z_{t+h}. \tag{22}$$

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.[36] Figure 2 plots $\mathrm{E}\left[\mathrm{IIRF}_g(x,h)\right]$, 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 position 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 Appendix Figure D7 the same quantities as a function of the state variable x.

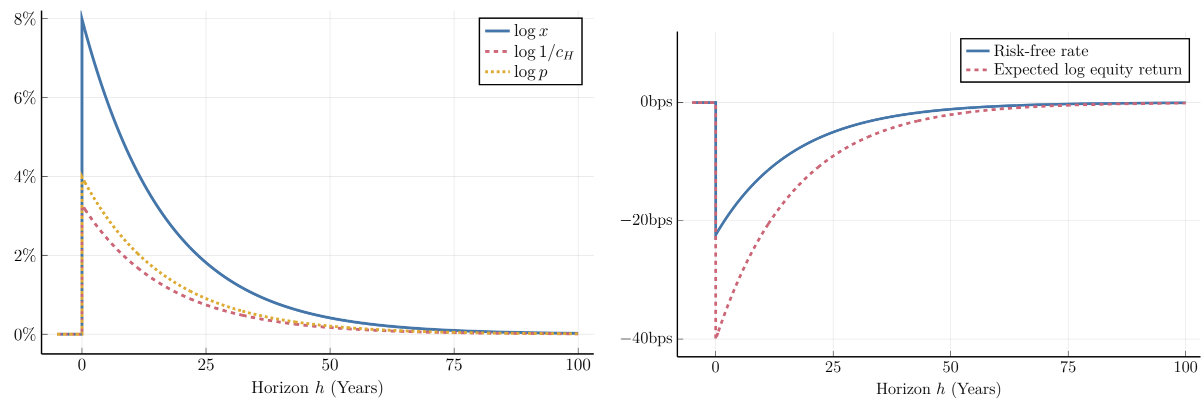
These impulse response functions reveal that aggregate shocks generate very persistent effects on equilibrium prices. In fact, as discussed in Appendix D.2, all infinitesimal impulse response

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

[36] Lemma 2 (stated and proven in Appendix D.2) implies that IIRF_g can be obtained as the solution of the linear PDE

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

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



(a) Wealth-to-consumption ratios

(b) Expected log returns

Figure 2: Infinitesimal impulse response functions

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.

functions decay at the same exponential rate, which corresponds to the "spectral gap" of the infinitesimal generator associated with the process $(x_t)_{t\in\mathbb{R}}$. In the calibrated model, this decay rate is approximately 0.05, which means that it takes more than a decade for the effect of an aggregate shocks on asset valuations to decay by half ($\log 2/0.05 \approx 13$ years). This low rate results from the combination of two mean-reverting forces for the share of wealth owned by entrepreneurs after an aggregate shock, which only act slowly: a mechanical force due to population renewal (death and population growth) and an economic force due to the equilibrium decline in equity returns, which decreases the growth rate of entrepreneurs relative to households.[37]

4.3 Examining the excess volatility of equity returns

Feedback loop. The calibrated model is able to match the high volatility of stock market returns despite the low volatility of aggregate income (Table 3). This is due to a feedback loop between asset prices and wealth inequality: on the one hand, an increase in asset prices increases wealth inequality (as entrepreneurs own levered positions in equity) while, on the other hand, an increase in wealth inequality increases asset prices (as entrepreneurs have a higher demand for assets). We can formalize this feedback loop through two equations. The first equation is the law of motion of

[37] 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.

the share of wealth owned by entrepreneurs (from Proposition 1):[38]

$$\sigma_x = x(\alpha_E - 1)\sigma_R, \tag{23}$$

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 (4) gives:

$$\sigma_R = \sigma + \sigma_p = \sigma + \partial_x \log p \times \sigma_x, \tag{24}$$

which says that, as long as entrepreneurs have a higher demand for assets (i.e. $\partial_x \log p > 0$), the volatility of asset returns increases with the volatility of the state variable. Combining these two equation allows us to solve for the volatility of stock market returns, σ_R:[39]

$$\underbrace{\sigma_R}_{\substack{\text{Return}\\ \text{volatility}}} = \underbrace{\frac{1}{1-(\alpha_E - 1)x\partial_x \log p}}_{\text{Multiplier} \geq 1} \times \underbrace{\sigma.}_{\substack{\text{Income}\\ \text{volatility}}} \tag{25}$$

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 $x\partial_x \log p$.[40] In the calibrated model, we have $\alpha_E = 2.0$, $\mathrm{E}[x] \approx 0.23$, and $\mathrm{E}[\partial_x \log p] \approx 1.77$, which gives a multiplier around 1.7. 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 (24) 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 4. *The volatility of asset valuations can be decomposed into two terms, which correspond to*

[38]This equation reflects the fact that, when an aggregate shock $\mathrm{d}Z_t$ hits the economy, the average wealth of entrepreneurs increases by $\alpha_E\sigma_R\,\mathrm{d}Z_t$ (in relative term) while the average wealth in the economy increases by $\sigma_R\,\mathrm{d}Z_t$ (in relative term). As a result, the share of wealth owned by entrepreneurs increases by $(\alpha_E - 1)\sigma_R\,\mathrm{d}Z_t$ (in relative term). Alternatively, the equation can be obtained by combining the volatility of x from Proposition 1 with the market clearing condition for equity (9').

[39]One way to understand this equation is that, following an aggregate income shock, the share of wealth owned by entrepreneurs increases via (23), which, in turn, increases valuations via (24), which then increases the share of wealth owned by entrepreneurs even more via (23)... 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 (25).

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

the present value of changes in future risk-free rates and in future (log) excess returns, respectively.

$$\sigma_p(x) = \underbrace{-\mathrm{E}\left[\int_0^\infty e^{-\int_0^t \frac{1}{p_s}\,\mathrm{d}s} \frac{\partial x_t}{\partial x_0} \partial_x r(x_t)\,\mathrm{d}t \Big| x_0 = x\right] \sigma_x(x)}_{\textit{Risk-free rate channel}}$$
$$\underbrace{-\mathrm{E}\left[\int_0^\infty e^{-\int_0^t \frac{1}{p_s}\,\mathrm{d}s} \frac{\partial x_t}{\partial x_0} \partial_x \left(\mu_R(x_t) - \frac{1}{2}\sigma_R^2(x_t) - r(x_t)\right) \mathrm{d}t \Big| x_0 = x\right] \sigma_x(x).}_{\textit{Excess return channel}} \quad (26)$$

where $\partial x_t / \partial x_0$ denotes the stochastic derivative of the process $(x_t)_{t \in \mathbb{R}}$ with respect to its value at $t = 0$.

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. In turn, 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.[41]

First, this decomposition is exact, which is useful as Campbell and Shiller (1988)'s log linearization can have large errors in non linear models (e.g. Pohl et al., 2018). Second, because of the continuous-time setup, each term can be computed analytically as a solution of a linear ODE.[42] Third, the decomposition is valid at each point of the state space x, and 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.

One can easily extend Proposition 4 to obtain a similar decomposition for the infinitesimal impulse response function of asset valuations, $\mathrm{IIRF}_{\log p}(x, h)$:[43]

$$\mathrm{IIRF}_{\log p}(x,h) \equiv \underbrace{-\mathrm{E}\left[\int_h^\infty e^{-\int_h^t \frac{1}{p_s}\,\mathrm{d}s} \frac{\partial x_t}{\partial x_0} \partial_x r(x_t)\,\mathrm{d}t \Big| x_0 = x\right] \sigma_x(x)}_{\text{Risk-free rate channel}}$$
$$\underbrace{-\mathrm{E}\left[\int_h^\infty e^{-\int_h^t \frac{1}{p_s}\,\mathrm{d}s} \frac{\partial x_t}{\partial x_0} \partial_x \left(\mu_R(x_t) - \frac{1}{2}\sigma_R^2(x_t) - r(x_t)\right) \mathrm{d}t \Big| x_0 = x\right] \sigma_x(x),}_{\text{Excess return channel}} \quad (27)$$

[41] This decomposition, which is new to my knowledge, holds in any asset pricing model in which the growth rate of cashflows and expected returns are 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, Brunnermeier and Sannikov, 2014, He and Krishnamurthy, 2013, Gârleanu and Panageas, 2015...). See the proof of Proposition 4 in Appendix A for more details.

[42] Lemma 2 from Appendix D.2 implies that, for any f, the function $u(x) \equiv \mathrm{E}\left[\int_0^\infty e^{-\int_0^t \frac{1}{p_s}\,\mathrm{d}s} \frac{\partial x_t}{\partial x_0} f(x)\,\mathrm{d}t \Big| x_0 = x\right]$ can be obtained as the solution of the linear ODE

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

[43] This can be obtained by combining the expression for the impulse response function (21), $\mathrm{IIRF}_{\log p}(x,h) = \mathrm{E}\left[\frac{\partial x_t}{\partial x_0}\partial_x \log p(x_t) | x_0 = x\right]$, with the expression for $\partial_x \log p(x)$ (42) obtained in the proof of Proposition 4.

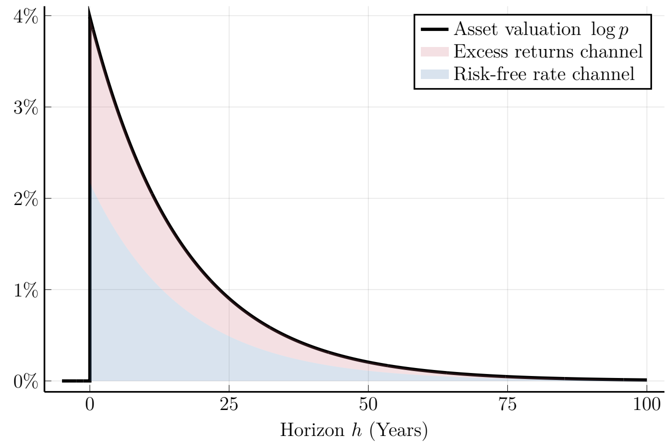


Figure 3: Decomposing the impulse response of asset valuations to aggregate shocks

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" given in (27). The graph can be interpreted as the first-order response to a one standard-deviation annual shock in aggregate income.

where the special case $h = 0$ corresponds to the decomposition for σ_p given in (26). Figure 3 plots the associated decomposition for the average infinitesimal impulse response as a function of the horizon, $h \to \mathrm{E}\left[\mathrm{IIRF}_{\log p}(x,h)\right]$. In terms of magnitude, I find that the risk-free rate channel and the excess return channel account for, respectively, 56% and 44% of the volatility of asset valuations. This is roughly consistent with the impulse responses for log expected returns plotted in Figure 2b, which show that the response in the risk free-rate and in excess-returns contribute equally to the response in log expected returns to aggregate shocks. As shown in Appendix Figure D8, 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.[44]

4.4 Impulse response of top wealth inequality

I now use the calibrated model to trace out the effect of aggregate shocks on top wealth shares over the entire horizon. This analysis complements my empirical results, that focused on their short-term responses (as standard errors from local projections become too large after a few years).

Impulse response of surviving entrepreneurs' wealth. I first characterize the effect of an aggregate shock on the average normalized wealth of "surviving" entrepreneurs. Here, and in the rest of the paper, 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). As discussed below, this object is a good starting point to understand the impulse response of top wealth shares.

[44]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 Appendix Figure D7).

Formally, I denote $\epsilon(x,h)$ the impulse response of the average normalized wealth of surviving entrepreneurs at horizon h starting from an economy in state x:[45]

$$\epsilon(x_t,h) \equiv \frac{\partial \mathrm{E}_t\left[\log \mathbb{E}_{t+h}\left[w_{i,t+h} | i \in \mathbb{I}_{E,t} \cap \mathbb{I}_{E,t+h}\right]\right]}{\partial Z_t}. \tag{28}$$

The next proposition characterizes this impulse response analytically.

Proposition 5. *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:*

$$\epsilon(x,h) = \sigma_{wE}(x) + \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\partial_x\left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)(x_t)\,\mathrm{d}t \Big| x_0 = x\right]\sigma_x(x). \tag{29}$$

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 entrepreneurs going forward (i.e., for $h > 0$). Note that this proposition implies that ϵ can be computed numerically using a version of Feynman-Kac formula.[46]

Using the fact that $\sigma_{wE} = (\alpha_E - 1)(\sigma + \sigma_p)$, one can rewrite this impulse response functions as the sum of two terms capturing, respectively, the effect of changes in asset income and changes in asset valuations:

$$\epsilon(x,h) = \underbrace{(\alpha_E - 1)\sigma}_{\substack{\text{Due to the response} \\ \text{in asset income}}} + \underbrace{(\alpha_E - 1)\sigma_p(x) + \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\partial_x\left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)(x_t)\,\mathrm{d}t \Big| x_0 = x\right]\sigma_x(x)}_{\substack{\text{Due to the response} \\ \text{in asset valuation}}}. \tag{30}$$

The first term of this decomposition accounts for the response of asset incomes to aggregate shocks. Indeed, an aggregate shock $\mathrm{d}Z_t$ increases the average income earned by entrepreneurs by $\alpha_E\sigma\,\mathrm{d}Z_t$ and the average income in the economy by $\sigma\,\mathrm{d}Z_t$. Hence, in the absence of any change in the valuation of assets (their price-to-income ratios p), this aggregate shock permanently increases the normalized wealth of entrepreneurs by $(\alpha_E - 1)\sigma\,\mathrm{d}Z_t$. This corresponds to the income term in (30).

The second term of this decomposition accounts for the response of asset valuations to aggregate shocks. On the one hand, the endogenous rise in asset valuations amplifies the initial

[45] The fact that it depends on the state of the economy purely though the value of the state variable at that time is proven in Proposition 5 below.

[46] Lemma 2 (stated and proven in Appendix D.2) implies that that the function $u(x,h) \equiv \mathrm{E}\left[\int_0^h \frac{\partial x_h}{\partial x_0} f(x_h) | x_0 = x\right]$ can be obtained as the solution of the linear PDE

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

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

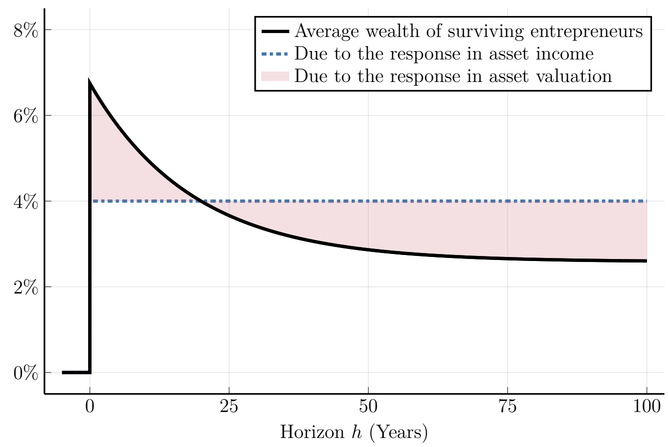


Figure 4: Decomposing the impulse response of the average wealth of surviving entrepreneurs

Notes: This figure plots the average impulse response of the (log) average (normalized) wealth of surviving entrepreneurs (i.e.$h \to \mathrm{E}\left[\epsilon(x,h)\right]$), as well as its decomposition into the effect of changes in asset income and changes in asset valuations (30). The graph can be interpreted as the first-order response to a one standard-deviation annual shock in aggregate income.

response of entrepreneurs' wealth to an aggregate shock at $h = 0$, by $(\alpha_E - 1)\sigma_p$. On the other hand, the resulting decline in future returns decreases their future growth rates at $h > 0$, by $\mathrm{E}\left[\frac{\partial x_t}{\partial x_0}\partial_x\left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)(x_t)\middle| x_0 = x\right]\sigma_x(x)$. The sum of these two forces corresponds to the valuation term in (30).

To visualize the overall impulse response, Figure 4 plots $\mathrm{E}\left[\epsilon(x,h)\right]$ as a function of the horizon h, as well as its decomposition (30) into the effect of changes in asset income and changes in asset valuations. At $h = 0$, the two channels are quantitatively similar: half of the relative growth of entrepreneurs' wealth is due to an increase in their income, $(\alpha_E - 1)\sigma$, while the other half is due to an increase in the valuation of their assets, $(\alpha_E - 1)\sigma_p$. As the horizon h increases, however, the effect of changes in asset valuations progressively declines and even becomes negative. In other words, the positive effect of higher asset valuations for entrepreneurs' wealth is more than compensated by the negative effect of lower returns going forward. I analyze in more details the effect of the rise in asset valuations in Appendix D.3.

Impulse response of top wealth shares. I now study the long-term response of top wealth shares to aggregate shocks. 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 an horizon of 100 years. Consistently with the empirical evidence presented in Section 2.2, higher top percentiles respond more to aggregate shocks. As discussed in Section 4.4, this reflects the fact that the relative proportion of entrepreneurs increases to one in the right tail of the distribution.[47] The key new takeaway of the figure is that top percentiles takes a very long time to mean revert, with a speed of mean

[47]Formally, it is easy to show that the instantaneous response of the share of wealth owned by a top percentile p (at $h = 0$) is given by $w_t(p)\sigma_{we} + (1 - w_t(p))\sigma_{wH}$, where $w_t(p)$ represents the share of wealth owned by entrepreneurs in the top percentile p at time t, which tends to σ_{wE} at $p \to 0$.

reversion that declines steeply in the right tail.[48] To understand why, it is useful to distinguish between two sources of mean reversion following an aggregate shock. The first source of mean reversion is that higher valuations reduce the relative growth rate of entrepreneurs. This force, which was at work in explaining the impulse response of the average wealth of surviving entrepreneurs, is not enough to fully bring back top wealth shares to their original values in the long-run, as $\epsilon(x, \infty) > 0$.

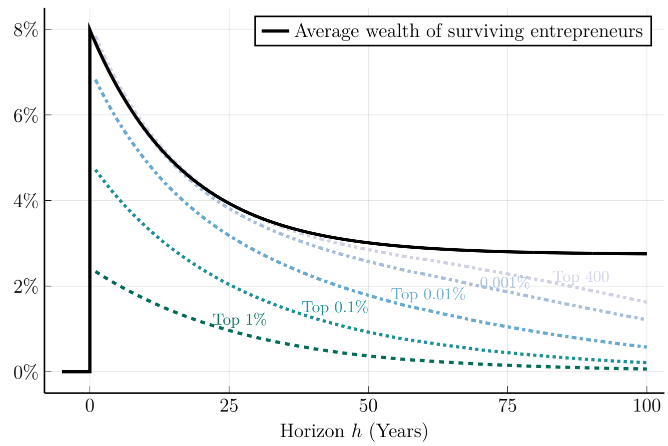


Figure 5: Impulse response of top wealth shares

Notes: This figure plots the infinitesimal impulse response for the average (normalized) wealth of surviving entrepreneurs; that is, $h \to \mathrm{E}[\epsilon(x, h)]$. 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.

The second, and more essential, source of mean reversion is demographics. As time passes, existing agents are gradually replaced by newborn agents, who are not directly impacted by an aggregate shock that happened before their birth. The higher the top percentile, the longer it takes for existing agents to be replaced by newborns, and, therefore, the longer it takes for aggregate shocks to dissipate. Consistently with this idea, Figure 5 shows that the impulse response of the top 0.001% and top 400 initially coincide with the impulse response of the average wealth of surviving entrepreneurs, $\mathrm{E}[\epsilon(x, h)]$, until some point at which newborns start reaching the top percentile. This analysis echoes Gabaix et al. (2016)'s finding that transition dynamics are slower in the right tail: in the context of a stochastic economy, similar forces imply that aggregate shocks have more persistent effects in the right tail.

Fluctuations of top wealth shares in the model versus data. 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 are the standard devi-

[48] In terms of magnitude, the horizon after which the initial effect of an aggregate shock is divided by three increases from 31 years for the top 1% to 61 years for the top 400.

ation of top wealth shares roughly corresponds to the area below its impulse response function.[49]

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	0.13	0.19	0.19
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 the data, there may be additional sources of fluctuations in top wealth shares. One interesting question is: how much fluctuations in top wealth shares can 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).[50] The results, reported in Table 4, reveal that the calibrated model can account for approximately 40% of the standard deviation of top wealth shares in the data. Said differently, the heterogeneous exposure of agents to aggregate shocks can explain a sizable fluctuations of top wealth shares, but it cannot fully explain the 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 in the data and in the model, after feeding the model with the realization for excess returns across the 20th century. This exercise, done in Appendix D.4, reveals that the model can explain the persistent decline in wealth inequality during the Great Depression, its rise immediately after WW2, and its

[49]Formally, the Clark-Ocone formula allows us to write the logarithm of the share of wealth owned by a top percentile p as a moving average of past aggregate shocks. This gives, using Ito's isometry:

$$\log S_t = \mathrm{E}[\log S_t] + \int_{-\infty}^{t} \mathrm{E}_s \left[\frac{\partial \log S_t}{\partial Z_s}\right] \mathrm{d}Z_s \implies \mathrm{Var}[\log S_t] = \mathrm{E}\left[\int_0^{\infty} \mathrm{E}_0 \left[\frac{\partial \log S_h}{\partial Z_0}\right]^2\right] \mathrm{d}h.$$

[50]This is because the naive estimate for the standard deviation of a persistent process suffers from a downward bias in finite sample. Note that 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).

rise during the dot-com bubble. However, the model cannot fully explain the decline in inequality in the 1940s, nor its rise beginning in the 1980s: overall, the model has a hard time reproducing the U-shape of top wealth inequality. Hence, 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 top wealth shares (beyond aggregate shocks), such as changes in taxes (Hubmer et al., 2021), changes in saving rates (Saez and Zucman, 2016, Mian et al., 2020), changes in labor income inequality (Rosen, 1981, Gabaix and Landier, 2008, Terviö, 2008, Straub, 2019), or changes in idiosyncratic shocks (Benhabib et al., 2019; Atkeson and Irie, 2022; Gomez, 2023).

5 Conclusion

This paper examines theoretically and empirically the joint dynamics of asset prices and wealth inequality in response to aggregate shocks. When an aggregate shock hits the economy, wealth inequality increases. As wealth is re-balanced towards agents with a higher demand for assets, asset valuations increase in equilibrium. This feedback loop between wealth inequality and asset prices magnifies the effect of an aggregate shock on top wealth inequality in the short-run while dampening it in the medium-run, as higher valuations are associated with lower asset returns going forward.

Overall, my paper makes three distinct contributions. 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. Second, I characterize the wealth distribution in a Markovian economy with aggregate shocks and I obtain a simple formula for its tail index in terms of the average logarithmic return of households at the top of the wealth distribution. Third, I build a heterogeneous-agent model that matches (and sheds light) on the impulse response of wealth inequality and asset prices to aggregate shocks.

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 between wealth inequality and asset prices could have effects on real quantities as well, 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.

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 Eco-*

nomic Studies, 2022, *89* (1), 45–86.

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.

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.

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.

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 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.

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.

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

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.

Maruyama, Gisirô and Hiroshi Tanaka, "Ergodic Property of N-Dimensional Recurrent Markov Processes," *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics*, 1959, *13* (2), 157–172.

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.

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.

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/.

__ **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.

Shwartz, Adam and Alan Weiss, *Large deviations for performance analysis: queues, communication and computing*, Vol. 5, CRC Press, 1995.

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 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.

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.

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.

Welch, Ivo and Amit Goyal, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 2008, *21* (4), 1455–1508.

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 $x_t = \left(\int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i\right) / \left(\int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i + \int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i\right)$. Applying Ito's lemma gives:

$$\mathrm{d}x_t = x_t(1-x_t)\left(\frac{\mathrm{d}\left(\int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i\right)}{\int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i} - \frac{\mathrm{d}\left(\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i\right)}{\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i} - \left(\frac{\mathrm{d}\left[\int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i\right]}{\int_{i\in\mathbb{I}_{Et}} W_{it}\,\mathrm{d}i} - \frac{\mathrm{d}\left[\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i\right]}{\int_{i\in\mathbb{I}_{Ht}} W_{it}\,\mathrm{d}i}\right)\frac{\mathrm{d}\left[\int_{i\in\mathbb{I}_{it}} W_{it}\,\mathrm{d}i\right]}{\int_{i\in\mathbb{I}_{it}} W_{it}\,\mathrm{d}i}\right). \tag{31}$$

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 (31) gives

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

Combining this expression with the law of motion of W_{it} for households (5) and entrepreneurs (7) gives the result. □

Proof of Proposition 2. Equation (14) implies the following law of motion for the logarithm of normalized wealth for individual i in group $j \in \{E, H\}$:

$$\mathrm{d}\log w_{it} = \left(\mu_{wjt} - \frac{1}{2}\sigma_{wjt}^2 - \frac{1}{2}\nu_{wjt}^2\right)\mathrm{d}t + \sigma_{wjt}\,\mathrm{d}Z_t + \nu_{wjt}\,\mathrm{d}B_{it}. \tag{32}$$

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

$$\log w_{it} = \log w_{is} + \int_s^t \left(\mu_{wju} - \frac{1}{2}\sigma_{wju}^2 - \frac{1}{2}\nu_{wju}^2 \right) \mathrm{d}u + \int_s^t \sigma_{wju}\,\mathrm{d}Z_u + \int_s^t \nu_{wju}\,\mathrm{d}B_{iu}. \tag{33}$$

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 (33), 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:

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

□

Proof of Proposition 3. In line with the typical approach in large deviations theory, the proof is structured in three distinct steps.[51] In the first step, we use the existence of cross-sectional moments of wealth of order lower than ζ_j to prove that the limit *superior* of $\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{jt}) / \log w$ 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 $\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{jt}) / \log w$ 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 $\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{jt}) / \log w$ is exactly $-\zeta_j$ for $j \in \{E, H\}$, which implies that the limit of $\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_t) / \log w$ is exactly $-\min(\zeta_H, \zeta_E)$. As the distribution of wealth is itself a random variable, all of these statement should be understood as holding at any point in time $t \in \mathbb{R}$ almost-surely (i.e. with probability one).

Step 1. This step proves

$$\limsup_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{jt})}{\log w} \leq -\zeta_j \text{ for } j \in \{E, H\}. \tag{34}$$

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 $0 \leq \xi < \zeta_j$. Applying Ito's lemma to the definition of $m_{jt}(\xi)$ gives:

$$\mathrm{d}m_{jt}(\xi) = \mathrm{d}\left(\frac{1}{|\mathbb{I}_{jt}|} \int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\,\mathrm{d}i\right) = m_{jt}(\xi) \left(\frac{\mathrm{d}\left(\int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\,\mathrm{d}i\right)}{\int_{i\in\mathbb{I}_{jt}} w_{it}^\xi\,\mathrm{d}i} - \frac{\mathrm{d}\,|\mathbb{I}_{jt}|}{|\mathbb{I}_{jt}|}\right). \tag{35}$$

[51] See, for instance, the proof of the Gartner-Ellis theorem in Shwartz and Weiss (1995).

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 (35) 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 (14) 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

$$\begin{aligned}\mathrm{d}m_{jt}(\xi) = &\left(\xi\mu_{wjt} + \frac{1}{2}\xi(\xi-1)\left(\sigma_{wjt}^2+\nu_{wjt}^2\right) - (\eta+\delta)\right)m_{jt}(\xi)\,\mathrm{d}t + \xi\sigma_{wjt}m_{jt}(\xi)\,\mathrm{d}Z_t \\ &+ (\eta+\delta)\left(\frac{\eta+\delta+\phi}{\eta+\delta}\right)^{\xi} e^{\frac{1}{2}\xi(\xi-1)\nu_0^2}\,\mathrm{d}t.\end{aligned} \tag{36}$$

Given this law of motion, Lemma (1) (stated an proven in Section C.2) implies that the process $(m_{jt}(\xi))$ remains finite as long as

$$\mathrm{E}\left[\xi\left(\mu_{wjt} - \frac{1}{2}\sigma_{wjt}^2\right) + \frac{1}{2}\xi(\xi-1)\nu_{wjt}^2\right] < \eta+\delta. \tag{37}$$

Combined with the definition of ζ_j (16), this implies that $m_{jt}(\xi)$ is finite for $\xi \in (0,\zeta_j)$.[52]

We now use this result to derive an upper bound on the limit superior of the tail probability.

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

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

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 (34).
Step 2. This step proves

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

We will show this using the law of large numbers, which, intuitively, disciplines the fraction of individuals in older cohorts who must be above a certain threshold. Formally, we start by rewriting the probability of wealth being higher than a certain threshold from Proposition 2

$$\begin{aligned}
\mathbb{P}_t(w_{it} \geq w | j \in \mathbb{I}_{jt}) &= \int_{s=-\infty}^{\infty} (\eta + \delta) e^{-(\eta+\delta)(t-s)} \left(1 - \Phi\left(\frac{\log w - \mu_{j,s\to t}}{\nu_{j,s\to t}}\right)\right) \mathrm{d}s \\
&= \int_{\alpha=0}^{\infty} |\log w| (\eta + \delta) e^{-(\eta+\delta)\alpha \log w} \left(1 - \Phi\left(\frac{\log w - \mu_{j,t-\alpha \log w \to t}}{\nu_{j,t-\alpha \log w \to t}}\right)\right) \mathrm{d}\alpha. \quad (39)
\end{aligned}$$

where the second line uses the change of variable $\alpha = (t - s)/\log w$.

We first tackle the simpler case $j = H$, for which $\mathrm{E}\left[\nu_{wHt}^2\right] = 0$. When $\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right] \leq 0$, $\zeta_H = \infty$ and (38) is trivial. Otherwise, we are in the case $\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right] > 0$. The strong 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 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 turn, 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 $\mu_{H,t-\alpha \log w \to t} \geq \log w$ for any $\alpha \geq \frac{1+\epsilon}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]}$. Combining this inequality with (39) implies that, for any $\log w \geq \max\left(0, \frac{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]}{1+\epsilon} a_0\right)$,

$$\begin{aligned}
\mathbb{P}_t(w_{it} \geq w | i \in \mathbb{I}_{Ht}) &\geq \int_{\frac{1+\epsilon}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]}}^{\infty} (\log w)(\eta + \delta) e^{-(\eta+\delta)\alpha \log w} \left(1 - \overline{\Phi}\left(\frac{\log w - \mu_{H,t-\alpha \log w \to t}}{\nu_0}\right)\right) \mathrm{d}\alpha \\
&\geq \int_{\frac{1+\epsilon}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]}}^{\infty} (\log w)(\eta + \delta) e^{-(\eta+\delta)\alpha \log w} \frac{1}{2} \mathrm{d}a \\
&\geq \frac{1}{2} e^{-(1+\epsilon)\frac{\eta+\delta}{\mathrm{E}\left[\mu_{wHt} - \frac{1}{2}\sigma_{wHt}^2\right]} \log w}.
\end{aligned}$$

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

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

Since this inequality holds for any $\epsilon > 0$, we can pass to the limit $\epsilon \to 0$ to obtain (38).

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

We now tackle the more complex case $j = E$, for which $\mathrm{E}\left[\nu^2_{wEt}\right] > 0$. The strong law of large numbers implies that $\frac{\mu_{E,t-a\to t}}{a} \to \mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma^2_{wEt} - \frac{1}{2}\nu^2_{wEt}\right]$ and $\frac{\nu^2_{E,t-a\to t}}{a} \to \mathrm{E}\left[\nu^2_{wEt}\right]$ as $a \to \infty$. Together with the asymptotic behavior of $\overline{\Phi}$, this implies that

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

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

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

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

Combining this inequality with (39) gives, for any $w \geq \underline{w}$,

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

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

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

Since $\alpha \in \left(0, 1/|\mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma^2_{wEt} - \frac{1}{2}\nu^2_{wEt}\right]|\right)$ and $\epsilon > 0$ were chosen arbitrarily, we get

$$\liminf_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | E \in \mathbb{I}_{Et})}{\log w} \geq -\inf_{\alpha\in\left(0,1/|\mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma^2_{wEt} - \frac{1}{2}\nu^2_{wEt}\right]|\right)} \left\{I(\alpha)\right\}. \tag{40}$$

The minimum for $I(\cdot)$ on $\left(0, 1/\left|\mathrm{E}\left[\mu_{wEt} - \frac{1}{2}\sigma_{wEt}^2 - \frac{1}{2}\nu_{wEt}^2\right]\right|\right)$, which is attained for

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

equals $I_E(\alpha_E^*) = \zeta_E$. Hence, (40) implies

$$\liminf_{w\to\infty} \frac{\log \mathbb{P}_t(w_{it} \geq w | E \in \mathbb{I}_{Et})}{\log w} \geq -\zeta_E. \tag{41}$$

Step 3. Combining the results proven 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 is the first part of the Proposition. Combining this result with (15) 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, if $\zeta_E < \zeta_H$, then

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

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

Proof of Proposition 4. I prove a slightly more general proposition. 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. Denote $p_t(x)$ the ratio of the asset value to its income, which solves the market pricing equation:

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

Differentiating with respect to x_0 gives

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

Rearranging,

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

Integrating forward gives

$$\partial_x \log p(x) = \mathrm{E}\left[\int_0^\infty e^{-\int_0^t \frac{1}{p_s}\,\mathrm{d}s}\frac{\partial x_t}{\partial x_0}\partial_x\left(g_D(x_t)-g_R(x_t)\right)\mathrm{d}t\Big|x_0=x\right]. \tag{42}$$

Multiplying both sides by σ_x gives

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

Applying this formula in the context of our model, where $g_D(x) = g-\phi-\frac{1}{2}\sigma^2$ and $g_R(x) = \mu_R - \frac{1}{2}\sigma_R^2$, gives

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

Adding and subtracting by $\partial_x r(x)$ in the right-hand-side gives the result. □

Proof of Proposition 5. Consider an entrepreneur $i \in \mathbb{I}_{Et}\cap\mathbb{I}_{Et+h}$. We can express the normalized wealth of the entrepreneur at time $t+h$ in terms of their wealth at time t:

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

Averaging across all entrepreneurs in $\mathbb{I}_{Et}\cap\mathbb{I}_{Et+h}$:

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

Taking logarithms and the expectations at time t

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

Differentiating with respect to an aggregate shock at time t:

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

which gives 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 $\epsilon(x,h)$. □

Online Appendix

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:

$$\log\left(\frac{W_{p,t+h}}{W_{p,t-1}}\right) - (h+1)\log R_{f,t} = \alpha_{p,h} + \beta_{p,h}\left(\log R_{M,t} - \log R_{f,t}\right) + \sum_{c\in\mathcal{C}} \gamma_c Z_{c,t-1} + \epsilon_{p,t+h},$$
$$\log\left(\frac{S_{p,t+h}}{S_{p,t-1}}\right) = a_{p,h} + b_{p,h}(\log R_{M,t} - \log R_{f,t}) + \sum_{c\in\mathcal{C}} \gamma_c Z_{c,t-1} + e_{p,t+h}, \tag{43}$$

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 (1) and (1′) discussed in the main text. The advantage of adding pre-determined 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. Note, however, that adding pre-determined variables means that there is no longer a one-to-one mapping between the response of the average wealth in top percentiles and the response of top percentile wealth shares; that is, $b_{p,h} \neq \beta_{p,h} - \beta_{100\%,h}$ in general.

By analogy with the omitted variable bias, it is particularly important to add variables correlated with expected excess stock returns and/or variables correlated with the expected growth of the dependent variable. For the first set of variables, I add the dividend-price ratio, which is a known predictor of excess returns, as well as a five-year moving 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 top percentiles 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 lagged growth of the dependent variable, which helps to capture low-frequency changes in the expected growth of top percentiles. 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.[53]

Table B1 reports the results. Overall, I find very similar estimates for $\beta_{p,3}$ and $b_{p,3}$ as the ones obtained in the baseline specification (Table 1). The intuition is that most of the variations in excess stock market returns comes from variations in *unexpected* excess stock market returns, which is

[53] 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 $t-1$.

uncorrelated with variables known at time $t-1$.[54]

Table B1: Exposure to stock market returns after controlling for pre-determined 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***	0.61***	0.79***	0.79***
	(0.09)	(0.11)	(0.13)	(0.19)	(0.26)
Lagged average wealth	0.04	−0.17	−0.11	0.00	−1.11
	(0.25)	(0.22)	(0.22)	(0.24)	(0.93)
Two-year lagged average wealth	−0.07	0.17	0.13	0.01	0.95
	(0.26)	(0.23)	(0.22)	(0.25)	(0.79)
Lagged cross-sectional variance of returns	−1.96**	−1.05	−0.45	0.20	−5.79
	(0.84)	(0.90)	(1.02)	(1.30)	(4.27)
Lagged dividend-price ratio	−0.11**	−0.09	−0.08	−0.13	−0.16
	(0.05)	(0.06)	(0.08)	(0.11)	(0.29)
Five-year average lagged excess returns	−0.82***	−0.75**	−0.65	−0.65	0.31
	(0.31)	(0.36)	(0.42)	(0.49)	(1.39)
Adjusted R^2	0.23	0.25	0.23	0.21	0.34
N	102	102	102	102	30
Panel B: Wealth share					
Excess returns		0.12**	0.23**	0.40***	0.34*
		(0.05)	(0.09)	(0.14)	(0.18)
Lagged wealth share		−0.03	−0.06	−0.08	−0.74
		(0.29)	(0.27)	(0.26)	(0.46)
Two-year lagged wealth share		−0.13	−0.16	−0.20	0.39
		(0.28)	(0.26)	(0.24)	(0.41)
Lagged cross-sectional variance of returns		1.24***	2.40***	3.59***	−4.42***
		(0.37)	(0.69)	(1.08)	(1.40)
Lagged dividend-price ratio		−0.02	−0.07**	−0.18***	−0.05
		(0.02)	(0.03)	(0.06)	(0.12)
Five-year average lagged excess returns		−0.04	0.02	−0.00	0.25
		(0.11)	(0.20)	(0.30)	(0.75)
Adjusted R^2		0.25	0.29	0.32	0.49
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 pre-determined variables; that is, equation (43) 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, 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

[54] 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. This seems to be a byproduct of the limited time frame of the data, as Appendix B.3 shows that the reaction of 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).

(average of data sources over a year rather than at a point in time), and that private assets may take time to response 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), one way to isolate the first effect is to add future treatments as controls in the local projection specification; that is,

$$\begin{aligned}\log\left(\frac{W_{p,t+h}}{W_{p,t-1}}\right)-(h+1)\log R_{f,t}&=\alpha_h+\sum_{i=0}^{h}\beta_{h,i}\left(\log R_{M,t+i}-\log R_{f,t+i}\right)+\epsilon_{p,t,t+h},\\ \log\left(\frac{S_{p,t+h}}{S_{p,t-1}}\right)&=a_{p,h}+\sum_{i=0}^{h}\beta_{h,i}\left(\log R_{M,t+i}-\log R_{f,t+i}\right)+e_{p,t+h}.\end{aligned}\tag{44}$$

Table B2 reports the result for $h=3$. One can see that the coefficient 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, and so there is little difference in controlling for future excess stock returns or not.

Finally, note that while some amount of mispecification in my local projection specifications is inevitable (excess stock market returns are very close, but not exactly the same as, unexpected stock market returns), these estimates are still informative as long as I use the exact same specification when comparing the model to the data (which I do in Appendix Figures D5 and D6).

B.2 Alternative data sources

Alternative series on top wealth shares. Due to data limitation, 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 time of publication by incorporating several methodological improvements described in Saez and Zucman (2020) and Saez and Zucman (2022).

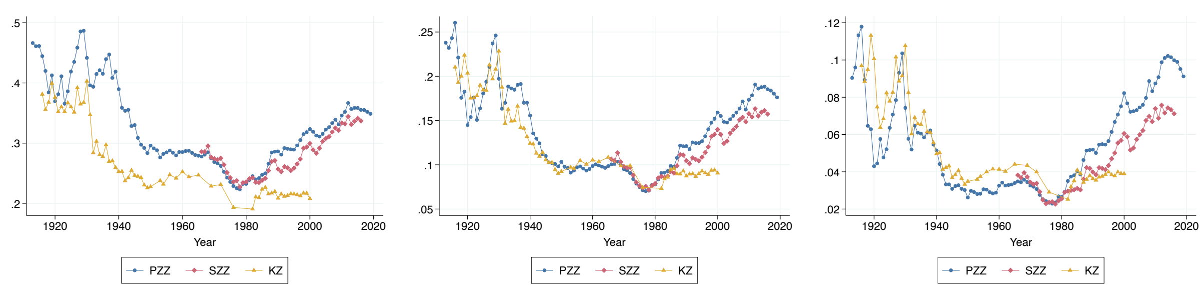
There are two major alternative series for top wealth shares available in the literature. The first alternative is the series constructed from income tax returns by Smith et al. (2023) from 1966 to 2016. While this series initially disagreed with Saez and Zucman (2016), subsequent updates in both series largely reconciled these difference, 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

Table B2: 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 (44) 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.

Zucman (2022) for a more thorough discussion of the difference between these three series. Figure B1 compares the evolution of top wealth shares across the tree series.



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

Figure B1: Alternative series for top wealth shares

Notes: The figure plots three alternative series for top wealth shares. The label "PZZ" 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 versions corresponding to the ones at publication dates.

To assess the extent to which my results depends on the data source for top wealth shares, I re-estimate the baseline specification (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) by 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 B3 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.

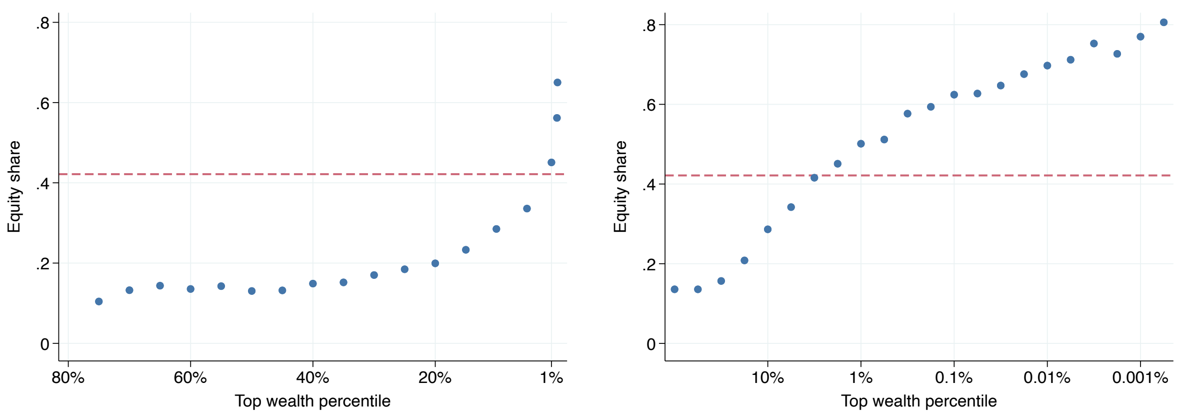
Table B3: Exposure to stock market returns using alternative series for top wealth shares

	Top 100%	Top 1%	Top 0.1%	Top 0.01%
	(1)	(2)	(3)	(4)
Panel A: Wealth data from 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
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

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.[55]

Figure B2a 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 the majority of the wealth distribution, but increases sharply within the top 1%. Figure B2b 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.



(a) Linearly-spaced top percentiles

(b) Loglinearly-spaced top percentiles

Figure B2: Equity share across the wealth distribution

Notes: Figure B2a plots the average equity share within 20 linearly spaced percentile bins in the wealth distribution. Figure B2b 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 networth: (`equity` + `bus`) / `net worth`. Data from SCF 1989-2019.

The first row of Table B4 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 B4), 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 B4 also decomposes the equity share into several subcomponents, revealing that the increase in the average equity share across the wealth distribution is mostly 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 half of their wealth in equity. The last row reports the fraction of income in each percentile that takes the form of labor income.

[55]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 investments in public and private equity.

Table B4: 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	0.18	0.16	0.14
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 networth: `(equity + bus) / networth`. "Entrepreneurs" are define 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

The growth of the average wealth in a top percentile can always be decomposed 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:[56]

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

I then estimate the baseline specification (1) after replacing the dependent variable by each of these

[56] 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.

two terms; that is,

$$\sum_{t \le s \le t+h} \text{Intensive term}_s - (h+1)\log R_{f,t} = \alpha^{\text{intensive}}_{p,h} + \beta^{\text{intensive}}_{p,h}(\log R_{M,t} - \log R_{f,t}) + \epsilon^{\text{intensive}}_{p,t+h},$$
$$\sum_{t \le s \le t+h} \text{Extensive term}_s = \alpha^{\text{extensive}}_{p,h} + \beta^{\text{extensive}}_{p,h}(\log R_{M,t} - \log R_{f,t}) + \epsilon^{\text{extensive}}_{p,t+h}. \tag{45}$$

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, $\beta^{\text{intensive}}_{p,h} + \beta^{\text{extensive}}_{p,h} = \beta_{p,h}$.

Figure B3 plots the resulting estimates for $\beta^{\text{intensive}}_{p,h}$ and $\beta^{\text{extensive}}_{p,h}$ 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.

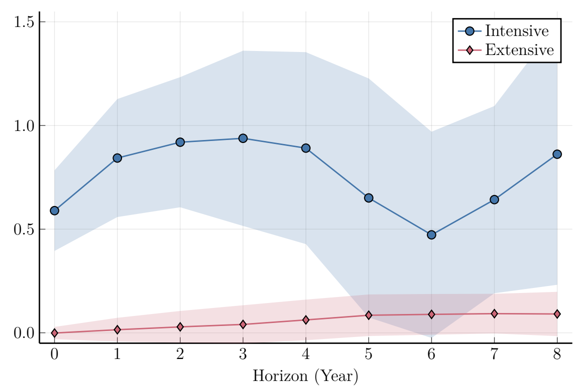


Figure B3: Decomposing the response of the average wealth in the top 400

Notes: The figure reports the estimates for $\beta^{\text{intensive}}_{p,h}$ and $\beta^{\text{extensive}}_{p,h}$ from the regression model (45) for $0 \le h \le 8$ as well as their 5%–95% confidence intervals using heteroskedasticity consistent standard errors. At each horizon $0 \le h \le 8$, the intensive and extensive estimates sum up exactly to the coefficients plotted in Figure 1e; that is, $\beta^{\text{intensive}}_{p,h} + \beta^{\text{extensive}}_{p,h} = \beta_{p,h}$.

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 (8) gives

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

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

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

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

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

where $\alpha_E(x) = \min\left(\alpha_E, \frac{1}{x}\right)$. This equation reflects the feedback loop discussed in (25). This can be used to express the volatility of χ and p using Ito's lemma:

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

Moreover, market clearing for the risky asset (9) gives an expression for $(\mu_R - r)(x)$:

$$\begin{aligned}1 &= x_t\alpha_E(x) + (1-x)\left(\frac{1}{\gamma}\frac{(\mu_R - r)(x)}{\sigma_R(x)^2} + \frac{1-\gamma}{\gamma}\frac{\sigma_{\chi}(x)}{\sigma_R(x)}\right) \\ \implies (\mu_R - r)(x) &= \frac{1 - x\alpha_E(x)}{1-x}\gamma\sigma_R(x)^2 + (\gamma-1)\sigma_{\chi}(x)\sigma_R(x), \end{aligned} \tag{46}$$

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 (46) from the expression for $\mu_R(x)$ in (4) gives an expression for $r(x)$.

Step 4. Plugging all of these quantities into the household's HJB equation (11) 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:

$$0 = F(\boldsymbol{\chi}_{i+1}) - \frac{\boldsymbol{\chi}_{i+1} - \boldsymbol{\chi}_i}{\Delta}. \tag{47}$$

Each update requires solving a non-linear equation (it corresponds to a fully implicit Euler method). Economically, it is equivalent to solve 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 enough 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 and, therefore the algorithm becomes like Newton-Raphson. In particular, the convergence is quadratic around the solution. I stop the iteration as soon as $F(\boldsymbol{\chi}_i)$ 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 an semi-implicit Euler method). I find that allowing multiple inner iterations and adjusting Δ dynamically are important to ensure convergence of this non-linear PDE. This solution method is useful to solve other asset pricing models globally. I uploaded it as an online package for Julia `https://github.com/matthieugomez/EconPDEs.jl`

C.2 Stability of linear functionals

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

Lemma 1. *Let $x_t \in \mathbb{R}$ be a continuous-time strong Markov process non-explosive, irreducible, positive recurrent with unique invariant probability measure. Denote* P *(resp.* E*) denotes the probability measure (resp. expectation) with respect to the invariant probability measure of x. Consider the process*

$$\mathrm{d}M_t = \Big(\mu(x_t)M_t + b(x_t)\Big)\,\mathrm{d}t + \sigma(x_t)M_t\,\mathrm{d}Z_t, \tag{48}$$

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

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

Proof. While this result is well known in the discrete-time case (e.g. Vervaat, 1979), 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 (1959). I then apply 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:

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

where A_n and B_n are i.i.d over time. Moreover, A_1 is positive a.s., B_1 is non negative a.s. with $\mathrm{P}(B_1 > 0) > 0$ and $\mathrm{E}[\log B_1] < \infty$. As proven by Vervaat (1979), X_n converges in distribution if $\mathrm{E}[\log A_1] < 0$ (i.e. if $\mathrm{E}\left[\int_{T_1}^{T_2} \left(\mu(x_u) - \frac{1}{2}\sigma(x_u)^2\right) \mathrm{d}u\right] \geq 0$), and converges a.s. to infinity if $\mathrm{E}[\log A_1] > 0$. Finally, as shown in Maruyama and Tanaka (1959), any integrable function f, $\mathrm{E}\left[\int_{T_1}^{T_2} f(x_u)\,\mathrm{d}u\right] \geq 0$ iff $\mathrm{E}\left[f(x)\right] \geq 0$. This gives the result. □

C.3 An analytical formula for the average wealth above a threshold

This following result, which is a corollary of Proposition 2, derives an analytical expression for the average (normalized) wealth above a threshold.

Corollary 1. *We have, for $q \in \mathbb{R}$,*

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

where $\Phi(\cdot)$, $\mu_{j,s\to t}$ and $\nu_{j,s\to t}$ are defined in Proposition 2.

Proof. Using the law of iterated expectations, we have:

$$\mathbb{E}_t\left[w_{it} 1_{w_{it} \geq q} | i \in \mathbb{I}_{jt}\right] = \int_{-\infty}^{t} (\eta+\delta) e^{-(\eta+\delta)(t-s)} \mathbb{E}_t\left[w_{it} 1_{w_{it} \geq w} | a_{it} = t-s, i \in \mathbb{I}_{jt}\right] \mathrm{d}s, \tag{49}$$

where a_{it} denotes the age of individual i at time t. We know from the proof of Proposition 2 that, within each cohort, log wealth is normally distributed with mean $\mu_{j,s\to t}$ and standard deviation $\nu_{j,s\to t}$. To conclude the proof, it is enough to show that $\mathbb{E}\left[e^Z 1_{Z\geq z}\right] = e^{\mu + \frac{1}{2}\nu^2}\left(1 - \Phi\left(\frac{z-\mu}{\nu} - \nu\right)\right)$ for a Gaussian random variable Z with mean μ and standard deviation ν. Indeed, we have

$$\begin{aligned}
\mathbb{E}\left[e^Z 1_{Z\geq z}\right] &= \int_z^\infty e^u \frac{1}{\sqrt{2\pi\nu^2}} e^{-\frac{1}{2}\frac{(u-\mu)^2}{\nu^2}} \mathrm{d}u \\
&= \int_{(z-\mu)/\nu}^\infty e^{(\mu+\nu v)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} \mathrm{d}v \qquad \text{(using the change of variable } v = (u-\mu)/\nu) \\
&= e^{\mu+\frac{1}{2}\nu^2} \int_{(z-\mu)/\nu}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v-\nu)^2} \mathrm{d}v \\
&= e^{\mu+\frac{1}{2}\nu^2} \int_{(z-\mu)/\nu-\nu}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \mathrm{d}y \qquad \text{(using the change of variable } y = v-\nu) \\
&= e^{\mu+\frac{1}{2}\nu^2}\left(1 - \Phi\left(\frac{z-\mu}{\nu} - \nu\right)\right).
\end{aligned}$$

□

C.4 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 in relation to 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:

$$\log\left(\frac{A_{t+1}+H_{t+1}}{A_t+H_t}\right) \approx \omega_p \log\left(\frac{A_{p,t+1}}{A_{p,t}}\right) + (1-\omega_p)\log\left(\frac{H_{p,t+1}}{H_{p,t}}\right). \tag{50}$$

Projecting this approximation on stock returns (as in specification 1) 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 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 the U.S. on average from 1992 to 2014 (see Appendix Table D5). 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 $\omega_p \approx 13.2\%$.[57]

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-

[57] Similarly, Appendix Table B4 reports that the share of labor income in the top 0.01% is 14% using data from the Survey of Consumer Finances.

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 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 die, so that this income on average depreciates at rate ϕ relative to the economy. Hence, both 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 decision only depend on total wealth. In particular, this distinction between 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 $+\infty$ $\delta < \phi$.[58] 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, and 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.

[58]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 $(\eta + \delta)/(\delta - \phi)$.

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 the log return of entrepreneurs and the growth rate of the economy separately 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 may potentially reflect 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 $t+1$

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

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:

$$\log\left(\frac{\overline{W}_{t+1}}{\overline{W}_t}\right) = \log \overline{R}_{t+1} + \log\left(1 + \frac{\overline{Y}_{t+1} - \overline{T}_{t+1} - \overline{C}_{t+1}}{\overline{W}_t}\right)$$
$$\approx \log \overline{R}_{t+1} + \frac{\overline{Y}_{t+1}}{\overline{W}_t} - \frac{\overline{T}_{t+1}}{\overline{W}_t} - \frac{\overline{C}_{t+1}}{\overline{W}_t}. \tag{51}$$

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.[59] I then obtain the consumption rate $\overline{C}_{t+1}/\overline{W}_t$ as a residual.

Table D5 reports the results. The average annual wealth growth of households in Forbes 400 over the 1982-2017 time period is 5.1% (in real term); 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 (51), 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 D5: 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 (51) (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 Stochastic derivative

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

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

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

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:

$$\frac{\partial x_t}{\partial x_0} = e^{-\kappa t}\frac{\phi(x_0)}{\phi(x_t)}\hat{M}_t, \tag{52}$$

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

$$\mathbb{T} : g \to \lim_{h\to 0}\frac{1}{h}\left(\mathrm{E}\left[\frac{\partial x_h}{\partial x_0}g(x_h)|x_0 = x\right] - g(x)\right). \tag{53}$$

Plugging the decomposition (52) into the definition for the IIRF (21) gives:

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

This implies that $\mathrm{IIRF}_g(x,h)$ decays with the horizon h at an exponential rate rate κ. Note that this decay rate does not depend on the initial state x nor on the function $g(\cdot)$.[60] After computing numerically κ as the (opposite) of the principal eigenvalue of the operator $\mathbb{T}$ defined in (52), I obtain $\kappa \approx 0.05$.

A generalization of Feynman-Kac formula. The following lemma, which is a generalization of the Feyman-Kac formula, provides a way to compute analytically a large class of expectations involving the stochastic derivative $\partial x_t/\partial x_0$. In particular, I use it to compute analytically the IIRF defined in (21), the Campbell-Shiller decomposition in Proposition 4, and the impulse response of entrepreneurs' wealth in Proposition 5.

Lemma 2. *Given a set of smooth functions f, g, v, the function*

$$u(x,h) \equiv \mathrm{E}\left[\int_0^h e^{-\int_0^t v(s)\,\mathrm{d}s}\frac{\partial x_t}{\partial x_0}f(x_t)\,\mathrm{d}t + e^{-\int_0^h v(s)\,\mathrm{d}s}\frac{\partial x_h}{\partial x_0}g(x_h)\Big|x_0 = x\right] \tag{54}$$

can be computed numerically as the solution of the linear PDE

$$\partial_h u(x,h) = f(x) + \Big(\partial_x\mu_x(x) - v(x)\Big)u(x,h) + \Big(\mu_x(x) + \sigma_x(x)\partial_x\sigma_x(x)\Big)\partial_x u(x,h) + \frac{1}{2}\sigma_x^2(x)\partial_{xx}u(x,h) \tag{55}$$

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

Proof. Equation (54) implies the following recurrence relation for $0 < \tau < h$

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

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

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

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

Applying Ito's lemma together with the law of motion of $\partial x_t / \partial x_0$ (22) implies that u solves the following PDE

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

The initial boundary condition can be obtained by taking (54) with $h = 0$. $\square$

I now briefly discuss how to solve this linear PDE using a finite difference method. 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 (53):

$$\mathbb{T} : u \to \lim_{h \to 0} \frac{1}{h} \left(\mathrm{E}\left[\frac{\partial x_h}{\partial x_0} u(x_h) | x_0 = x \right] - u(x) \right) = (\partial_x \mu_x) u + \Big(\mu_x + \sigma_x \partial_x \sigma_x\Big) \partial_x u + \frac{1}{2} \sigma_x^2 \partial_{xx} u.$$

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:

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

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

$$\frac{\boldsymbol{u}^{j+1} - \boldsymbol{u}^j}{\Delta h} = \boldsymbol{f} + (\mathbb{T} - \mathrm{Diag}(\boldsymbol{v}))\, \boldsymbol{u}^{j+1},$$

where $\boldsymbol{f} = (f(x_i))_{0 \le i \le N-1}$ and $\boldsymbol{v} = (v(x_i))_{0 \le i \le N-1}$ denote the discretized versions of the functions f and v on the grid $\boldsymbol{x}$. Indeed, this equation corresponds to the PDE (55), discretized with respect to the horizon h and with respect to the state variable x.

D.3 Decomposing the response of the average wealth of surviving entrepreneurs

Remember the expression (30) obtained in the main text for the impulse response of the average wealth of entrepreneur wealth:

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

The following proposition further decomposes the second term in this decomposition; that is, the effect of changes in asset valuations on entrepreneurs' wealth.

Proposition 6. *The effect of changes in asset valuations (defined in (30)) can be decomposed into two terms:*

$$\begin{aligned}
&\underbrace{(\alpha_E - 1)\sigma_p(x) + \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\partial_x\left(\mu_{wE} - \frac{1}{2}\sigma_{wE}^2\right)(x_t)\,\mathrm{d}t\middle| x_0 = x\right]\sigma_x(x)}_{\substack{\textit{Effect of changes}\\ \textit{in asset valuation}}} \\
&\approx \underbrace{\mathrm{E}\left[(\alpha_E - 1)\frac{\partial x_h}{\partial x_0}\partial_x \log p(x_h)\middle| x_0 = x\right]\sigma_x(x)}_{\textit{Revaluation effect} > 0} \\
&\quad \underbrace{+\,\mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0}\partial_x\left(r + \alpha_E\left(\frac{1}{p} - r_t\right) - \frac{1}{2}\alpha_E(\alpha_E - 1)\sigma_R^2\right)(x_t)\,\mathrm{d}t\middle| x_0 = x\right]\sigma_x(x).}_{\textit{Accumulation effect} < 0}
\end{aligned}$$

This proposition decomposes the effect of changes in asset valuations on the normalized wealth of entrepreneurs 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 that they accumulate. The fact that the second term is negative reflects the fact that higher asset valuations mean that entrepreneurs receive less income per unit of wealth. The approximation is obtained by neglecting the effect of aggregate shocks on the equity exposure of entrepreneurs α_E (which is accurate as α_E remains constant over most of the state space (6)).

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:

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

In other words, the endogenous response of asset valuations to aggregate shocks plays an amplifying effect on entrepreneurs' wealth in the short-run but a dampening effect in the longer-run.

Proof of Proposition 6. Using the expression for μ_{wE} and σ_{wE} given in (14), we have:

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

Plugging this formula into the expression for $\epsilon(x, h)$ obtained in Proposition 5, we get:

$$\begin{aligned} \epsilon(x,h) &= (\alpha_E - 1)(\sigma + \sigma_p(x)) \\ &+ \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_t} \partial_x \left(r + \alpha_E \left(\frac{1}{p} - \phi - r \right) - \frac{1}{2}\alpha_E(\alpha_E - 1)\sigma_R^2 \right)(x_t)\, \mathrm{d}t \Big| x_0 = x \right] \sigma_x(x) \\ &+ \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_t} \partial_x \left((\alpha_E - 1)\left(g - \frac{1}{2}\sigma^2 + \mu_p - \frac{1}{2}\sigma_p^2 \right) \right)(x_t)\, \mathrm{d}t \Big| x_0 = x \right] \sigma_x(x). \end{aligned} \tag{56}$$

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

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

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 (56) gives

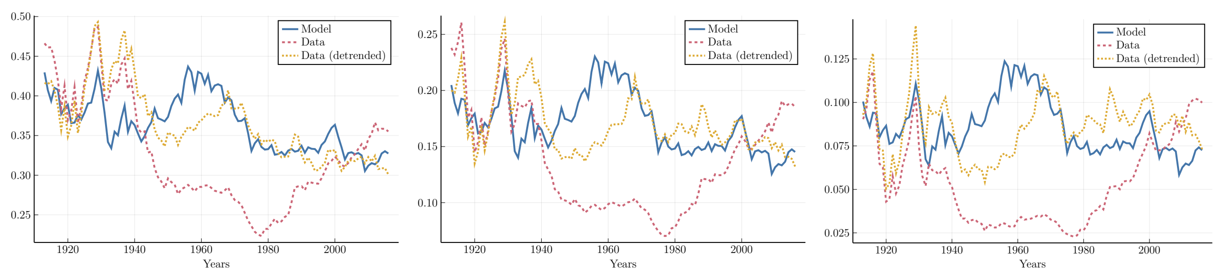
$$\begin{aligned} \epsilon(x,h) = &(\alpha_E(x) - 1)\sigma + (\alpha_E(x) - 1)\mathrm{E}\left[\frac{\partial x_h}{\partial x_0} \partial_x \log p(x_h) | x_0 = x \right] \sigma_x(x) \\ &+ \mathrm{E}\left[\int_0^h \frac{\partial x_t}{\partial x_0} \partial_x \left(r + \alpha_E \left(\frac{1}{p} - \phi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

The third term becomes zero under the approximation $\alpha_E \approx \alpha_E(x_0)$, which gives the result. □

D.4 Historical dynamics of top wealth shares in the model and in 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 be explained by this mechanism? 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 D4 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 cycles 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 this 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.[61] One can see that the two series remain very close over the time sample — in particular, it is hard to know whether the remaining discrepancy reflects some other economic forces or simply some measurement error in realized top wealth shares.[62]



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

Figure D4: Historical dynamics of top wealth shares (model versus 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) as well as a detrended version after adjusting for the structural break in the logarithmic growth of top wealth shares in 1979.

D.5 Additional tables and figures

Table D6 reports the average level of top wealth shares in the data and in the calibrated model. One can see that the model matches very well the average level of top wealth shares across top percentiles. 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 the fact that the tail index of the wealth distribution is slightly smaller in the model (1.43) relative to the data (1.5).

Figure D5 compares the estimates obtained from local projections of excess stock market re-

[61] 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.

[62] An alternative way to "detrend" realized top wealth shares would be to use the Hodrick-Prescott filter. The goal of this filter is to isolates business cycle dynamics. However, this filter is not adapted to our purpose as aggregate shocks generate long lived fluctuations in top wealth shares (see Figure 5).

Table D6: Average level of top wealth shares (model versus 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).

turns on the average wealth in top percentile p at horizon h, in the model and in the data. The "data" estimates correspond exactly to the ones plotted in Figure 1. To facilitate the comparison between the 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.[63] 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.[64]

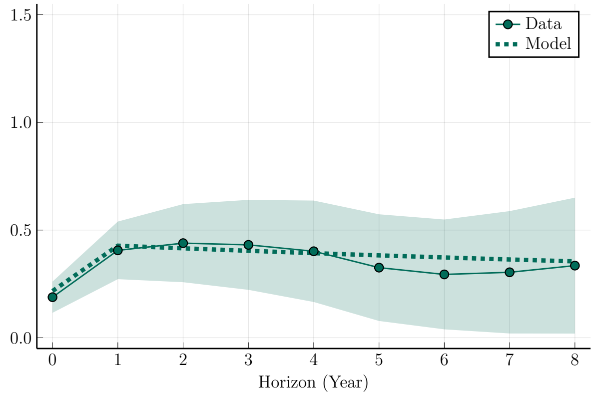
As a complement to the impulse response functions plotted in the main tex (Figure 2), Figure D7 reports important economic quantities as a function of the state variable x. More precisely, Figure D7a plots the drift and volatility of the state variable, x, as well as its associated stationary density represented as a shaded area. Figure D7b plots the price-to-income ratio $p(x)$ and the wealth-to-consumption ratio $c_H(x)$ of households. Finally, Figure D7c plots the risk-free rate and the expected log stock market return (i.e., the return on levered equity).

Table D7 reports the sensitivity of asset price moments with respect to three parameters relating to entrepreneurs (their subjective discount rate ρ_E, their exposure to aggregate risk α_E, and their proportion in the population π) and three parameters relating to households (their subjective discount rate ρ, their risk aversion γ, and their elasticity of intertemporal substitution ψ). One important takeaway is that household preferences that tend to increase the standard deviation of returns also tend to increase the average interest rate.

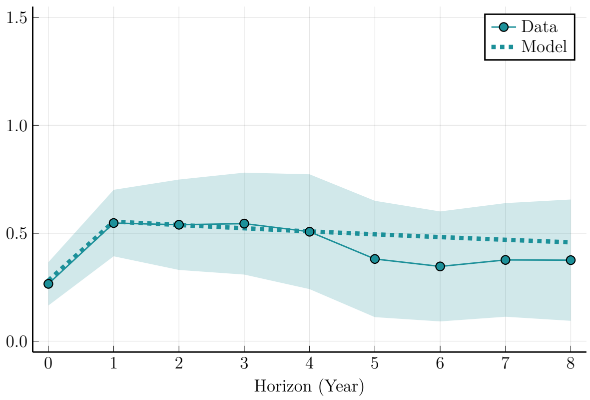
Figure D8 plots the decomposition for $\sigma_p(x)$ in terms of a "risk-free rate channel" and an "excess return channel" defined in Proposition 4 as a function of x.

[63]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 means that this bias is small in the calibrated model.

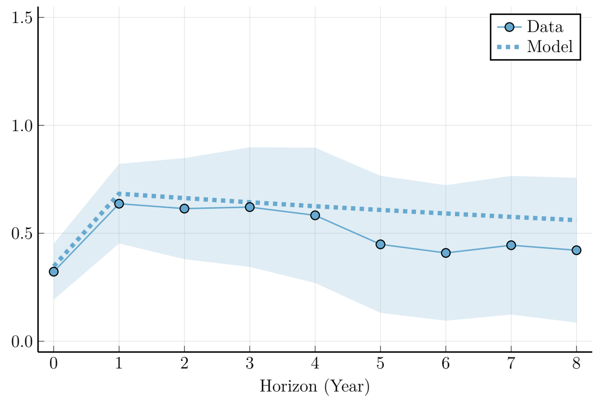
[64]This is important to capture the fact that, in the data, the effect of stock market returns is lower at time $h=0$ relative to $h=1$.



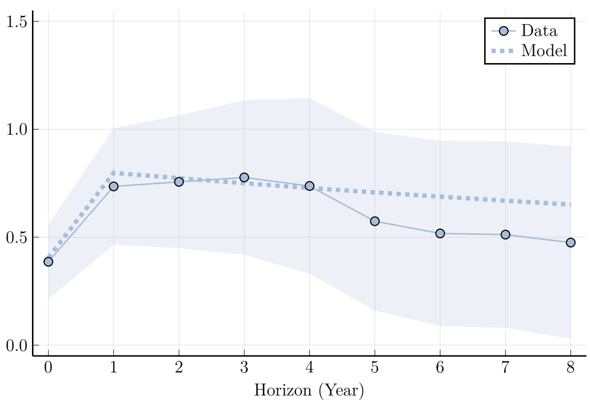
(a) Top 100%



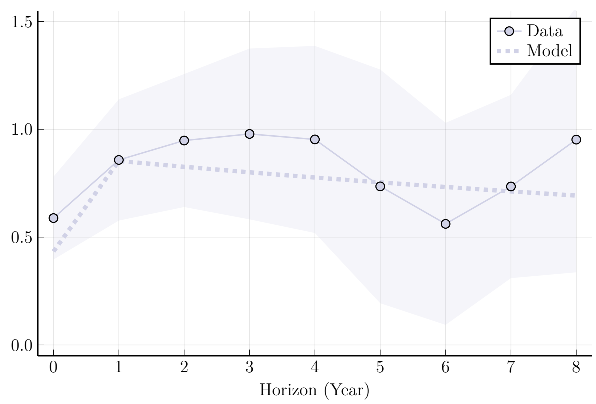
(b) Top 1%



(c) Top 0.1%



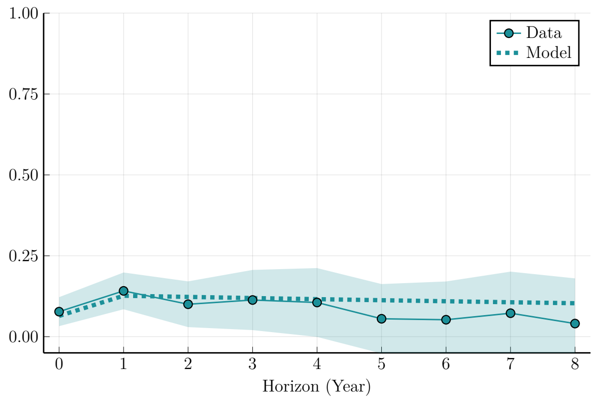
(d) Top 0.01%



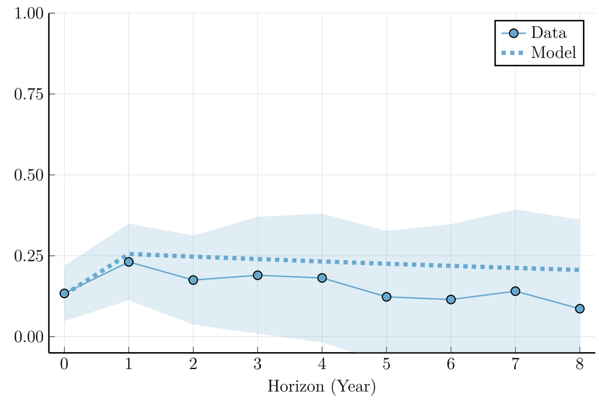
(e) Top 400

Figure D5: Response of wealth in top percentiles to excess stock returns (model versus data)

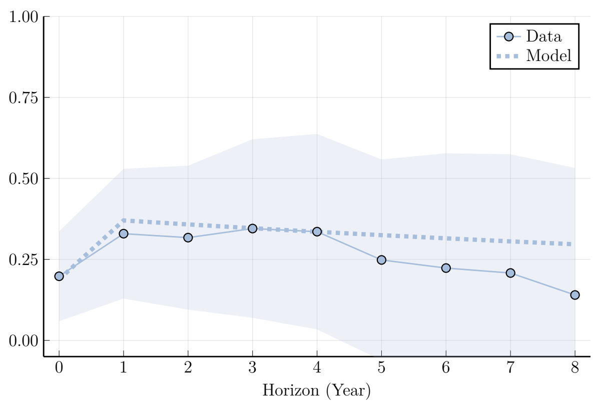
Notes: The figure reports the estimates for $\beta_{p,h}$ estimated via the regression (1) for $0 \leq h \leq 8$, as well as their 5%–95% confidence intervals using heteroskedasticity consistent standard errors. Each figure corresponds to a different top percentile. Figure D5a corresponds to all U.S households ($p = 100\%$). Figures D5b-D5d correspond to the top 1%, 0.1%, 0.01% using data from Saez and Zucman (2016). Figure D5e corresponds to Forbes 400. 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.



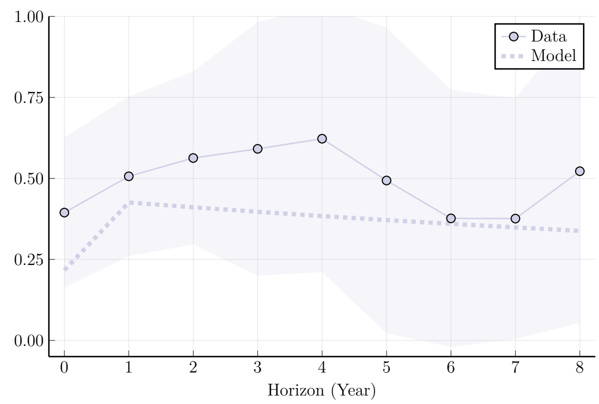
(a) Top 1%



(b) Top 0.1%



(c) Top 0.01%



(d) Top 400

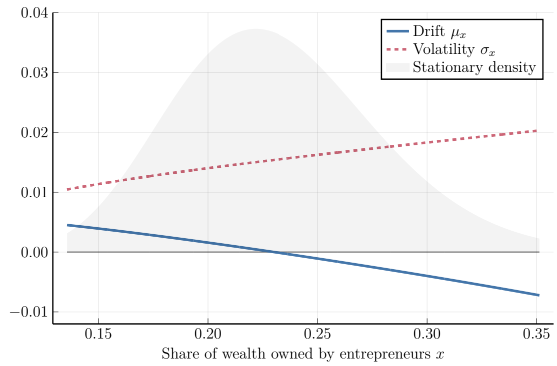
Figure D6: Response of top wealth shares to excess stock returns (model versus data)

Notes: The figure reports the estimates for $b_{p,h}$ 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. Figures D6a-D6c correspond to the top 1%, 0.1%, 0.01% using data from Saez and Zucman (2016). Figure D6d corresponds to Forbes 400. 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.

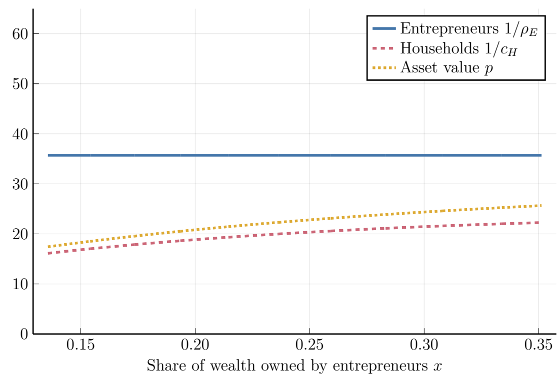
Table D7: Sensitivity analysis

	Entrepreneurs' parameters			Households' preferences		
	ρ_E	α_E	π	ρ	$1/\gamma$	ψ
Average risk-free rate	0.022	0.038	0.004	0.015	0.040	0.008
Standard deviation risk-free rate	−0.010	0.018	−0.001	0.008	0.021	0.001
Average equity return	0.025	−0.019	−0.014	0.003	−0.006	0.002
Standard deviation equity return	−0.161	0.284	0.000	0.092	0.182	0.001

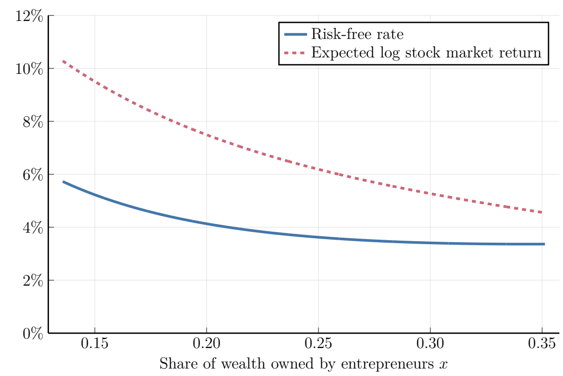
Notes: The table reports the effect of changing each parameter by 1% on asset price moments. More precisely, for each moment m_i and parameter θ_j, the table reports $\frac{m_i((1+\varepsilon)\theta_i)-m_i((1-\varepsilon)\theta_i)}{2\varepsilon}$ with $\varepsilon = 0.5$.



(a) Drift and volatility of the state variable



(b) Wealth-to-consumption ratios



(c) Expected log returns

Figure D7: 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, $r + \lambda(\mu_R - r) - \frac{1}{2}\lambda^2\sigma_R^2$ (19). The bounds of the x-axis correspond to the 1% and the 99% quantiles of the state variable.

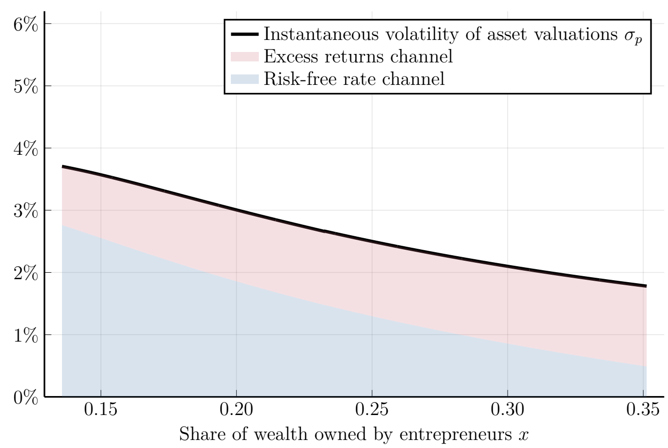


Figure D8: Decomposing the impulse response of asset valuations to aggregate shocks

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" given in (26). The bounds of the x-axis correspond to the 1% and the 99% quantiles of the state variable.