Counterfactual Wealth Distributions*

MATTHIEU GOMEZ †

First version: March 2024
This version: April 2024
[latest version]

Abstract

This paper studies the response of the wealth distribution to small changes in the economic environment. I express the first-order response of top wealth shares as a weighted average response of individual wealth. This allows me to obtain closed-form formulas for the impulse response function of top wealth shares to perturbations in individual wealth dynamics. This analytical approach helps to isolate the key forces and empirical moments that determine the response of wealth inequality to broad economic shocks. It can also serve as a simple accounting framework to trace out the dynamics of top wealth shares under various counterfactual scenarios. I leverage these results to revisit existing issues in the literature, such as the effect of transitory and permanent shocks in asset returns on wealth inequality.

*I am grateful to seminar participants at Columbia and Penn for comments.
†Columbia University; mg3901@columbia.edu

Introduction

An important goal for policy making is to understand what drives movements in the wealth distribution or how different economic policies would affect wealth inequality. One common way to answer these questions is to build models of the wealth distribution and use them to simulate the effect of policy changes on wealth inequality. One drawback of these models is that they can be fairly complex, making it challenging to discern the forces within the models that are responsible for specific outcomes.

This paper develops a tractable and transparent framework to analyze the dynamics of top wealth shares under various counterfactual scenarios. I show that the first-order response of top wealth shares can be expressed as the weighted average response of individual wealth. This allows me to derive closed-form formulas for the dynamics of top wealth shares in response to perturbations in individual wealth dynamics in a wide range of random growth models.

This analytical approach helps to isolate the key forces and measurable moments that shape how the wealth distribution responds to broad economic changes. This is useful for more structural approaches as it can function as a "diagnostic" tool for existing models of wealth inequality. This is also useful for more reduced-form approaches as it can serve as a simple accounting framework to compute wealth inequality counterfactuals. More precisely, I show how these analytical results can be combined with data on the wealth trajectory of individuals to trace out the dynamics of top wealth shares after changes in the economic environment.

The paper is organized in four parts. In the first part, I show that, in a wide range of models, the first-order response of the average wealth in a top percentile is simply given by the average wealth response of individuals in the top percentile. Put differently, while economic shocks typically generate composition changes in the top percentile, these changes are second-order for the response of the average wealth in top percentiles. Intuitively, this comes from the fact that composition changes are the product of the relative mass of individuals entering the top percentile and the average impact of these entrants on the average wealth in the top percentiles: as both terms are first-order in the size of the perturbation, their product is second-order. The key assumption is that wealth of each individual is differentiable in the size of the perturbation.

One important subtlety is that the response of the average wealth in a top percentile at horizon h is equal to the average wealth response at horizon h of individuals *who will be in the top percentile* at horizon h, and *not* the individuals initially in the top percentile (at $h = 0$). While composition changes in top percentiles due to a small deviation in the economy are second-order, composition changes due to the normal passage of time are not. Still, this does not complicate the application of this formula too much as we can always take these composition changes as given.

To examine the accuracy of this first-order approximation, I also derive a second-order approximation for the response of the average wealth in top percentiles. This second-order term, that captures the effect of composition changes, depends on the heterogeneity in the response of individual wealth and the mass of individuals around the percentile threshold. Intuitively, the higher the heterogeneity in wealth deviations, and the higher the density of people around the threshold, the higher the effect of composition changes. Still, holding other things equal, one can expect the first-order approximation to be more accurate for distributions that are highly unequal, as this reduces the extent to which a perturbation in the economic environment induces re-ranking.

Second, I characterize the deviation of individual wealth in terms of perturbation in wealth dynamics in a large class of random growth models. This allows me to obtain simple formula for the impulse response of top wealth shares to changes in the dynamics of individual wealth in a wide range of random-growth models. This is very useful since a large part of the inequality literature uses such processes to model the dynamics of the wealth distribution.

I first focus on the simple case in which individual wealth follows a geometric random walk with population renewal (e.g. death or population growth). In this case, I obtain a straightforward formula for the dynamics of top wealth shares in response to a uniform change in individual growth rates. In particular, the long-run response of top wealth shares is simply equal to the change in growth rates times the difference between the (wealth-weighted) average age of individuals in the top percentile and individuals in the overall economy. A similar result holds for the more general case of a type-specific change in growth rates: in this case, the long-run response depends on the average number of years spent in that type for individuals across the wealth distribution.

I then extend these results to the more general case in which individual wealth follows a geometric random walk with an additive term. In this case, I show that one can obtain similar formulas after extending the concept of age to be an index discounting previous periods based on how significantly the additive force affects wealth dynamics. For instance, for an individual that receives a bequest at birth, this index depends on the ratio of the bequest to the individual's initial wealth.

I discuss how these analytical findings can serve as diagnostic tools for structural researchers. One general result is that deviations in top wealth shares are determined by two things: the direct effect of the perturbation on the growth rate of wealth and the elasticity of terminal wealth to wealth shocks over time (which depends on how persistent wealth shocks are). One diagnostic tool for models of wealth inequality interested in running counterfactuals could be to use the analytical findings to distinguish between these two effects. In particular, I show that existing models of the wealth distribution report very different values for this elasticity of individual wealth to changes in growth rate, which implies that these models predict very different responses for the exact same counterfactual.

In the third part of the paper, I argue that these analytical results can be used as diagnostic tools for existing models of top wealth inequality. I use off-the-shelf models of inequality from the recent literature and I compute the effect of a permanent increase in the rate of wealth accumulation by 1 percentage point on top wealth shares in each model. There are two main key takeaways: the first is that there is a significant heterogeneity across models for the behavior of top wealth shares in response to this counterfactual. The second is that my analytical formulas, which are first-order approximations, provide a very good quantification of these counterfactual changes. Hence, these analytical results can be used as a diagnostic tool to better understand how existing models of wealth inequality obtain such different counterfactuals.

Finally, in the last part of the paper, I leverage my analytical results to develop a simple accounting framework to compute counterfactuals on the wealth distribution. I propose a simple model of wealth accumulations that is broadly consistent with the data. The key implication of the model is that the elasticity of a household's current wealth to their prior wealth can be directly read from the relative importance of labor income and bequest in their wealth accumulation.

I then combine this model with data on the lifetime wealth trajectory of individuals at the top of the wealth distribution, from the Survey of Consumer Finances and the Forbes 400 list. I argue that my accounting framework provides credible and transparent answers to important questions in the inequality literature such as: What is the role of labor income inequality and bequests for wealth inequality? How would top wealth shares evolve if the average return on wealth were to increase? How does the wealth distribution respond to transitory or permanent shocks in the returns of specific assets?

Literature review. There is a growing literature trying to understand the effect of various changes in economic environment on wealth inequality. One strand of the literature has turned to large macro models that are able to capture the rich heterogeneity of the micro data (e.g. De Nardi, 2004, Kaymak and Poschke, 2016, Benhabib et al., 2019, Peter, 2019, Hubmer et al., 2021, Guvenen et al., 2023...). One drawback of these models is that they can be fairly complex, making it challenging to discern the forces within the models that are responsible for specific outcomes.

Another strand of the literature relies on more stylized and analytically tractable "random growth" (e.g. Wold and Whittle, 1957, Benhabib et al., 2011, Jones, 2015, Piketty and Zucman, 2015, Moll et al., 2022). While these models are able to obtain closed-form formula for the effect of changes in Pareto inequality, this is only informative about the asymptotic behavior of top wealth shares, both with respect to time (long-run) and with respect to wealth (right-tail). I will stress that these results can be very uninformative about the deviation of top wealth shares at a given time or at a given top percentile.

Finally, the paper contributes a more reduced-form literature that computes counterfactuals on wealth inequality. One widespread approach, started with Saez and Zucman (2016), is to use an accounting decomposition of wealth in top percentiles assuming that the set of households in the top percentile remains the same over time (see Martínez-Toledano, 2020 or Mian et al., 2020 for applications). The accounting framework discussed in the last section of the paper can be seen as a version of this approach that does not require this assumption — while the difference between the two does not matter much in the short-run (say, a few years), it can lead to dramatically different results in the longer-run (say, a few decades). My accounting approach is closer, in spirit, to the long-run counterfactual done by Feiveson and Sabelhaus (2018) and Ozkan et al. (2023), who ask how wealth inequality would change if the lifetime trajectory of the individuals in the top 0.01% were to change. This insight is related to Gomez and Gouin-Bonenfant (2024), who show that the long-run effect of a perturbation on Pareto inequality depends on its effect on the entire trajectory of individuals who end up in the top percentile.

My analytical approach provides closed-form formulas for the transition dynamics of top wealth shares in response to (small) changes in the growth rate of individual wealth. This improves on the existing literature, that typically resort to numerical method (Kolmogorov-Forward equation) or simulations to compute the response of top wealth shares to aggregate shocks. Moreover, my results on the transition path of top wealth shares complements Gabaix et al. (2016), Luttmer (2016), and Atkeson and Irie (2022), who characterize analytically the convergence of densities according to global norms (e.g. the integral of the absolute difference, or the difference in moments). For the case of a mechanical effect in the growth rate of wealth, I show that one key statistic is the gradient of age with respect to wealth. This relates this paper to Luttmer (2011) who computes this gradient in a model of firm growth and compares it with the data. Finally, the analytical approach in this paper relies on similar tools as Gomez (2023), who studies the change in top wealth shares over a small period of time (rather than deviations in top wealth shares due to a small perturbation).

My results are closely related to Saez and Zucman (2019), who studies the effect of an increase in billionaire taxation on the average wealth of future billionaires. In particular, they show that, in a model in which wealth follows a random-walk, the infinitesimal effect of an increase in wealth tax can be expressed as the wealth-weighted average number of years that fortunes have been exposed to the wealth tax, which is similar to the type of results I obtain in Section 2. Relative to this paper, my contribution is to generalize this framework to other types of perturbations, characterize the condition under which composition changes are first-order (depending on the baseline economy and the type of the perturbation), and examine their second-order effects, both analytically and quantitatively.

At a general level, this paper contributes to the growing work employing first-order approximations to characterize economies with a large degree of micro heterogeneity — see Chetty (2009) and Kleven (2021) for a review of this approach in public finance, Costinot and Rodríguez-Clare (2014) in trade, and Baqaee and Rubbo (2023) or Auclert et al. (2018) in macroeconomics.

Roadmap. The paper is organized in three sections. Section 1 characterizes the response of the average wealth in the top percentile in terms of the average response of individual wealth in the top percentile. Section 2 characterizes the response of individual wealth in terms of perturbation in the dynamics of individual wealth. Combining these results allow me to obtain simple formula for the effect of perturbation in individual wealth on top wealth shares. Section 3 presents a reduced-form approach to compute counterfactuals using data on the trajectory of individuals across the wealth distribution.

1 Counterfactual Deviations in Top Wealth Shares

In this section, I study perturbations in the distribution of wealth around a baseline economy. Section 1.1 presents a general decomposition for the deviation in the average wealth in a top percentile in terms of an intensive margin (the average wealth deviation of individuals in the top percentile) and an extensive margin (deviation in the composition of individuals in the top percentile). Section 1.2 and Section 1.3 then derives simple expressions for these terms at the first and second-order, respectively.

1.1 Deviation in the Average Wealth in the Top Percentile

Consider an economy indexed by time $t \in \mathbb{R}$ with a continuum of individuals indexed by i. Assume that the cumulative distribution function of wealth is absolutely continuous and increasing, with a finite average.[1] Denote W_{it} the wealth of individual i at time t. Consider a given top percentile p (say, the top $p = 1\%$), and denote Q_t the quantile corresponding to the top percentile p at time t:

$$\mathbb{P}(W_{it} \geq Q_t(p)) = p \tag{1}$$

Denote $\overline{W}_t$ the average wealth in the top percentile p:

$$\overline{W}_t = \mathbb{E}[W_{it} | W_{it} \geq Q_t]. \tag{2}$$

Consider a change in the economic environment compared to a baseline scenario (e.g. deviation in technology or asset prices) indexed by θ, starting from time $t = 0$. For any economic variable Y_t in the baseline economy (such as individual wealth, quantile, or average wealth in the top), denote $Y_t(\theta)$ its value in the perturbated economy indexed by θ and $\Delta Y_t = Y_t(\theta) - Y_t(0)$ its difference between the counterfactual and the baseline economies. It is important to keep in mind that Δ represents a vertical deviation (i.e. with respect to a perturbation) rather than a horizontal change (i.e. a change with respect to time). To visualize this, Figure 1 plots an example for the wealth path of some individual i over time, in the baseline economy and the perturbated one.

I am interested in the effect of this deviation on the average wealth in the top percentile. I will focus on the response of the average wealth in the top percentile for two reasons. First, it is typically the object measured and reported in existing models of inequality. Second, requirement for differentiability of the average wealth are typically lower relative to quantiles. Finally, note

[1] This is only to simplify the exposition. In reality, one only needs this assumption to hold around the quantile.

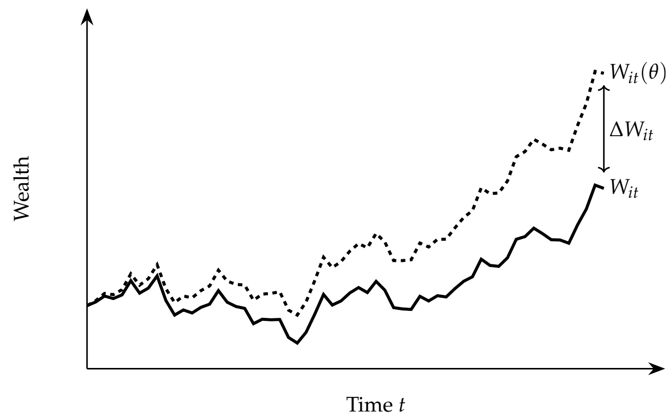


Figure 1: Baseline (solid) and perturbated (dashed) wealth path for some individual i

that the knowledge of the response of all top wealth shares is equivalent to the knowledge of the response of the whole distribution. By definition, the deviation in the average wealth in top percentile is:

$$\Delta \overline{W}_t = \mathbb{E}[W_{it} + \Delta W_{it} | W_{it} + \Delta W_{it} \geq Q_t + \Delta Q_t] - \mathbb{E}[W_{it} | W_{it} \geq Q_t]. \tag{3}$$

The key subtlety is the perturbation typically changes not only the wealth of individuals initially the top percentile, but also the set of individuals in the top percentile (in the baseline versus perturbated economy). More precisely, adding and subtracting (3) by $\mathbb{E}[W_{it} + \Delta W_{it} | W_{it} \geq Q_t]$ gives that the deviation in the average wealth in the top percentile is the sum of two terms:

$$\begin{aligned} \Delta \overline{W}_t = & \underbrace{\mathbb{E}\left[W_{it} + \Delta W_{it} | W_{it} \geq Q_t\right] - \mathbb{E}\left[W_{it} | W_{it} \geq Q_t\right]}_{\text{Intensive margin}} \\ & + \underbrace{\mathbb{E}\left[W_{it} + \Delta W_{it} | W_{it} + \Delta W_{it} \geq Q_t + \Delta Q_t\right] - \mathbb{E}\left[W_{it} + \Delta W_{it} | W_{it} \geq Q_t\right]}_{\text{Extensive margin}}. \end{aligned} \tag{4}$$

The first term corresponds to the average wealth change for the set of individuals in the top percentile in the baseline economy. The second term reflect the effect of changes in the composition of households in the top percentile: they correspond to the counterfactual wealth of individuals that enter the top percentile minus the counterfactual wealth of individuals that exit it. Note that, by definition, the second term is always non-negative.

1.2 First-Order Approximation

We now assume that individual wealth is differentiable with respect to the size of the deviation θ.

Assumption 1. *The function $\theta \to W_{it}(\theta)$ is in $\mathcal{C}^1$ for all individuals.*[2]

[2] Here, and in the rest of the paper, I use $\mathcal{C}^k$ to denote the set of functions that are differentiable $k-$th times with

When this assumption is satisfied, we denote $\mathrm{d}W_{it} = W'_{it}(\theta)\,\mathrm{d}\theta$. This assumption restricts the set of economic deviations to the ones that generate smooth changes in individual wealth. For instance, this assumption allows for deviations in the economic environment that affect the size of individual wealth jumps; however, it rules out deviations that affect the probability of this jump happening. One way to think about this assumption is that it allows for deviations that change the realization of wealth across different state of natures, not for deviations that change the probability of these states happening.[3]

While the first assumption ensures that deviations in wealth are first-order in θ, the second assumption ensures that the mass of individuals that enter or exit is first-order in θ. Together, these two assumptions ensure that the extensive margin in (4) is second-order in θ.

Proposition 1 (First-Order Approximation)**.** *At the first-order in θ, the deviation in the average wealth in a top percentile is:*

$$\mathrm{d}\overline{W}_t = \mathbb{E}\left[\mathrm{d}W_{it} | W_{it} \geq Q_t\right].$$

This proposition says that the first-order deviation in the average wealth in a top percentile is simply given by the average wealth deviation of individuals in the top percentile. Relative to the expression for the non-infinitesimal deviation (4), this proposition says that one can abstract from the extensive margin: the first-order effect of a shock on the average wealth in a top percentile can be computed *as if* the set of individuals in the top percentile remained the same in the perturbated economy. This is very convenient, both analytically (as one does not need to take into account re-ranking effects), but also empirically (as one only needs data on the effect of the perturbation on the wealth of individuals currently in the top)

Composition changes over time. This proposition characterizes the entire impulse response function of the wealth distribution, from its instantaneous response $t = 0$ to its long-run response $t \to \infty$. Despite the simplicity of the formula, note that there is an important subtlety: what matters for the response of the average wealth in the top percentile at some horizon t is the average wealth response of individuals who will be in the top percentile at time t, not the response of those who are initially in the top at $t = 0$. Mathematically:

$$\mathrm{d}\overline{W}_t = \mathbb{E}\left[\mathrm{d}W_{it} | W_{it} \geq Q_t\right] \neq \mathbb{E}\left[\mathrm{d}W_{it} | W_{i0} \geq Q_0\right]. \tag{5}$$

The impulse response of the average wealth in a top percentile is *not* the average impulse response of individuals initially in the top percentile. The difference between the two typically grows with

continuous derivatives.

[3]Still, this more general case could be handled using similar techniques as in Gomez (2023). However, in this case, composition changes would become first-order, as some individuals experience discontinuous changes in wealth in response to small changes in the economic environment

the horizon t, since the population in the top percentile typically renews over time.[4] One way to summarize this point is that, while composition changes in the top percentile induced by the perturbation are first-order in θ, composition changes induced by the normal passage of time are not.[5] We will see many examples of this distinction below.

Quantiles. As shown in the proof of Proposition 1 in Appendix A, the idea that composition changes do not matter at the first order also works for quantiles; that is,

$$\mathrm{d}Q_t = \mathbb{E}\left[\mathrm{d}W_{it} | W_{it} = p\right].$$

In the rest of the paper, however, I focus on the average wealth in top percentiles shares in the rest of the talk as they are typically easier to work with, both theoretically and empirically.

Logarithm. I now extend Proposition 5 to derive a similar formula for the deviation in the *log* average wealth in top percentiles in terms of the deviations in *log* individual wealth. Dividing Proposition 1 by $\overline{W}_t$ and rearranging gives:

$$\mathrm{d}\log \overline{W}_t = \mathbb{E}^{W_{it}}\left[\mathrm{d}\log W_{it} | W_{it} \geq Q_t\right]. \tag{6}$$

In words, the log deviation for the average wealth in the top percentile is the wealth-weighted average log deviation of individuals in the top percentile. Working with logarithms is particularly useful to express the deviation in top wealth shares, as $S_t(p) = p\overline{W}_t(p)/\overline{W}_t(100\%)$ implies that the *log* deviation in the top wealth share is the difference between the *log* deviation for the average wealth in the top and in the economy, i.e. $\mathrm{d}\log S_t(p) = \mathrm{d}\log \overline{W}_t(p) - \mathrm{d}\log \overline{W}_t(100\%)$.

1.3 Second-Order Approximation

I now examine the response of the average wealth in a top percentile beyond the first-order. I find that the second-order term depends on the cross-sectional variance of wealth deviations as well as the shape of the wealth distribution. The more unequal the wealth distribution is, the smaller the effect of composition changes, and, therefore, the better the first-order approximation is.

One can integrate the expression obtained in Proposition 1 to obtain an expression for a non-

[4]Formally, a decomposition similar to (4) gives the following relationship between the two:

$$\mathrm{d}\overline{W}_t = \underbrace{\mathbb{E}\left[\mathrm{d}W_{it} | W_{i0} \geq Q_0\right]}_{\substack{\text{Wealth deviation for} \\ \text{individuals initially in top}}} + \underbrace{\mathbb{E}\left[\mathrm{d}W_{it} | W_{it} \geq Q_t\right] - \mathbb{E}\left[\mathrm{d}W_{it} | W_{i0} \geq Q_0\right]}_{\substack{\text{Difference between the wealth deviation of individuals} \\ \text{that entered and those that exited between 0 and } t\text{ .}}}$$

[5]One reason is that the horizon t is not-infinitesimal, in contrast with the size of the perturbation θ. However, even for infinitesimal horizons, Gomez (2023) show that composition changes are first-order with respect to the horizon whenever wealth paths are not differentiable with respect to time.

infinitesimal change in policy θ

$$\Delta \overline{W}_t = \int_{\vartheta=0}^{\theta} \mathbb{E}\left[\mathrm{d}W_{it}(\vartheta) | W_{it}(\vartheta) \geq Q_t(\vartheta) \right] \mathrm{d}\vartheta. \tag{7}$$

This expression differs from the first-order approximation obtained in Proposition 1 as the set of individuals in the top percentile (i.e., for which $W_{it}(\vartheta) \geq Q_t(\vartheta)$) changes along the policy path.

One can approximate the integral (7) using the trapezoid rule to obtain a *second-order* approximation for the counterfactual change in top shares:

$$\Delta \overline{W}_t = \frac{1}{2} \Big(\mathbb{E}\left[\Delta W_{it} | W_{it} \geq Q_t \right] + \mathbb{E}\left[\Delta W_{it} | W_{it} + \Delta W_{it} \geq Q_t + \Delta Q_t \right] \Big) + o(\theta^2). \tag{8}$$

This second-order approximation expresses the change in the average wealth in the top percentile as the average of two terms: the average wealth deviation of households who are in the percentile in the baseline economy, $\mathbb{E}\left[\Delta W_{it} | W_{it} \geq Q_t \right]$ and the average wealth deviation of households who are in top percentile in the perturbated economy, $\mathbb{E}\left[\Delta W_{it} | W_{it} + \Delta W_{it} \geq Q_t + \Delta Q_t \right]$. It is straightforward to rewrite (8) to obtain:[6]

$$\Delta \overline{W}_t = \underbrace{\mathbb{E}\left[\Delta W_{it} | W_{it} \geq Q_t \right]}_{\text{Intensive margin}} + \underbrace{\frac{1}{2} \Big(\mathbb{E}\left[\Delta W_{it} | W_{it} + \Delta W_{it} \geq Q_t + \Delta Q_t \right] - \mathbb{E}\left[\Delta W_{it} | W_{it} \geq Q_t \right] \Big) + o(\theta^2)}_{\text{Extensive margin}}.$$

The first term, which is first-order in θ, is simply the intensive margin defined in (4). The second term, which is second-order in θ, is half the difference between the average deviation of people that enter the top percentile in the perturbated economy, relative to the baseline economy, and people that exit it. It can be seen as a second-order approximation of the extensive margin defined in (4). The next proposition gives an analytical expression for this term in terms of the variance of the deviation in wealth across individuals.

Proposition 2 (Second-Order Approximation). *Assume that the function $\theta \to W_{it}(\theta)$ is in $\mathcal{C}^2$ and that $W \to (\mathbb{E}[\partial W_{it}], \mathbb{E}[\partial W_{it}^2])$ is continuous. Then, the deviation in the average wealth in a top percentile is:*

$$\Delta \overline{W}_t = \underbrace{\mathbb{E}[\Delta W_i | W_{it} \geq Q_t]}_{\textit{Intensive margin}} + \underbrace{\frac{1}{2} \frac{g_t(Q_t)}{p} \mathbb{V}[\Delta W_i | W_{it} = Q_t] + o(\theta^2)}_{\textit{Extensive margin}},$$

[6] Comparing this term with the exact extensive margin defined in (4) implies that the reminder — the term denoted by $o(\theta^2)$ — is:

$$\mathbb{E}\left[W_{it} + \frac{1}{2}\Delta W_{it} \Big| W_{it} + \Delta W_{it} \geq Q_t + \Delta Q_t \right] - \mathbb{E}\left[W_{it} + \frac{1}{2}\Delta W_{it} \Big| W_{it} \geq Q_t \right].$$

This term, which is third-order in θ, corresponds to the average wealth of people that enter the top percentile (averaged before and after the perturbation) minus the average wealth of people that exit the top percentile. Relative to the second-order term, this term captures the asymmetry of the deviation in average wealth for entering and exiting agents: we can expect this term to be close to zero when the density of agents is close to constant around the threshold and the distribution of wealth deviation is symmetric.

where $\mathbb{V}$ denotes the cross-sectional variance of the deviation in individual wealth.

This proposition provides a formula for the deviation in average wealth that takes into account the second-order effect of composition changes. One nice result is that the deviation in the average wealth in the top percentile still only depend on the wealth deviation of individuals in the top percentile, even at the second-order. Second, it makes it easier to quantify second-order effects due to composition changes, both theoretically and empirically. In particular, note that the second-order term increases with the density of households at the percentile threshold and with the variance of wealth deviations at the percentile threshold, as both increase the importance of composition effects. In Appendix B.2, I use this insight to argue that second-order terms are likely to be small for wealth distributions that are as unequal as in the data.

2 Application to Random Growth Models of Inequality

The previous section expressed deviations of the average wealth in top percentiles in terms of deviations in the wealth of top individuals. I now use this result to characterize the impulse response of top wealth shares to shocks in individual wealth.

I focus on random growth models inequality, which are commonly used in the literature as they can generate realistic distributions for highly skewed economic variables such as wealth, income, firm size or city size. Focusing on statistical models of wealth allows me to maintain a fairly large degree of generality while abstracting from specific economic considerations (as discussed below, a lot of different economic models can give rise to random-growth models). Section 2.1 studies the case where individual wealth follows a geometric random walk with some probability of wealth dissipation (e.g. death). Section 2.2 studies the case where individual wealth follows a geometric random walk plus some additive shock (e.g. labor income). Finally, Section 2.3 discusses my results in the context of the existing literature.

2.1 Geometric Random Walk with Dissipation Shocks

Setup. In this section, I assume that individual wealth follows a random walk until they die (or, more generally, suffer a dissipation shock), in which case their wealth resets to an initial value. This type of statistically process has often been used to study the distribution of wealth (e.g. Piketty and Zucman, 2015, Jones, 2015, Moll et al., 2022, Gomez and Gouin-Bonenfant, 2024), income (e.g. Jones and Kim, 2018), and firm size (e.g. Luttmer, 2011) — I relegate the reader to these papers for micro-foundations. More precisely, I assume that individual wealth evolves as a geometric random walk

$$W_{it+1} = G_{it+1} W_{it}, \tag{9}$$

until they suffer a "dissipation shock", in which case wealth is reset to some random value $Y_{it+1} > 0$. One natural interpretation of this dissipation shock is death or population growth, but the concept is more general.[7] Importantly, and this is an key generalization relative to existing random-growth models, I allow the joint distribution of the initial wealth Y_{it}, the growth rate of wealth, G_{it} and of the dissipation shock to be arbitrarily correlated over time, and, potentially, to depend on time t and individual i.[8]

In this model, individual wealth is positive almost sure. Hence, taking logs and iterating backward on (9) allows us to write log individual wealth as the cumulative sum of past shocks:

$$\log W_{it} = \sum_{s=t-a_{it}+1}^{t} \log G_{is}, \tag{10}$$

where a_{it} denotes the age of the individual (i.e. the number of periods since the last dissipation shock). Combining this equation with (6) implies the following formula for the log deviation in the average wealth in a top percentile.

Proposition 3. *Consider an arbitrary perturbation path in the realization of growth rates in* (9) $\mathrm{d}\log G_{is}$ *for all i and $s \leq t$. The resulting deviation of the average wealth in the top percentile is:*

$$\mathrm{d}\log \overline{W}_t = \mathbb{E}^{W_{it}}\left[\sum_{s=t-a_{it}+1}^{t} \mathrm{d}\log G_{is} \middle| W_{it} \geq Q_t\right]. \tag{11}$$

where $\mathbb{E}^{W_{it}}$ denotes the wealth-weighted expectation with respect to the wealth distribution.

In words, this proposition says that the log deviation of the average wealth in the top percentile is the wealth-weighted average of the sum of the past perturbation of growth rates experienced by individuals in the top percentile. One key point is that the deviation in $\log \overline{W}_t$ depends on *all* the previous perturbations in the growth rate of these top individuals, not just the ones that happened when they already were in the top percentile.[9]

Uniform perturbation. The expression obtained in Proposition 3 is very general. The next corollary considers the particular case in which there is a uniform increase in individual growth rates by $\mathrm{d}g$ started τ periods ago. This increases the log wealth of individual i by $\min(a_{it}, \tau)\,\mathrm{d}g$, where $\min(a_{it}, \tau)$ corresponds to the number of periods in which individual i experienced these higher growth rates. The resulting deviation for the average wealth in a top percentile is a weighted average of these individual deviations.

[7] We will discuss more precisely how to treat bequests and dynasties in Section 3. See Moll et al. (2022) for other interpretations.

[8] The only important condition for the results below is that (i) a perturbation to log wealth at time s transmit one-to-one to time t in the absence of dissipation shock and (ii) the resulting distribution of individual wealth at time t satisfies the smoothness assumptions described in the previous section.

[9] This is related to our discussion of (5) in the previous section.

Corollary 1. *Consider a* uniform *perturbation* $\mathrm{d}g$ *in the growth rate of individual wealth from* $t-\tau$ *to* t. *The resulting deviation in the average wealth in the top percentile at* t *is:*[10]

$$\mathrm{d}\log \overline{W}_t = \mathbb{E}^{W_{it}}\left[\min(a_{it}, \tau) | W_{it} \geq Q_t\right] \mathrm{d}g. \tag{12}$$

The key takeaway is that the distribution of ages in the top percentile acts as a "sufficient statistic" for the effect of a uniform increase in growth rates on the average wealth in a top percentile. Put differently, while the exact specification of the model (e.g.the distribution of growth rates G_{it} and dissipation shocks, or the extent to which they vary over time, between ages, or across types of agents) matters for the counterfactual effect of a change in individual growth rate on the wealth distribution, it only matters through the effect that it has on the joint distribution of wealth and age in the baseline economy.

Note that, in the absence of dissipation shocks (e.g. infinite-horizon model), the relative change in the average wealth in top percentiles after τ period would simply equal $\tau\,\mathrm{d}g$; that is, it would grow without bounds with the horizon. Dissipation shocks, hence, are key to generate finite long-run elasticities of top wealth shares to increases in individual growth rates.[11]

It is particularly intuitive to consider the instantaneous effect ($\tau = 1$) and long-run effect ($\tau = \infty$) of this uniform increase in growth rates . Corollary 12 implies

$$\begin{aligned} \mathrm{d}\log \overline{W}_t &= \mathrm{d}g && \text{for } \tau = 1 \\ \mathrm{d}\log \overline{W}_t &= \mathbb{E}^{W_{it}}\left[a_{it} | W_{it} \geq Q_t\right] \mathrm{d}g && \text{for } \tau = \infty. \end{aligned}$$

In words, the short-run elasticity of the average wealth in the top percentile is one while its long-run one is given by the average age in the top percentile. The higher the average age in the top, the longer individuals in the top benefit from these higher growth rates, and, therefore, the larger the resulting deviation in their wealth. The flipside of this observation is that, the higher the average in the top, the longer it takes for top wealth shares to converge to their long-term values.

This result is related Gabaix et al. (2016), who study the speed of convergence of random-growth models between two steady states. My contribution, relative to this paper, is to characterize the transition dynamics of the average wealth in a top percentile following a uniform change in growth rates (they focus on characterizing the transition dynamics of the Laplace transform of the wealth density). Note that Corollary 1 implies that the speed of convergence of the average wealth in a top percentile is inversely related to the average age of individuals in the top. Hence, this result implies a tight link between two important observations in the inequality literature: Gabaix et al. (2016)'s emphasis that standard random growth models generate slow transition dynamics that are too slow relative to the data and Luttmer (2011)'s emphasis that they imply an

[10] This follows directly from setting $\mathrm{d}\log G_{is} = 1_{s \geq t-\tau}\,\mathrm{d}g$ in Proposition 3.

[11] The idea that dissipation shocks (e.g. overlapping generation models) are important to generate finite elasticities of capital to interest rates is well-known in the literature. See, for instance, Piketty and Saez (2013) or Moll et al. (2022).

average age in top percentile that is too high relative the data (see Atkeson and Irie, 2022 for a related intuition). I discuss in more detail the gradient of age in random-growth models in Section 2.4 below.

I now briefly discuss the implication of these results for the log deviation of the top wealth *share* $S_t \equiv p\overline{W}_t(p)/\overline{W}_t(100\%)$. As discussed in the previous section, taking logarithms and differentiating gives that the log deviation of the top wealth share is simply given by the difference in the log deviation of the average wealth in the top percentile and the log deviation of the average wealth in the economy. In particular, Corollary 1 implies

$$\mathrm{d}\log \overline{S}_t = \Big(\mathbb{E}^{W_{it}}\left[\min(a_{it},\tau)|W_{it} \geq Q_t\right] - \mathbb{E}^{W_{it}}\left[\min(a_{it},\tau)\right]\Big)\,\mathrm{d}g. \tag{13}$$

The evolution of top wealth shares simply depends on the (wealth-weighted) age distribution in the top percentile relative to the economy. Note that the short-run response of top wealth shares ($\tau = 1$) is zero, which reflects the fact that everyone in the economy benefits similarly from a change in growth rates in the first year. However, the long-run response of top wealth shares is positive as long as the average age in the top percentile is higher than the average age in the economy. Finally, note that, the higher the average age of people in the economy, the longer it takes for top wealth shares to converge to their long-run values.

Type-specific perturbation. I now discuss another special case that is commonly encountered in the inequality literature: an increase in the growth rate of only one type of agents (e.g. Luttmer, 2012 or Gabaix et al., 2016). In this case, we can still obtain a relatively simple formula for the deviation of the average wealth in the top percentile.

Corollary 2. *Consider a set of* type-specific *perturbation* $\mathrm{d}g$ *in the growth rate of individual wealth from* $t-\tau$ *to* t*. The resulting deviation in the average wealth in the top percentile at* t *is:*[12]

$$\mathrm{d}\log \overline{W}_t = \mathbb{E}^{W_{it}}\left[\sum_{\max(t-\tau,t-a_{it}+1)}^{t} \theta_{is}\middle| W_{it} \geq Q_t\right]\mathrm{d}g. \tag{14}$$

where θ_{is} *is a dummy variable denoting whether agent* i *at time* s *belongs to that type*

In words, the effect of an increase in the growth rate of agents of type j depends on the (wealth-weighted) number of periods that individuals in the top percentile have spent in this specific type in the previous τ periods. In particular, the short-run elasticity of the average wealth in the top percentile to a type-specific increase in growth rates is the fraction of individuals in this top percentile that are currently of this type. Its long-run elasticity is the average number of periods that individuals in the top percentile have spent in this type over their lifetimes.

Note that the average wealth in a top percentile responds more to a uniform increase in individual growth rates (Corollary 1) than to a type-specific one (Corollary 2). However, the opposite

[12] This follows directly from setting $\mathrm{d}\log G_{is} = \theta_{is}1_{s\geq t-\tau}\,\mathrm{d}g$ in Proposition 3.

may hold for top wealth *shares*; in particular, an increase in the growth rate of the type of agents making it to the top of the distribution will tend to generate faster and larger changes in top wealth *shares* relative to a uniform increase in the growth rate of all agents. This distinction is typically very important quantitatively.

2.2 Geometric Random Walk with Additive Shocks

While the model discussed in the previous section gave simple formulas for the dynamics of top wealth shares, it may a bit too stylized to be taken to the data: on the one hand, it assumes perfect persistence of wealth within each lifetime, while, on the other hand, assuming zero persistence at when the dissipation shock hits.

I now discuss a more flexible process which may be more realistic. More precisely, I assume that wealth evolves geometrically with some additive force (a Kesten, 1973 process):

$$W_{it+1} = G_{it+1} W_{it} + Y_{it+1}, \tag{15}$$

where G_{it}, Y_{it} are positive random variables. This type of process has been used to analyze the distribution of city size (e.g. Gabaix, 1999) as well as the distribution of wealth (e.g. Benhabib et al., 2011 and Gabaix et al., 2016).[13] As above, I allow the distribution of (G_{it+1}, Y_{it+1}) to be serially correlated, and to potentially depend on the time t or the individual i.[14]

Differentiating (15) with respect to an arbitrary sequence of perturbations for the growth rate of wealth $(\mathrm{d}\log G_{it})_{s\leq t}$ gives the following recurrence relation for the deviation in log wealth:[15]

$$\mathrm{d}\log W_{it+1} = \left(1 - \frac{Y_{it+1}}{W_{it+1}}\right)\left(\mathrm{d}\log W_{it} + \mathrm{d}\log G_{it+1}\right).$$

Solving this system backward gives the perturbation of individual wealth at time t as the cumulated sum of all previous shocks in the growth rate of wealth that happened in the past:

$$\mathrm{d}\log W_{it} = \sum_{s=-\infty}^{t}\left(\prod_{u=s}^{t}\left(1 - \frac{Y_{iu}}{W_{iu}}\right)\right)\mathrm{d}\log G_{is}. \tag{16}$$

Combining this equation with (6) gives the following formula for the log deviation of the average wealth in the top percentile.

Proposition 4. *Consider an arbitrary perturbation path in the realization of growth rates in* (15) $\mathrm{d}\log G_{is}$

[13] In Benhabib et al., 2011, this process models the law of motion of wealth across cohorts, which is a mix of bequest from the previous cohort, W_{it}, and savings due to human capital Y_{it}. In Gabaix et al., 2016, this process models the law of motion of financial wealth of infinite-horizon agents who are assumed to consume a fixed fraction of their wealth and who receive some labor income every period Y_{it}.

[14] Again, the only condition is that the resulting distribution of individual wealth at time t satisfies the smoothness assumptions described in the previous section.

[15] It would be straightforward to extend these formulas to consider deviations in labor income $(\mathrm{d}\log Y_{it})$.

for all i and $s \leq t$. The resulting deviation of the average wealth in the top percentile is

$$\mathrm{d}\log \overline{W}_t = \mathbb{E}^{W_{it}}\left[\sum_{s=-\infty}^{t}\left(\prod_{u=s}^{t}\left(1-\frac{Y_{iu}}{W_{iu}}\right)\right)\mathrm{d}\log G_{is}\middle| W_{it} \geq Q_t\right]. \tag{17}$$

As in Proposition 11, this proposition says that the log deviation of the average wealth in the top percentile is the wealth-weighted average of the sum of the past perturbation of growth rates experienced by individuals currently in the top percentile. The important distinction, however, is that the extent to which past wealth shocks are discounted over time depends on the relative importance of the additive force Y_{it} in the level of wealth W_{it}. Intuitively, the higher the magnitude of Y_{it} relative to wealth, the lower the effect of past shocks in wealth on current wealth.

2.3 Discussion

General statistical processes. So far, we have focused on the case where individual wealth follows a geometric random walk with dissipation shocks (Section 2.1) or with additive shocks (Section 2.2). In in Appendix C.1, I generalize such analysis to very general statistical models. The key idea is that the deviation of individual wealth can always be rewritten as the cumulative sum of past perturbations (moving-average representation). Different statistical models for individual wealth simply lead to different expressions for this moving average representation. I discuss this point in more details in Appendix C.1.

Deviation in tail indices. I now briefly emphasize the advantage of my approach relative to the existing literature on inequality, which tends to focus on analytical characterization for tail indices. When wealth follows a random growth model of the type discussed in Section 2.1 or 2.2, one can often show that, under additional condition of stationarity, the distribution of wealth has a right Pareto tail; that is,[16]

$$\log \overline{W}(p) \sim -\frac{1}{\zeta}\log p \quad \text{as } p \to 0. \tag{18}$$

where ζ denotes the tail index of the wealth distribution. Consider a permanent perturbation in the distribution of these growth rates starting from time $t = 0$. In this case, one can typically show that the perturbated wealth distribution remains Pareto and obtain an analytical formula for the resulting deviation in the tail index (see Appendix C.2 for examples). This analytical characterization implies the following characterization for top wealth shares:[17]

$$\lim_{t\to\infty} \mathrm{d}\log \overline{W}_t(p) \sim (-\,\mathrm{d}\log\zeta)\log \overline{W}_t(p) \quad \text{as } p \to 0. \tag{19}$$

[16]Here, and in the rest of the paper, I use $f(p) \sim g(p)$ as $p \to 0$ to denote $\lim_{p\to 0} f(p)/g(p) = 1$.
[17]Formally, (18) implies $\lim_{t\to\infty} \mathrm{d}\log \overline{W}_t(p) \sim -(\zeta^{-2}\,\mathrm{d}\zeta)\log p \sim -(\zeta^{-1}\,\mathrm{d}\zeta)\log \overline{W}(p)$ as $p \to 0$.

The key takeaway is that, while a deviation in the tail index is informative on the asymptotic behavior of top wealth shares, it only characterizes the asymptotic deviation in top wealth shares in the long-run ($t \to \infty$) and in the right tail ($p \to 0$).[18] This is much less precise than the expressions obtained in Proposition 3 or Proposition 4, that characterize the deviation of the average wealth in the top at *any* given time horizon t and at *any* given top percentile p. In fact, I discuss in Appendix C.2 how focusing on the response of the tail index can be very uninformative about the actual response of top wealth shares, even when one restricts oneself to the long-run response of top percentile shares. Moreover, my analytical results characterize the response of top wealth shares to the wide range of perturbations that affect top wealth shares without affecting tail indices (e.g. transitory perturbations or perturbations that only impact individuals for a finite amount of time periods) and they are also valid in much more general models (without requiring that the wealth distribution is stationary or obeys a Pareto shape). Of course, all these advantages come at a cost: this analytical framework can only characterize small deviations in the wealth distribution away from a baseline.

2.4 A Diagnostic Tool for Models of Inequality

I now use these analytical results as a "diagnostic tool" to examine counterfactuals implied by existing models of top wealth inequality. The two key takeaways of this section are that: (i) my (first-order) approach provides a good approximation to the (non-infinitesimal) deviation of top wealth shares in each model and (ii) it provides a clear interpretation of what drives the difference across these models.

I focus on examining three models of wealth inequality, which are all particular instances of the model with dissipation shocks discussed in Section 2.1: Gabaix et al. (2016), Moll et al. (2022) and Gomez and Gouin-Bonenfant (2024). To be able to compare these models more easily, I examine the effect of the same experiment across these models: I increase the rate of capital accumulation by 1 percentage point in each model. For each model, I report the long-run log deviation for the average wealth in the top 0.01%, for the average wealth in the economy, and for the difference between the two (i.e., the log deviation in the top 0.01% wealth share). I also report the results from my first-order formula, which, in this context, simply corresponds to the average number of periods since the last dissipation shock in each group.

The first takeaway from Table 1 is that the first-order approximation captures well the deviation in the average wealth in the top 0.01%, the average wealth for the economy (top 100%), and, therefore, the difference between the two (the top 0.01% wealth share). The non-infinitesimal deviation in top wealth shares is typically larger than the results from the first-order approximation,

[18] If the wealth distribution is *exactly* Pareto, then there is a one-to-one mapping between the tail index and the *level* of top wealth shares. However, most existing models of wealth inequality only imply that the wealth distribution has a Pareto *tail*, consistently with the data. One exception is models where individual wealth follows a random walk with a reflecting barrier at some lower level of wealth.

Table 1: Long-Run Semi-Elasticity of Wealth to Individual Growth Rates (x100)

	Non-infinitesimal deviation			First-order deviation		
	Top 0.01%	Top 100%	Difference	Top 0.01%	Top 100%	Difference
Moll et al. (2022)	149	31	117	134	26	108
Gabaix et al. (2016)	116	44	71	85	37	48
Gomez and Gouin-Bonenfant (2024)	24	10	14	22	10	12

Notes: The table reports the effects of a uniform 1 percentage point increase in the growth rates of individual wealth for the log average wealth in the top 0.01% (Column 1), top 100% (Column 2), and for the top 0.01% wealth share (Column 3), which is the difference between the two. The table also reports the wealth-weighted average effect of a 1pp. uniform increase for individual wealth, which corresponds to the wealth weighted average age in each top percentile (i.e. number of periods since last dissipation shock), see Proposition 3.

reflecting the fact that the re-ranking of individuals always has a positive effect on inequality.[19]

The second takeaway is that, while all of these models are calibrated to match the shape of the U.S. distribution, they imply very different joint distribution for age and wealth, and, therefore, for the response of top wealth shares to an increase in the growth rate of individual wealth.[20] The average "age" of households in the top 0.01% (number of periods since dissipation shock) ranges from 22 in Gomez and Gouin-Bonenfant (2024) to 134 in Moll et al. (2022). Hence, a given increase in the growth rate of individuals will lead to much larger responses for the average wealth in the top percentile in Moll et al. (2022) relative to Gomez and Gouin-Bonenfant (2024): more precisely, I find that a 1pp increase in individual growth rate increases the wealth share of the top 0.01% by 1.17 log points in Moll et al. (2022) but only 0.14 log points in Gomez and Gouin-Bonenfant (2024).

Hence, despite the fact that these models match similar cross-sectional moments about the wealth distribution (e.g. level of top wealth shares), they have vastly different implications on the relationship between wealth and age, and therefore, on the effect of a given increase in individual growths rate on inequality.

How comes different models lead to drastically different age gradient across the wealth distribution? I explore this question in Appendix C.2. I show that, in a large class of random-growth models, the average age increases linearly with log wealth, with a slope equal to the inverse of the derivative of the cumulant generating function (CGF) of the growth rate of wealth taken at the tail index. Intuitively, the more convex the CGF is, the higher the dispersion in growth rates across individuals, and, therefore, the less time it takes for the most successful individuals to reach top percentiles. In particular, a model in which the growth rate of individual wealth has a high skewness or kurtosis (potentially because some individuals experience higher growth rates over long periods of time, a la Luttmer, 2011) is a model in which the luckiest households reach the right tail of the distribution pretty quickly, and, therefore, in which top wealth shares respond less to a given increase in individual growth rates.

One remaining question is: given that these models disagree so much on the gradient of "age"

[19] Formally, the extensive margin in (4) is positive.

[20] Note that Gabaix et al. (2016) is calibrated to match the distribution of income rather than wealth.

across the wealth distribution (i.e. the extent to which wealth comes from old money or new money), and that this gradient play a key role for counterfactuals, can we discipline this gradient using actual microdata? I explore this question in the next section.

3 An Accounting Framework to Compute Counterfactuals

The previous sections of the paper developed analytical tools to quantify the effect of changes in individual wealth dynamics on the dynamics of top wealth shares. I now use these results to develop a simple accounting framework to compute the response of top wealth shares to various counterfactual scenarios. Section 3.1 presents the framework, Section 3.2 discusses the data, and Section 3.3 presents some applications.

3.1 A Model of Wealth Accumulation

I consider the following law of motion of wealth for household i between t and $t+1$ is:

$$W_{it+1} = R_{it+1}W_{it} + B_{it+1} + Y_{it+1} - C_{it+1}, \tag{20}$$

where $R_{it+1} > 0$ denotes the return on wealth, $Y_{it+1} > 0$ denotes labor income, $B_{it+1} > 0$ denotes bequests received, and C_{it+1} denotes consumption. With some probability, individuals die, at which point their wealth is redistributed as inheritance to their children.

Linearizing this equation and solving backward gives the deviation in log wealth at some point t as the cumulative sum of past perturbations in returns, labor income, and bequests.

Proposition 5. *The effect of a sequence of deviation in past returns, labor income, and bequests on current wealth is:*

$$\mathrm{d}\log W_{it} = \sum_{s=t-a_{it}+1}^{t} \left(\prod_{u=s+1}^{t} \left(1 - \frac{B_{iu} + Y_{iu} - \xi_{iu} C_{iu}}{W_{iu}}\right)\right) dy_{is} \tag{21}$$

where $\mathrm{d}y_{is}$ *denotes the direct effect of the perturbations on the relative wealth at time* s*:*

$$dy_{is} \equiv \left(1 - \frac{B_{is} + Y_{is} - C_{is}}{W_{is}}\right) \mathrm{d}\log R_{is} + \frac{B_{is}}{W_{is}} \mathrm{d}\log B_{is} + \frac{Y_{is}}{W_{is}} \mathrm{d}\log Y_{is}$$

and ξ_{iu} *denotes one minus the elasticity of consumption to wealth*

$$\xi_{iu} \equiv 1 - \frac{W_{iu}}{C_{iu}} \frac{\partial C_{iu}}{\partial W_{iu}}$$

This proposition expresses the deviation in consumption at some present time as a sum of the direction deviations in previous wealth. That the persistence of these perturbations on wealth depend on the relative importance of labor and bequest on wealth, as well as the marginal propensity to consume out of wealth. Assuming that the remaining perturbation are not causally related, this expression allows one to trace out the effect of returns, labor income or bequests using the historical path of the labor income and bequest to wealth ratios of households.

To visualize this equation, Figure 2 plots the elasticity of current wealth to an exogenous increase in returns s periods before on the household's own wealth (in blue) as well as on the household's parent wealth (in red). There is a normal decay over time due to the importance of labor income relative to financial wealth. There is also a discrete discounting that happens when receiving bequest: receiving a large inheritance decreases the effect of the household own returns received prior to this inheritance while increasing the effect of the parents own returns. Finally, note that the model implies that the persistence of returns is very small for individuals at the bottom of the wealth distribution, but very high for individuals at the top of the wealth distribution (for which labor income represents an infinitesimal fraction of current wealth.)

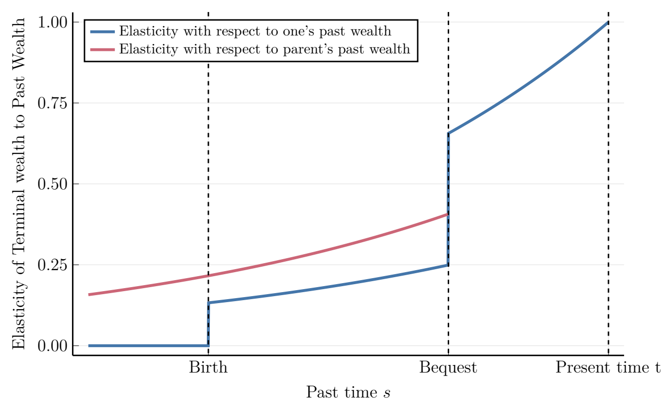


Figure 2: Elasticity of present wealth (time t) to a shock in prior wealth (time s) in the model.

Notes: The figure plots the elasticity of wealth of household i at time t with respect to wealth of household j at time $s < t$, $\partial \log W_{it} / \partial \log W_{js}$, as a function of s. In blue is the case $j = i$ (household's own wealth), and in red is the case where j leaves a bequest to i (parents' wealth). The sum of the area under these curves can be interpreted as the elasticity of current wealth to a uniform change in past growth rates.

Assumption on ξ. The case $\xi_{iu} = 0$ corresponds to the case in which the marginal propensity to consume is equal to the average consumption as a proportion of wealth. The case $\xi_{iu} = 1$ corresponds to the case in which the marginal propensity to consume is equal to zero. Whenever households consumes proportionally to wealth plus human capital (total wealth), one can show that ξ_{iu} corresponds exactly to the ratio of human capital in total wealth; that is, $\xi_{iu} \in (0, 1)$ indexes the relative importance of human capital at time u.

In the rest of the paper I will assume that $\xi_{iu} = 0$ for simplicity. Given the discussion above, this assumption describes well agents who heavily discount their future labor income relative to their current financial wealth. In particular, this assumption holds true either if (i) human capital and future inheritance represents a small share of financial wealth or (ii) households heavily discount all future labor income and inheritance due to their idiosyncratic risk (which captures well the behavior of hand-to-mouth households). Hence, this reduced-form assumption captures well the behavior of households with a high level of wealth (for which human capital accounts for a negligible part of their wealth) or the behavior of hand-to-mouth agents (who consume their wealth and labor income every period).

3.2 Data and Methodology

Proposition 5 allows us to compute the effect of changes in returns on changes in household wealth as long as we can observe the historical path of their labor income, inheritance and wealth. I now discuss the data I use to construct a measure of the lifetime path of labor income, wealth, bequest received by households in top percentiles for the U.S. from a combination of the Survey of Consumer Finances (SCF) from 1989 to 2016 and Forbes 400 in years.

Since the SCF does not report the previous income and wealth for individuals currently in top percentiles, I use a "synthetic cohort" approach where I construct the average income and wealth across top percentiles within each cohort, and I then construct the lifetime path of each agent by assuming that agents remain in their relative percentiles over their lifetimes (as in the "pseudo-panel" approach of Feiveson and Sabelhaus, 2019). I set the labor income of households in Forbes 400 to zero. I define the age of a household as the age of the household head minus 25.

Finally, I construct a measure of bequest received across top percentiles by using the total bequest distributed over their lifetimes as reported in SCF (which reports the amount and the year of the three biggest inheritance or in-vivo transfer received over the household's lifetime). Finally, to compute the average labor income to wealth ratio of parents, I assume that, with probability half, they are in the same top percentile (within their cohorts) as their children, while, with probability half, they are randomly drawn from the population. For households in Forbes, I classify individuals manually into three categories: self-made, heirs, and in-between. I assume that the inheritance to wealth ratio at the time of inheritance is zero for self-made fortunes (60% of observations), 1 for heirs (24% of observations), and 0.5 for people in between (16% of observations).

To visualize the data construction, Table 2 reports the implied long-run elasticity of wealth to changes in individual growth rate, computed recursively from (5). If individuals receive no inheritance and if labor income is an infinitesimal fraction of their wealth, this corresponds literally to the age of individuals, as in the model with dissipation shocks discussed in Section 2.1. Relative to this baseline, the presence of labor income tends to decrease the persistence of wealth shocks while the presence of bequests tends to increase them.

The first column of the table reports the elasticity from (5) that one would obtain without taking into account inheritance of labor income; i.e., defining the age of the household as the household head's age minus 25, as discussed above. Note that the gradient of average age across the distribution is almost zero. This almost zero gradient implies that without bequests and labor income, changes in individual returns would not affect top wealth shares. The second column of the table adjusts for bequests, i.e. reports the elasticity obtained from (5) while still setting $Y_{iu} = 0$. This increases the elasticity of wealth at the top since richer households tend to receive larger

Table 2: Long-Run Semi-Elasticity of Wealth to Returns Implied by the Data

	Household age (Baseline)	Taking into account bequests	Taking into account bequests & labor income
Top 100%	33	37	26
Top 10%	34	38	28
Top 1%	35	38	31
Top 0.1%	36	40	35
Top 0.01%	37	43	38
Top 400	39	55	46

Notes: The table reports summary statistics on the wealth-weighted distribution of age in top percentiles of the U.S. economy. Data from the Survey of Consumer Finances and Forbes 400 from 1989 to 2016.

bequests. The third column of the table then also accounts for the relative importance of labor income by reporting the full formula (5). This tends to decrease the average elasticity of wealth at the bottom since labor income constitutes a large fraction of their wealth every period. Overall, I find that the gradient of "age" becomes much steeper once one accounts for inheritances and labor income.

I now briefly compare my estimates for the long-run elasticity of individual wealth to changes in individual growth rates to the range of estimates implied by the existing literature discussed in Section 2.4. In average, my estimates are lower then existing calibrations. One way to interpret this evidence is that Moll et al. (2022) and Gabaix et al. (2016) effectively overestimate the role of dynastic wealth in top percentiles while Gomez and Gouin-Bonenfant (2024) underestimate it relative to the U.S. data.

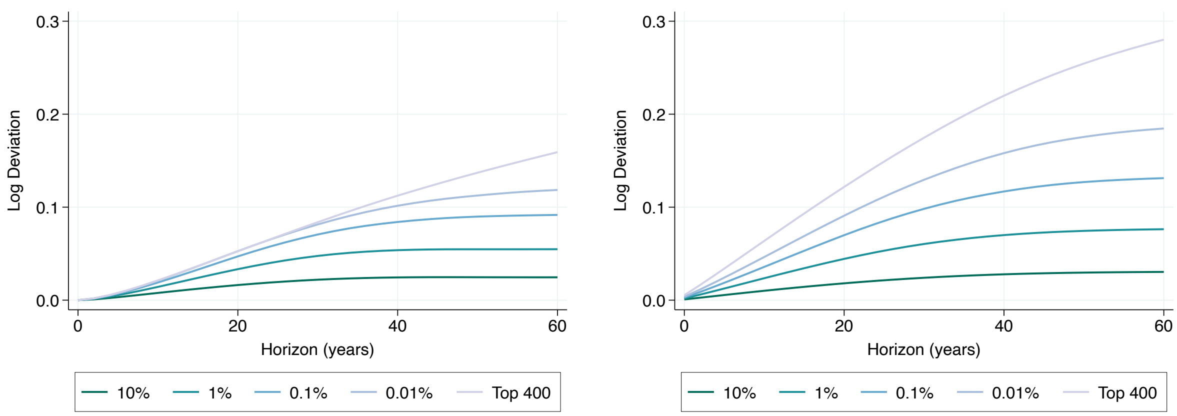
3.3 Results

I now use this methodology and data to compute the dynamics of top wealth shares under a number of counterfactual scenarios.

Permanent changes in asset returns. As summarized by Piketty and Zucman (2015), most models of wealth inequality imply that an increase in the average return of wealth increases wealth inequality (or, at least, the thickness of its right tail). However, there is still substantial uncertainty on how large the effect of an increase in the average return on wealth on top wealth inequality actually is. I now use (5) to compute the response of top wealth shares to a permanent increase in asset returns.

Figure 3a plots the result for a 1pp uniform increase in the growth rate of wealth. In the first few years, the effect of an increase in the return on wealth is almost zero. As discussed in Section 2.1, this occurs because all individuals in the economy experience this higher growth rate; hence, top wealth *shares* do not react. Still, top wealth shares start increasing after a few years. There is "fanning-out" of top wealth shares which reflects the fact that agents at the top of the wealth distribution benefit relatively more from higher growth rates, as reported in Table 2. The underlying reason is that the effect of a given increase in asset returns is much more persistent for agents at the top of the wealth distribution than for the average agent in the economy. Overall, I find that, after 60 years, a 1pp increase in the return on individual wealth increases the share of wealth owned by the top 0.01% by 0.1 log points. Note, however, that this increase is relatively small compared to the overall rise in top wealth inequality since 1980, which amounts to approximately 1 log point (Saez and Zucman, 2022)

Figure 5a plots the increase in the average return of only one type of assets, public and private equity. Relative to the previous experience, this increases the return of each agent by 1pp times the share of wealth invested in equity. Relative to the previous experiment, we observe an immediate increase in top wealth shares, reflecting the fact that individuals at the top of the wealth distributions disproportionately hold equity. We can see that, relative to the previous experiment, top wealth share increase faster and with a bigger magnitude. Overall, a 1pp increase in the return on equity increases the top 0.01% by 0.2 log points, which is twice as much as the effect of a 1pp increase in the overall return on wealth.



(a) 1pp. Increase in Returns Across Asset Classes

(b) 1pp. Increase in Return on Equity

Figure 3: Effect of a Permanent Increase in Returns on Top Wealth Shares

Notes: The figures report the gradient of age and number of years spent in the high-growth category in three models of inequality in the literature. GLLM (2016) refers to Gabaix et al. (2016), MRR (2022) to Moll et al. (2022), and GG (2024) to Gomez and Gouin-Bonenfant (2024).

As pointed out in Piketty (2014), one contributor to wealth inequality, beyond the average return on assets, is also the fact that richer households tend to own asset classes with higher average returns. One open question, however, is to assess magnitude of this effect. Put differently, how much lower would top wealth shares be if the average return was equalized across equity, housing, and debt? My framework provides a simple way to answer this question. Figure 4 plots the result of computing a counterfactual where I assume that the average return of every asset class (debt, equity, and housing) gets equalized starting from some time $t = 0$. I find that, over the long-run, the wealth share of the top 0.01% would decrease by 0.4 log points after sixty years (i.e.

a decrease of approximately 30%). The drop is even more pronounced for households in Forbes 400, who almost hold 100% of their wealth in equity (see Gomez, 2017 for more evidence).

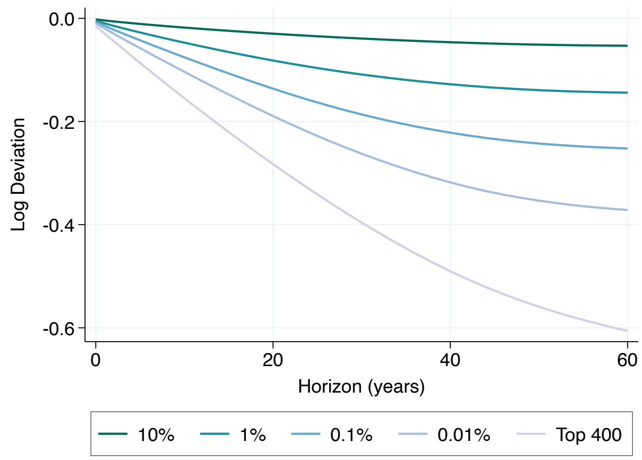
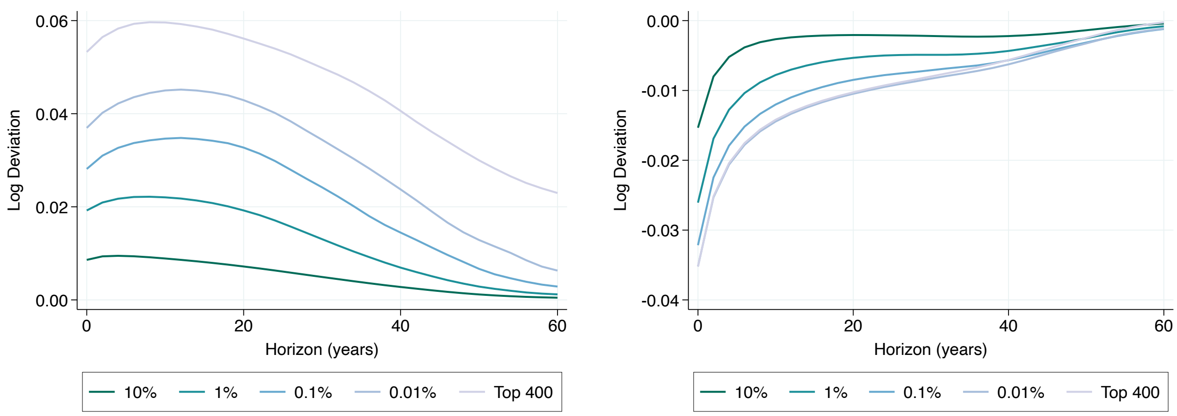


Figure 4: Effect of Equalizing Returns Across Asset Classes on Top Wealth Shares

One-time shock in asset returns. I now use the methodology to compute the impulse response of top wealth shares to a one-time increase in asset returns (as opposed to a permanent increase in returns). One key result is that these shocks tend to have very persistent effect on top wealth shares.

Figure 5a reports the effect of a one-time 10% increase in equity returns. The short-run impact at $t = 0$ is large, which reflects the fact that households in top percentiles are more exposed to equity returns than the average household in the economy. My estimates are consistent with the reduced-form exposures from Kuhn et al., 2020 and Gomez, 2017. The key contribution of this methodology is to evaluate the longer-run effect of these asset returns on top wealth shares. In particular, one striking fact is how persistent the response is for Forbes 400. This reflects the fact that a substantial portion of individuals in Forbes 400 inherit their wealth; as a result, the impact of increased equity returns extend beyond the lifetimes of the original equity holders.

Figure 5b reports the effect of a one-time 10% increase in housing returns. Top wealth shares drop, which reflects the fact that top households hold less housing than the rest of the distribution (see, for instance, Kuhn et al., 2020, Greenwald et al., 2022, Martínez-Toledano, 2020, . . .). However, the effect of an increase in housing returns is much less persistent than the effect of equity returns. This reflects the fact that wealth shocks tend to be much less persistent at the bottom of the wealth distribution, as a larger fraction of the wealth for the average household in the economy comes from capitalized labor income, which dampens the effect of past returns on capital on current wealth (see Proposition 5).



(a) 10pp. Increase in Equity Returns

(b) 10pp. Increase in Housing Returns

Figure 5: Effect of a One-Time Increase in Asset Returns on Top Wealth Shares

4 Conclusion

The starting point of this paper is the observation that, for a wide range of counterfactuals, the response of the average wealth in a top percentile is equal, at the first-order, to the average wealth deviation of individuals in the top percentile. I leverage this simple insight to obtain clean formulas for the impulse response of top wealth shares in a wide range of random-growth models. I then use these results as a diagnostic tool to analyze counterfactuals implied by existing models of wealth inequality. I also use them as as an accounting framework to compute the dynamics of top wealth shares under various counterfactual scenarios.

Overall, this paper provides a flexible and transparent methodology to analyze the effects of economic shocks on wealth inequality. This methodology could help both reduced-form and structural approaches to generate credible responses of the wealth distribution to counterfactual changes in the economic environment.

References

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, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models," *Econometrica*, 2021, *89* (5), 2375–2408.

_ , Matthew Rognlie, and Ludwig Straub, "The intertemporal keynesian cross," Technical Report, National Bureau of Economic Research 2018.

Baqaee, David and Elisa Rubbo, "Micro propagation and macro aggregation," *Annual Review of Economics*, 2023, *15*, 91–123.

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

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

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

Chetty, Raj, "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods," *Annu. Rev. Econ.*, 2009, *1* (1), 451–488.

Costinot, Arnaud and Andrés Rodríguez-Clare, "Trade theory with numbers: Quantifying the consequences of globalization," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 197–261.

de Saporta, Benoite, "Tail of the Stationary Solution of the Stochastic Equation $Y_{n+1} = aY_n + b_n$ with Markovian coefficients," *Stochastic Processes and their Applications*, 2005, *115*, 1954–1978.

Feiveson, Laura and John Sabelhaus, "How does intergenerational wealth transmission affect wealth concentration?," 2018.

_ **and** _ , "Lifecycle patterns of saving and wealth accumulation," 2019.

Gabaix, Xavier, "Zipf's Law for Cities: an Explanation," *Quarterly Journal of Economics*, 1999, pp. 739–767.

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

Gaillard, Alexandre, Christian Hellwig, Philipp Wangner, and Nicolas Werquin, "Consumption, Wealth, and Income Inequality: A Tale of Tails," 2023.

Gomez, Matthieu, "Asset prices and wealth inequality," *Working Paper*, 2017.

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

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

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

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, Sergio Ocampo, and Daphne Chen, "Use it or lose it: Efficiency and redistributional effects of wealth taxation," *The Quarterly Journal of Economics*, 2023, *138* (2), 835–894.

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, "Pareto and Piketty: The macroeconomics of top income and wealth inequality," *The Journal of Economic Perspectives*, 2015, *29* (1), 29–46.

_ **and Jihee Kim**, "A Schumpeterian model of top income inequality," *Journal of political Economy*, 2018, *126* (5), 1785–1826.

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.

Kesten, Harry, "Random difference equations and renewal theory for products of random matrices," 1973.

Kleven, Henrik J, "Sufficient statistics revisited," *Annual Review of Economics*, 2021, *13*, 515–538.

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, "On the mechanics of firm growth," *The Review of Economic Studies*, 2011, *78* (3), 1042–1068.

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

_ , "Further notes on micro heterogeneity and macro slow convergence," *Working Paper*, 2016.

Martínez-Toledano, Clara, "House price cycles, wealth inequality and portfolio reshuffling," *WID. World Working Paper*, 2020, (2020/02).

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.

Nardi, Mariacristina De, "Wealth inequality and intergenerational links," *The Review of Economic Studies*, 2004, *71* (3), 743–768.

Ozkan, Serdar, Joachim Hubmer, Sergio Salgado, and Elin Halvorsen, "Why Are the Wealthiest So Wealthy? A Longitudinal Empirical Investigation," *A Longitudinal Empirical Investigation (March 9, 2023)*, 2023.

Peter, Alessandra, "Owning Up: Closely Held Firms and Wealth Inequality," *JMP Stan*, 2019.

Piketty, Thomas, *Capital in the twenty-first century*, Harvard University Press, 2014.

_ **and Emmanuel Saez**, "A theory of optimal inheritance taxation," *Econometrica*, 2013, *81* (5), 1851–1886.

_ **and Gabriel Zucman**, "Wealth and Inheritance in the Long Run," *Handbook of Income Distribution*, 2015, *2*, 1303–1368.

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** — , “Progressive wealth taxation,” *Brookings Papers on Economic Activity*, 2019, *2019* (2), 437–533.

— **and** — , “Top wealth in america: A reexamination,” Technical Report, National Bureau of Economic Research 2022.

Touchette, Hugo, “The large deviation approach to statistical mechanics,” *Physics Reports*, 2009, *478* (1-3), 1–69.

Wold, Herman OA and Peter Whittle, “A model explaining the Pareto distribution of wealth,” *Econometrica, Journal of the Econometric Society*, 1957, pp. 591–595.

Appendix

A Proofs

Proof of Proposition 1. In the first step of the proof, I first relate the deviation of densikty of wealth to the deviations in individual wealth. For any function $f \in \mathcal{C}^1$ defined on $\mathbb{R}$ that decays quickly enough with wealth, the cross-sectional average of $f(W_{it})$ in the counterfactual economy can be expressed in two ways:

$$\int_{\mathbb{R}} f(W) \left(g_t(W) + \Delta g_t(W) \right) dW = \int_{\mathbb{R}} \mathbb{E}[f(W_{it} + \Delta W_{it}) | W_{it} = W] g_t(W) dW,$$

where g_t denotes the density of wealth at time t in the economy and d denotes the usual differential operator with respect to wealth (*not* the differential operator with respect to the size of the pertubation). Subtracting both sides by the cross-sectional average of $f(W_{it})$ in the baseline economy gives:

$$\int_{\mathbb{R}} f(W) \Delta g_t(W) dW = \int_{\mathbb{R}} \mathbb{E}[f(W_{it} + \Delta W_{it}) - f(W_{it}) | W_{it} = W] d\mathbb{P}(W_{it} \leq W).$$

Dividing by θ and taking the limit $\theta \to 0$ gives

$$\int_{\mathbb{R}} f(W) \, \mathrm{d}g_t(W) dW = \int_{\mathbb{R}} f'(W) \mathbb{E}[\mathrm{d}W_{it} | W_{it} = W] g_t(W) dW. \tag{22}$$

Integrating by parts gives

$$\int_{\mathbb{R}} f'(W) (-\, \mathrm{d}G_t(W)) dW = \int_{\mathbb{R}} f'(W) \mathbb{E}[\mathrm{d}W_{it} | W_{it} = W] g_t(W) dW.$$

Since this holds for any continuous function f', this must hold at every point, which implies

$$\mathrm{d}G_t(W) = -\mathbb{E}[\mathrm{d}W_{it} | W_{it} = W] g_t(W) \tag{23}$$

where G_t denotes the cumulative distribution function for wealth at time t. Note that this equation is very similar in spirit to Kolmogorov-Forward equation, except that instead of characterizing the derivative of the CDF with respect to time t, it characterizes the derivative of the CDF with respect to the perturbation size θ.

In the second step of the proof, I use this result to obtain the deviation in the average wealth in the top percentile. Applying the implicit function theorem on the definition of the top quantile (1) gives:

$$\mathrm{d}Q_t = -\frac{1}{g_t(Q_t)} \int_{Q_t}^{\infty} \mathrm{d}g_t(W) \, \mathrm{d}W. \tag{24}$$

Differentiating the definition of the average wealth in the top percentile (2) gives:

$$\begin{aligned} \mathrm{d}\overline{W}_t(\theta) &= \frac{1}{p}\int_{Q_t}^{\infty} W\,\mathrm{d}g_t(W)dW - \frac{1}{p}Q_t g_t(Q_t)\,\mathrm{d}Q_t \\ &= \frac{1}{p}\int_{Q_t}^{\infty}(W - Q_t)\,\mathrm{d}g_t(W)dW \\ &= -\frac{1}{p}\int_{Q_t}^{\infty}\mathrm{d}G_t(W)dW. \end{aligned}$$

where the last line uses integration by part. Substituting out $\mathrm{d}G_t(W)$ using (23) gives the result

$$\mathrm{d}\overline{W}_t = \frac{1}{p}\int_{Q_t}^{\infty}\mathbb{E}[\mathrm{d}W_{it}|W_{it} = W]g_t(W)dW.$$

Additionally, combining (24) with (23) gives a similar formula for the infinitesimal deviation in quantiles:

$$\mathrm{d}Q_t = \mathbb{E}[\mathrm{d}W_{it}|W_{it} = W].$$

□

Proof of Proposition 2. Differentiating (22) at the second order in θ gives

$$\int_{\mathbb{R}} f(W)\,\mathrm{d}^2 g_t(W)dW = \int_{\mathbb{R}}\left(f'(W)\mathbb{E}[\mathrm{d}^2W_{it}|W_{it} = W] + \frac{1}{2}f''(W)\mathbb{E}[(\mathrm{d}W_{it})^2|W_{it} = W]\right)g_t(W)dW.$$

Integrating by parts the right hand side gives

$$\int_{\mathbb{R}} f(W)\,\mathrm{d}^2 g_t(W)dW = \int_{\mathbb{R}} f(W)\left(-\partial_W\left(\mathbb{E}[\mathrm{d}^2W_{it}|W_{it} = W]g_t(W)\right) + \frac{1}{2}\partial_{WW}\left(\mathbb{E}[(\mathrm{d}W_{it})^2\,|W_{it} = W]g_t(W)\right)\right)dW.$$

Since this holds for any continuous function f, this must hold pointwise; hence,

$$\mathrm{d}^2 g_t(W) = -\partial_W\left(\mathbb{E}[\mathrm{d}^2W_{it}|W_{it} = W]g_t(W)\right) + \partial_{WW}\left(\mathbb{E}[(\mathrm{d}W_{it})^2\,|W_{it} = W]g_t(W)\right) \tag{25}$$

Now, differentiating the deviation of top wealth shares at the second-order in θ gives

$$\begin{aligned} \mathrm{d}^2\overline{W}_t &= \frac{1}{p}\int_{Q_t}^{\infty}(W - Q_t)\,\mathrm{d}^2 g_t(W)dW - \frac{1}{p}\mathrm{d}Q_t\int_{Q_t}^{\infty}\mathrm{d}g_t(W)dW \\ &= \frac{1}{p}\int_{Q_t}^{\infty}(W - Q_t)\,\mathrm{d}^2 g_t(W)dW - \frac{g_t(Q_t)}{p}\left(\frac{1}{g_t(Q_t)}\int_{Q_t}^{\infty}\mathrm{d}g_t(W)dW\right)^2. \end{aligned} \tag{26}$$

Combining with (23) and (25) gives

$$\mathrm{d}^2\overline{W}_t = \frac{1}{p}\int_{Q_t}^{\infty}(W-Q_t)\left(-\partial_W\left(\mathbb{E}[\mathrm{d}^2W_{it}|W_{it}=W]g_t(W)\right)+\partial_{WW}\left(\mathbb{E}[(\mathrm{d}W_{it})^2\,|W_{it}=W]g_t(W)\right)\right)dW$$
$$-\frac{g_t(Q_t)}{p}\left(\frac{1}{g_t(Q_t)}\int_{Q_t}^{\infty}\left(-\partial_W\left(\mathbb{E}[\mathrm{d}W_{it}|W_{it}=W]g_t(W)\right)\right)dW\right)^2$$

Integrating by parts gives

$$\begin{aligned}\mathrm{d}^2\overline{W}_t &= \frac{1}{p}\int_{Q_t}^{\infty}\mathbb{E}[\mathrm{d}^2W_{it}|W_{it}=W]g_t(W)dW+\frac{g_t(Q_t)}{p}\mathbb{E}\left[(\mathrm{d}W_{it})^2\,|W_{it}=W\right]\\ &-\frac{g_t(Q_t)}{p}\mathbb{E}\left[\mathrm{d}W_{it}|W_{it}=W\right]^2\\ &=\mathbb{E}[\mathrm{d}^2W_{it}|W_{it}\geq Q_t]+\frac{g_t(Q_t)}{p}\mathbb{V}[\mathrm{d}W_i|W_{it}=Q_t].\end{aligned}$$

I conclude the proof by plugging the expressions $\mathrm{d}\overline{W}_t$ and $\mathrm{d}^2\overline{W}_t$ obtained above into a second-order Taylor approximation for $\Delta\overline{W}_t$:

$$\begin{aligned}\Delta\overline{W}_t &= \frac{\mathrm{d}\overline{W}_t}{\mathrm{d}\theta}\theta+\frac{1}{2}\frac{\mathrm{d}^2\overline{W}_t}{\mathrm{d}\theta^2}\theta^2+o(\theta^2)\\ &=\mathbb{E}\left[\frac{\mathrm{d}W_{it}}{\mathrm{d}\theta}|W_{it}\geq Q_t\right]\theta+\frac{1}{2}\left(\mathbb{E}\left[\frac{\mathrm{d}^2W_{it}}{\mathrm{d}\theta^2}|W_{it}\geq Q_t\right]+\frac{g_t(Q_t)}{p}\mathbb{V}\left[\frac{\mathrm{d}W_i}{\mathrm{d}\theta}|W_{it}=Q_t\right]\right)\theta^2+o(\theta^2)\\ &=\mathbb{E}[\Delta W_{it}|W_{it}\geq Q_t]+\frac{1}{2}\frac{g_t(Q_t)}{p}\mathbb{V}[\Delta W_i|W_{it}=Q_t]+o(\theta^2).\end{aligned}$$

□

Proposition 5. Consider a one-time perturbation in return, income and bequest at time $s<t$. Given (20), this creates a perturbation in wealth

$$\mathrm{d}\log W_{is}=\left(1-\frac{Y_{is}+B_{is}-\xi_{is}C_{is}}{W_{is}}\right)\mathrm{d}\log R_{is}+\frac{Y_{is}}{W_{is}}\mathrm{d}\log Y_{is}+\frac{B_{is}}{W_{is}}\mathrm{d}\log B_{is},$$

where I assumed that consumption is chosen after R_{is} is realized but before labor income and bequests are received. The effect of this deviation on wealth at time t is given by (assuming returns, labor income, and bequests do not react to the wealth level):

$$\begin{aligned}\frac{\partial W_{it}}{\partial W_{is}} &= R_{it}\frac{\partial W_{it-1}}{\partial W_{is}}-\frac{\partial C_{it}}{\partial W_{is}}\\ &=\left(R_{it}-\frac{\partial C_{it}}{\partial W_{it-1}}\right)\frac{\partial W_{it-1}}{\partial W_{is}},\end{aligned}$$

which implies

$$\frac{\partial W_{it}}{\partial W_{is}} = \prod_{s+1}^{t} \left(R_{iu} - \frac{\partial C_{iu}}{\partial W_{iu}} \right).$$

Equivalently, in logs

$$\begin{aligned}
\frac{\partial \log W_{it}}{\partial \log W_{is}} &= \prod_{s+1}^{t} \left(R_{iu} - \frac{\partial C_{iu}}{\partial W_{iu}} \right) \left(1 - \frac{Y_{iu} + B_{iu} - C_{iu}}{W_{iu}} \right) \\
&\approx \prod_{s+1}^{t} \left(1 - \frac{Y_{iu} + B_{iu} - C_{iu} + \frac{\partial C_{iu}}{\partial W_{iu}}}{W_{iu}} \right)
\end{aligned}$$

where the approximation error decreases for small period of times and converges to zero in the continuous-time lime. □

B Appendix for Section 1

B.1 Case where Individual Wealth is Non-Differentiable in Perturbation

In the main text, we have focused on the common case in which the deviation in wealth due to the perturbation is differentiable in the perturbation size θ (Assumption 1). In reality, some perturbation may generate non-differentiable or event discontinuous changes in individual wealth. For instance, a change in wealth tax may change the composition of individuals in a country. A change in corporate taxes may change the decision of being a worker and an entrepreneur, generating discontinuous changes in wealth.

Still, one can still characterize the first-order response of the average wealth in the top percentile as long as the set of individuals for which $\theta \to W_{it}(\theta)$ is not differentiable is, itself, infinitesimal. Under this assumption, the average wealth in the top percentile is still differentiable in θ and we have:

$$\mathrm{d}\overline{W}_t = \mathrm{d}\mathbb{E}\left[(W_{it}(\theta) - Q_t)^+ \right].$$

We can get something closer to Proposition 1 by distinguishing between the set of individuals with differentiable and non differentiable wealth. Formally, denote $\mathcal{ND}$ the set of individuals for which ΔW_{it} is not differentiable between 0 and θ. The first-order response in the average wealth in the top percentile is:

$$\begin{aligned}
\mathrm{d}\overline{W}_t &= \mathbb{E}\left[\mathrm{d}W_{it} | W_{it} \geq Q_t \right] \\
&+ \left(\lim_{\theta \to 0} \frac{1}{\Delta\theta} \mathbb{E}\left[(W_{it}(\theta) - Q_t)^+ | W_{it} < Q_t \right] - \lim_{\theta \to 0} \frac{1}{\Delta\theta} \mathbb{E}\left[(Q_t - W_{it}(\theta))^+ | W_{it}(\theta) \geq Q_t \right] \right).
\end{aligned}$$

B.2 Second-Order Approximation for Log Deviations

In Section 1, we derived a simple expression (6) for the deviation of the log average wealth in a top percentile in terms of the wealth-weighted average of individual wealth. I now derive a similar expression that is valid at the second-order. Dividing Proposition 2 by $\overline{W}_t$ implies the following second-order approximation for the average wealth in a top percentile:

$$\frac{\Delta \overline{W}_t}{\overline{W}_t} = \underbrace{\mathbb{E}^{W_{it}}\left[\frac{\Delta W_{it}}{W_{it}}\Big| W_{it} \geq Q_t\right]}_{\text{Intensive margin } O(\theta)} + \underbrace{\frac{1}{2}\frac{g_t(Q_t)Q_t^2}{p\overline{W}_t}\mathbb{V}\left[\frac{\Delta W_i}{W_{it}}\Big| W_{it} = Q_t\right]}_{\text{Extensive margin } O(\theta^2)} + o(\theta^2). \tag{27}$$

Note that this expression is very similar to Gomez (2023) for the growth rate of the average wealth in the top percentile in continuous-time. In particular, note that, when the wealth distribution in the baseline economy has a Pareto tail, we have $g_t(Q_t)Q_t/(p\overline{W}_t) \to \zeta - 1$ where ζ is the tail exponent of the wealth distribution. Hence, holding other things equal, second-order effects are lower when the baseline wealth distribution is higher. Intuitively, a higher level of inequality reduces both churning around the percentile threshold as well as the effect of this churning on the average wealth beyond the top percentile. Quantitatively, $\zeta \approx 1.5$ for the U.S. wealth distribution, which implies that while the first-term (the intensive margin) is equal to the average relative deviation in wealth, the second-term (the extensive margin) is equal to a fourth of the variance of the relative deviation in wealth. Hence, we can expect this second-order effects to be relatively small.

One can rewrite (27) to obtain a second-order approximation for the deviation in the log average wealth in a top percentile:[21]

$$\Delta \log \overline{W}_t = \underbrace{\log \mathbb{E}^{W_{it}}\left[e^{\Delta \log W_{it}}\Big| W_{it} \geq Q_t\right]}_{\text{Intensive margin } O(\theta)} + \underbrace{\frac{1}{2}\frac{g_t(Q_t)Q_t^2}{p\overline{W}_t}\mathbb{V}\left[\Delta \log W_{it}\right]}_{\text{Extensive margin } O(\theta^2)} + o(\theta^2). \tag{28}$$

Note that the "intensive" margin $\log \mathbb{E}^{W_{it}}[e^{\Delta \log W_{it}} | W_{it} \geq Q_t]$ is typically larger than the change in the average log wealth $\mathbb{E}^{W_{it}}[\Delta \log W_{it} | W_{it} \geq Q_t]$ (the first-order term in (6)) due to Jensen inequality: since log is a concave function, the logarithm of the average deviation is always higher than

[21] More precisely, this can be derived from:

$$\begin{aligned}
\Delta \log \overline{W}_t &= \log\left(1 + \frac{\Delta \overline{W}_t}{\overline{W}_t}\right) \\
&= \log\left(1 + \mathbb{E}^{W_{it}}\left[\frac{\Delta W_{it}}{W_{it}}\Big| W_{it} \geq Q_t\right] + \frac{1}{2}\frac{g_t(Q_t)Q_t^2}{p\overline{W}_t}\mathbb{V}\left[\frac{\Delta W_i}{W_{it}}\right] + o(\theta^2)\right) \\
&= \log\left(1 + \mathbb{E}^{W_{it}}\left[\frac{\Delta W_{it}}{W_{it}}\Big| W_{it} \geq Q_t\right]\right) + \frac{1}{2}\frac{g_t(Q_t)Q_t^2}{p\overline{W}_t}\mathbb{V}\left[\frac{\Delta W_{it}}{W_{it}}\right] + o(\theta^2) \\
&= \log \mathbb{E}^{W_{it}}\left[e^{\Delta \log W_{it}}\Big| W_{it} \geq Q_t\right] + \frac{1}{2}\frac{g_t(Q_t)^2 Q_t}{p\overline{W}_t}\mathbb{V}\left[\Delta \log W_{it}\right] + o(\theta^2).
\end{aligned}$$

the average of the log deviation.

B.3 Comparison with Differentiated Kolmogorov-Forward Equation

I now briefly discuss my results relative to an alternative approach, which would study the effect of perturbations in the economy on the wealth distribution by differentiating the Kolmogorov-Forward equation (as in, e.g. Auclert et al., 2021). Suppose that, in the baseline economy, the distribution of wealth evolves according to

$$\mathbb{P}_{t+1} = \mathbb{A}_{t+1}\mathbb{P}_t \tag{29}$$

where $\mathbb{P}_t$ denotes the CDF of the wealth distribution and $\mathbb{A}_{t+1}$ denotes a linear operator encoding how the distribution of individual wealth at time t is related to the distribution of wealth at time $t+1$.

Differentiating (30) with respect to a perturbation of wealth dynamics gives:

$$\mathrm{d}\mathbb{P}_{t+1} = \mathbb{A}_{t+1}\,\mathrm{d}\mathbb{P}_t + (\mathrm{d}\mathbb{A}_{t+1})\,\mathbb{P}_t \tag{30}$$

Solving this recurrence relationship backward gives:

$$\mathrm{d}\mathbb{P}_t = \sum_{s=-\infty}^{t}\left(\prod_{u=s+1}^{t}\mathbb{A}_s\right)\mathrm{d}\mathbb{A}_s\mathbb{P}_{s-1} \tag{31}$$

The intuition for this equation is as follows: the deviation in the law of motion of wealth at some time s generates an instantaneous change in the distribution of wealth at time s given by $\mathrm{d}\mathbb{A}_s\mathbb{P}_{s-1}$. The effect of this deviation at time s on the distribution at time t is then given by the operator $\left(\prod_{u=s+1}^{t}\mathbb{A}_s\right)$, which depends on the baseline law of motion for individual wealth from s to t.

This equation is very useful to compute first-order deviations in the wealth distribution in terms of first-order deviations in the evolution of wealth across periods. However, it is not suited to characterize top wealth shares (which is the object we observe in the data) or to construct counterfactual distributions using reduced-form evidence from the data.

C Appendix for Section 2

C.1 General Model of Wealth Dynamics

We can generalize the results obtained in the main text to much more general statistical processes for wealth. More precisely, under regularity conditions, the deviation for log wealth admits a moving-average (MA) representation; that is,

$$\mathrm{d}\log W_{it} = \sum_{s=-\infty}^{t} \frac{\partial \log W_{it}}{\partial \log W_{is}} \mathrm{d}\epsilon_{is}.$$

where $\mathrm{d}\epsilon_{is}$ denotes the direct effect of the perturbation on log wealth at time s. The process $\partial \log W_{it} / \partial \log W_{is}$ is called the "first-derivative process" for log wealth, and it reflects the extent to which shocks in wealth persist over time. The assumption for this moving average representation to hold is that this process decays quickly enough.

Combining this equation with Proposition 1 gives the resulting change of the average wealth in a top percentile as the cumulative sum of past perturbations in relative wealth times the first-derivative process for log wealth:

$$\mathrm{d}\log \overline{W}_t = \mathbb{E}^{W_{it}} \left[\sum_{s=-\infty}^{t} \frac{\partial \log W_{it}}{\partial \log W_{is}} \mathrm{d}\epsilon_{is} | W_{it} \geq Q_t \right].$$

Different statistical processes for wealth then lead to different expressions for this first-derivative process. Figure 2 plots the first-derivative process for the two models discussed above as a function of s. Both decay over time, reflecting the fact that the persistence of wealth shocks is less than one. In the first model, this is due to the dissipation shocks, while, in the second model, this is due to the presence of additive shocks.

Note that, in a model in which portfolio returns increase with wealth (as in the model with non-homothetic preferences in Gaillard et al., 2023), the first-derivative process for log wealth may be higher than one for some time, before decaying due to death or some other negative force.

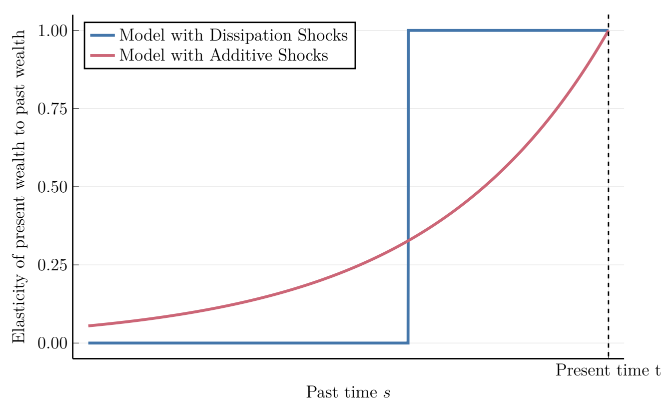


Figure C1: Elasticity of Present Wealth (time t) to Past Wealth (time s) in Random-Growth Models.

Notes: The figure plots the elasticity of wealth at time t with respect to wealth at time $s < t$, $\partial \log W_{it} / \partial \log W_{is}$, as a function of s. In the model where wealth follows a geometric random walk with dissipation shocks (9), this elasticity is given by $\partial \log W_{it} / \partial \log W_{is} = 1_{s \geq t - a_{it}}$. In the model where wealth follows a geometric random walk with additive shocks (15), this elasticity is given by $\partial \log W_{it} / \partial \log W_{is} = \prod_{u=s+1}^{t} (1 - Y_{is} / W_{is})$. The area under these curves can be interpreted as the elasticity of current wealth to a uniform change in past growth rates.

C.2 Comparing my Analytical Results to Existing Results for Tail Indices

I now briefly compare my approach to the existing literature that focuses on characterizing tail indices.

Tail index. I now add a certain number of assumption to the model to ensure that the distribution in the baseline economy is stationary with a Pareto tail. Assume that individual wealth follows the law of motion (9), that the probability of death faced by each individual is constant over time and equal to δ, and that the distribution of $\log G_{it}$ is i.i.d. Then, it is well known that the distribution of wealth has a Pareto tail with tail index ζ given by the unique positive number satisfying:

$$\log \mathbb{E}\left[G_{it}^{\zeta}\right] = \delta.$$

Similarly, if individual wealth follows the law of motion (15), that the distribution of G_{it} is i.i.d. with $\mathbb{E}[\log G_{it}] < 0$, and that the distribution of Y_{it} is i.i.d with thin tails, then it is well known that the distribution of wealth has a Pareto tail with tail index ζ given by the unique positive number satisfying:

$$\log \mathbb{E}\left[G_{it}^{\zeta}\right] = 0.$$

Deviation in tail indices. In both cases, consider a *permanent* perturbation in the law of motion of individual wealth. Differentiating the expressions above tells us that, if the perturbated wealth distribution is also Pareto, then the resulting deviation in the tail index is:

$$\mathrm{d}\log\zeta = -\frac{\mathbb{E}^{G_{it}^{\zeta}}\left[\mathrm{d}\log G_{it}\right]}{\mathbb{E}^{G_{it}^{\zeta}}\left[\log G_{it}\right]}.$$

where $\mathbb{E}^{G_{it}^{\zeta}}$ denotes the distribution of growth rates "tilted" by G_{it}^{ζ}.

Physical interpretation of tilted distribution. As discussed in Touchette (2009), the distribution of growth rates that is exponentially tilted by G_{it}^{ζ} can be interpreted as the distribution of prior growth rates conditional on being the top. Indeed, combining the previous equation with (11) implies

$$\mathbb{E}^{W_{it}}\left[\sum_{s=t-a_{it}+1}^{t} \mathrm{d}\log G_{is} | W_{it} \geq Q_t\right] \sim \frac{\mathbb{E}^{G_{it}^{\zeta}}\left[\mathrm{d}\log G_{it}\right]}{\mathbb{E}^{G_{it}^{\zeta}}\left[\log G_{it}\right]} \log \overline{W}_t \quad \text{as } p \to 0.$$

In the particular case of a uniform increase in growth rate, this implies

$$\mathbb{E}^{W_{it}}\left[a_{it} | W_{it} \geq Q_t\right] \sim \frac{\log \overline{W}(p)}{\mathbb{E}^{G_{it}^{\zeta}}\left[\log G_{it}\right]} \text{as } p \to 0.$$

This expression shows that the lifetime average "speed" of individuals in the right tail is asymptotically equal to the average growth rate of wealth in the distribution that is tilted by G_{it}^{ζ}; that is, $\mathbb{E}^{G_{it}^{\zeta}}[\log G_{it}]$.

One can easily see that $\mathbb{E}^{G_{it}^{\zeta}}[\log G_{it}]$ can also be seen as the derivative of the CGF function of log growth rates at ζ; that is,

$$\mathbb{E}^{G_{it}^{\zeta}}[\log G_{it}] = \frac{\partial \log \left[G_{it}^{\zeta}\right]}{\partial \zeta}.$$

Now, it is well known that the CGF can be written as a Taylor expansion of cumulants; that is

$$\log \mathbb{E}[G_{it}^{\zeta}] = \zeta\mu + \frac{1}{2}\zeta^2\sigma^2 + \frac{1}{6}\zeta^3 \cdot \text{skewness} \cdot \sigma^3 + \frac{1}{24}\zeta^4 \cdot \text{kurtosis} \cdot \sigma^4 + \dots$$

where μ, σ, skewness, kurtosis to the mean, standard deviation, skewness, and kurtosis of $\log G_{it}$. Hence, the 'speed' of individuals reaching the right tail of the wealth distribution can be rewritten as the derivative of that expression with respect to ζ:

$$\begin{aligned}\lim_{p\to 0} \frac{\log \overline{W}_p}{\mathbb{E}^{W_{it}}[a_{it}|W_{it} \geq Q_t]} &= \partial_\zeta \log \left[G_{it}^{\zeta}\right] \\ &= \mu + \zeta\sigma^2 + \frac{1}{2}\zeta^2 \cdot \text{skewness} \cdot \sigma^3 + \frac{1}{6}\zeta^3 \cdot \text{kurtosis} \cdot \sigma^4 + \dots\end{aligned}$$

Hence, for a given tail index, a higher mean, variance, skewness and kurtosis lead to a faster wealth path to top percentiles. Note that, in the case in which log growth rates are serially correlated over time, these cumulants should be understood as the derivatives of the scaled cumulating generating function.[22]

Serial correlation. All these results can be easily extended to models with serial correlations between individual growth rates (or models with type-specific distributions of growth rates), which is an important pattern in the data. In this type of model, Saporta (2005) and Beare and Toda (2022) show that the tail index is characterized by

$$\lim_{T\to\infty} \frac{1}{T} \log \mathbb{E}\left[\left(\prod_{t=1}^{T} G_{it}^{\zeta}\right)\right] = \delta.$$

Differentiating this expression gives:

$$\mathrm{d}\log\zeta = -\lim_{T\to\infty} \frac{\mathbb{E}^{\prod_{t=1}^{T} G_{it}^{\zeta}}\left[\sum_{t=1}^{T} \mathrm{d}\log G_{it}\right]}{\mathbb{E}^{\prod_{t=1}^{T} G_{it}^{\zeta}}\left[\sum_{t=1}^{T} \log G_{it}\right]}.$$

[22] That is, $\zeta \to \lim_{T\to\infty} \frac{1}{T} \log \mathbb{E}\left[\prod_{t=1}^{T} G_{it}^{\zeta}\right]$.

This type of expression is obtained by Gomez and Gouin-Bonenfant (2024). Note that this equation implies that perturbating uniformly the growth rate of all agents in the economy must generates a change in the tail index that is as big, in magnitude, as perturbating the growth rate of only one type of agents. Yet, we discussed in Section 2 how an increase in the growth rate of agents that end up at the top of the wealth distribution typically generates much larger changes in top wealth shares relative to a uniform increase in the growth rate of all agents in the economy. This type of distinction cannot be picked up by tail indices, as they only characterize the asymptotic deviation in top wealth shares as $p \to 0$. Overall, focusing on changes in tail indices can sometimes be very unhelpful to characterize the actual movements in top wealth shares.

C.3 Diagnostic Tools for Other Counterfactuals

Relative to Table 1 in the main text, Table C1 examines the effect of a different counterfactual, which corresponds to an increase in the growth rate of only a subset of agents. I consider an increase in the growth rate of the agent with the higher average growth rate. Note, however, that the interpretation of this "high-type" varies across models so one cannot really do direct comparison across models — the goal of this table is simply to show that my first-order approximations performs well for this type of counterfactuals too.

Table C1: Long-Run Semi-Elasticity of Wealth to Growth Rate of High-Type

	Non-infinitesimal deviation			First-order effect		
	Top 0.01%	Top 100%	Difference	Top 0.01%	Top 100%	Difference
Moll et al. (2022)	149	8	141	134	5	129
Gabaix et al. (2016)	37	5	32	36	5	31
Gomez and Gouin-Bonenfant (2024)	15	0	15	15	0	15

Notes: The table reports the effect of a 1pp. increase in the high-type in each model, as well as the first-order effect computed using the analytical results derived in Section 2.