Uncovering the Dynamics
of the Wealth Distribution

Thomas Blanchet[*]

Job Market Paper

This version: September 30, 2022

☞ GET LATEST VERSION

Abstract

I introduce a new way of decomposing the evolution of the wealth distribution using a simple continuous time stochastic model, which separates the effects of mobility, savings, labor income, rates of return, demography, inheritance, and assortative mating. Based on two results from stochastic calculus, I show that this decomposition is nonparametrically identified and can be estimated based solely on repeated cross-sections of the data. I estimate it in the United States since 1962 using historical data on income, wealth, and demography. I find that the main drivers of the rise of the top 1% wealth share since the 1980s have been, in decreasing level of importance, higher savings at the top, higher rates of return on wealth (essentially in the form of capital gains), and higher labor income inequality. I then use the model to study the effects of wealth taxation. I derive simple formulas for how the tax base reacts to the net-of-tax rate in the long run, which nest insights from several existing models, and can be calibrated using estimable elasticities. In the benchmark calibration, the revenue-maximizing wealth tax rate at the top is high (around 12%), but the revenue collected from the tax is much lower than in the static case.

1 Introduction

Wealth inequality has sharply increased in the United States. By combining income tax returns with macroeconomic balance sheets, Saez and Zucman (2020a) find that the share of wealth owned by the top 1% has increased by more than 10 pp. since the late 1970s.[1] Kuhn, Schularick, and Steins (2020) find a similar trend using survey data (Figure 1).[2]
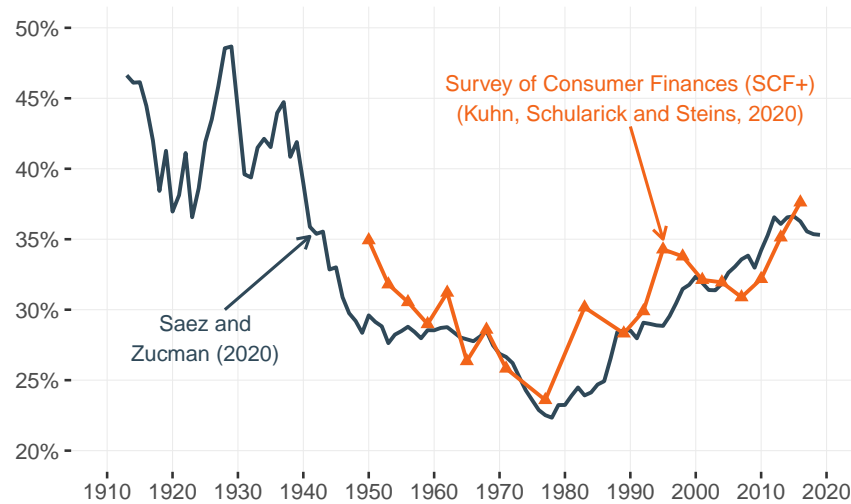
Figure 1: Top 1% Wealth Share in the United States

But despite this growing amount of data documenting the historical evolution of the distribution of wealth, our understanding of the drivers of this trend remains elusive. Is it solely a consequence of the rise of labor income inequality? Or is it also the result of higher rates of return? What about capital gains? Does it have anything to do with the decline of the estate tax? Does it reflect changes to the distribution of saving rates? These are basic questions, and if we had long-run longitudinal data on income and wealth, they would be fairly easy to answer. After all, they simply ask how observable phenomena affect the process of wealth accumulation in direct, mechanical ways. But because longitudinal wealth data is scarce and limited, in practice, answering them has remained a challenge. That we don't have a straightforward understanding of such proximate causes of wealth inequality is a problem in and of itself, but it also impedes our ability to answer several related questions. It makes it more difficult to

---

[1]Saez and Zucman (2020a) is a revision of Saez and Zucman (2016), which accounts for heterogeneous returns, as well as the more important role of private business wealth at the top found by Smith, Zidar, and Zwick (2022). See also Saez and Zucman (2020b).

[2]Other estimates, notably Smith, Zidar, and Zwick (2022), also find a similar increase of the top 1% share, although they also find a somewhat more muted increase than Saez and Zucman (2020a) for the top 0.01% and narrower top groups.

adequately calibrate economic models of the wealth distribution, which are used to investigate the deeper, underlying causes of rising inequality. It limits our understanding of the effect of widely discussed policies, such as wealth and inheritance taxation.

Contribution   This paper addresses this issue by introducing a new way to decompose the evolution of wealth inequality in terms of individual-level factors, and in spite of the lack of panel data. The decomposition accounts for demography, inheritance, assortative mating, labor income, rates of return, consumption, and, importantly, mobility. It does so in a way that only requires repeated cross-sections, and therefore can be applied to historical data. The approach is tractable and allows for a transparent, visual identification of the key parameters.

I estimate this decomposition in the United States since the 1960s, and in doing so, I establish the main direct drivers of the rise of wealth inequality over that period. To that end, I use historical microdata on the distribution of income and wealth (Saez and Zucman, 2020a), which I combine with a large set of data from censuses, surveys, demographic tables, and macroeconomic accounts. For example, I construct measures of assortative mating as well as age-specific marriage and divorce rates since the 1960s to estimate their effect on the wealth distribution. I also microsimulate the entire demographic history of the United States since the mid-19th century to statistically reconstruct intergenerational linkages, and combine them with distributional parameters for intergenerational wealth transmission as a way to measure the effect of inheritance and of the estate tax.

Finally, I develop a way to incorporate wealth taxation within the framework of this paper, derive a simple formula for how the wealth distribution would react to any wealth tax in the long run, and use my empirical estimates to calibrate it for the United States. The framework of the paper presents several advantages here as well. A typical difficulty when analyzing wealth or capital taxation is that the standard models lead to a menagerie of "corner solutions," and the associated policy recommendations are often extreme, fragile, and hard to interpret. Small changes in parameters (e.g., an intertemporal elasticity of substitution just above or below one) lead to diametrically opposite results (e.g., an optimal tax of 0% or 100%) (Straub and Werning, 2020).[3] In light of this situation, Saez and Stantcheva (2018) have argued for a simpler framework — one that centers the same equity-efficiency trade-off that governs the theory of optimal labor income taxation (Piketty and Saez, 2013b) and, in practice, dominates policy considerations. In this framework, the elasticity of capital supply with respect to the net-of-tax rate directly dictates the optimal level of capital taxation.[4] But finding a way to

---

[3]This is for example the case in the reanalysis of the model of Judd (1985) by Straub and Werning (2020) in the case with no government spending.

[4]As long as this elasticity is finite, positive capital taxes are desirable. We can interpret earlier arguments that capital should never be taxed (Chamley, 1986; Judd, 1985) as the consequence of assuming an infinite elasticity

model this elasticity in a realistic yet tractable way, without facing the same degeneracy issues as earlier models, remains difficult. Here, because the model is stochastic and features mobility, it organically leads to well-behaved steady-states and finite elasticities of wealth with respect to the net-of-tax rate, under a wide range of economic behaviors, and without the need to resort to *ad hoc* modeling devices such as wealth in the utility function (e.g., Saez and Stantcheva, 2018). Because the continuous-time formalism is highly tractable, I can obtain simple formulas that combine and extend insights from earlier studies of progressive wealth taxation (e.g. Jakobsen et al., 2020; Saez and Zucman, 2019).

Overview of the Methodology   This paper uses a stochastic, continuous-time framework, and its methodology rests on two results from stochastic calculus. First, I use Gyöngy's (1986) theorem to prove that two factors drive the dynamics of the wealth distribution: the mean and the variance of savings, conditional on wealth. This remains true even in models that feature heterogeneous agents and complex shocks: Gyöngy's (1986) theorem shows that this complexity can be marginalized out, and therefore that the evolution of the wealth distribution can, under very general conditions, be summarized by a single stochastic differential equation (SDE). This SDE captures the two forces that shape the distribution of wealth: the average savings of the different parts of the distribution (a.k.a. the *drift*), and the mobility across the distribution (a.k.a. the *diffusion*).

Having reduced the dynamic of wealth to this simple model, I apply the Kolmogorov (1931) forward equation, a standard result of stochastic calculus, which is known to be extremely useful in understanding how wealth distributions get their power-law shape (Gabaix, 2009). This equation relates how wealth evolves stochastically at the individual level with how the distribution of wealth evolves deterministically in the aggregate. Traditionally, this equation is used deductively: starting from a set of parameters for how individual wealth evolves, we apply it to determine how the distribution changes. This paper argues that it is possible to use that equation *inductively*. That is, looking at how the distribution changes, we can infer the parameters (the mean and the variance of savings) that govern the underlying individual wealth accumulation process.

That we can infer the individual wealth accumulation process — including a mobility parameter — from cross-sectional data may *prima facie* be counterintuitive. This paper provides a simple explanation for that fact. Basically, mobility is a force that spreads out observations and, therefore, flattens the wealth density locally. Consequently, the changes it exerts on the distribution depend on how flat it is to begin with. In contrast, the local effect exerted by the drift (i.e., how much wealth grows at a given point *on average*) is akin to a local translation

(Piketty and Saez, 2013a; Saez and Stantcheva, 2018).

4

and is independent of the shape of the density. This distinction is not a mere technicality: it is the central piece of the mechanism that allows steady-state distributions to emerge in the first place, and it has empirically measurable consequences. By looking at how the local evolution of the wealth density relates to its current shape, I can therefore estimate both parameters.

I estimate the full model, while also accounting for income and the role played by demography, assortative mating, and inheritance. I show that the model correctly reproduces the past evolution of wealth inequality. When I compare my findings to the (albeit scarce and limited) longitudinal survey data on wealth that exists in the United States, I find that my results are consistent. I use the model to decompose the evolution of wealth inequality and to generate simple counterfactuals showing how wealth inequality would have evolved if certain factors (e.g., labor income inequality, capital gains) had stayed at their 1980 levels.

Finally, I include wealth taxation into the model, and derive a simple formula for how the wealth distribution would react to an arbitrary wealth tax in the long run, which connects the field of capital taxation to the theories of the wealth distribution. This formula depends on three factors: mobility across the wealth distribution, tax avoidance, and savings responses. Mobility is built into the framework and calibrated using this paper's estimate. For tax avoidance and savings responses, I rely on reduced-form expressions that depend on simple, estimable behavioral elasticities, which I calibrate using quasi-experimental evidence from the literature (Brülhart et al., 2016; Seim, 2017; Jakobsen et al., 2020; Zoutman, 2018; Ring, 2020; Londoño-Vélez and Ávila-Mahecha, 2021). A comprehensive tool to simulate the effect of arbitrary wealth taxes in the United States, using this paper's formulas, is available online at `https://thomasblanchet.github.io/wealth-tax/`.

Findings   After applying the decomposition to the United States since 1962, I find that the main drivers of the rise of the top 1% wealth share since the 1980s have been, in decreasing level of importance, higher savings at the top, higher rates of return on wealth (essentially in the form of capital gains), and higher labor income inequality. Other factors have played a minor role.

Notably, the wealth accumulation process that I estimate features a rather large heterogeneity of savings for a given level of wealth and, therefore, a large degree of mobility across the distribution. That amount of mobility is comparable to what I independently find in longitudinal survey data from the Survey of Consumer Finances (SCF) and the Panel Study of Income Dynamics (PSID), and in a panel of the 400 richest Americans from Forbes. This finding cuts against the "dynastic view" of wealth accumulation, and suggests that existing models of the wealth distribution underestimate the extent of wealth mobility. This has several implications.

First, the large heterogeneity of savings implies that high levels of wealth inequality can be

sustained even if the wealthy consume, on average, a sizeable fraction of their wealth every year. Second, high mobility offers a straightforward answer to a puzzle pointed out by Gabaix et al. (2016): that the usual stochastic models that can explain steady-state levels of inequality are unable to account for the speed at which inequality has increased in the United States.[5] But high mobility leads to faster dynamics and, therefore, can account for the dynamic of wealth we observe in practice.[6]

The study of wealth taxation also yields new insights. First, the degree of mobility across the wealth distribution — which explicitly appears in the formula I derive — generates a mechanical effect that fixes many degeneracy issues from deterministic models. And it is a crucial determinant of the response of the wealth stock to a tax. Indeed, without mobility, an annual wealth tax would affect the same people repeatedly, and absent behavioral responses, the tax base would eventually go to zero. With mobility, new, previously untaxed wealth continuously enters the tax base, which avoids that phenomenon. Saez and Zucman (2019) study this effect, but within a very restrictive setting — my formula extends and operationalizes their idea to more complex and realistic cases, and this affects some of the conclusions.[7] In light of this result, the significant mobility that I find suggests that a wealth tax could raise more revenue, all other things being equal. Second, even in the simplest model, the reduced-form elasticity of wealth with respect to the net-of-tax rate is nonconstant — much higher for low rates than higher ones. (This directly results from the fact that the complete formula depends on both the average and the marginal tax rate.) As a consequence, the revenue-maximizing rate for a linear wealth tax above $50m is high (12% in my benchmark calibration), yet the revenue raised from said tax in the long run is quite low, only about a quarter of what the tax would raise in the absence of response.[8] Third, my simulations suggest that the effects of an annual wealth tax differ in fundamental ways from those of an inheritance tax of seemingly comparable magnitude. I show in a simple model that this distinction is driven by the lifetime of generations: the longer people live, the less potent the estate tax compared to a wealth tax.

---

[5]Gabaix et al. (2016) primarily focus on income inequality, but their formal findings apply to wealth as well.

[6]This solution differs from the solutions proposed by Gabaix et al. (2016), which involve introducing temporary changes to the drift to accelerate dynamics at the beginning of a transition.

[7]Saez and Zucman (2019) consider a wealth tax at a constant *average* rate above a threshold. My formulas apply to arbitrary wealth taxes, including the more common case of a constant *marginal* rate within a bracket.

[8]If the elasticity were constant, a 12% revenue-maximizing rate would be associated with a revenue reduction of 60% compared to the inelastic case, as opposed to 75% here.

2 Literature Review

2.1 Models of the Wealth Distribution

The conventional way of studying the determinants of wealth inequality is to construct a model of an economy where people face various individual shocks. The prototypical model for this line of work comes from Aiyagari (1994) and Huggett (1993), who study the distribution of wealth in Bewley (1977) type models in which people face idiosyncratic uninsurable labor income shocks. In these models, people accumulate wealth for precautionary or consumption smoothing purposes. But these motives quickly vanish as wealth increases, so they cannot rationalize large wealth holdings. As a result, these models notoriously underestimate the real extent of wealth inequality.

The literature, then, has searched for ways to extend these models and match the data. One possibility involves additional saving motives for the rich, such as a taste for wealth (Carroll, 1998) or bequests (De Nardi, 2004). Another solution is to introduce idiosyncratic stochastic rates of returns in the form of entrepreneurial risk (Quadrini, 2000; Cagetti and De Nardi, 2006; Benhabib, Bisin, and Zhu, 2011). Yet another option introduces heterogeneous shocks to the discount rate (Krusell and Smith, 1998).

Recent contributions have zoomed in on specific features and issues. Following the finding from Bach, Calvet, and Sodini (2020) (in Sweden) and Fagereng et al. (2020) (in Norway) that the wealthy enjoy higher returns, several papers have examined this mechanism in the United States. Xavier (2021) finds that heterogeneous returns play a critical role in explaining the current levels of wealth inequality. Cioffi (2021) embeds a portfolio choice model in a wealth accumulation model. In this paper, because of heterogeneous portfolio composition, households have different exposures to aggregate risks. The wealthy invest more in high-risk, high-return assets like equity. As a result, Cioffi (2021) finds that abnormally high stock market returns have been a critical driver of the rise in wealth inequality in the United States. Hubmer, Krusell, and Smith (2018), on the other hand, use a different model and conclude that changes in tax progressivity, rather than changes in rates of return, explain the rise in wealth inequality.

In general, models of the wealth distribution seek to explain two facts: that wealth inequality is high (that is, higher than income inequality) and that its top tail is shaped like a power law. Existing models differ across many dimensions, but when it comes to explaining these facts, they usually share a common mechanism: namely, that wealth accumulates through a series of individual multiplicative random shocks, with frictions at the bottom. Assuming that the shocks have adequate properties, such models can generate both high wealth inequality and power-law tails. In discrete time, we can formulate this idea using the theory of Kesten (1973) processes. Assume that the wealth w_{it} of person i evolves according to a linear recurrence equation with

random coefficients: $w_{i,t+1} = a_{it}w_{it} + b_{it}$, where a_{it} is a multiplicative shock (typically reflecting stochastic returns or preference shocks) and b_{it} is a friction term (typically reflecting labor income). Then, under very general conditions, wealth admits a steady-state distribution with a power-law tail, whose fatness is determined by the properties of the multiplicative shock a_{it} (Kesten, 1973; Grincevićius, 1975; Vervaat, 1979; Goldie, 1991; Grey, 1994). We can formulate a continuous-time version of this mechanism using stochastic calculus (Gabaix, 2009).[9] This formalism presents major advantages in terms of flexibility and tractability, notably because of the Kolmogorov (1931) equation, which establishes a straightforward relationship between the distribution of the individual random shocks and the evolution of the wealth distribution in the aggregate.
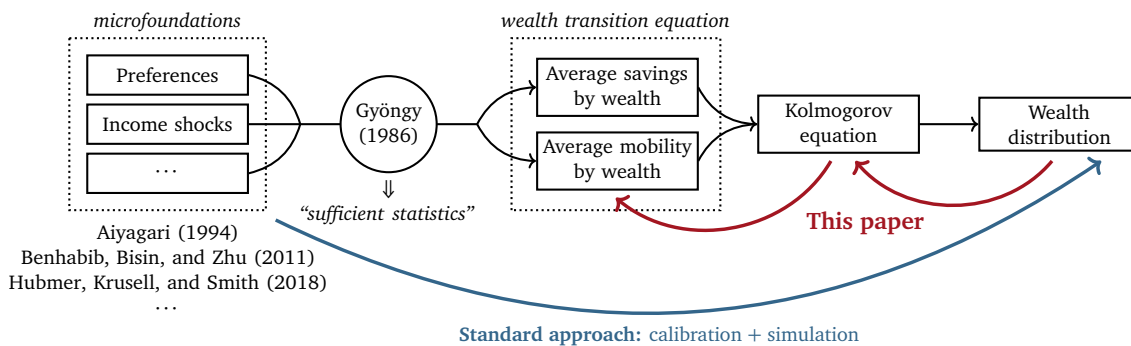

Figure 2: The Different Approaches for Studying the Wealth Distribution

We can understand the relation of the current paper to this literature with the help of Figure 2. The "standard approach," represented by the blue arrow at the bottom, takes a *deductive* path. Starting from a set of microfoundations, it solves and simulates the model to arrive at a wealth distribution. Assuming the model matches the data, we can use it to study policies and generate counterfactuals. Implicitly or explicitly, that approach can be broken down into several steps. First, microfoundations lead to a wealth transition equation that effectively determines the dynamics of the wealth distribution. For every level of wealth, two parameters characterize this equation: the average, and the variance of wealth growth (i.e., mobility). That these two parameters are sufficient to characterize the dynamics of the distribution can be rigorously proven using Gyöngy's (1986) theorem. Then, I can feed these parameters into the Kolmogorov (1931) equation, which describes how the wealth density evolves.

This paper takes the reverse path, as characterized by the two red arrows in Figure 2. I start from the data on the evolution of the wealth distribution and then use the Kolmogorov (1931) equation to *infer* the parameters of the wealth transition equation. Many different microfounda-

---

[9]Saporta and Yao (2004) rigorously study the the continuous time version of Kesten (1973) processes. In this paper I consider simpler — and less literal — continuous time analogs to Kesten (1973) processes.

tions can lead to the same wealth transition equation; therefore, in my approach, the complete model of the economy remains unknown. Yet knowing just the wealth transition equation already opens many possibilities. It is sufficient to study many mechanisms, counterfactuals, and policies. And when more complete models remain necessary, the approach makes it easier to discriminate between them.

We can, in particular, compare this paper to two recent studies that also use a primarily empirical approach within the framework of stochastic wealth accumulation models. Benhabib, Bisin, and Luo (2019) fit a model of stochastic wealth accumulation to data for the United States. They start from an explicitly microfounded model of consumption and use the method of simulated moments to identify the parameters that best replicate observed data. These parameters determine the shape of the utility function and the process for the rate of return. There are four main differences between their approach and mine. First, they estimate consumption by estimating a two-parameter utility function, while I estimate a nonparametric profile of saving rates by wealth. Their approach is more structural and tightly parameterized; my approach is more flexible and reduced form. Second, they use direct information on wealth mobility in a given year, from the 2007–2009 panel of the SCF, to estimate their model; I validate my estimate of mobility against the SCF, but I do not use it directly. Third, their model is nonlinear and estimated using the method of moments; my approach works by fitting a univariate linear equation for each level of wealth, making the source of identification highly explicit and easy to check graphically. Fourth, their primary goal is to replicate the levels rather than the trends in wealth inequality, so their main model assumes that the economy is at its steady state. In a separate exercise, they additionally replicate trends. My approach directly reproduces both levels and trends by construction — and in fact, I use the trends as an essential source of identification.

Gomez (2022) decomposes the evolution of top wealth shares in the United States while accounting for the role of demography and mobility. Like this paper, Gomez (2022) decomposes changes in the wealth distribution that are caused by the first moment of wealth growth rates (average savings) and by the second moment (mobility).[10] Like this paper, Gomez (2022) finds an important role for mobility. His methodology, however, requires longitudinal data, for which he uses the Forbes 400 ranking of the wealthiest Americans.[11] In contrast, I use historical data on income and wealth to directly infer this parameter, allowing me to analyze the rise of wealth inequality since 1962. In contrast, the Forbes 400 rankings started in 1982. My analysis uses the entire wealth distribution, not only the top of the tail. Finally, I also account for labor

---

[10]The decomposition in Gomez (2022) also accounts for higher moments of the distribution of wealth growth rates (i.e., skewness, kurtosis, etc.), but he finds that these have a negligible impact on the wealth distribution.

[11]Gomez (2022) also applies his methodology without direct panel data but with a separate estimate of the variance of wealth growth. However, estimating this variance still requires some form of longitudinal data.

income, inheritance taxation, and assortative mating. Gomez (2022) and this paper provide complementary evidence of the dynamics of wealth accumulation and find a similar result: that mobility across the wealth distribution is sizeable and plays a crucial role in shaping the wealth distribution.

2.2 Synthetic Saving Rates

Another way of using wealth distribution data to study the drivers of wealth inequality is to construct *synthetic saving rates* for the different parts of the wealth distribution (e.g., Saez and Zucman, 2016; Kuhn, Schularick, and Steins, 2020; Garbinti, Goupille-Lebret, and Piketty, 2021). Synthetic savings are constructed as if each bracket of the wealth distribution was a single infinitely lived individual: for example, if the average wealth of the top 1% is $10m at the start of the year and $11m at the end, then the top 1% had $1m in synthetic savings. As such, synthetic savings are a composite measure that combines the effects of average wealth growth with the effect of mobility. They do not distinguish between them.

Since my approach explicitly accounts for mobility, it lets me disentangle the two effects. In the spirit of Gomez (2022), I show that synthetic savings are the sum of several terms: one that depends on average wealth growth, one that depends on mobility, and additional terms that account for other factors, such as demography. This decomposition makes it possible to model synthetic savings more realistically.

Notably, I find that the way synthetic saving rates combine the effect of mobility with other factors is endogenous to the wealth distribution itself. That is, synthetic savings will differ depending on whether inequality is high or low, even if people's actual behavior is the same. To eliminate this endogeneity, we would have to assume zero mobility in the wealth distribution.

2.3 Wealth and Capital Taxation

A well-known, early result in the theory of optimal taxation is that, in a standard neoclassical model, the optimal linear tax rate on capital is zero, even from the perspective of workers who own no capital (Chamley, 1986; Judd, 1985).[12] This result has been overturned in more sophisticated models which introduce, for example, uncertainty (Aiyagari, 1994), incomplete markets (Farhi, 2010), heterogeneous altruism (Farhi and Werning, 2013), tax progressivity (Saez, 2013) or capital accumulation by workers (Bassetto and Benhabib, 2006). More recently,

---

[12]Another famous rationale for not taxing capital comes from Atkinson and Stiglitz (1976). In their model, there is no heterogeneity of wealth conditional on income. Therefore, assuming income can be taxed, taxing wealth provides no additional equity gains while also distorting intertemporal choices. This justification for not taxing capital would not apply to my model since wealth is heterogeneous conditional on income.

Straub and Werning (2020) have reanalyzed the models of Chamley (1986) and Judd (1985), and have overturned many of their conclusion: they show that significant capital taxes are in fact often desirable within these two models.

Overall, the main issue with the theory of capital taxation is that its models tend to behave erratically. Many of its results focus on edge cases and corner solutions, which are highly dependent on the specific primitives of the economy, can be hard to interpret, and imply unrealistic behavior. As a result, the theory has remained of limited use for guiding policy. For example, a common interpretation of the zero-tax result of Chamley (1986) and Judd (1985) is that their model implicitly assumes the capital stock to be infinitely elastic.[13] This assumption makes capital taxation infinitely costly, so the equity gain from taxation is never worth the efficiency cost. Such a setting arguably constitutes an extreme edge case. In more realistic models featuring a finite elasticity, the usual equity-efficiency trade-off would determine the desirability of capital taxation (Piketty and Saez, 2013a; Saez and Stantcheva, 2018). In line with this view, Saez and Stantcheva (2018) ultimately argued that the long-run elasticity of the capital stock to the net-of-tax rate is a sufficient statistic for the optimal design of capital taxation and developed formulas for optimal tax rates, similar to those that exist for labor income (Saez, 2001; Piketty and Saez, 2013b).

However, the value of that elasticity remains elusive. In the *short run*, several empirical papers have used quasi-experimental settings to estimate it: Seim (2017) in Sweden, Londoño-Vélez and Ávila-Mahecha (2021) in Colombia, Brülhart et al. (2016) in Switzerland, Zoutman (2018) in the Netherlands and Jakobsen et al. (2020) in Denmark. With some exceptions, these elasticities tend to be small. This is consistent with the view that a government trying to raise revenue with a one-off, unexpected wealth tax could choose a very high marginal rate. But the policy-relevant parameter is the *long-run* elasticity, which is far more uncertain and likely larger. Indeed, the short-run elasticity only captures tax avoidance or short-run saving responses. But over time, wealth taxes also keep people from accumulating wealth, either through mechanical (lower post-tax rates of return) or behavioral effects (lower savings). This process slowly erodes the tax base. But because it takes a long time to materialize, it is hard to cleanly identify it in the data.

This paper provides a useful framework for addressing this issue. We can start from the empirically estimated wealth transition equation, which reproduces the true evolution of the wealth distribution. Then we can modify that equation to incorporate a wealth tax, as well as an arbitrary set of behavioral responses. For this paper, I focus on two effects (tax avoidance and a decrease in savings), but the framework is flexible enough to accommodate many

---

[13]See Auerbach and Hines (2001), Piketty and Saez (2013a), and Saez and Stantcheva (2018), although this interpretation is not universally accepted, see Straub and Werning (2020).

other extensions. Using the modified equation for the evolution of wealth, we can simulate counterfactual evolutions of the wealth distribution and therefore estimate how the tax base would react to any wealth tax. We can fully simulate the model to observe transitory dynamics. However, if we focus on the long run, we can obtain simple analytical formulas that characterize the alternative steady-state. In my model, the long-run elasticity of capital supply remains finite in all cases because of mobility, but setting the mobility parameter to zero recovers the infinite elasticity as in Chamley (1986) and Judd (1985).

We can connect the approach in this paper to two recent papers that try to assess the long-run effects of wealth taxes. Jakobsen et al. (2020) use their short-run elasticity estimates to calibrate a structural model of savings at the top. They indeed find a higher elasticity in the long run. Saez and Zucman (2019) consider the problem of taxing the very top of the wealth distribution (billionaires) using data from the Forbes rankings. These two papers provide models that shed different lights on the problem. Jakobsen et al. (2020) model wealth accumulation using a deterministic model of intertemporal choice. This model features standard preferences over a consumption path and a taste for end-of-life wealth (i.e., bequests). They use it to derive analytical expressions linking the long-run elasticity of wealth to the short-run elasticity and preference parameters. This model emphasizes the role of behavioral savings responses, but as it is deterministic, it does not account for the role of mobility. In contrast, Saez and Zucman (2019) only focus on mobility. Because they look at billionaire wealth, they assume that the role of consumption is negligible. In that model, mobility is the sole determinant of wealth elasticity in the long run. If it is low, then a wealth tax ends up taxing the same people again and again: as their wealth mechanically goes down, so does the tax base. Therefore, the elasticity of taxable wealth is high, and the ability to tax wealth is limited in the long run. However, if mobility is high, the tax base often gets renewed. Individuals are subjected to the tax for shorter periods, with new, previously untaxed wealth entering the tax base regularly: as a result, the elasticity is lower.

My model combines the mechanisms from both papers and accounts for tax avoidance. Combining the mechanisms as such substantively alters some of the conclusions. In Jakobsen et al. (2020), there is a mechanical effect of wealth taxes, and because the model is deterministic, this mechanical effect becomes infinite in the long run. Their model can generate a finite elasticity of capital supply overall, because an infinite behavioral effect compensates for the infinite mechanical effect in the opposite direction. However, looking at these effects separately in the long run still runs into degeneracy problems.[14] In comparison, because I introduce mobility, my model remains well-behaved at the steady state even if I shut down behavioral responses. In Saez and Zucman (2019), the wealth tax can only consist of a constant *average* tax

---

[14] Jakobsen et al. (2020) circumvent the issue by looking at a long but finite horizon (30 years).

rate over a threshold. In comparison, my model allows for arbitrary tax schedules, including the more standard case of a constant *marginal* rate above a threshold. Far above the tax threshold, the marginal and the average tax rate are similar, so their model behaves similarly to the mechanical component of my mine. Close to the threshold, however, the distinction between them matters. In particular, I find that under a purely mechanical model, even confiscatory wealth tax rates would raise a non-negligible revenue from people just above the tax threshold who recently entered the tax base. And so the Laffer curve never goes to zero. As a result, unlike Saez and Zucman (2019), I conclude that the mechanical model remains insufficient to characterize revenue-maximizing tax rates: doing so requires a behavioral response.

3 Theory

Time is continuous, indexed by t. Individuals are indexed by i. Individual wealth w_{it} evolves stochastically. I model this evolution using an Itô process, which can flexibly model most continuous-time stochastic processes. Such a process is locally characterized by two parameters: the *drift* μ_{it}, and the *diffusion* σ_{it}^2. Over a small period of time $\mathrm{d}t$, the change $\mathrm{d}w_{it}$ in the value of the wealth w_{it} of individual i at time t has mean $\mu_{it}\,\mathrm{d}t$ and variance $\sigma_{it}^2\,\mathrm{d}t$. This process is represented in the form of a SDE and commonly written as follows:

$$\mathrm{d}w_{it} = \mu_{it}\,\mathrm{d}t + \sigma_{it}\,\mathrm{d}B_{it}$$

This section will explain how the parameters for drift and diffusion at the individual level can be connected to the aggregate evolution of the wealth distribution (while accounting for heterogeneity, demography, etc.), and therefore how they can be inferred from changes in the shape of the wealth distribution.

3.1 Income and Consumption

3.1.1 Evolution of Individual Wealth

Individual i has idiosyncratic consumption, labor income, and rate of return on their wealth. Similarly to Carroll's (1997) interpretation of Friedman's (1957) permanent income hypothesis, labor income is the sum of a permanent (i.e., slow-moving) component y_{it}, and of transitory shocks with variance v_{it}^2 — with y_{it} and v_{it}^2 both being arbitrary bounded stochastic processes. Consumption and rates of return follow a similar model, with slow-moving components c_{it} and r_{it}, and transitory shocks with variances γ_{it}^2 and ϕ_{it}^2. I express all monetary quantities as a fraction of average national income, which grows at rate g_t. As a result, wealth w_{it} evolves

according to the SDE:[15]

$$dw_{it} = \underbrace{\left(y_{it} + (r_{it} - g_t)w_{it} - c_{it}\right)}_{\mu_{it} \text{ (drift)}} dt + \underbrace{\left(v_{it}^2 + \phi_{it}^2 w_{it}^2 + \gamma_{it}^2\right)^{1/2}}_{\sigma_{it} \text{ (diffusion)}} dB_{it} \tag{1}$$

That is, individual wealth is a stochastic process with a drift μ_{it} equal to $y_{it} + (r_{it} - g_t)w_{it} - c_{it}$ and a diffusion σ_{it}^2 equal to $v_{it}^2 + \phi_{it}^2 w_{it}^2 + \gamma_{it}^2$. Note that the economy's growth rate g_t appears in the drift term due to the normalization of all the quantities by the economy's average income.

3.1.2 Evolution of the Distribution of Wealth

We can directly relate the evolution of individual wealth described by equation (1) to the overall distribution of wealth. A standard result of stochastic calculus states that, if a large number of stochastic processes follow the same SDE, then the probability density function (PDF) that describes the distribution of their value at a given time follows a partial differential equation (PDE) known as the Kolmogorov (1931) forward equation (Gabaix, 2009). This result provides a direct way to connect the evolution of individual wealth (as in equation (1)) with the aggregate distribution of wealth (as is observed in historical data). However, I need to account for the possibility that the drift μ_{it} and the diffusion σ_{it}^2 vary across individuals i. To that end, I will apply a result from Gyöngy (1986), which states that it is possible to average out the heterogeneity of individual wealth processes and still retrieve the same wealth distribution in the aggregate. After applying that result, I use the Kolmogorov (1931) forward equation to connect equation (1) to the evolution of the wealth distribution.

Reduction to a Single Equation using Gyöngy's (1986) Theorem In simple terms, Gyöngy's (1986) theorem states the following. Consider a large number of stochastic processes w_{it}, each following a SDE with their own drift μ_{it} and diffusion σ_{it}^2. Then the PDF describing the distribution of the value of these processes will behave exactly as if their drift μ_{it} and their diffusion σ_{it}^2 were replaced by the conditional expectations $\mu_t(w) = \mathbb{E}[\mu_{it}|w]$ and $\sigma_t^2(w) = \mathbb{E}[\sigma_{it}^2|w]$. This result, plus some basic rules of stochastic calculus, makes it possible to reduce the arbitrarily complex nature of individual wealth accumulation into a single SDE that characterizes the entire wealth distribution. I can state the following result.

Proposition 1. *Let $y_t(w)$, $r_t(w)$ and $c_t(w)$ be the average labor income, rate of return and consumption, conditional on wealth w at time t. Similarly, let $v_t^2(w)$, $\phi_t^2(w)$ and $\gamma_t^2(w)$ be the*

---

[15]This formulation assumes that the transitory shocks are uncorrelated. We could account for correlated shocks by including additional covariance terms, as in Bienaymé's identity. To simplify the exposition, I focus on the uncorrelated case.

variance of labor income, rates of return and consumption, conditional on wealth w at time t.
Then, the stochastic process governed by the SDE with deterministic coefficients:

$$dw_{it} = \underbrace{(y_t(w_{it}) + (r_t(w_{it}) - g_t)w_{it} - c_t(w_{it}))}_{\mu_t(w)\ (drift)} dt + \underbrace{(v_t^2(w_{it}) + \phi_t^2(w_{it})w_{it}^2 + \gamma_t^2(w_{it}))^{1/2}}_{\sigma_t(w)\ (diffusion,\ or\ mobility)} dB_{it}$$

has the same marginal distribution as the process described by equation (1).

Proof. See Appendix A.1. □

Proposition 1 reduces the dynamics of the wealth distribution to a single SDE, in which everyone with the same wealth w faces the same drift $\mu_t(w)$ and the same diffusion $\sigma_t^2(w)$. In that equation, the diffusion $\sigma_t^2(w)$ becomes easily interpretable as a mobility parameter. If it were equal to zero, everyone with the same wealth would face the same wealth growth, so there would be no movement across the distribution. But when that parameter is not zero, the approach can account for the mobility across the wealth distribution, which is a sizeable phenomenon (Gomez, 2022). From now on, I will therefore refer to $\sigma_t^2(w)$ as a mobility parameter.

The Kolmogorov Forward Equation   Using the reduced SDE of Proposition 1, I can now apply the Kolmogorov forward equation. The density f_t which describes the distribution of wealth at time t obeys the PDE:

$$\partial_t f_t(w) = -\partial_w[\mu_t(w)f_t(w)] + \frac{1}{2}\partial_w^2[\sigma_t^2(w)f_t(w)] \tag{2}$$

This equation describes the evolution of a quantity that is observable in the historical data (the density of wealth f_t) while connecting it to parameters that characterize wealth accumulation at the individual level: the drift $\mu_t(w)$, and the diffusion $\sigma_t^2(w)$. Thus, it directly connects individual economic behavior with the distribution of wealth.

3.1.3   Interpretation of the Different Effects

The previous section derived equation (2) using results from stochastic calculus. This section provides a more direct and intuitive understanding of the equation. The purpose is twofold. First, it provides an understanding of the central equation that does not require familiarity with stochastic calculus. Second, it gives a more transparent basis for how and why the decomposition introduced in this paper works.

Integration of the Kolmogorov Forward Equation   For empirical purposes, it is useful to rewrite equation (2) in its integrated version, which involves the cumulative distribution function

(CDF) of wealth, $F_t(w)$. After re-arranging terms, we get:

$$\underbrace{-\frac{\partial_t F_t(w)}{f_t(w)}}_{\substack{\text{local change in the}\\\text{wealth distribution}}} = \underbrace{\mu_t(w)}_{\substack{\text{local effect of average}\\\text{change in wealth}}} - \underbrace{\frac{1}{2}\partial_w \sigma_t^2(w)}_{\substack{\text{local effect of the}\\\text{mobility gradient}}} - \underbrace{\frac{1}{2}\sigma_t^2(w)\frac{\partial_w f_t(w)}{f_t(w)}}_{\substack{\text{local effect}\\\text{of mobility}}} \qquad (3)$$

Each of these terms captures a different mechanism. I discuss them in turn below. To fix ideas, let us consider $w = \$1\text{bn}$ and that the population is normalized to one, so that $1 - F_t(w)$ is the number of billionaires, and $-\partial_t F_t(w)$ is the change in the number of billionaires.

Local Effect of Average Change in Wealth   Assume that wealth growth is a deterministic function of wealth: everyone with the same wealth experiences the same wealth change, so there is no mobility ($\sigma_t \equiv 0$). Consider the case where wealth growth is positive at the top, so that the number of billionaires increases.

Over a short period, the number of people crossing the $1bn threshold will be proportional to (i) $f_t(w)$, the number of people that were initially at the threshold, and (ii) $\mu_t(w)$, the pace at which their wealth increases. Therefore, we get $-\partial_t F_t(w) = \mu_t(w) f_t(w)$, which corresponds to equation (3) when $\sigma_t \equiv 0$.

This formula is known as a transport equation. If wealth growth is uniform ($\mu_t \equiv \mu$), then it translates the entire wealth distribution by a factor $\mu\,dt$ over a period dt. The general formulation makes it possible to consider non-uniform wealth growth.[16]

Local Effect of Mobility   Now, consider the opposite thought experiment. Everyone, even if they have the same wealth, experiences a different wealth growth. But they are as likely to go up or down, so, on average, wealth growth is zero ($\mu_t \equiv 0$). Furthermore, assume that the amplitude of wealth variations is uniform ($\sigma_t \equiv \sigma$). Under these conditions, the number of billionaires will still change.

Indeed, some people just below $1bn will see their wealth increase and become billionaires. This flow is proportional to (i) $f_t(w^-)$, the number of people right below $1bn, and (ii) the amplitude σ^2 by which their wealth varies. Conversely, some people just above $1bn will see their wealth decrease and drop out of the list of billionaires. This flow is proportional to (i) $f_t(w^+)$, the number of people right above $1bn and (ii) the amplitude σ^2 by which their wealth varies.

In general, these two flows will not cancel out because there are not as many people right below and right above $1bn. This difference between the population on each side of

---

[16]Note that with the change of variable $p = F_t(w_t)$, we get $\partial_t w_t(p) = \mu_t(p)$. Hence, $\mu_t(p)$ is the growth of the pth quantile.

the threshold is effectively captured by the *derivative* of the density f_t. Mathematically, this derivative appears from writing the difference between the two flows, and then taking the limit as $w^+ \to w$ and $w^- \to w$. After applying the correct proportionality factor, we get $-\partial_t F_t(w) = \frac{1}{2}\sigma_t^2(w)\partial_w f_t(w)$, which corresponds to equation (3) when $\mu_t \equiv 0$ and $\sigma_t \equiv \sigma$. Unlike the equation modeling the effect of average wealth growth, in this equation, the effect on the number of billionaires depends not on the *value* of the density but on its *gradient*. This formula is known as a diffusion equation and is best understood as a transformation that flattens wealth density.

To make sense of this effect, consider two extremes. First, if the wealth density is flat at the top, then the number of people who cross the \$1bn threshold from both sides will cancel out. Hence, even though wealth changes at the individual level, the overall effect on the distribution is nil. Now, assume that, on the contrary, the wealth density is infinitely steep, so that there are no billionaires but many people just below \$1bn. Some people will see their wealth increase and become billionaires. But since there is initially no one above \$1bn, there can be no countervailing flow of people leaving the group. And therefore, the number of billionaires will increase very fast.

Local Effect of the Mobility Gradient If the amplitude of wealth mobility is not uniform, there is a third effect on the wealth distribution. Indeed, assume that there is more variation in wealth growth above \$1bn than below: then, downward mobility will exceed upward mobility. Thus, even with no average growth and a flat density, people who drop out of the billionaire group will outnumber those who enter it. This phenomenon creates an additional effect on $-\partial_t F_t(w)$, which depends on how mobility varies throughout the distribution. Hence, it is a function of the mobility gradient $\partial_w \sigma_t^2(w)$, and is equal to $-\frac{1}{2}f_t(w)\partial_w\sigma_t^2(w)$.

3.1.4 Phase Portrait and The Dynamics of Wealth Inequality

Consider the case where the drift and mobility parameters are constant over time: $\mu_t(w) \equiv \mu(w)$ and $\sigma_t^2(w) \equiv \sigma^2(w)$. Equation (3) is best represented as a curve that relates the current level to the current evolution of inequality, as in Figure 3. This curve — a phase portrait — lets us picture the dynamics of wealth inequality in a simple way.

The term $\partial_w f_t(w)/f_t(w)$ on the right-hand side of equation (3) measures the relative slope of the density. In the upper tail of the wealth distribution, the density is sloping downward, so $\partial_w f_t(w)/f_t(w) < 0$. This value is a good proxy for top wealth inequality. If it is high, then the density decreases at a slow rate: i.e., the tail is fat, and inequality is high. Conversely, if its value is low, then inequality is low as well.[17] We can analogously interpret the term $-\partial_t F_t(w)/f_t(w)$

---

[17]If wealth follows a Pareto distribution with coefficient α, then $\partial_w f_t(w)/f_t(w) = -(1+\alpha)/w$. So, as $\alpha \to 1$,
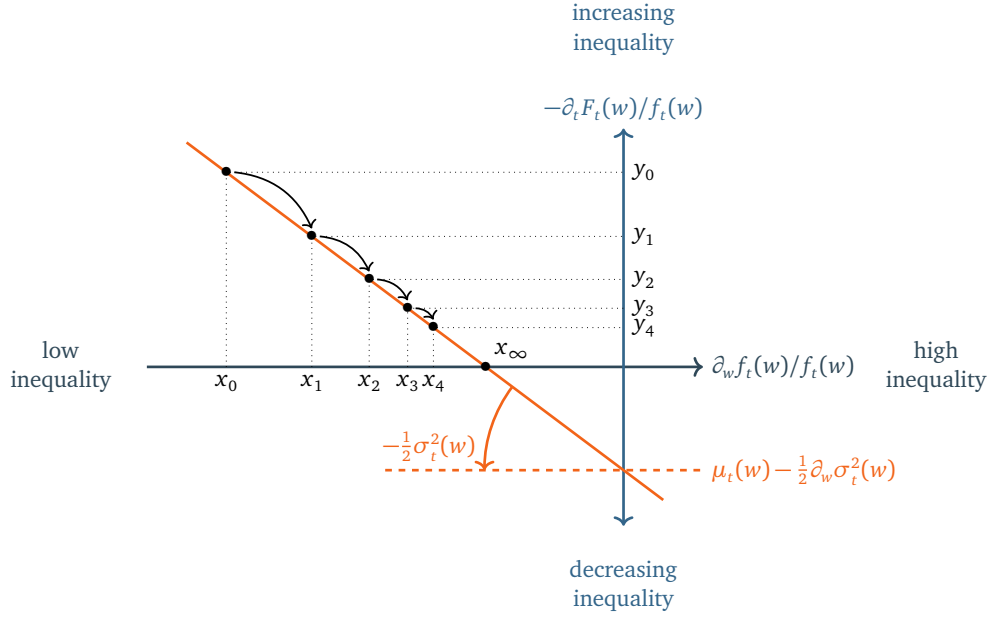
Note: This diagram is a phase portrait of the dynamics of inequality for a given wealth level w located towards the top of the wealth distribution. These dynamics can be pictured in a two-dimensional space where each axis represents a quantity associated to the wealth distribution, whose PDF is $f_t(w)$ and whose CDF is $F_t(w)$. The x-axis is equal to $\partial_w f_t(w)/f_t(w)$ and is a proxy for inequality levels (high values mean high inequality). The y-axis is equal to $-\partial_t F_t(w)/f_t(w)$ and is a proxy for inequality changes (high values mean increasing inequality). The diagram represents the transition from a low inequality level (x_0) to a high inequality level (x_∞), with constant $\mu(w)$ and $\sigma^2(w)$. During this transition, the system moves alongside the orange line with slope $-\frac{1}{2}\sigma^2(w)$ and intercept $\mu(w)-\frac{1}{2}\partial_w\sigma^2(w)$. When the system reaches the x-axis, where $-\partial_t F_t(w)/f_t(w) = 0$, it is a its steady-state.

Figure 3: Phase Portrait: Convergence to a Higher-Inequality Steady State

on the left-hand side of equation (3) as a change in the fatness of the top tail of the distribution. When it exceeds zero, the tail becomes fatter, and inequality increases.[18]

In Figure 3, the first quantity $\partial_w f_t(w)/f_t(w)$ (inequality level, in blue) is on the x-axis and the second quantity $-\partial_t F_t(w)/f_t(w)$ (inequality change, in green) is on the y-axis. Equation (3) states that, with stable parameters, the data points representing the current state of inequality must lie alongside a straight line (in orange), with intercept $\mu_t(w) - \frac{1}{2}\partial_w\sigma_t^2(w)$ and slope $-\frac{1}{2}\sigma_t^2(w)$. Where this line crosses the x-axis, inequality is neither increasing nor decreasing: this is the distribution's steady state.

---

inequality goes to infinity and $\partial_w f_t(w)/f_t(w)$ increases to zero.

[18]If wealth follows a Pareto distribution with coefficient α_t, then $-\partial_t F_t(w)/f_t(w) = -(w\log w)(\partial_t\alpha_t/\alpha_t)$. Therefore, $-\partial_t F_t(w)/f_t(w) > 0$ is associated with a decrease in α_t, which corresponds to an increase in inequality.

Dynamics of Wealth Inequality and Convergence to a Steady State   Consider the transition to a high inequality steady state. On Figure 3, start from the inequality level x_0. This level is below the steady-state level x_∞, so inequality will change. We can decompose this change into (i) the effect of average wealth growth and of the mobility gradient, which the intercept captures, and (ii) the effect of mobility, which the slope captures. As we can see on the graph, (i) decreases inequality, whereas (ii) increases it. At first, the effect of (ii) is stronger, so inequality increases overall by y_0. In the next period, inequality is higher, equal to x_1. A higher inequality means a fatter tail, hence a flatter density. Having a flatter density weakens the effect of (ii). The effect of (i), on the other hand, remains unchanged. Overall, inequality still increases, but by a lower amount, y_1. The process repeats. Inequality keeps increasing, which flattens the density, weakens the effect of (ii) but leaves the effect of (i) unchanged. Asymptotically, we reach x_∞. At this point, the effect of (ii) has been weakened to the point that it perfectly counterbalances (i). We have reached the steady state.

Rationale for the Steady State   This framework provides a robust justification for the emergence of a steady state — one that doesn't preclude but also doesn't require behavioral responses. Without mobility, the only way to get a steady-state distribution of wealth is for behavioral responses (and general equilibrium effects) to lead to a point where saving rates as a proportion of wealth become identical throughout the distribution. Any deviation from this situation leads to a degenerate steady state in the long run. In contrast, here, the distribution eventually stabilizes because of the mechanical interaction between mobility and drift, so the model remains well-behaved for a wider range of economic behaviors.

Distinction Between Drift and Mobility   This framework shows how drift and mobility have different impacts on the distribution and why that distinction matters. In a model without mobility, the red line in Figure 3 would be flat. The only way to account for inequalities of wealth that are neither increasing nor decreasing at a constant rate is to assume that the intercept $\mu_t(w)$ changes with each period (which usually implies some form of behavioral change). In contrast, once we introduce mobility, it becomes possible to account for any linear downward-slopping evolution of inequality in the phase portrait of Figure 3 with a much more parsimonious model that only involves two time-invariant parameters: one for drift and one for mobility.

3.2   Other Processes Affecting the Wealth Distribution

I account for other phenomena impacting wealth distribution besides drift and diffusion (which capture individual income and consumption). The three phenomena that I consider in this

paper are birth and death, inheritance, and assortative mating. I first introduce these effects in equation (2) and then gives a version of equation (3) that include them as well.

Births and Deaths At any time t, a fraction δ_t of people die. Let g_t be the density of their wealth. Simultaneously, a fraction β_t of people appears with a random initial endowment drawn from a distribution with density h. The total population grows at a rate $n_t = \beta_t - \delta_t$. This process impacts $\partial_t f_t(w)$, i.e., the change in the wealth density at wealth w and time t, by:

$$\underbrace{\zeta_t(w)}_{\text{demography}} \equiv \underbrace{\beta_t h(w)}_{\text{injection}} - \underbrace{\delta_t g_t(w)}_{\text{deaths}} - \underbrace{n_t f_t(w)}_{\text{normalization}}$$

and the equation (2) becomes:

$$\partial_t f_t(w) = \underbrace{-\partial_w[\mu_t(w)f_t(w)] + \frac{1}{2}\partial_w^2[\sigma_t^2(w)f_t(w)]}_{\text{income and consumption}} + \underbrace{\zeta_t(w)}_{\text{demography}}$$

Inheritance I model inheritance as a jump process. With a probability $\pi_t(w)$, people see their wealth jump from w to $w + \lambda$ where λ is the amount of inheritance received, net of taxes. Let $s_t(\lambda|w)$ be the density of the value of the inheritance, conditional on the value of wealth, and conditional on receiving an inheritance. We can model the jump process as a death with rate $\pi_t(w)$ and as an injection with rate $\int \pi_t(w - \lambda)f_t(w - \lambda)s_t(\lambda|w - \lambda)\, d\lambda$. So, the effect of inheritance on $\partial_t f_t(w)$, i.e., the change in the wealth density at wealth w and time t, is:

$$\xi(w) \equiv \int \pi_t(w - \lambda)f_t(w - \lambda)s_t(\lambda|w - \lambda)\, d\lambda - \pi_t(w)f_t(w)$$

and the equation (2) with both demography and inheritance is:

$$\partial_t f_t(w) = \underbrace{-\partial_w[\mu_t(w)f_t(w)] + \frac{1}{2}\partial_w^2[\sigma_t^2(w)f_t(w)]}_{\text{income and consumption}} + \underbrace{\zeta_t(w)}_{\text{demography}} + \underbrace{\xi_t(w)}_{\text{inheritance}}$$

Marriages, Divorces and Assortative Mating Finally, I account for the effect of marriages and divorces. I adopt the convention that wealth is split equally among spouses. At any time t, a fraction θ_t of people get married. Let $p_t(w)$ be the density of their wealth as individuals, and $q_t(w)$ be the density of their wealth as couples. The strength of assortative mating is captured by the relationship between p_t and q_t: at the limit, if people always choose a spouse with identical wealth, then $p_t(w) = 2q_t(2w)$. The effect of marriages on the wealth distribution is therefore $2\theta_t q_t(2w) - \theta_t p_t(w)$. We can construct an opposite process for divorces. Let $\chi_t(w)$ be

combined effect of marriages and divorces on $\partial_t f_t(w)$. Then the equation (2) with the effect of demography, inheritance, marriages, and divorces becomes:

$$\partial_t f_t(w) = \underbrace{-\partial_w[\mu_t(w)f_t(w)] + \frac{1}{2}\partial_w^2[\sigma_t^2(w)f_t(w)]}_{\text{income and consumption}} + \underbrace{\zeta_t(w)}_{\text{demography}} + \underbrace{\xi_t(w)}_{\text{inheritance}} + \underbrace{\chi_t(w)}_{\substack{\text{marriages} \\ \text{and divorces}}} \quad (4)$$

Integrated Version   Rewrite equation (4) in an integrated form, similar to (3). Define $F_t(w) = \int_{-\infty}^{w} f_t(s)\,ds$, $Z_t(w) = \int_{-\infty}^{w} \zeta_t(s)\,ds$, $\Xi_t(w) = \int_{-\infty}^{w} \xi_t(s)\,ds$ and $X_t(w) = \int_{-\infty}^{w} \chi_t(s)\,ds$. After integrating equation (4) and re-arranging terms, we get:

$$-\frac{\partial_t F_t(w)}{f_t(w)} + \frac{Z_t(w)}{f_t(w)} + \frac{\Xi_t(w)}{f_t(w)} + \frac{X_t(w)}{f_t(w)} = \mu_t(w) - \frac{1}{2}\partial_w \sigma_t^2(w) - \frac{1}{2}\sigma_t^2(w)\frac{\partial_w f_t(w)}{f_t(w)} \quad (5)$$

This equation is similar to (3), with additional correction terms on the left-hand side to account for demography, inheritance, and assortative mating. The fundamental dynamics of wealth inequality remain similar to those pictured in the phase portrait in Figure 3, except that the y-axis needs to be adapted to include the correction terms. This equation will serve as the basis for the decomposition introduced in this paper.

3.3   Decomposition of the Different Effects

I will use equation (5) to decompose the various factors affecting the distribution of wealth. All the terms in the equation can be directly observed or separately estimated, except those related to consumption. Therefore, these unobserved parameters have to be estimated as a type of "residual." This section explains how.

Estimating Equation for the Complete Model   Let us go back to the equation (5). Separate the drift $\mu_t(w)$ and the diffusion $\sigma_t^2(w)$ into their observed component (income) and their unobserved component (consumption):

$$\underbrace{\mu_t(w)}_{\substack{\text{total drift} \\ \text{(average} \\ \text{wealth change)}}} = \underbrace{\tilde{\mu}_t(w)}_{\substack{\text{mean income} \\ \text{(observed)}}} - \underbrace{c_t(w)}_{\substack{\text{mean} \\ \text{consumption} \\ \text{(unobserved)}}} \qquad \underbrace{\sigma_t^2(w)}_{\substack{\text{total mobility} \\ \text{(variance of} \\ \text{wealth change)}}} = \underbrace{\tilde{\sigma}_t^2(w)}_{\substack{\text{variance of} \\ \text{income} \\ \text{(observed)}}} + \underbrace{\gamma_t^2(w)}_{\substack{\text{variance of} \\ \text{consumption} \\ \text{(unobserved)}}}$$

Move the observed components of drift and mobility to the left-hand side of equation (5). The complete equation for the left-hand side becomes:

$$\text{LHS}_t(w) \equiv \underbrace{-\frac{\partial_t F_t(w)}{f_t(w)}}_{\substack{\text{inequality} \\ \text{change}}} + \underbrace{\frac{Z_t(w)}{f_t(w)} + \frac{\Xi_t(w)}{f_t(w)} + \frac{X_t(w)}{f_t(w)}}_{\substack{\text{demography, inheritance,} \\ \text{marriages and divorces}}} - \underbrace{\tilde{\mu}_t(w) + \frac{1}{2}\partial_w \tilde{\sigma}_t^2(w) + \frac{1}{2}\tilde{\sigma}_t^2(w)\frac{\partial_w f_t(w)}{f_t(w)}}_{\text{drift and mobility induced by income}}$$

All the components of $\text{LHS}_t(w)$ are either directly observable or separately estimable. We can therefore rewrite (5) in its final form:

$$\text{LHS}_t(w) = \underbrace{c_t(w) - \frac{1}{2}\partial_w \gamma_t^2(w)}_{\text{intercept}} - \underbrace{\frac{1}{2}\gamma_t^2(w)}_{\text{slope}} \times \frac{\partial_w f_t(w)}{f_t(w)} \tag{6}$$

Identification   Equation (6) provides the basis for estimating the parameters. It shows a relationship similar to that shown in Figure 3. We require three assumptions to get point estimates of the mean $\{c_t(w)\}_{w\in\mathbb{R}}$ and the variance $\{\gamma_t^2(w)\}_{w\in\mathbb{R}}$ of consumption conditional on wealth over a period $[t_0, t_1]$.

Assumption 1. *For all w, we can observe (or separately estimate)* $\text{LHS}_t(w)$ *and* $\partial_w f_t(w)/f_t(w)$ *at k distinct times $t_k \in [t_0, t_1]$ with $k \geq 2$.*

Assumption 2. *For all w, the parameters $c_t(w)$ and $\gamma_t^2(w)$ are constant over time, i.e., $c_t(w) \equiv c(w)$ and $\gamma_t^2(w) \equiv \gamma^2(w)$ for $t \in [t_0, t_1]$.*

Assumption 3. *The wealth distribution is changing over time, in the sense that, for all w, we observe the distribution for at least to periods $r, s \in [t_0, t_1]$ such that $\partial_w f_r(w)/f_r(w) \neq \partial_w f_s(w)/f_s(w)$.*

These three assumptions ensure we can meaningfully fit a line through the different points as in Figure 3. Assumption 1 states that we need to observe the wealth distribution and its evolution during two different periods, a consequence of the fact that we need at least two points to be able to fit a line. In practice, and in the presence of statistical noise, more than two points are preferable to get robust estimates. Assumption 2 ensures that the parameters that govern the relationship (6) remain stable over the period of time under consideration. Finally, Assumption 3 states that the distribution of wealth must change over the period where we observe it. This assumption comes from the fact that we are using local variations in the flatness of the density to disentangle the drift from mobility. This strategy only works if the flatness does, in fact, vary. If these three assumptions are satisfied, then we can proceed with the estimation in two steps:

Step 1 Estimate $c_t^*(w) = c_t(w) - \frac{1}{2}\partial_w \gamma_t^2(w)$ and $\gamma_t^2(w)$ for every w using a line-fitting method.

Step 2 Using the estimate for γ_t^2, estimate the mobility gradient $\partial_w \gamma_t^2$, and use it to get the estimate of $c_t(w) = c_t^*(w) + \frac{1}{2}\partial_w \gamma_t^2(w)$.

Ex-ante Estimate of the Effect's Magnitude   I use variations in the flatness of the density over time to disentangle the effect of drift from that of mobility. But are these variations large enough to provide reliable empirical estimates? Some rough but simple calculations can give a general idea of the magnitude of the effects at play. Between its high point in the 1970s and today, the Pareto coefficient of wealth in the upper tail went from $\alpha \approx 2$ to $\alpha \approx 1.5$. Assume a mobility parameter $\sigma^2 \approx 0.16w^2$ at the top, which matches the estimates in this paper (and separate estimates from the SCF and the PSID). The effect of mobility under these conditions is equal to $\frac{1}{2}\sigma^2(1+\alpha)w$. A useful way to interpret this number is to say that mobility has the same effect on the distribution as an average wealth growth of $\frac{1}{2}\sigma^2(1+\alpha)$, which went from $\frac{1}{2} \times 0.16 \times (1+2) = 24\%$ to $\frac{1}{2} \times 0.16 \times (1+1.5) = 20\%$ of wealth — a 4% change. The change in the effect of mobility that is attributable to the flattening of the wealth density at the top is, therefore, sizable — equivalent to the mechanical effect that a permanent 4% wealth tax would have.

Potential Limitations   The main limitation of the method is that it requires the parameters for the drift $c(w)$ and the mobility $\gamma^2(w)$ induced by consumption to be constant (or at least not to have a downward or upward trend) over sufficiently long periods. Estimates may be biased if this is not the case. For example, assume that the drift decreases over time. On the phase portrait in Figure 3, the linear relationship (in orange) moves down over time. As a result, the observed data points on the phase portrait will lie on a curve that could take many different shapes, but, in general, will appear more downward-sloping than the true relationship. If we were to estimate the decomposition in that context, we would therefore overestimate the diffusion parameter.

In practice, there are two ways to address this issue. The first one is to check that the empirical phase portrait to which we fit the line remains relatively close to linearity. While this does not guarantee that the assumption of a constant $c(w)$ and $\gamma^2(w)$ is verified, it can identify some problematic cases. The second one is to verify that the parameters estimated from the decomposition are consistent with external estimates of savings and mobility. I perform both checks in this paper's empirical section, suggesting that the simple model with a constant $c(w)$ and $\gamma^2(w)$ since the 1980s works well.

4 Estimation for the United States

4.1 Data

I estimate the parameters in decomposition (6) for the United States. I start in 1962, when sufficiently detailed microdata on the distribution of income and wealth in the United States becomes available. I primarily rely on the tax-based microdata from Saez and Zucman (2020a), which I complement in a number of ways. Using the data collected, I perform microsimulations of mortality, birth, inheritance, marriage, and divorce, which I use to estimate the impact of the corresponding phenomenons on the wealth distribution. I briefly review the data and methodology below and provide more details in Appendix B.

Income and Wealth For income and wealth, I rely on the Distributional National Accounts (DINA) tax-based microdata from Saez and Zucman (2020a). These data distribute the entirety of national income and wealth, as measured by the national accounts, every year since 1962, to adult individuals (20 and older). They infer the distribution of wealth from capital income flows, following the capitalization method of Saez and Zucman (2016).[19] This data does not distribute capital gains since they are not part of national income. I incorporate them in the data by assuming a constant capital gains rate by asset class, as in Robbins (2018). The information on age in the DINA microdata is also limited, so I replace it with information from the Survey of Consumer Finances (SCF). I match DINA and SCF observations one-to-one based on their rank in the wealth distribution. I use this to estimate a rank in the age distribution by sex for each observation and attribute to them the age that matches this rank according to official demographic data.

Demography and Intergenerational Linkages I estimate the entire demographic history of the United States since 1850 by year, age, and sex, including population counts, mortality rates, female and male fertility rates by birth order. I start the estimation in 1850, long before the income and wealth data starts (in 1962), because I use the demographic data to simulate intergenerational linkages between parents and children. So if a centenarian dies in 1962, I must be able to retrace that person's entire fertility history since their birth and retrace the mortality history of that person's children.

I construct this data by collecting and harmonizing data from official sources (United States Census Bureau), historical databases (Human Mortality Database, Human Life Table Database, Human Fertility Database, Human Fertility Collection) and academic publications (Haines,

---

[19]Their latest revision (Saez and Zucman, 2020a) accounts for heterogeneous returns.

1998). To make projections for the future, I rely on the forecast (medium variant) of the World Population Prospects (United Nations, 2019). For male fertility rates, which are not a standard demographic parameter, I combine the female fertility rates with the joint age distribution of mixed-sex couples in the IPUMS census microdata (Ruggles et al., 2022).

First, I use this data to simulate deaths, assuming that people die at random according to their age and sex-specific mortality rate. I also account for births by assuming that people enter the population at age 20 with a constant exogenous wealth distribution estimated from the data.[20,21] Second, I simulate intergenerational linkages. For every person in the data, I assume this person had children according to their sex, age, year, and birth-order-specific fertility rates. Then I assume these children experience mortality in line with their sex, year, and age-specific mortality rates. This methodology generates a distribution for the age and sex of the direct descendants of every person in the sample, which I use to distribute inheritances.

Inheritance and Estate Taxation   When someone dies, I assume that their wealth is transmitted to their spouse (if any), without estate tax, or to their children, after payment of the estate tax. To calculate the estate tax, I collect complete statutory schedules of the federal estate tax over the second half of the 20th century. I assume that wealth is split equally among descendants, as is the norm in the United States (Menchik, 1980). I account for the possibility that wealthier people are more likely to inherit and receive larger inheritances. I use the SCF to estimate the relative probability of receiving an inheritance, as well as the rank in the inheritance distribution, as a function of the rank in the wealth distribution, conditional on age. Within a given age group, I simulate international wealth transmission according to these parameters.

Marriages, Divorces and Assortative Mating   I collect data on the aggregate rate of divorce and marriage from the National Vital Statistics System (NVSS) (Center for Disease Control, 2022). In addition, I reconstruct age and sex-specific rates using population data disaggregated by marital status from IPUMS census microdata (Ruggles et al., 2022).

I determine the extent of assortative mating by estimating the joint distribution of the ranks in the wealth distribution at the time of marriage using the Survey of Income and Program Participation (SIPP) panel (2013–2016). I also consider the impact of assortative mating on divorce by estimating the distribution of the share of wealth owned by each couple member before they get divorced using the same data.

---

[20]People aged 20 have very low levels of wealth, so in practice, this is close to assuming that people start with zero wealth.

[21]The birth rate is estimated here as a residual between population growth and the crude death rate, so in effect, it also incorporates the effect of immigration.

Then, I simulate the process of marriage and divorce by randomly selecting people to get wedded in a given year according to age and sex-specific marriage rates, and then marry them to one another to reproduce the joint distribution of wealth ranks at marriage in the SIPP. Similarly, I simulate the effect of divorce by randomly selecting people to get separated according to their age and sex-specific divorce rates, and split the couple's wealth among both spouses according to the SIPP data.

4.2 Estimation

Wealth Distribution   First, I normalize the distribution of wealth by the average national income per adult. Then, I transform it using the inverse hyperbolic sine function (asinh).[22] This transformation makes the distribution of wealth easier to manipulate empirically. And because we operate in a continuous time framework, it creates no practical difficulty: Itô's lemma establishes a direct correspondence between the parameters of the process for w_{it} and for asinh(w_{it}).[23]

I select a range of values (from -1 to 2000 times average income) which the wealth microdata consistently covers over the entire period.[24] I divide this range into bins of equal size, each representing 0.1 units of asinh(wealth).[25] Figure 4a plots the density of wealth, as estimated by the frequency of these bins. I display two years: 1978, which has the lowest inequality, and 2019, which has the highest inequality and is also the most recent data available. I use the logarithm of the density, so changes in the top tail of the wealth distribution are more clearly visible. It appears linear in the top tail, which follows from the fact large wealth holdings follow a power law.[26] Note that $\partial_w \log f_t(w) = \partial_w f_t(w)/w$, i.e., the derivative of the logarithm of the density is equal to the quantity of interest on the right-hand side of equation (6). Hence, our interest lies in the slopes of the lines, which I estimate by running locally weighted linear regressions through the values of Figure 4a.[27] The top tail has driven most of the changes in the wealth distribution, and indeed this is where we observe most of the variation. In 1978, when inequality was at its lowest, the density at the top was quite steep, with a slope around -2. By 2019, inequality had increased dramatically, leading to a fatter tail and a flatter density, with a slope around -1.5.

---

[22]I.e., asinh $: x \mapsto \log(x + \sqrt{x^2 + 1})$.

[23]Itô's lemma states that, if a process x_t follows the SDE $dx_t = \mu_t\, dt + \sigma_t\, dB_t$, then $\phi_t(x_t)$ follows the SDE $d\phi_t(x_t) = \left[\partial_t \phi_t(x_t) + \mu_t \partial_x \phi_t(x_t) + \frac{1}{2}\sigma_t^2 \partial^2 \phi_t(x,t)\right] dt + \sigma_t \partial_x \phi_t(x)\, dB_t$. See Appendix Section C.3 for details.

[24]The range goes from -0.9 to 8.3 on the asinh scale.

[25]This represents 91 bins.

[26]At the top, the wealth distribution is approximately Pareto, and the inverse hyperbolic sine is approximately logarithmic, so the distribution of transformed wealth in approximately exponential.

[27]In the benchmark specification, I use a rectangular kernel and a bandwidth of 1.5. See Appendix C.4 for robustness checks.
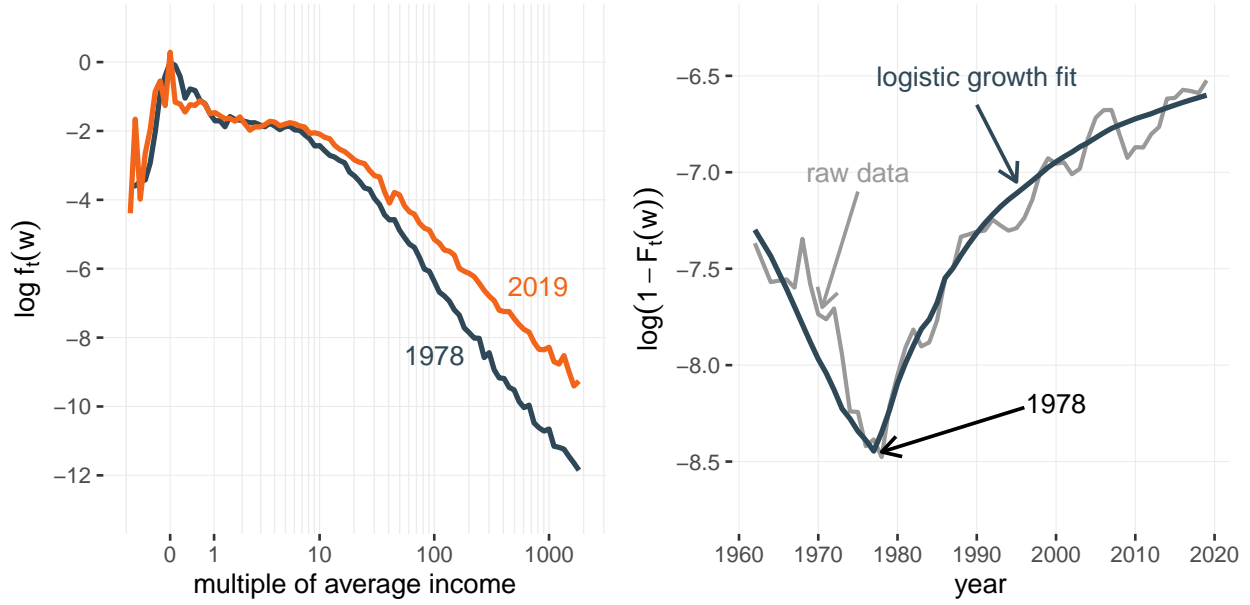
(a) Log Density in 1978 and 2019   (b) Log CCDF at 500 Times Average Income

Source: Own computation using the Distributional National Accounts (DINA) microdata from Saez and Zucman (2020a). *Note:* Wealth is always expressed as a multiple of national income. Densities in Figure 4a are estimated as histograms with 91 bins of size 0.1 on the asinh(wealth) scale, ranging from −1 to 2000 times average income (−0.9 to 8.3 on the asinh scale).

Figure 4: Distribution of Wealth and Its Evolution

Changes in the Distribution of Wealth Over Time For each wealth bin, I estimate $(1 - F_t(w))/f_t(w)$, which is the relevant measure for the evolution of the wealth distribution as it appears on the left-hand side of equation (6). Let us begin with $\log(1 - F_t(w))$, the logarithm of the numerator (see Figure 4b for the bin corresponding to 500 times the average income). The raw data shows two clear trends, one on each side of the year 1978, which correspond to the decreasing part and the increasing part of the U-shaped evolution of wealth inequality. I use a parametric approximation to filter out the short-run variations around these trends (which are not my focus here). Using nonlinear least squares, I fit a logistic growth model for each bin, separately on each side of the year 1978.[28,29] From this parametric approximation, I estimate the time derivative of $\log(1 - F_t(w))$. Finally, I retrieve the quantity of interest using the fact that $\partial_t F_t(w) = (1 - F_t(w))\partial_t \log(1 - F_t(w))$.

---

[28]The logistic curve can be written as $L : t \mapsto L(t) = \frac{x_\infty}{1 + (x_\infty/x_0 - 1)\exp(-\rho t)}$ where (x_0, x_∞, ρ) are the three parameters which capture, respectively, the initial value, the asymptotic value, and the rate of convergence.

[29]This model is attractive for three reasons: (i) empirically, it fits the data well, (ii) it captures the features that we expect, i.e., that of a process that grows at first, and then settles to a steady state, and (iii) if we locally approximate the distribution with an exponential distribution with rate parameter λ_t, then equation (3) collapses to a logistic differential equation for λ_t, whose solution is the logistic curve function. Note that this approximation can only be valid locally and does not provide a global solution to the partial differential equation (3).

Other Processes   I separately estimate the effects of income, demography, inheritance, marriage, and divorce, using the microdata and, when needed, microsimulations based on the microdata. For income, I directly calculate the mean and the variance within each wealth bin and use this to estimate the drift and the mobility induced by income. For demography, inheritance, marriage, and divorce, I simulate the processes in the microdata and use the difference between the CDFs before and after simulation to estimate the effects.

Estimation of Drift and Mobility   Having estimated all the observable components of equation (6), we can plot the empirical counterpart to the phase portrait in Figure 3, for every wealth bin. The result, for each year and for the bin corresponding to a wealth level of 500 times the average income, is shown in Figure 5a. Due to the inverse hyperbolic sine transform of wealth and to the inclusion of effects besides drift and mobility, the interpretation of the parameters is not as straightforward as in Figure 3. But the fundamental linear relationship should hold.

Two facts stand out. First, there has been a structural break between 1962–1977 and 1978–2019. Indeed, it is impossible to account for wealth's evolution during both periods by assuming the same linear relationship on the phase portrait: the underlying accumulation process (i.e., the parameters of the propensity to consume) must have changed. Second, within each of these periods, the relationship between the left-hand side and the right-hand side of equation (6) is indeed linear. Therefore, a parsimonious model, with a constant mean and variance of consumption by wealth, can account for the trajectory of wealth since 1978, and, separately, for the trajectory between 1962 and 1977. If we focus, for example, on the post-1978 period, we can see the dynamics described in Figure 3 at play. We start in 1978, with a low but rapidly increasing inequality level. But as inequality goes up, the pace at which it increases slows down progressively.

Note that, while there is unmistakable evidence that the intercept of the linear relationship (which captures drift) has changed between periods, there is no clear sign that the slope (which captures mobility) is different.[30] In light of this, and to get more robust estimates, I assume the same mobility parameter over both periods and only let the drift vary. I apply the same model within all wealth bins: for each of them, I fit two linear relationships with the same slope. I use Deming (1943) regressions to account the presence of error terms on both sides of equation (6). Appendix C.1 provides details of the procedure, alongside robustness checks. I extract the coefficients from these regressions and transform them so that they can be interpreted in terms of the mean and variance of consumption (see Step 2 in Section 3.3, as well as Appendix C.3 for additional adjustments to account for the inverse hyperbolic sine transform of wealth). This finally allows me to plot Figure 5b, the profile of the mean and the variance of consumption

---

[30]This is partly the result of a smaller sample size over 1962–1977.

(a) Empirical Phase Portrait

(b) Consumption by Wealth
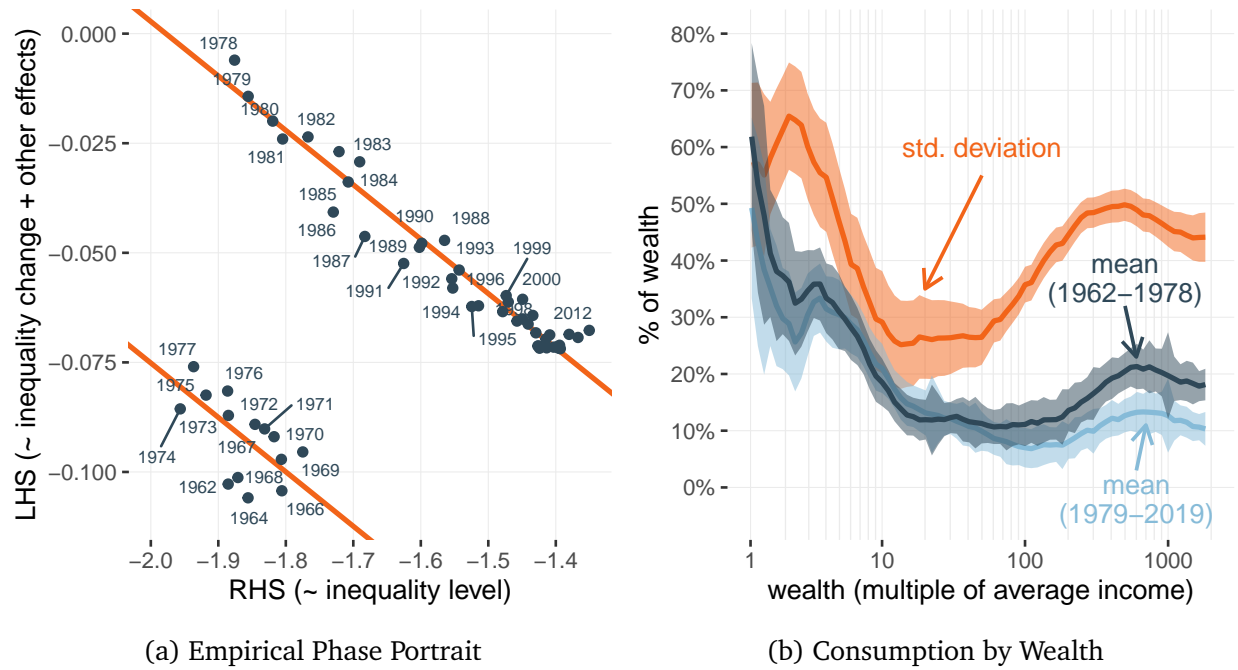
Source: Author's estimations. *Note:* This scatter plot shows the empirical counterpart to the phase portrait described by equation (6), for a level of wealth that correspond to 500 times the average national income (around $40m in 2022). The two linear lines, fitted for 1962–1977 and 1978–2019, correspond to the relationship defined by equation (6). See main text and Appendix C for details on the estimation procedure.

Source: Author's estimations. *Note:* This graph shows the average and the standard deviation of consumption by wealth, as estimated from the slope and the intercept of the linear relationships in the left panel, for every wealth bin. Parameters have been adjusted to account for the data transformations, as explained in Appendix C.3. Areas around the lines indicate 95% confidence intervals, estimated using a bootstrap procedure described in Appendix C.2.

Figure 5: Estimation of Consumption by Wealth

by wealth. This figure also displays 95% confidence intervals, calculated using a bootstrap procedure, which accounts for the presence of error terms on both sides of equation (6), as well as autocorrelations across years and wealth bins, described in Appendix C.2.

Several findings emerge from Figure 5b. First, the variance of consumption is large, which implies a significant role for mobility in the wealth distribution. Second, on average, people consume a significant fraction of their wealth, even at the top, and even in periods of increasing wealth inequality. This matters, in particular, for our understanding of the wealth distribution in the steady state. Significant consumption levels at the top — in general exceeding income — create a tendency for large wealth holdings to reverse toward the mean. In the long run, the reversion towards the mean counterbalances mobility's effect, making it possible for a

steady-state distribution to emerge.[31] If consumption at the top were too low, then wealth at the top would grow without bounds, and so would inequality. Note that at no point did I restrict the parameter values to force the existence of a steady state: a nondegenerate steady state arises naturally from the data in a model with constant drift and constant mobility. Finally, we see that changes in the average consumption between 1962–1978 and 1979–2019 are most significant at the top of the distribution (i.e., wealth above 50 times the average income), which aligns with the view that top wealth holders have been the primary drivers of rising wealth inequality.
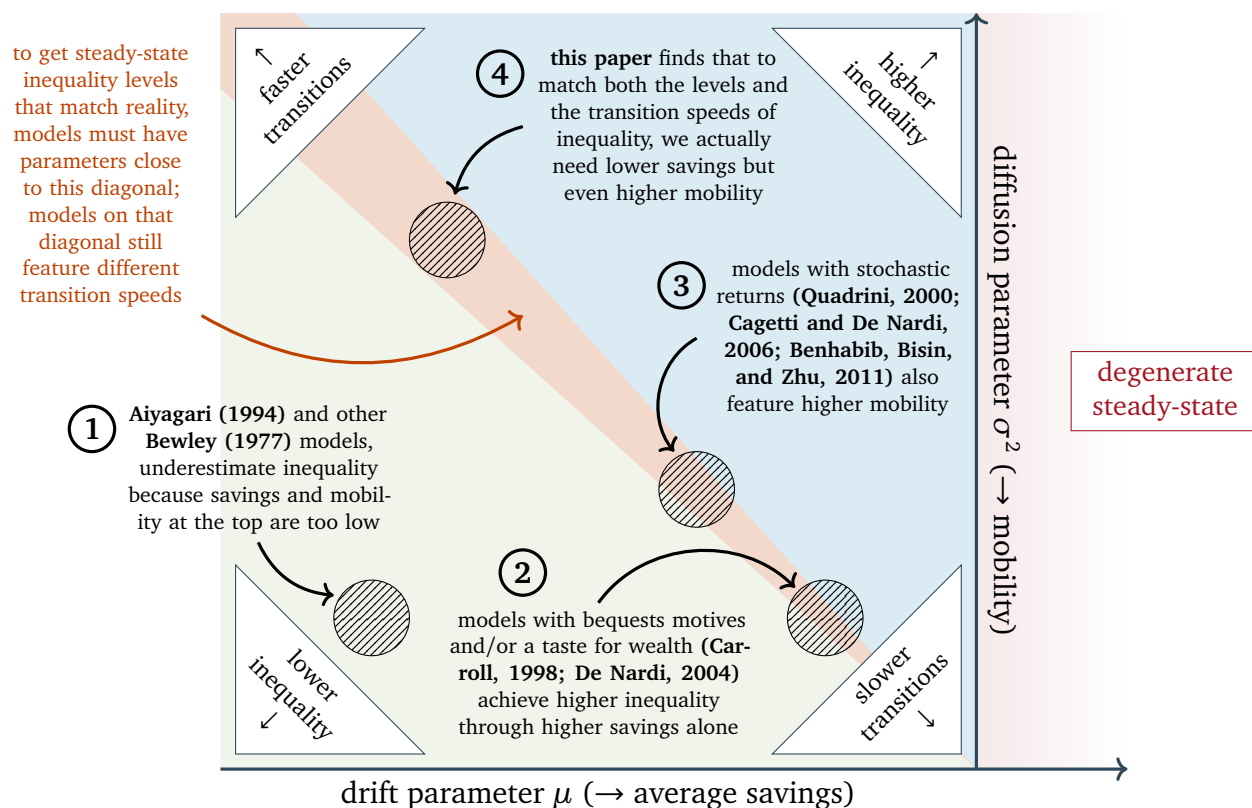

to get steady-state inequality levels that match reality, models must have parameters close to this diagonal; models on that diagonal still feature different transition speeds

↖ faster transitions

④ **this paper** finds that to match both the levels and the transition speeds of inequality, we actually need lower savings but even higher mobility

↑ higher inequality

③ models with stochastic returns **(Quadrini, 2000; Cagetti and De Nardi, 2006; Benhabib, Bisin, and Zhu, 2011)** also feature higher mobility

① **Aiyagari (1994)** and other **Bewley (1977)** models, underestimate inequality because savings and mobility at the top are too low

lower inequality ↙

② models with bequests motives and/or a taste for wealth **(Carroll, 1998; De Nardi, 2004)** achieve higher inequality through higher savings alone

slower transitions ↓

diffusion parameter σ^2 (→ mobility)

degenerate steady-state

drift parameter μ (→ average savings)

Figure 6: Estimated Parameters and Their Relationship to the Literature

Relation to the Rest of the Literature How do these estimates relate to the rest of the literature, and what can we learn from them? We can simplify the situation by focusing solely on the top of the distribution and on the two key parameters (the drift and the mobility) while ignoring the other, less important phenomenons (mobility gradient, demography, etc.) Figure 6 summarizes the situation. The various models of the literature can be schematically represented

---

[31]The presence of demographic effects also contributes to the existence of a nondegenerate steady-state.

on a two-dimensional plane, where the x-axis corresponds to the amount of drift (μ), and the y-axis corresponds to the amount of mobility (σ^2). In this representation, all models with a nondegenerate steady-state lie on the top left quadrant, pictured in Figure 6. The bottom half is not meaningful because it implies negative mobility; the top right quadrant implies an infinite steady-state inequality because there is no reversion towards the mean at the top.[32]

What wealth distribution is implied by the different points? At the steady-state, the derivative of the wealth distribution with respect to time is zero, and therefore equation (3) becomes:

$$0 = \mu(w) - \frac{1}{2}\sigma^2(w)\frac{\partial_w f(w)}{f(w)} \tag{7}$$

This equation characterizes a set of straight diagonal lines for every distribution of wealth, passing through the origin of the plane. Each of them is an "isoinequality" line, defining the set of parameter values that lead to the same steady-state distribution of wealth. These isoinequality lines indicate that models can attain any long-run inequality level, either using a high-savings/low mobility regime (points close to the origin) or using a low-savings/high mobility regime (points far away from the origin).

The set of lines that roughly match the inequality levels typically seen in the United States is colored in orange in Figure 6. Combinations of parameters above this line correspond to higher inequality; combinations that lie below, to lower inequality. Importantly, models on the same isoinequality line still differ when it comes to dynamics. High-savings/low-mobility regimes feature slow transitions between steady-states, while the opposite holds for low-savings/high-mobility regimes.[33]

We can now study where the different models in the literature stand and compare them to this paper's estimate. Start the from Aiyagari (1994) and similar Bewley (1977) models (item 1, Figure 6). These models notoriously underestimate inequality for two reasons. First, people in these models accumulate wealth only for precautionary or consumption smoothing motives, so they have no reasons to accumulate the type of large wealth holdings we observe in practice. Second, because everyone earns the same rate of return, mobility at the top is only the result of labor income shocks. Since labor income is small compared to wealth at the top of the distribution, there is limited mobility as well. These facts put these models squarely in the

---

[32]The drift term μ is normalized by the economy's growth rate, so it is still possible to have a nondegenerate steady-state if people at the top experience positive wealth growth on average as long as that growth remains below the economy's growth rate. Demography and the mobility gradient are other phenomenons that make it possible to sustain a steady state with positive drift at the top. In any case, it remains true that the emergence of a steady-state requires limited wealth growth at the top.

[33]Figure 3 can demonstrate this. Increasing the slope of the line while keeping the same point of intersection with the x-axis leads to the same steady-state but with higher derivatives of the distribution with respect to time (on the y-axis), and therefore faster transitions.

bottom left corner of Figure 6. To fix this problem, a second set of models (item 2, Figure 6) introduced additional saving motives (Carroll, 1998; De Nardi, 2004) such as a taste for wealth, or for bequests. These models manage to match observed inequality levels by increasing savings at the top but do not fundamentally change the extent of wealth mobility. This puts them within the area of realistic steady-state inequality levels (orange diagonal) by moving them to the right of Aiyagari (1994) models in the bottom right corner. A third set of models (item 3, Figure 6) also introduces idiosyncratic stochastic returns (Quadrini, 2000; Cagetti and De Nardi, 2006; Benhabib, Bisin, and Zhu, 2011). This increases mobility at the top because people with the same initial wealth may now move up or down the distribution depending on whether they get high or low returns. That being said, mobility remains quite limited because it is only the result of heterogeneous labor and capital income. Conditional on wealth, however, consumption remains essentially homogeneous because of consumption smoothing. This limited amount of mobility implies slow dynamics, as was identified by Gabaix et al. (2016) for income inequality. This paper (item 4, Figure 6) finds that to match the dynamics of inequality that we observe, we need even higher mobility (and consequently lower savings).[34]

We can also the synthetic savings method (Saez and Zucman, 2016; Kuhn, Schularick, and Steins, 2020; Garbinti, Goupille-Lebret, and Piketty, 2021) to this paper. In equation (3), define the synthetic saving as $\tilde{\mu}_t(w) \equiv \mu_t(w) - \frac{1}{2}\sigma_t^2(w)\partial_w f_t(w)/f_t(w)$, and then apply the change of variable $w = Q_t(p)$, where $0 < p < 1$ is a fractile and $Q_t = F_t^{-1}$ is the quantile function. We get:

$$\partial_t Q_t(p) = \tilde{\mu}_t(Q_t(p))$$

which indeed corresponds to the traditional definition of synthetic savings. Note, however, that the definition of $\tilde{\mu}_t(w)$ depends on the distribution of wealth, so a more accurate formula would be $\partial_t Q_t(p) = \tilde{\mu}_t(Q_t(p), \partial_p Q_t(p))$. Synthetic saving rates methods can either choose to ignore the dependency on $\partial_p Q_t(p)$, or explicitly eliminate it by setting $\sigma_t(w) = 0$.

4.3 Validation

We can assess the validity and consistency of the model in two different ways. First, we can look at its *internal consistency*. (If we simulate the evolution of the wealth distribution using the estimated parameters, do we reproduce the observed data?) Second, we can look its *external consistency*. (Are the estimated parameters consistent with external observations?) In this section, I address both questions.

---

[34]This is a parsimonious alternative to the solutions suggested by Gabaix et al. (2016), which involve the introduction of additional short-run dynamics at the beginning of transition periods.

Replication of Observed Wealth Inequality Dynamics Starting from the distribution of wealth in 1962, and assuming that all the factors which affect the wealth distribution remain at their observed value, I can use the mean and variance of the propensity to consume estimated from the model to simulate the evolution of the wealth distribution. An elementary requirement for the general validity of the approach is that the evolution of the simulated wealth distribution matches the one observed in reality.

(a) Top 1% Share   (b) Growth Incidence Curves
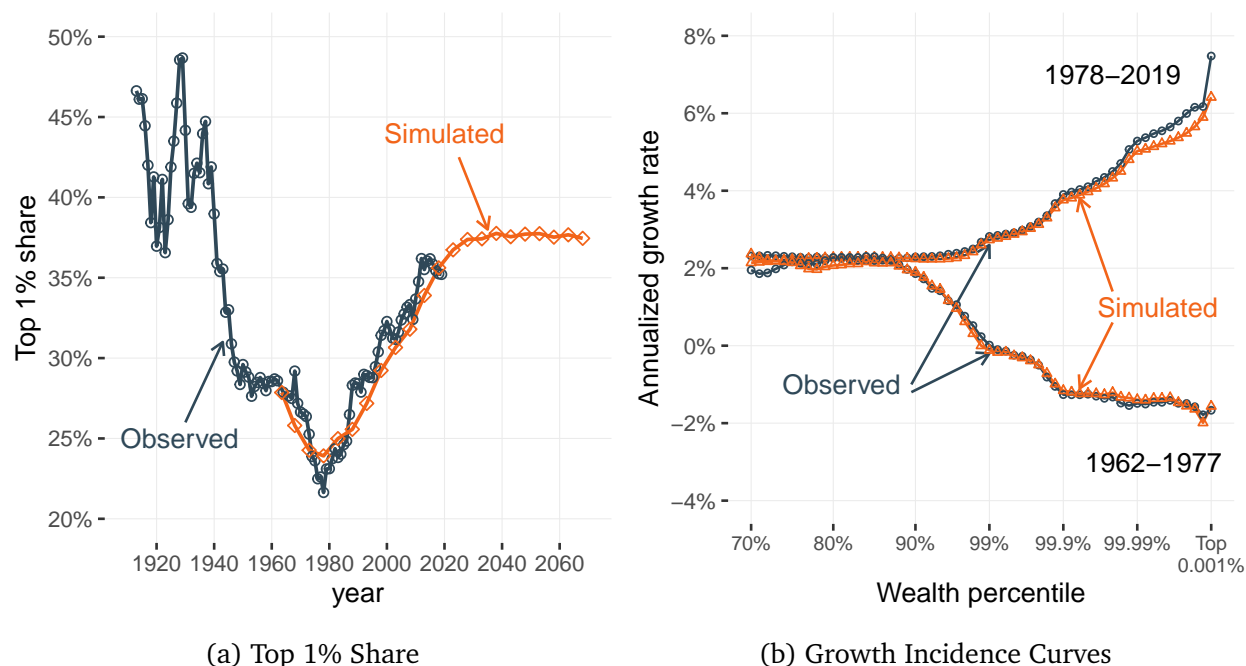
Note: The simulation of the model involves randomly simulated values: to filter out the resulting statistical noise, I simulate the model five times and take the median of the simulations. See main text for details. After 2019, the simulation use the demographic projection (medium variant) from the World Population Prospects (United Nations, 2019) and otherwise assumes that economic parameters remain fixed at their latest observed values.

Figure 7: Comparison of the Model with Observed Dynamics

Figure 7 confirms that this is the case. Figure 7a compares the evolution of the top 1% wealth share in real and simulated data, and shows that we reproduce both the decrease in wealth inequality over 1962–1978, and the increase over 1979–2019. Figure 7b goes further by showing the growth incidence curves (GICs) for the top 30% (a group that has consistently owned about 90% of total wealth since the 1960s).[35] Again, we reproduce the observed growth rates for every percentile (and for fractions of a percentile within the top 1%), both over

---

[35]Average wealth below the 70th percentile is very low, even zero or negative for some percentiles, and therefore it is not meaningful to calculate their growth rates.

1962–1977 and over 1978–2019.

Consumption   To evaluate the external validity of the model, I now compare the model's estimates of consumption to external data. It is important to note that direct evidence on the value of these parameters is scant — which is, in fact, a central motivation for the indirect approach taken in this paper. Nonetheless, we can use two surveys to shed some light on these values. First, there is the SCF, which is usually cross-sectional but had a panel wave between 2007 and 2009. Then there is the PSID, which has been recording wealth every five years since 1984.

		Model		SCF	PSID
		1962–1977	1978–2019	2007–2009	1984–2019
Mean (% of wealth)	50–90%	39%	31%	32%	38%
	90–99%	15%	15%	17%	14%
	Top 1%	13%	9%	21%	12%
Std. Dev. (% of wealth)	50–90%	50%		42%	61%
	90–99%	27%		31%	41%
	Top 1%	37%		35%	31%

Source: Author's calculations using the Survey of Consumer Finances (SCF) and the Panel Study of Income Dynamics (PSID). *Notes:* All numbers are yearly values: data from the SCF and the PSID, which are calculated between several years, are rescaled by a factor Δt (for the means) and $\sqrt{\Delta t}$ (for the standard deviation) where Δt is the number of years, to make estimates comparable. I winsorize the bottom and the top 2.5% of survey-based consumption estimates to limit the impact of measurement error. For the model, values may differ from Table 2 because they are population-weighted, rather wealth-weighted.

Table 1: Distribution of the Propensity to Consume: Comparison With Other Sources

Because these surveys record wealth and income longitudinally, I can use them to estimate the consumption of each respondent. In principle, I can calculate consumption as the difference between income and the variation of wealth between two consecutive interviews. In practice, doing so involves considerable difficulties, and estimates based on that approach only exist to give rough orders of magnitudes.[36] Table 1 nonetheless provides the mean and variance

---

[36]First, the surveys do not record the income earned between interview years, so I have to assume that the respondent's incomes have not changed between interviews. Second, some forms of income, especially accrued capital gains, are not always properly recorded in the surveys, so changes in asset prices might affect the results. Third, since we are calculating consumption as a residual, we are more sensitive to measurement error, which can be a significant issue in surveys. Fourth, surveys record their data at time intervals that are less frequent than this paper's yearly data. I rescale the survey estimates of the mean and the variance of consumption by a factor Δt, where Δt is the time between interviews, to annualize all numbers and make them comparable. But that doesn't

of consumption in the surveys, as a fraction of wealth, for three brackets, estimated so as to be as comparable as possible to the model's parameters. Overall, we observe broadly similar numbers. In particular, this exercise confirms our two notable findings: significant consumption levels (including at the top) and important mobility throughout the distribution. One of the largest discrepancies between the model and the surveys concern the mean consumption of the top 1% which is twice as high in the SCF than according to the model. This difference could be explained by the fact that the survey was conducted over the great recession and that the consumption estimate in the SCF might be polluted by asset price declines.

Wealth Mobility   Another way to check the external validity of my approach is to compare the mobility implied by the model with the mobility we find in the survey data. This approach is less detailed, but more robust than attempts to estimate consumption, because it does not require information on income. Figure 8 compares wealth mobility in the model with the SCF and the PSID. On the x-axis, I group observations according to their wealth rank, and on the y-axis, I plot the distribution of the wealth ranks for each group in the following survey wave, using the median rank and the interquartile range. Then I estimate comparable quantities using the model. For the SCF (Figure 8a), I estimate mobility over two years to match the frequency of the survey and show results up to the top 0.1%. For the PSID, the interval is five years, and given the smaller sample size and lack of oversampling at the top, I only go up to the top 5%.

Once again, the model is broadly consistent with the panel survey data. There is a fair amount of persistence in the wealth rank over time. On average, an observation in wave $n+1$ remains close to its rank in wave n. But there is noticeable variability around this central tendency, which shows that there is still significant movement in the wealth distribution over time. The magnitude of this variability, as shown by the interquartile ranges in Figure 8, is similar in the data and the model.

5   The Drivers of Wealth Inequality

5.1   Decomposition of Wealth Growth

I can use equation (6) of the model to get a straightforward decomposition of the growth of any part of the wealth distribution. This decomposition is similar to that of Gomez (2022), with the difference being that it is estimated directly from comprehensive wealth data in the United States, and accounts for more factors. To understand the decomposition, define

---

fix the fact that we are discretizing the underlying continuous-time process and that the larger time steps lead to coarser approximations.

(a) Panel SCF (2007–2009)   (b) PSID (1984–2019)
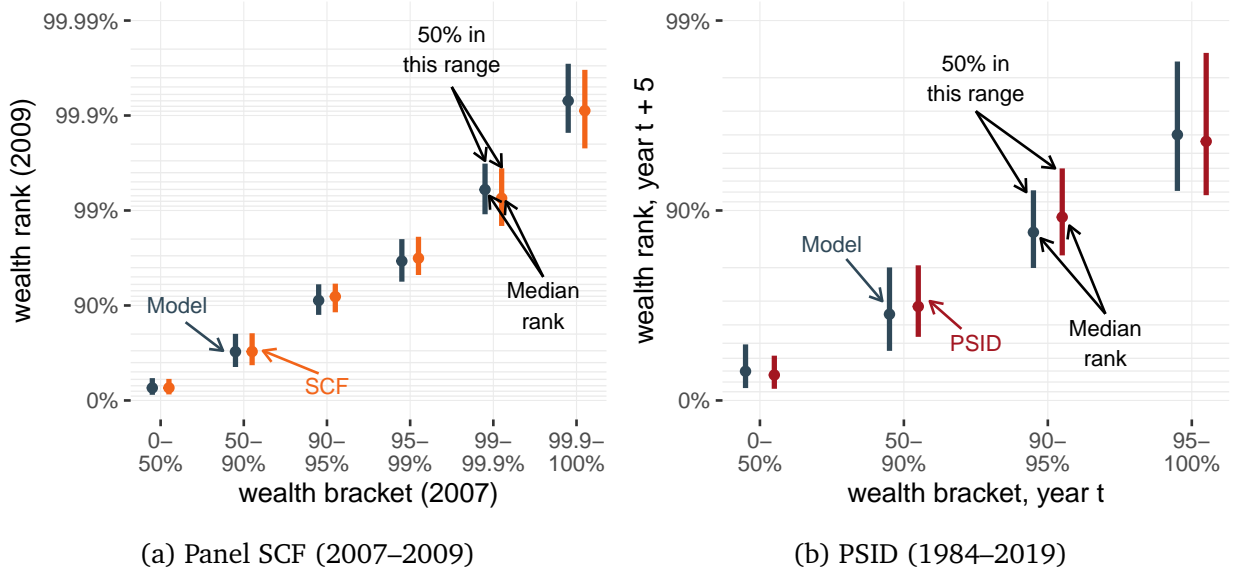
Source: Own computation using the Panel Survey of Consumer Finances (SCF) (2007–2009) and the Panel Study of Income Dynamics (PSID) (1984–2019).

Figure 8: Comparison of Mobility in the Model with Panel Survey Data

$Q_t(p)$, the p-*th* quantile of the wealth distribution at time t. By definition, $F_t(Q_t(p)) = p$. Taking the derivative of this expression with respect to time, and letting $w = Q_t(p)$ we get that $\partial_t Q_t(p) = -\partial_t F_t(w)/f_t(w)$. Therefore, the left-hand side of equation (6) is equal to the variation of the p-th quantile. Let $p = 99\%$ and let $W_t(p)$ be the average wealth of the top 1%. We can write the rate of growth of $W_t(p)$ as the average of the growth of the individual fractiles that make up the top 1%, i.e., $\partial_t W_t(p) = \frac{1}{1-p} \int_p^{+\infty} \partial_t Q_t(r) \, dr$. Therefore, if we average the effects on the right-hand side of equation (6), we can decompose the growth of the top percentile of wealth.

Table 2 shows this decomposition separately for the period of decreasing inequality (1962–1978) and increasing inequality (1979–2019). For each period, I express the total rate of growth of the top wealth percentile as the sum of its different components. One virtue of this presentation is that it puts all the effects we consider here on the same scale — although they are all conceptually very distinct. We can say, for example, that demography over 1979–2019 had the same effect as a 1.8 pp. decrease in the rate of return would have had.

Two facts stand out in Table 2. First, drift and mobility dominate the other factors by far. Each of them have an opposite effect, so when combined, they partially cancel out. But this should not make us overlook that these effects are each very sizable on their own and that the trajectory of wealth inequality is effectively determined by temporary imbalances between the two. Second, the increase in the drift — which accounts for most of the change in the growth

36

Effect	1962–1978	1979–2019	Difference
Demography	−1.6%	−1.8%	−0.2%
Inheritance	1.6%	1.1%	−0.6%
Marriages & Divorces	−0.07%	−0.02%	+0.05%
Drift	−10.1%	−4.1%	+6.1%
incl. Labor income	2.2%	2.2%	−0.007%
incl. Capital income	4.7%	4.5%	−0.3%
incl. Capital gains	−1.8%	1.0%	+2.9%
incl. Consumption	−15.2%	−11.8%	+3.5%
Mobility	9.4%	9.0%	−0.4%
incl. Income	0.6%	0.4%	−0.2%
incl. Consumption	8.8%	8.6%	−0.2%
Equals: Average annual growth	−0.8%	4.2%	+5.0%

Source: Author's calculations. *Notes:* Growth is adjusted for inflation. The growth numbers correspond to the average annual growth of the top 1% of the wealth distribution over each period. Values differ from Table 1 because they are wealth-weighted, rather population-weighted.

Table 2: Decomposition of Wealth Growth in the Top 1%

rate of the top 1% — is primarily driven by two factors: an increase in capital gains and a decrease in average consumption. The increase in capital gains comes largely from the fact that wealth holders experienced capital losses over 1962–1978. On the other hand, regular capital incomes cannot account for the rise of wealth inequality since they have been lower during 1979–2019 than during 1962–1978.

While Table 2 provides a valuable overview of the drivers of top wealth growth, it is limited in its ability to tell us how the different factors have truly impacted wealth inequality. Indeed, many of the terms in equation (6) are endogenous to the distribution of wealth itself, and wealth evolves through several feedback loops between each side of equation (6). For this reason, we cannot simply calculate a counterfactual growth rate for the wealth of the top 1% by changing specific terms in the decomposition presented in Table 2. The next section goes deeper by performing counterfactual simulations where we change some parameters.
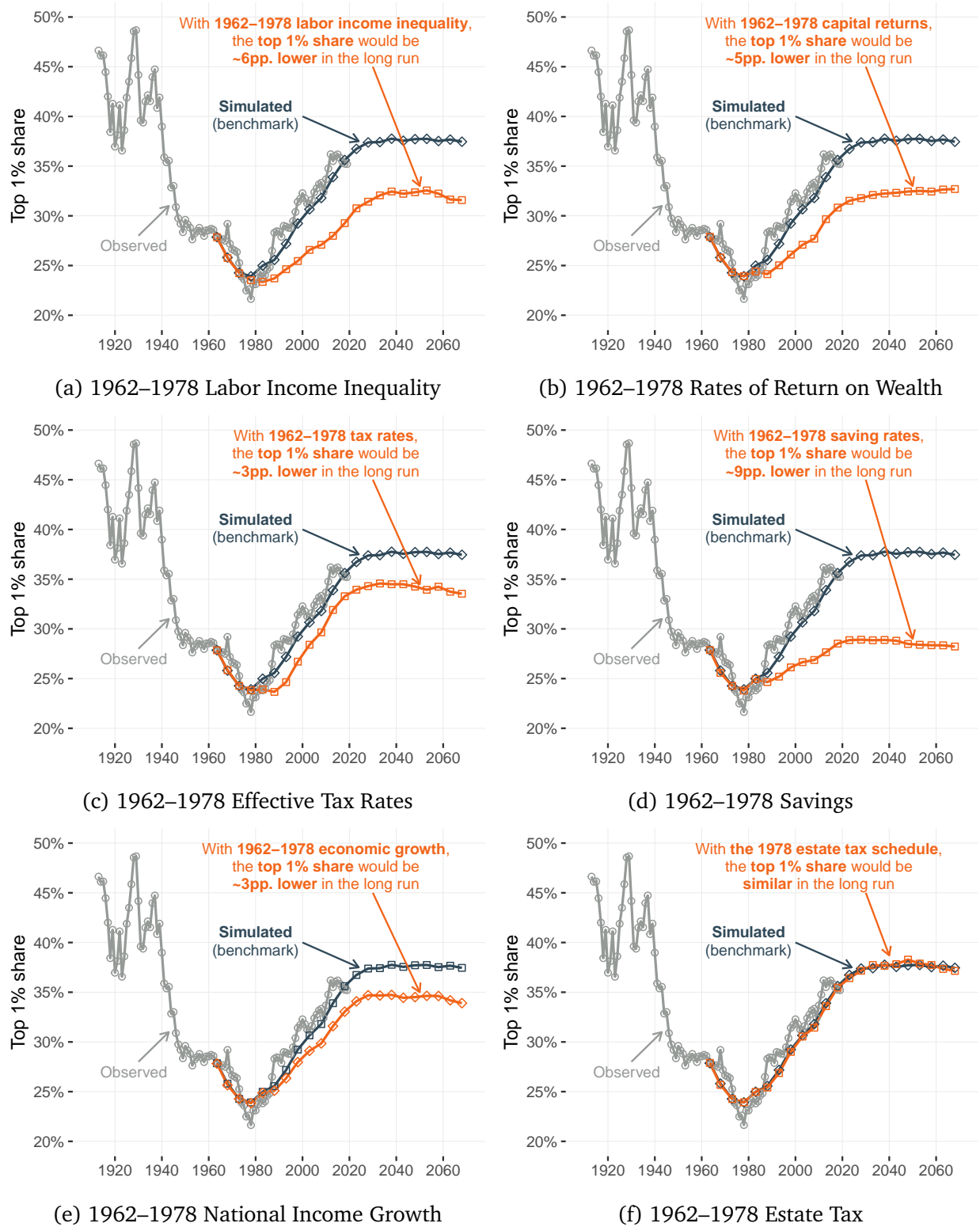
5.2 Counterfactuals

In this section, I change various parameters of the wealth accumulation process and observe how the distribution of wealth would have evolved under these different circumstances. I stress that these counterfactuals exist to explore the direct, proximate consequences of the

effects under study. In particular, when I consider changes in, say, income or taxation, I do so while explicitly keeping the consumption unchanged, so that I can explore one specific channel at a time. Therefore, this exercise should not be interpreted as a full counterfactual, which would incorporate both direct and indirect effects. It remains a powerful way to explore many mechanisms and can be used to clarify the facts that more exhaustive models would need to match.

Benchmark   Figure 9 shows the evolution of the top 1% wealth share according to the model under different scenarios. I run the simulation from the beginning of the data (in 1962) to the year 2070. In each case, I compare the result to a benchmark scenario, which estimates the future evolution of wealth assuming the characteristics of the economy are held fixed after 2019. Specifically, I assume that economic growth remains at its 2010–2019 average and that the distribution of labor income and capital rates of return remain at their 2019 value. I simulate the effect of demography in the future using the projections (medium variant) from the World Population Prospects (United Nations, 2019), and otherwise assume that the correlation between age and wealth remains constant after 2019. In this benchmark scenario, the top 1% wealth stabilizes around 37–38% in the long run.

Labor Income Inequality   In Figure 9a, I estimate what the distribution of wealth would look like today if the distribution of labor income had stayed the same after 1978 as it was over the 1962–1978 period. That is, I give people with a given rank in the wealth distribution after 1978 the average mean and variance of labor income from people with the same rank over 1962–1978. This implies that the distribution of labor income is held fixed. I find that the increase in the top 1% wealth since the 1980s would have been significantly lower under those conditions and would be about 6pp. lower in the long run. The distribution of labor income has thus been an important driver of rising wealth inequality, but remains far from explaining the full rise.

Rates of Return of Wealth   In Figure 9b, I maintain the rates of return on capital at their average 1962–1978 value. I find an effect on the top 1% wealth share that is similar to that of labor income. Importantly, capital gains drive the entire effect. Indeed, as shown in Table 2, regular capital returns have actually been lower since 1979 than during 1962–1978. But overall wealth returns have been higher, particularly due to the large capital losses observed during the 1970s. The effect seen in Figure 9b is in fact similar to what we would obtain by assuming zero capital gains since 1978.

(a) 1962–1978 Labor Income Inequality

(b) 1962–1978 Rates of Return on Wealth

(c) 1962–1978 Effective Tax Rates

(d) 1962–1978 Savings

(e) 1962–1978 National Income Growth

(f) 1962–1978 Estate Tax

Note: See main text for details. The benchmark simulation use the demographic projection (medium variant) from the World Population Prospects (United Nations, 2019) and otherwise assumes that economic parameters remain fixed at their latest observed values. The counterfactual projections change a parameter but leave the others constant. Each model simulation involves randomly simulated values: to filter out the resulting statistical noise, I simulate the model five times and take the median of the simulations.

Figure 9: Counterfactual Evolution of Wealth Inequality

Taxation   In Figure 9c, I focus my attention on taxation. Note that tax rates do not explicitly intervene in my decomposition since its parameters depend directly on the post-tax distribution of labor income and post-tax rates of return. To explore the effects of taxation, I therefore construct alternative distributions of labor income and capital return, for which I assume that pretax distributions evolved as they did in real life, but for which average effective tax rates by wealth percentile are maintained at their average 1962–1978 level. Because tax progressivity has decreased since the 1980s (Saez and Zucman, 2020b), the wealthiest people in this scenario end up with less after-tax labor income and lower after-tax rates of return. As a result, they accumulate less wealth. Figure 9c shows the outcome of this process. I find that this direct effect of taxation is real but also significantly more muted than the overall effect of labor income or rates of return. This aligns with the fact that the rise in post-tax income inequality is mostly explained by rising pretax inequality and that the decrease in tax progressivity only plays a secondary role.

Savings   Figure 9d considers the role of the changes in savings estimated by the model. Of the different effects I consider, this one plays the most prominent role. Assuming the same average consumption as in 1962–1978, the top 1% wealth share would be about 9pp. lower in the long run. This implies that savings have played a large role in explaining today's levels of wealth inequality: the wealthy are wealthier today than in the 1960s and 1970s in large part because they have saved more.[37]

Economic Growth   Economic growth also plays an important role in equation (6). Indeed, people at the top of the wealth distribution derive most of their income from capital, so the rate at which they accumulate wealth follows the rate of return r_t. On the other hand, people at the bottom of the wealth distribution derive most of their income from labor, so their ability to accumulate wealth follows the growth of labor income which matches g_t in the long run. This mechanism explicitly manifests itself through the fact that wealth accumulation in equation (1) depends on $r_{it} - g_t$. This dependence of wealth inequality on $r - g$ was emphasized by Piketty (2014) and Piketty and Zucman (2015). In Figure 9e, I explore the role that the slowdown of growth since the 1980s had on the wealth distribution. To that end, I increase economic growth by a constant factor over 1979–2019 to match the average growth over 1962–1978. I find a limited but noticeable effect: had the economic growth not slowed down, the top 1% wealth share would be about 3pp. lower.

---

[37]Naturally, such changes in consumption could themselves be the result of changes to income and taxation. But the current exercise maintains the distinction between the direct and the indirect effects of income and taxes on wealth inequality.

Estate Taxation   Estate taxation has undergone many reforms, leading to a top marginal rate that is half as high today as it was in the 1960s. What role has estate taxation played in the increase in wealth inequality? I explore this issue in Figure 9f, in which I freeze the estate tax schedule after 1978. I find virtually no impact of this change on the wealth distribution. Two factors account for this finding. First, the evolution of the overall progressivity of the estate tax over the 20th century has actually been more ambiguous than what the trajectory of the top marginal tax rate would suggest (see Figure B.3 in appendix). The very high marginal tax rates of the 1960s did not kick in until extremely high levels of wealth. Following the reforms in the 1980s, estates in the order of tens of millions of dollars were actually taxed more heavily. Estates needed to reach hundreds of millions of dollars to benefit from the reforms. It wasn't until the 2000s that the estate tax became unambiguously less progressive.

Second, estate taxation has an intrinsically weak impact on the wealth distribution. Weaker, say, than an annual wealth tax of seemingly comparable magnitude. As an illustration, running the model with a radical estate tax (100% tax on estates above \$100k) would only reduce the top 1% wealth share by 1.5pp. in the long run. A naive view would suggest that taxing wealth every year at a net-of-tax rate $1 - \tau$ is equivalent to taxing wealth every n years at a rate $(1 - \tau)^n$. This exercise shows this is not the case: taxing wealth once every generation, as the estate tax does, fundamentally alters the nature of the tax. The typical lifetime of a generation is the main determinant of this fact: the longer generations live, the more different the estate tax is from a wealth tax. Section 6.3 develops a simplified model where this finding can be derived analytically.

Other Effects   The other effects considered in the model (demography, assortative mating) have had a negligible impact on the distribution of wealth.

6   The Taxation of Wealth

In this section, I use the model of this paper to assess the long-run effect of wealth taxes at the top of the distribution. Understanding this effect is crucial to understanding capital taxation in general and wealth taxation in particular, as recent studies have argued that the long-run elasticity of wealth with respect to the net-of-tax rate is a sufficient statistic for optimal capital taxation (Saez and Stantcheva, 2018; Piketty and Saez, 2013a). However, while short-run responses to wealth taxation can be measured empirically (e.g., Brülhart et al., 2016; Seim, 2017; Jakobsen et al., 2020; Zoutman, 2018; Ring, 2020; Londoño-Vélez and Ávila-Mahecha, 2021), the long-run responses are more elusive.

This paper provides a powerful framework for addressing that question. Indeed, it can

estimate counterfactual steady-state wealth distributions, in which we assume different rates of return or different savings. This problem is, effectively, equivalent to estimating a counterfactual wealth distribution under a wealth tax (which decreases post-tax rates of return) with or without behavioral responses (which modify saving rates). A key advantage of this approach — which is often absent from studies of optimal capital taxation — is the presence of mobility. Not only is mobility a desirable feature on its own, but it also ensures that the model is well-behaved under a wide range of economic behaviors, because it naturally leads to nondegenerate steady-states for the wealth distribution. This makes it a realistic, easy and tractable way to explore the equity-efficiency trade-offs that are involved in wealth taxation. In contrast, many models without mobility generate infinite responses of capital supply to taxation; under these conditions, the efficiency concerns always dominate the equity concerns, which leaves no room to discuss trade-offs.

One way to use this paper's model is to fully simulate the evolution of the wealth distribution under various wealth tax assumptions, similarly to what was done in Section 5.2. This solution is the most complete but also the most computationally demanding. This section pursues an alternative path, where I focus on the long run and use a slightly simplified model. In these conditions, it becomes possible to derive analytical formulas for how the wealth distribution would eventually react to any wealth tax. This makes it easy to draw Laffer curves for wealth taxation and determine what would be the revenue-maximizing wealth tax rate at the top.[38]

6.1 Theoretical Results

6.1.1 Wealth Taxation Without Behavioral Responses

Consider the introduction of a nonlinear wealth tax $\tau(w)$ on wealth w. I will first assume that there is no behavioral response following the introduction of the wealth tax, so the evolution of wealth is simply characterized by $dw_{it} = (\mu(w_{it}) - \tau(w_{it}))\,dt + \sigma_t(w_{it})\,dB_{it}$. I can state the following result.

Proposition 2. *Let f be the steady-state density of wealth without a wealth tax. Then the steady-state density f^* with a wealth tax is equal to the steady-state density without a wealth tax, reweighted by a factor $\theta(w)$:*

$$f^*(w) \propto \theta(w)f(w) \qquad where \qquad \theta(w) = \exp\left\{-\int_{-\infty}^{w} \frac{2\tau(s)}{\sigma^2(s)}\,ds\right\}$$

---

[38]This section does not develop a complete theory of optimal wealth taxation, which would be beyond the scope of this paper. Instead, it focuses on the observable responses of the wealth distribution to a wealth tax. I leave the issue of integrating this framework with a complete optimal tax model for future research.

In particular, if $\tau(w) = \tau(w - w_0)_+$ (i.e., the wealth tax is linear with rate τ above a threshold w_0), and if $\sigma(w) = \sigma w$ for $w \geq w_0$ (i.e., diffusion is proportional to wealth above w_0) then $\theta(w)$ simplifies to:

$$\forall w > w_0 \qquad \theta(w) = \exp\left\{-\frac{2\tau}{\sigma^2}\left(\frac{w_0}{w} - 1\right)\right\}\left(\frac{w}{w_0}\right)^{-2\tau/\sigma^2}$$

and $\theta(w) = 1$ otherwise.

Proof. See Appendix A.2. □

That result makes it possible to estimate how the tax base would react to a wealth tax in the long run, effectively by reweighting the steady-state distribution of untaxed wealth using the function θ. The setting mentions the introduction of a new wealth tax where there previously was none, but we could apply the same result to an increase or a decrease of an existing wealth tax by redefining τ as a change in the rate of the wealth tax.[39]

The result emphasizes the role of mobility, as explained by Saez and Zucman (2019). The impact on the tax base depends on $\tau(w)/\sigma^2(w)$ and not just $\tau(w)$. Doubling the parameter $\sigma(w)$ quadruples the parameter $\sigma^2(w)$, which implies that a tax rate four times as high would lead to the same change in the tax base. The intuition is the same as in Saez and Zucman (2019): high mobility means that people only get taxed for a short period and that new, previously untaxed wealth keeps entering the tax base. As a result, the tax base does not react too much to wealth taxation. When mobility goes to zero, however, the same wealth from the same people is taxed repeatedly so that the tax base eventually goes to zero.

Proposition 2 carries one important difference compared to the result of Saez and Zucman (2019). In the second part of the result, the reweighting factor is the product of two terms: $\exp\{-2\tau(w_0/w - 1)/\sigma^2\}$ and $(w/w_0)^{-2\tau/\sigma^2}$. The first impacts the distribution near the threshold w_0 while the second impacts the distribution away from the threshold. The discrete-time formula of Saez and Zucman (2019) only contains an equivalent of the second term, $(w/w_0)^{-2\tau/\sigma^2}$. This difference is because they consider the taxation of billionaires not at a marginal rate τ, but at an average rate τ. This distinction is important because the long-run mechanical effect of a wealth tax explicitly depends on the average tax rate and not, like behavioral responses, on the marginal rate. At the very top of the distribution, the distinction between the average and the marginal rate becomes negligible, hence the term $(w/w_0)^{-2\tau/\sigma^2}$ similar to the formula in Saez and Zucman (2019). But close to the threshold w_0, the distinction does matter and, in fact, has

---

[39]This result does not consider how the government uses the wealth tax. We can adapt the formula to include the redistribution of a lump-sum amount and then solve an equation numerically to ensure that we redistribute as much as we tax. In practice, the impact would be negligible as long as we focus on the top of the distribution. Indeed, much more wealth is taxed in that part of the distribution than would be redistributed. The lump sum rebate would have an impact at the bottom, however. Section C.5 in appendix addresses this question.

important consequences for the overall response of the tax base to the tax, as we will see in Section 6.2.

6.1.2 Behavioral Responses

Tax Evasion and Tax Avoidance   People can react to a wealth tax by hiding some of their wealth, either through tax evasion or tax avoidance. Assume that, in response to a marginal tax rate $\tau'(w)$, people only report a fraction $\alpha(w) = [1 - \tau'(w)]^\varepsilon$ of their wealth. The parameter ε is the elasticity of declared wealth to the marginal net-of-tax rate $1 - \tau'(w)$. For a small rate $\tau'(w) \ll 1$, people react by approximately hiding a fraction $\tau'(w)\varepsilon$ of their wealth. When $\varepsilon = 0$, people truthfully report all of their wealth. As ε goes to infinity, people start hiding all of their wealth to avoid paying the tax. With tax avoidance, people that own w in wealth pay $\tau(w)\alpha(w)$ instead of $\tau(w)$. In effect, this is equivalent to having a wealth tax with a lower rate. Therefore, the results for the purely mechanical model hold with minimal modifications.

Consumption   People may also react to a wealth tax by accumulating less wealth. Changes to savings have different implications than tax evasion. Indeed, tax evasion affects both the dynamic of wealth and the tax base. Savings, on the other hand, affect the dynamic of wealth but do not directly reduce the tax base.

Theory provides few constraints regarding how a wealth tax ought to affect saving rates, given the many settings and mechanisms we could consider. In broad terms, there can be a substitution effect that decreases savings (because a wealth tax makes deferring consumption more expensive). And there can be an income effect that increases savings (because a wealth tax makes people poorer, and they compensate by investing more). In the case of, say, labor supply, the widely accepted view that substitution effects dominate. No such consensus exists for savings.

The following reduced-form specification can nonetheless account for the overall effect in a direct and intuitive way. Assume that, in response to a tax with marginal rate $\tau'(w)$ on wealth, people increase their consumption by a factor $\beta(w) = [1 - \tau'(w)]^{-\eta}$. The parameter η captures the elasticity of consumption with respect to the marginal net-of-tax rate $1 - \tau'(w)$.[40] The drift in the dynamic of wealth now includes an additional term $c(w)[1 - \beta(w)]$ where $c(w)$ is the average propensity to consume. The savings behavioral response amplifies the impact of the wealth tax.

---

[40]I will ignore the cases where $\eta < 0$ (i.e., income effects dominate), even though they are a theoretical possibility and even though some studies find this result (Ring, 2020). Indeed, it is problematic to assume in a taxation context that the tax base responds positively to the tax. Moreover, the elasticity has to change sign at some point, otherwise, a 100% wealth tax would correspond to infinite savings. However, if true, it would imply that wealth tax rates could be higher.

Complete Model   Behavioral responses change the drift term, which is analogous to a change in the effective rate of the wealth tax. Therefore Proposition 2 can be directly extended to account for behavioral effects. The reweighting factor in the full model becomes:

$$\theta(w) = \exp\left\{-\int_{-\infty}^{w} \frac{2\tau(s)\alpha(s)}{\sigma^2(s)}\,\mathrm{d}s - \int_{-\infty}^{w} \frac{2c(s)[1-\beta(s)]}{\sigma^2(s)}\,\mathrm{d}s\right\}$$

6.2   Calibrations for the United States

Baseline Wealth Model   To illustrate the formulas of this section, I will consider the case of a linear tax on estates above \$50m. I consider a model of wealth accumulation with a mobility parameter and an average propensity to consume that match this paper's estimates for the United States over 1979–2019. I directly use the 2019 data as an estimate of the steady-state wealth distribution, given that the simulation in Section 5.2 suggest that wealth inequality is close to its long-run value.

Behavioral Elasticities   To calibrate ε and η, I rely on the recent empirical literature that exploit various quasi-experimental settings to assess behavioral reactions to a wealth tax.

Several of these papers present bunching evidence (Seim, 2017; Londoño-Vélez and Ávila-Mahecha, 2021; Jakobsen et al., 2020). Bunching provides the cleanest estimates of pure tax avoidance elasticity. Indeed, the true value of wealth in the short run tends to follow unpredictable asset movements so it would be hard for a household to precisely bunch at kink points. Seim (2017) finds an elasticity of 0.5 in Sweden, and Jakobsen et al. (2020) find elasticities that are even lower in Denmark. Londoño-Vélez and Ávila-Mahecha (2021) find a higher estimate (2–3) in Colombia.

As their main identification strategy, Jakobsen et al. (2020) pursue a difference-in-difference approach that exploit various tax reforms. This allows them to compute elasticities that incorporate dynamic and saving responses over larger time spans. Over an 8-year time frame, they find a sizable elasticity at the top of about 18 with respect to the net-of-tax rate. The authors argue that most (90%) of it can be attributed to a behavioral effect (as opposed to a mechanical effect). Assuming that the elasticities cumulate multiplicatively over time, this would correspond to a yearly behavioral elasticity of 1.4 for both the saving and tax avoidance response. Zoutman (2018), also using a difference-in-difference strategy, finds a much higher elasticity of almost 14. Seim (2017) also analyzes saving responses to a wealth tax but does not find any. Ring (2020) exploits geographic discontinuities in the exposure to wealth taxation to estimate savings responses and actually finds an *increase* in savings in response to wealth taxation. Brülhart et al. (2016) find a much higher overall elasticity (23–34) in Switzerland

using both between canton variations of the tax rate and within variation in the Bern canton. They also look at bunching evidence but find much lower effects there.

Note that the tax avoidance elasticity is not a pure structural parameter, but also results from how strongly a wealth tax is enforced. For the baseline calibration, I will consider a limited tax avoidance response ($\varepsilon = 1$), which is around the values found by Seim (2017), Londoño-Vélez and Ávila-Mahecha (2021) and Jakobsen et al. (2020). I will also consider a medium savings response ($\eta = 1$), in line with Jakobsen et al. (2020), and around the median of existing studies. I will also consider an alternative scenario with a higher saving response ($\eta = 3$) and a higher tax avoidance response ($\varepsilon = 3$).[41]

(a) Laffer Curves

(b) Revenue-Maximizing Tax Rates
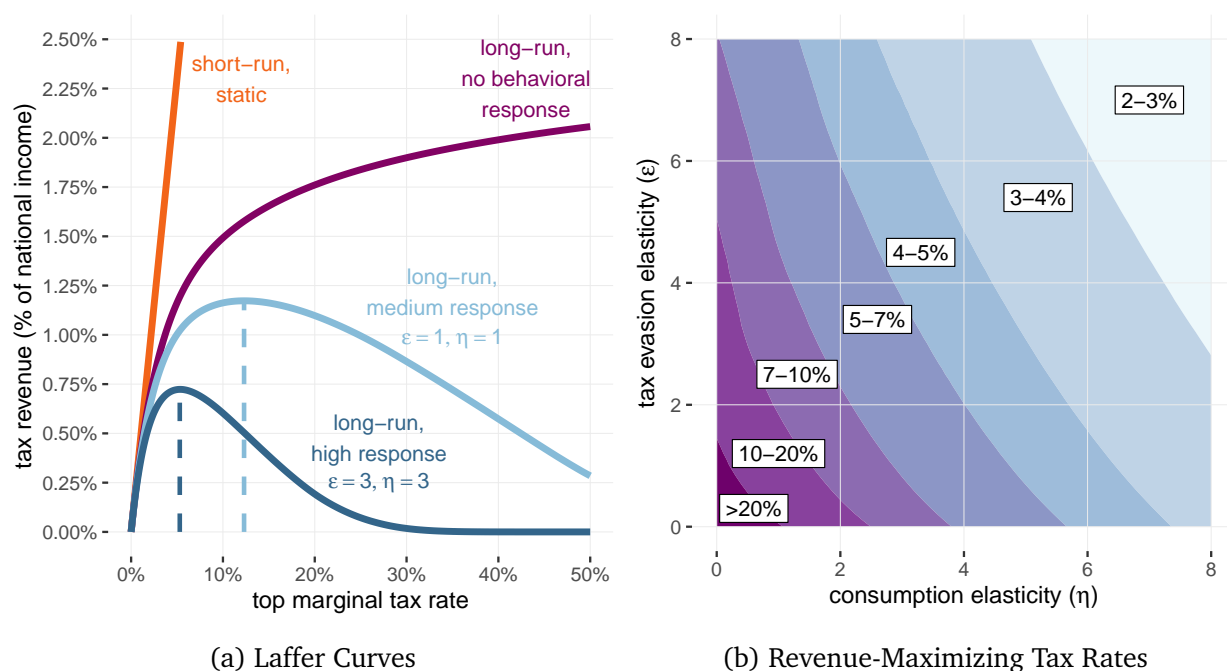
Figure 10: Effects of a Wealth Tax on Estates Above $50m

Results Figure 10a shows the Laffer curves for the wealth tax, i.e., the amount of tax revenue raised as a function of the tax rate. As a point of reference, we can start from the short-run, static revenue estimate (in orange). This estimate assumes the tax base to be entirely nonreactive, and as a result, the wealth tax can raise a very large amount of revenue. The second curve (in purple) accounts for how the wealth tax progressively eats away its tax base in the long run, even without behavioral responses. As argued by Saez and Zucman (2019), mobility is the central factor determining that curve's shape. In a model without mobility (e.g., Jakobsen et al.,

---

[41]Very large behavioral responses, as found by Brülhart et al. (2016), are of limited interest for the current exercise, since they make long-run dynamic effects negligible compared to the immediate static effect of behavioral responses.

2020) this curve would be equal to zero for all positive tax rates. This dynamic mechanical effect is sizable and significantly lowers tax revenue in the long run. However, it is not strong enough to produce the usual, inverted U-shaped Laffer curve: the tax base never goes to zero, even with a 100% tax rate. As a consequence, that model is insufficient to generate guidance on the revenue-maximizing rate.[42] The reason behind the result is that the dynamic mechanical effect depends on the average tax rate and not, like behavioral responses, on the marginal rate. Hence, close to the exemption threshold, the impact of the tax remains limited even with very high rates. So, in theory, it remains possible to keep raising revenue by increasing the tax rate *ad infinitum*. To get a more realistic and actionable model, we need to incorporate behavioral responses, which I do for the two last curves (in blue). In the benchmark calibration, the revenue-maximizing rate remains high ($\approx 12\%$), but the amount of revenue raised is only a fraction of what a static scoring finds.

There is still a fair amount of uncertainty regarding the value of behavioral elasticities, so in Figure 10b, I estimate the revenue-maximizing rates for a wide range of elasticity values. A notable result here is that the consumption elasticity has a stronger impact than the tax avoidance elasticity. This is because tax avoidance actually has an ambiguous effect on the tax base. On the one hand, it directly lowers the tax base since people underreport their assets. But on the other hand, it increases the post-tax rate of return, allowing people to accumulate more, which grows the tax base in the long run.

6.3   Comparison with Inheritance Taxation

How does the annual taxation of wealth compare to inheritance taxation? In simulations, I found in Section 5.2 that the estate tax has a limited impact on the distribution. To illustrate and clarify this finding, I develop a simple model that compares the mechanical effects of an annual wealth tax to the effects of an inheritance tax. Assume that wealth w evolves according to the following SDE, where wealth is taxed at a constant rate τ, and μ is the rate of wealth growth without tax:

$$\mathrm{d}w = (\mu - \tau)w\,\mathrm{d}t + \sigma w\,\mathrm{d}B_t$$

Assume a reflecting barrier at the bottom, so that wealth cannot go below w_0. This is the simplest model that generates a Pareto distribution of wealth, with a PDF of the form $f(w) = \frac{\alpha w_0^\alpha}{w^{\alpha+1}}$ (Gabaix, 2009).

Assume that people die at a Poisson rate δ. When they do, their only child pays the estate

---

[42]Saez and Zucman (2019), who consider a somewhat analogous model, do obtain an estimate for the revenue-maximizing rate. This is because in their model, top wealth holders are taxed at an average, not marginal, rate τ (see also Section 6.1.1 for a discussion).

tax, inherits their wealth, and replaces them in the distribution. Let $(1-\chi)^{1/\delta}$ be the net-of-tax rate of the estate tax. (This specification makes the rates τ and χ *a priori* comparable.) At the steady-state, the wealth density f satisfies the Kolmogorov (1931) equation with the Poisson process included, and with the time the derivative set to zero:

$$0 = -\underbrace{(\mu-\tau)\partial_w w f(w)}_{\text{drift}} + \underbrace{\frac{1}{2}\sigma^2\partial_w^2 w^2 f(w)}_{\text{mobility}} - \underbrace{\delta f(w)}_{\text{deaths}} + \underbrace{\frac{\delta}{(1-\chi)^{1/\delta}}f\left(\frac{w}{(1-\chi)^{1/\delta}}\right)}_{\text{births}}$$

The steady-state still follows a Pareto distribution. Substituting f with its value $f(w) = \frac{\alpha w_0^\alpha}{w^{\alpha+1}}$, we get the following equation for the Pareto coefficient α, which is a measure of inequality (higher values mean less inequality):

$$0 = \mu + \frac{1}{2}(\alpha-1)\sigma^2 - \tau - \frac{\delta}{\alpha}[1-(1-\chi)^{\alpha/\delta}] \tag{8}$$
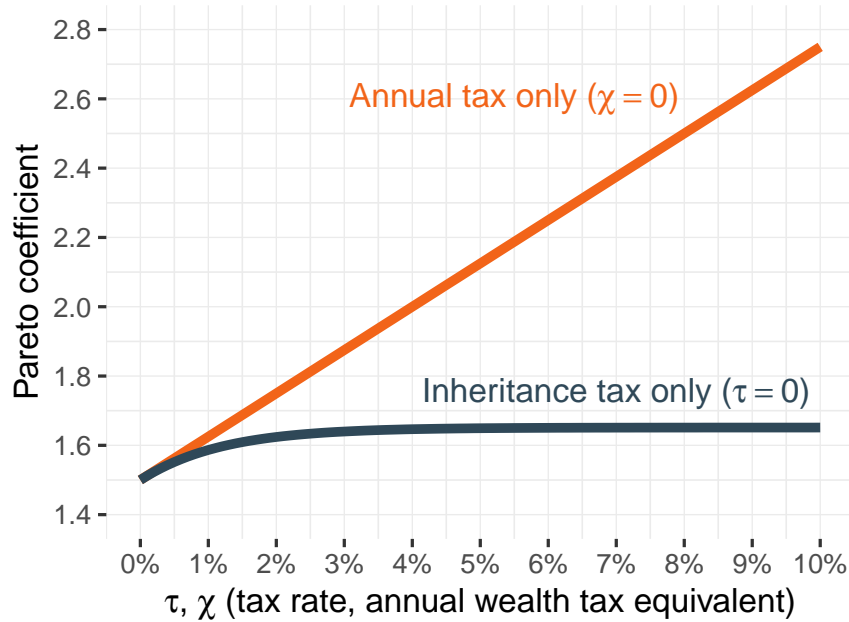

Figure 11: Pareto Coefficients under an Annual Wealth vs. an Inheritance Tax of Comparable Magnitude

Consider the following calibration: $\mu = -0.04$, $\sigma = 0.4$ and $\delta = 1/50$. We can solve this equation numerically to get the steady-state Pareto coefficient under any value of τ and χ. Figure 11 shows the results from this exercise. It compares the Pareto coefficients for two scenarios: an annual wealth tax (without inheritance tax) and an inheritance tax (of comparable magnitude, without an annual wealth tax). With an annual wealth tax, the Pareto coefficient increases linearly with the tax rate (therefore, inequality decreases). With an inheritance tax,

not only does the Pareto coefficient increase much less overall, but as the rate increases, the Pareto coefficients quickly reach an asymptote. Hence, an inheritance is incapable, no matter how large, of reducing inequality to the same extent that an annual wealth tax can.

Equation (8) provides the intuition behind that result. For a small estate tax rate χ, a first-order approximation gives $\frac{\delta}{\alpha}[1-(1-\chi)^{\alpha/\delta}] \approx \chi$, and therefore the estate tax has an effect similar to a wealth tax τ. But for higher rates χ, this is no longer true. In practice, the approximation does not work unless χ is extremely small because the ratio α/δ is high. As the estate tax rate χ increases, the Pareto coefficients converge to a maximum value α_1, which we obtain by solving (8) for $\chi = 1$:

$$\alpha_1 = \frac{1}{2}\left\{\alpha_0 + \sqrt{\alpha_0^2 + \frac{8\delta}{\sigma^2}}\right\}$$

where $\alpha_0 = 1 - 2(\mu - \tau)/\sigma^2$ is the Pareto coefficient in the absence of inheritance tax. The estate tax is, therefore, the least efficient when the ratio $8\delta/\sigma^2$ is small, in which case the Pareto coefficient for $\chi = 1$ is almost the same as the coefficient for $\chi = 0$. This can happen for two reasons: δ being low and σ^2 being high. When σ^2 is high, there is a lot of mobility within generations, so estate taxation is ineffective because wealth quickly recovers from it. But this effect also exists for annual wealth taxes (see Section 6.1.1), so it does not drive the gap between annual and inheritance taxes. When δ is low, people die and children inherit their wealth at a low rate, so estate taxation is ineffective because it happens too rarely. We can also see this effect by looking directly at (8): taking the estate tax rate χ to 100% has the same effect as an annual wealth tax with a rate δ/α. So for $\alpha \approx 1.5$, a 100% estate tax, which occurs every 50 years on average, is similar to a 3% annual wealth tax, and cannot get higher.[43]

7   Conclusion

This paper introduces a new way of decomposing the evolution of wealth inequality. This decomposition accounts for the different processes affecting the wealth distribution, including mobility. It only requires repeated cross-sections and can therefore be applied to available historical data.

Applying this decomposition to the United States, I find that the rise of wealth inequality

---

[43]Intuitively, δ creates a cap on the potency of the estate tax: when it is low, wealth is rarely transmitted, and so its effect is limited. The parameter α modulates that cap. When $\alpha \to 1$, inequality tends to infinity. Only one dynasty owns meaningful wealth, and the estate tax is maximally efficient because it can repeatedly affect that one dynasty. On the other hand, when $\alpha \to +\infty$, inequality tends to zero. Everyone owns a similar amount of wealth. The estate tax struggles to discriminate between the richest and the poorest in these conditions.

has been driven by higher savings at the top, higher rates of capital gains, and higher labor income inequality, with other factors playing a minor role. Applying the framework to the study of wealth taxation, I derive simple formulas that describe how the wealth distribution would react to a wealth tax in the long run. I use them to estimate revenue-maximizing tax rates for a linear tax on high estates. The website `https://thomasblanchet.github.io/wealth-tax/` provides a simulator to apply these formulas in the United States.

The approach developed in this paper offers several venues for future research. One would be to extend the decomposition of this paper alongside several dimensions. Recent work (Blanchet, Saez, and Zucman, 2022) has started to estimate distributional national accounts — including national wealth — disaggregated by race. This paper's methodology could be applied to such data and provide new insights that combine our understanding of the dynamics of the racial wealth gap (Derenoncourt et al., 2022) with our understanding of the dynamics of wealth inequality. To the extent that wealth can be meaningfully divided between household members (Frémeaux and Leturcq, 2020), the same could be done for gender.

In future research, the study of capital taxation that is sketched out in this paper could also be further expanded into a full-fledged theory of optimal capital taxation, one that integrate our theories of the wealth distribution with various social objectives (Saez and Stantcheva, 2016) and with the broader effects of wealth taxation on the macroeconomy (Gaillard and Wangner, 2021).

References

Aiyagari, S. R. (1994). "Uninsured Idiosyncratic Risk and Aggregate Saving". *The Quarterly Journal of Economics* 109.3, pp. 659–684. DOI: 10.2307/2118417. URL: `https://academic.oup.com/qje/article-lookup/doi/10.2307/2118417` (cit. on pp. 7, 8, 10, 30–32).

Atkinson, A.B. and J.E. Stiglitz (1976). "The design of tax structure: Direct versus indirect taxation". *Journal of Public Economics* 6.1-2, pp. 55–75. DOI: 10.1016/0047-2727(76)90041-4. URL: `https://linkinghub.elsevier.com/retrieve/pii/0047272776900414` (cit. on p. 10).

Auerbach, Alan and James Hines (2001). *Taxation and Economic Efficiency*. Tech. rep. w8181. Cambridge, MA: National Bureau of Economic Research, w8181. DOI: 10.3386/w8181. URL: `http://www.nber.org/papers/w8181.pdf` (cit. on p. 11).

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini (2020). "Rich Pickings? Risk, Return, and Skill in Household Wealth". *American Economic Review* 110.9, pp. 2703–2747. DOI: 10.1257/aer.20170666. URL: `https://pubs.aeaweb.org/doi/10.1257/aer.20170666` (cit. on p. 7).

Bassetto, Marco and Jess Benhabib (2006). "Redistribution, taxes, and the median voter". *Review of Economic Dynamics* 9.2, pp. 211–223. DOI: 10.1016/j.red.2006.02.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1094202506000056 (cit. on p. 10).

Benhabib, Jess, Alberto Bisin, and Mi Luo (2019). "Wealth Distribution and Social Mobility in the US: A Quantitative Approach". *American Economic Review* 109.5, pp. 1623–1647. DOI: 10.1257/aer.20151684. URL: https://pubs.aeaweb.org/doi/10.1257/aer.20151684 (cit. on p. 9).

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu (2011). "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents". *Econometrica* 79.1, pp. 123–157. DOI: 10.3982/ECTA8416. URL: http://doi.wiley.com/10.3982/ECTA8416 (cit. on pp. 7, 8, 30, 32).

Bewley, Truman (1977). "The permanent income hypothesis: A theoretical formulation". *Journal of Economic Theory* 16.2, pp. 252–292. DOI: 10.1016/0022-0531(77)90009-6. URL: https://linkinghub.elsevier.com/retrieve/pii/0022053177900096 (cit. on pp. 7, 30, 31).

Blanchet, Thomas, Emmanuel Saez, and Gabriel Zucman (2022). *Real-Time Inequality*. Tech. rep. w30229. Cambridge, MA: National Bureau of Economic Research, w30229. DOI: 10.3386/w30229. URL: http://www.nber.org/papers/w30229.pdf (cit. on p. 50).

Brülhart, Marius et al. (2016). *Taxing Wealth: Evidence from Switzerland*. Tech. rep. w22376. Cambridge, MA: National Bureau of Economic Research, w22376. DOI: 10.3386/w22376. URL: http://www.nber.org/papers/w22376.pdf (cit. on pp. 5, 11, 41, 45, 46).

Cagetti, Marco and Mariacristina De Nardi (2006). "Entrepreneurship, Frictions, and Wealth". *Journal of Political Economy* 114.5, pp. 835–870. DOI: 10.1086/508032. URL: https://www.journals.uchicago.edu/doi/10.1086/508032 (cit. on pp. 7, 30, 32).

Carroll, Christopher (1998). *Why Do the Rich Save So Much?* Tech. rep. w6549. Cambridge, MA: National Bureau of Economic Research, w6549. DOI: 10.3386/w6549. URL: http://www.nber.org/papers/w6549.pdf (cit. on pp. 7, 30, 32).

Carroll, Christopher D. (1997). "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis". *The Quarterly Journal of Economics* 112.1, pp. 1–55. DOI: 10.1162/003355397555109. URL: https://academic.oup.com/qje/article-lookup/doi/10.1162/003355397555109 (cit. on p. 13).

Center for Disease Control (2022). *National Vital Statistics System*. URL: https://www.cdc.gov/nchs/nvss/index.htm (cit. on p. 25).

Chamley, Christophe (1986). "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives". *Econometrica* 54.3, p. 607. DOI: 10.2307/1911310. URL: https://www.jstor.org/stable/1911310?origin=crossref (cit. on pp. 3, 10–12).

Cioffi, Riccardo A (2021). "Heterogeneous Risk Exposure and the Dynamics of Wealth Inequality". URL: https://rcioffi.com/files/jmp/cioffi_jmp2021_princeton.pdf (cit. on p. 7).

De Nardi, Mariacristina (2004). "Wealth Inequality and Intergenerational Links". *The Review of Economic Studies* 71.3, pp. 743–768. DOI: 10.1111/j.1467-937X.2004.00302.x. URL: https://academic.oup.com/restud/article-lookup/doi/10.1111/j.1467-937X.2004.00302.x (cit. on pp. 7, 30, 32).

Deming, W. Edwards (1943). *Statistical adjustment of data*. New York: Wiley (cit. on p. 28).

Derenoncourt, Ellora et al. (2022). *Wealth of Two Nations: The U.S. Racial Wealth Gap, 1860-2020*. Tech. rep. w30101. Cambridge, MA: National Bureau of Economic Research, w30101. DOI: 10.3386/w30101. URL: http://www.nber.org/papers/w30101.pdf (cit. on p. 50).

Fagereng, Andreas et al. (2020). "Heterogeneity and Persistence in Returns to Wealth". *Econometrica* 88.1, pp. 115–170. DOI: 10.3982/ECTA14835. URL: https://www.econometricsociety.org/doi/10.3982/ECTA14835 (cit. on p. 7).

Farhi, Emmanuel (2010). "Capital Taxation and Ownership When Markets Are Incomplete". *Journal of Political Economy* 118.5, pp. 908–948. DOI: 10.1086/657996. URL: https://www.journals.uchicago.edu/doi/10.1086/657996 (cit. on p. 10).

Farhi, Emmanuel and Iván Werning (2013). "Estate Taxation with Altruism Heterogeneity". *American Economic Review* 103.3, pp. 489–495. DOI: 10.1257/aer.103.3.489. URL: https://pubs.aeaweb.org/doi/10.1257/aer.103.3.489 (cit. on p. 10).

Frémeaux, Nicolas and Marion Leturcq (2020). "Inequalities and the individualization of wealth". *Journal of Public Economics* 184, p. 104145. DOI: 10.1016/j.jpubeco.2020.104145. URL: https://linkinghub.elsevier.com/retrieve/pii/S0047272720300098 (cit. on p. 50).

Friedman, Milton (1957). *A Theory of the Consumption Function*. National bureau of economic research 63. Princeton (N. J.): Princeton university press (cit. on p. 13).

Gabaix, Xavier (2009). "Power Laws in Economics and Finance". *Annual Review of Economics* 1.1, pp. 255–294. DOI: 10.1146/annurev.economics.050708.142940. URL: https://www.annualreviews.org/doi/10.1146/annurev.economics.050708.142940 (cit. on pp. 4, 8, 14, 47).

Gabaix, Xavier et al. (2016). "The Dynamics of Inequality". *Econometrica* 84.6, pp. 2071–2111. DOI: 10.3982/ECTA13569. URL: https://www.econometricsociety.org/doi/10.3982/ECTA13569 (cit. on pp. 6, 32).

Gaillard, Alexandre and Philipp Wangner (2021). "Wealth, Returns, and Taxation: A Tale of Two Dependencies". URL: https://www.ssrn.com/abstract=3966130 (cit. on p. 50).

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty (2021). "Accounting for Wealth-Inequality Dynamics: Methods, Estimates, and Simulations for France". *Journal of the European Economic Association* 19.1, pp. 620–663. DOI: 10.1093/jeea/jvaa025. URL: https://academic.oup.com/jeea/article/19/1/620/5846044 (cit. on pp. 10, 32).

Goldie, Charles M. (1991). "Implicit Renewal Theory and Tails of Solutions of Random Equations". *The Annals of Applied Probability* 1.1. DOI: 10.1214/aoap/1177005985. URL: https://projecteuclid.org/journals/annals-of-applied-probability/volume-1/issue-1/Implicit-Renewal-Theory-and-Tails-of-Solutions-of-Random-Equations/10.1214/aoap/1177005985.full (cit. on p. 8).

Gomez, Matthieu (2022). "Decomposing the Growth of Top Wealth Shares". *Econometrica*, p. 61. URL: https://www.matthieugomez.com/files/topshares.pdf (cit. on pp. 9, 10, 15, 35).

Grey, D. R. (1994). "Regular Variation in the Tail Behaviour of Solutions of Random Difference Equations". *The Annals of Applied Probability* 4.1. DOI: 10.1214/aoap/1177005205. URL: https://projecteuclid.org/journals/annals-of-applied-probability/volume-4/issue-1/Regular-Variation-in-the-Tail-Behaviour-of-Solutions-of-Random/10.1214/aoap/1177005205.full (cit. on p. 8).

Grincevićius, A. K. (1975). "One limit distribution for a random walk on the line". *Lithuanian Mathematical Journal* 15.4, pp. 580–589. DOI: 10.1007/BF00969789. URL: http://link.springer.com/10.1007/BF00969789 (cit. on p. 8).

Gyöngy, I. (1986). "Mimicking the one-dimensional marginal distributions of processes having an ito differential". *Probability Theory and Related Fields* 71.4, pp. 501–516. DOI: 10.1007/BF00699039. URL: http://link.springer.com/10.1007/BF00699039 (cit. on pp. 4, 8, 14).

Haines, Michael R. (1998). "Estimated Life Tables for the United States, 1850–1910". *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 31.4, pp. 149–169. DOI: 10.1080/01615449809601197. URL: http://www.tandfonline.com/doi/abs/10.1080/01615449809601197 (cit. on p. 24).

Hubmer, Joachim, Per Krusell, and Anthony A Smith (2018). "A Comprehensive Quantitative Theory of the U.S. Wealth Distribution". URL: http://www.econ.yale.edu/smith/hks_v18.pdf (cit. on pp. 7, 8).

Huggett, Mark (1993). "The risk-free rate in heterogeneous-agent incomplete-insurance economies". *Journal of Economic Dynamics and Control* 17.5-6, pp. 953–969. DOI: 10.1016/0165-1889(93)90024-M. URL: https://linkinghub.elsevier.com/retrieve/pii/016518899390024M (cit. on p. 7).

Human Fertility Collection (2019). Max Planck Institute for Demographic Research in Rostock, Germany and Vienna Institute of Demography in Vienna, Austria. URL: https://www.fertilitydata.org/.

Human Fertility Database (2019). Max Planck Institute for Demographic Research in Rostock, Germany and Vienna Institute of Demography in Vienna, Austria. URL: https://www.humanfertility.org/.

Human Life Table Database (2019). Max Planck Institute for Demographic Research in Rostock, Germany, Department of Demography at University of California at Berkeley, USA and Institut d'études démographiques (INED) in Paris, France. URL: https://www.lifetable.de/.

Human Mortality Database (2019). Max Planck Institute for Demographic Research in Rostock, Germany, Department of Demography at University of California at Berkeley, USA. URL: https://www.mortality.org.

Jakobsen, Katrine et al. (2020). "Wealth Taxation and Wealth Accumulation: Theory and Evidence From Denmark*". *The Quarterly Journal of Economics* 135.1, pp. 329–388. DOI: 10.1093/qje/qjz032. URL: https://academic.oup.com/qje/article/135/1/329/5584349 (cit. on pp. 4, 5, 11, 12, 41, 45, 46).

Judd, Kenneth L. (1985). "Redistributive taxation in a simple perfect foresight model". *Journal of Public Economics* 28.1, pp. 59–83. DOI: 10.1016/0047-2727(85)90020-9. URL: https://linkinghub.elsevier.com/retrieve/pii/0047272785900209 (cit. on pp. 3, 10–12).

Kesten, Harry (1973). "Random difference equations and Renewal theory for products of random matrices". *Acta Mathematica* 131.0, pp. 207–248. DOI: 10.1007/BF02392040. URL: http://projecteuclid.org/euclid.acta/1485889791 (cit. on pp. 7, 8).

Kolmogorov, A. (1931). "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung". *Mathematische Annalen* 104.1, pp. 415–458. DOI: 10.1007/BF01457949. URL: http://link.springer.com/10.1007/BF01457949 (cit. on pp. 4, 8, 14, 48).

Krusell, Per and Anthony A. Smith Jr. (1998). "Income and Wealth Heterogeneity in the Macroeconomy". *Journal of Political Economy* 106.5, pp. 867–896. DOI: 10.1086/250034. URL: https://www.journals.uchicago.edu/doi/10.1086/250034 (cit. on p. 7).

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins (2020). "Income and Wealth Inequality in America, 1949–2016". *Journal of Political Economy* 128.9, pp. 3469–3519. DOI: 10.1086/708815. URL: https://www.journals.uchicago.edu/doi/10.1086/708815 (cit. on pp. 2, 10, 32).

Londoño-Vélez, Juliana and Javier Ávila-Mahecha (2021). "Enforcing Wealth Taxes in the Developing World: Quasi-Experimental Evidence from Colombia". *American Economic Review: Insights* 3.2, pp. 131–148. DOI: 10.1257/aeri.20200319. URL: https://pubs.aeaweb.org/doi/10.1257/aeri.20200319 (cit. on pp. 5, 11, 41, 45, 46).

Menchik, Paul L. (1980). "Primogeniture, Equal Sharing, and the U.S. Distribution of Wealth". *The Quarterly Journal of Economics* 94.2, p. 299. DOI: 10.2307/1884542. URL: https://academic.oup.com/qje/article-lookup/doi/10.2307/1884542 (cit. on p. 25).

Piketty, Thomas (2014). *Capital in the twenty-first century*. Trans. by Arthur Goldhammer. Cambridge, Massachusetts London: The Belknap Press of Harvard University Press (cit. on p. 40).

Piketty, Thomas and Emmanuel Saez (2013a). "A Theory of Optimal Inheritance Taxation". *Econometrica* 81.5, pp. 1851–1886. DOI: 10.3982/ECTA10712. URL: http://doi.wiley.com/10.3982/ECTA10712 (cit. on pp. 4, 11, 41).

— (2013b). "Optimal Labor Income Taxation". *Handbook of Public Economics*. Vol. 5. Elsevier, pp. 391–474. DOI: 10.1016/B978-0-444-53759-1.00007-8. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780444537591000078 (cit. on pp. 3, 11).

Piketty, Thomas and Gabriel Zucman (2015). "Wealth and Inheritance in the Long Run". *Handbook of Income Distribution*. Vol. 2. Elsevier, pp. 1303–1368. DOI: 10.1016/B978-0-444-59429-7.00016-9. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780444594297000169 (cit. on p. 40).

Quadrini, Vincenzo (2000). "Entrepreneurship, Saving, and Social Mobility". *Review of Economic Dynamics* 3.1, pp. 1–40. DOI: 10.1006/redy.1999.0077. URL: https://linkinghub.elsevier.com/retrieve/pii/S1094202599900777 (cit. on pp. 7, 30, 32).

Ring, Marius Alexander Kalleberg (2020). "Wealth Taxation and Household Saving: Evidence from Assessment Discontinuities in Norway". URL: https://www.ssrn.com/abstract=3716257 (cit. on pp. 5, 41, 44, 45).

Robbins, Jacob A (2018). "Capital Gains and the Distribution of Income on the United States". URL: https://users.nber.org/~robbinsj/jr_inequ_jmp.pdf (cit. on p. 24).

Ruggles, Steven et al. (2022). *IPUMS USA: Version 12.0*. Version Number: 12.0 Type: dataset. DOI: 10.18128/D010.V12.0. URL: https://usa.ipums.org (cit. on p. 25).

Saez, Emmanuel (2001). "Using Elasticities to Derive Optimal Income Tax Rates". *The Review of Economic Studies* 68.1, pp. 205–229. DOI: `10.1111/1467-937X.00166`. URL: `https://academic.oup.com/restud/article-lookup/doi/10.1111/1467-937X.00166` (cit. on p. 11).

— (2013). "Optimal progressive capital income taxes in the infinite horizon model". *Journal of Public Economics* 97, pp. 61–74. DOI: `10.1016/j.jpubeco.2012.09.002`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0047272712001016` (cit. on p. 10).

Saez, Emmanuel and Stefanie Stantcheva (2016). "Generalized Social Marginal Welfare Weights for Optimal Tax Theory". *American Economic Review* 106.1, pp. 24–45. DOI: `10.1257/aer.20141362`. URL: `https://pubs.aeaweb.org/doi/10.1257/aer.20141362` (cit. on p. 50).

— (2018). "A simpler theory of optimal capital taxation". *Journal of Public Economics* 162, pp. 120–142. DOI: `10.1016/j.jpubeco.2017.10.004`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0047272717301688` (cit. on pp. 3, 4, 11, 41).

Saez, Emmanuel and Gabriel Zucman (2016). "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data". *The Quarterly Journal of Economics* 131.2, pp. 519–578. DOI: `10.1093/qje/qjw004`. URL: `https://doi.org/10.1093/qje/qjw004` (cit. on pp. 2, 10, 24, 32).

— (2019). "Progressive Wealth Taxation". *Brookings Papers on Economic Activity* Fall, pp. 437–533 (cit. on pp. 4, 6, 12, 13, 43, 46, 47).

— (2020a). "The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts". *Journal of Economic Perspectives* 34.4, pp. 3–26. DOI: `10.1257/jep.34.4.3`. URL: `https://pubs.aeaweb.org/doi/10.1257/jep.34.4.3` (cit. on pp. 2, 3, 24, 27).

— (2020b). *Trends in US Income and Wealth Inequality: Revising After the Revisionists*. Tech. rep. 27921. Cambridge, MA: National Bureau of Economic Research. DOI: `10.3386/w27921`. URL: `http://www.nber.org/papers/w27921.pdf` (cit. on pp. 2, 40).

Saporta, Benoîte de and Jian-Feng Yao (2004). "Tail of a linear diffusion with Markov switching". *Comptes Rendus Mathematique* 339.9, pp. 643–646. DOI: `10.1016/j.crma.2004.09.022`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1631073X0400439X` (cit. on p. 8).

Seim, David (2017). "Behavioral Responses to Wealth Taxes: Evidence from Sweden". *American Economic Journal: Economic Policy* 9.4, pp. 395–421. DOI: `10.1257/pol.20150290`. URL: `https://pubs.aeaweb.org/doi/10.1257/pol.20150290` (cit. on pp. 5, 11, 41, 45, 46).

Smith, Matthew, Owen Zidar, and Eric Zwick (2022). "Top Wealth in America: New Estimates under Heterogeneous Returns". *The Quarterly Journal of Economics*, qjac033. DOI: `10.1093/qje/qjac033`. URL: `https://academic.oup.com/qje/advance-article/doi/10.1093/qje/qjac033/6678447` (cit. on p. 2).

Straub, Ludwig and Iván Werning (2020). "Positive Long-Run Capital Taxation: Chamley-Judd Revisited". *American Economic Review* 110.1, pp. 86–119. DOI: `10.1257/aer.20150210`. URL: `https://pubs.aeaweb.org/doi/10.1257/aer.20150210` (cit. on pp. 3, 11).

United Nations (2019). *World Population Prospects, 2019 Revision*. OCLC: 1142478963. URL: `https://population.un.org/wpp/Publications/` (cit. on pp. 25, 33, 38, 39).

Vervaat, Wim (1979). "On a stochastic difference equation and a representation of non–negative infinitely divisible random variables". *Advances in Applied Probability* 11.4, pp. 750–783. DOI: `10.2307/1426858`. URL: `https://www.cambridge.org/core/product/identifier/S0001867800033024/type/journal_article` (cit. on p. 8).

Xavier, Inês (2021). "Wealth Inequality in the US: the Role of Heterogeneous Returns". URL: `https://www.ssrn.com/abstract=3915439` (cit. on p. 7).

Zoutman, Floris (2018). "The Elasticity of Taxable Wealth: Evidence from the Netherlands" (cit. on pp. 5, 11, 41, 45).

Appendix to "Uncovering the Dynamics of the Wealth Distribution"

Thomas Blanchet

This version: September 30, 2022

☞ GET LATEST VERSION

Contents

A Proofs Omitted from Main Text

A.1 Proposition 1

The proof of proposition 1 is a straightforward consequence of Gyöngy's (1986) theorem and of the rules of Itô calculus. Let us start by providing the full statement of Gyöngy's (1986) theorem.

Theorem (Gyöngy, 1986). *Let X_t be a n-dimensional stochastic process satisfying:*

$$dX_t = \alpha_t \, dt + \beta_t \, dB_t$$

where α_t and β_t are bounded and nonanticipative $n \times 1$ and $n \times m$ stochastic processes, respectively, $\beta_t \beta_t'$ is uniformly positive definite, and B_t is a m-dimensional Wiener process. Then there is a Markov process Y_t satisfying:

$$dY_t = a_t(Y_t) \, dt + b_t(Y_t) \, dB_t$$

where X_t and Y_t have the same marginal distributions for each t. We can construct Y_t by setting:

$$a_t(y) = \mathbb{E}[\alpha_t | X_t = y] \qquad\qquad b_t(y) = \mathbb{E}[\beta_t \beta_t' | X_t = y]^{1/2}$$

Define $\mu_{it} = (y_{it} + (r_{it} - g_t)w_{it} - c_{it})$ and $\sigma_{it} = \left(v_{it}^2 + \phi_{it}^2 w_{it}^2 + \gamma_{it}^2\right)^{1/2}$. Recall that the wealth of person i evolves according to $dw_{it} = \mu_{it} \, dt + \sigma_{it} \, dB_{it}$. Using Gyöngy's (1986) theorem, we can write $dw_{it} = \mu_t(w_{it}) \, dt + \sigma_t(w_{it}) \, dB_{it}$ where $\mu_t(w) = \mathbb{E}[\mu_{it} | w_{it} = w]$ and $\sigma_t^2(w) = \mathbb{E}[\sigma_{it}^2 | w_{it} = w]$. To simplify notations, consider all expectations conditional on $w_{it} = w$. Write $z_{it} = \mu_{it} \, dt + \sigma_{it} \, dB_{it}$. For the drift term, we have directly $\mathbb{E}[z_{it}] = \mathbb{E}[\mu_{it}] \, dt = \mu_t(w) \, dt$. For the diffusion term:

$$
\begin{aligned}
\mathrm{Var}(z_{it}) &= \mathbb{E}[(z_{it} - \mu_t(w) \, dt)^2] \\
&= \mathbb{E}[(\mu_{it} \, dt - \mu_t(w) \, dt + \sigma_{it} \, dB_{it})^2] \\
&= \underbrace{\mathbb{E}[(\mu_{it} \, dt - \mu_t(w) \, dt)^2]}_{=\, 0 \text{ because } (dt)^2 = 0} + \underbrace{\mathbb{E}[\sigma_{it}^2 \, dB_{it}^2]}_{\substack{=\, \mathbb{E}[\sigma_{it}^2] \, dt \\ \text{because } dB_{it}^2 = dt}} + \underbrace{2\, \mathbb{E}[\sigma_{it}(\mu_{it} - \mu_t(w)) \, dB_{it} \, dt]}_{=\, 0 \text{ because } dB_{it} \, dt = 0}
\end{aligned}
$$

Therefore, $\mu_t(w) \, dt = \mathbb{E}[z_{it}]$ and $\sigma_t^2(w) \, dt = \mathrm{Var}(z_{it})$. $\qquad\square$

Note that the proof rests on an approximation made possible by this paper's continuous-time formalism. Namely, it allows us to identify the average variance of individual shocks $\mathbb{E}[\sigma_{it}^2]$ with the overall variance of wealth growth $\mathrm{Var}(z_{it})$. This is because over an interval of time dt, the part of the variance that is explained by w is of order $(dt)^2$ while the rest of the variance is

of order dt, and so as $dt \to 0$, the former is negligible compared to the latter.

A.2 Proposition 2

Recall that wealth evolves according to $dw_{it} = (\mu(w) - \tau(w))\,dt + \sigma_t(w)\,dB_{it}$ where $\tau(w)$ is the wealth tax. Let f^* be the steady-state density of wealth with the wealth tax. It has to obey the Kolmogorov forward equation with the time derivative terms set to zero:

$$0 = -\partial_w[(\mu(w) - \tau(w))f^*(w)] + \frac{1}{2}\partial_w^2[\sigma^2(w)f^*(w)]$$

Solving this differential equation, we can write:

$$f^*(w) \propto \exp\left\{-2\int_0^w \frac{\sigma(s)\sigma'(s) - \mu(s)}{\sigma^2(s)}\,ds\right\}\exp\left\{-\int_0^w \frac{2\tau(s)}{\sigma^2(s)}\,ds\right\}$$

Note that the steady-state density f without the wealth tax corresponds to the case $\tau(w) \equiv 0$, and therefore $f^*(w) \propto f(w)\theta(w)$ where $\theta(w) \equiv \exp\left\{-\int_{w_0}^w \frac{2\tau(s)}{\sigma^2(s)}\,ds\right\}$, which gives the first part of the result. For the second part, assume that $\tau(w) = \tau(w - w_0)_+$ and that $\sigma(w) = \sigma w$ for $w > w_0$. Then $\theta(w)$ simplifies to:

$$\theta(w) = \exp\left\{-\frac{2\tau}{\sigma^2}\int_{w_0}^w \frac{(s - w_0)_+}{s^2}\,ds\right\} = \exp\left\{-\frac{2\tau}{\sigma^2}\left(\frac{w_0}{w} - 1\right)\right\}\left(\frac{w}{w_0}\right)^{-2\tau/\sigma^2}$$

which proves the second part of the result. $\qquad\square$

B Detailed Data Construction

B.1 Income and Wealth

For income and wealth data, I primarily rely on the DINA microdata from Piketty, Saez, and Zucman (2018), which are based on the Internal Revenue Service (IRS) individual public-use microdata files, and which have been continuously updated to reflect methodological changes and improvements (Saez and Zucman, 2020a; Saez and Zucman, 2020b). These files are annual (except for 1963 and 1965) since 1962. Each observation corresponds to an adult individual (20 or older), and each variable corresponds to an item of the national accounts distributed to the entire adult population. These files distribute the entirety of the income and wealth of the United States.

This data has several advantages. It provides distributional estimates that are consistent with macroeconomic aggregates. It has rather large samples (from about 175 000 in the 1960s to about 300 000 today), with oversampling of the richest. And because it is based on tax data, it captures the top tail of the distribution well. But it does have some drawbacks. First, it has limited socio-demographic information: in particular, age information is only available in the form of very broad age groups. Second, the data does not include capital gains because they are not part of national income as defined by the national accounts. For these reasons, I make some adjustments and imputations to these data, using the SCF and national accounts.

I use post-tax national income as my income concept of reference. It corresponds to income after all taxes and transfers. It also distributes government expenditures and the income of the corporate sector to individuals so as to sum up to net national income.
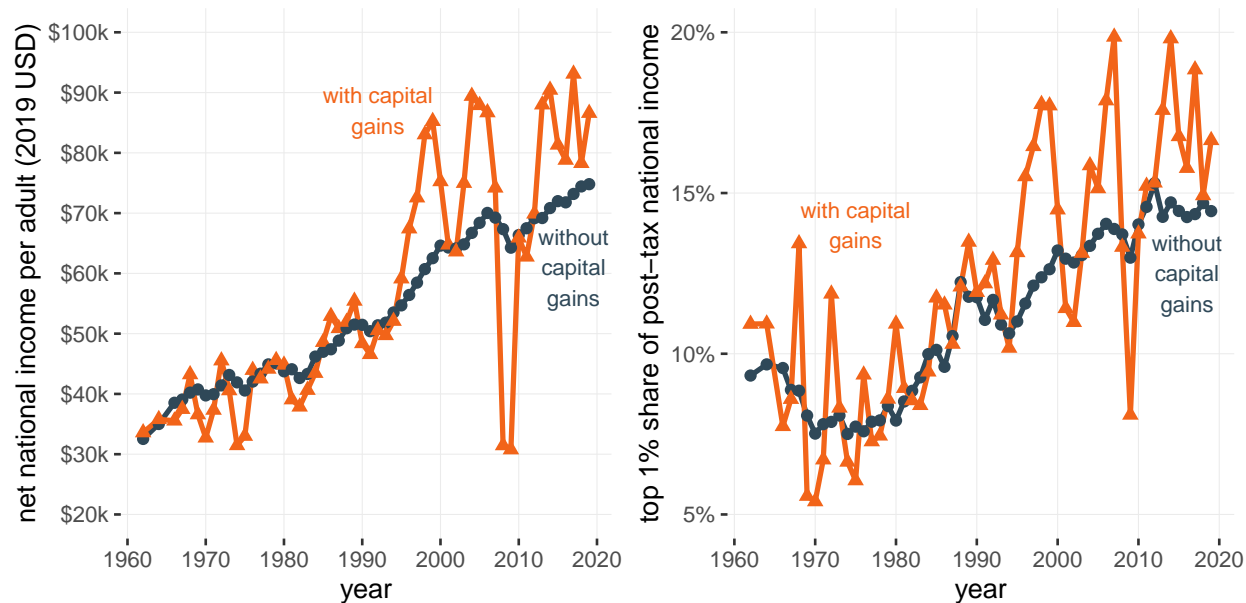
(a) Net National Income and Capital Gains

(b) Top 1% Post-tax National Income Share, with and without Capital Gains

Source: Author's computations using the DINA microdata and table TSD1 (online appendix) from Piketty, Saez, and Zucman (2018). *Note:* The unit of analysis is the adult individual (20 or older). Income is split equally between members of couples. Capital gains are estimated assuming a constant rate of capital gains by asset type.

Figure B.1: The Impact of Capital Gains on National Income and Its Distribution

Capital Gains We can measure capital gains when they accrue to individuals or when they are realized. For our purpose, accrued capital gains are more useful than realized ones, because they are the ones that reconcile changes in the value of the balance sheet with national income

and savings. On the other hand, whether a capital gain is realized now or later is the result of various tax and economic incentives that are not relevant here and do not correspond to any meaningful economic aggregate.

The DINA data only records taxable capital gains, which is essentially a measure of realized capital gains. These are a poor proxy for accrued capital gains (Alstadsæter et al., 2016). Instead, I estimate them individually using the capitalization approach of Robbins (2018). I retrieve the capital gains rate by year and asset type from the national accounts (Piketty, Saez, and Zucman, 2018, table TSD1 in appendix). Then, I assume that for a given asset type, everyone gets the same rate of capital gains. By construction, these micro-level capital gains estimates are consistent with macro totals. Their distribution follows the logic of the Saez and Zucman (2016) capitalization method.[1] Robbins (2018) provides a thorough discussion of why that measure is more appropriate to analyze the impact of asset price changes on inequality and the economy.

National income including capital gains can be quite volatile (figure B.1a), but on average their inclusion matters on several fronts. Robbins (2018) shows that their inclusion overturns certain stylized facts about the United States economy (such as the long-run decline of saving rates) and strengthens others (such as the rising capital share and increase in income inequality). As shown in figure B.1b, capital gains were dampening the top 1% share of post-tax national income during most of the 1970s, but since then, they have consistently increased it.

Wealth by Age The age information in the DINA data is very limited, so I cannot use it. Instead, I import it from the SCF and demographic estimates using constrained statistical matching. I calculate the rank in the wealth distribution in both the DINA and the SCF data, and the rank in the age distribution by sex and household type (single or couple) in the SCF data. Then, I match the DINA observations one by one to SCF observations based on their wealth rank to give them a rank in the age distribution.[2] Finally, I use the population structure from the demographic data to attribute an age to every DINA observation. By construction, the method preserves the wealth distribution in DINA, the population by age and sex from demographic sources, and the copula between wealth and age from the SCF.

---

[1]Although the income measure in the DINA data does not include capital gains, it does distribute income from the corporate sector to the owners of capital, which partly accounts for changes in asset prices. My capital gains measure is net of retained earnings, so there is no double counting.

[2]Note that both datasets are weighted, so that observations end up being duplicated and partially matched to one another. When the samples contain M and N observations, respectively, the resulting dataset contains at most $M + N - 1$ observations.

B.2 Demography

I compute the entire demography of the United States from 1850 to 2100. Although the income and wealth data does not start until 1962, the model requires demographic data that starts much earlier. Indeed, I need to simulate how wealth gets transmitted from one generation to the next. Therefore, if a supercentenarian dies in the 1960s, I have to be able to simulate their entire life history to know how many live children they have and how old they are. For all years and all ages, I estimate data on the population structure by age and sex, mortality (i.e., life tables), fertility (for both sexes), and intergenerational ties (age and sex of children). Sometimes, data is only available by age group (e.g., of five years) or a subset of years (e.g., every ten years). Whenever necessary, I interpolate estimates with a monotonic cubic spline (Fritsch and Carlson, 1980) to get data for every single year and age.

Population by Age and Sex   Before 1900, I directly estimate the population pyramid using the decennial census microdata from the IPUMS USA database (Ruggles et al., 2022). From 1900 to 1932, I use the National Intercensal Tables from the United States Census Bureau. From 1933 to 2016, I use population estimates from the Human Mortality Database.[3] After 2016, I use the projections from the World Population Prospects (United Nations, 2019).

Life Tables   Before 1900, I use the historical life tables from Haines (1998). From 1900 to 1932, I use the Human Life Table Database, and from 1933 to 2016, life tables from the Human Mortality Database. After 2016, I rely on projections from the World Population Prospects (United Nations, 2019). All tables are broken down by sex.

Age-Specific Fertility Rates by Birth Order   I estimate age-specific fertility rates by birth order for both sexes. For women, they are directly available from 1933 to 2016 from the Human Fertility Database. From 1917 to 1932, I use data from the Human Fertility Collection. That same source provides fertility rates until going back to 1895–1899, but without the breakdown by birth order. Therefore, before 1917, I assume that the birth order composition remains constant. Before 1895, there is no age-specific data available, so I use the data on the total fertility rate and rescale the age profile from 1895–1899 to that value.[4]

Unlike female fertility rates, male fertility rates are not a standard demographic indicator, so they are not directly available from any source. To estimate them, I combine the age-specific female fertility rates with the joint distribution of the age of opposite-sex couples since 1850

---

[3]See https://www.mortality.org/hmd/USA/DOCS/ref.pdf for detailed primary sources.

[4]See Gapminder: https://www.gapminder.org/news/children-per-women-since-1800-in-gapminder-world/

calculated using the decennial census microdata from the IPUMS USA database (Ruggles et al., 2022).

Age and Sex of Children I simulate the distribution of the number, age, and sex of living children for each year after 1962 (when income and wealth data starts). I do this for every age and both sexes, which allows me to realistically model how wealth gets transmitted from one generation to the next. To that end, I combine all the data above. I make every person have children randomly over their past lifetime according to year, age, and sex-specific fertility rate. Because I have the breakdown by birth order, I can take into account how the decision to have another child depends on the number of children that one already has. Then, I make each child go through life and die randomly according to their year, age, and sex-specific mortality rate. As a result, I can tie every individual in the database to fictitious descendants that are, on average, representative of the true composition of descendants.

B.3 Inheritance and Estate Tax

Part of the inheritance process is determined by the demographics and the distribution of wealth, while other parts have to be modeled separately. I assume that people die at random, conditional on their age and sex, so the distribution of inheritances corresponds to the distribution of wealth, weighted by mortality rates. I then assume that the wealth of decedents is either redistributed to their spouse (if any) or to their descendants (if they have no living spouse), after payment of the estate tax. The demography gives the age and sex of decedents (see section B.2). I assume that inheritance is split equally between children, as is the norm in the United States (Menchik, 1980).

While the demographic aspect of inheritance is endogenously determined by demography, I still need to model separately how wealth gets distributed for a given age and sex. This facet of the problem captures intergenerational wealth mobility, in the sense that wealthier people might also have wealthier parents and thus inherit more. There are two aspects to this question: the extensive margin (how likely are you to receive an inheritance in a given year?) and the intensive margin (how much inheritance do you receive?) To address this question, I use data from the SCF, which has been recording inheritance consistently since 1989. Because the probability of receiving an inheritance in a given year is very low overall (about half a percent, see figure B.2a), I have to pool all the 1989–2016 waves to get sufficient sample sizes.
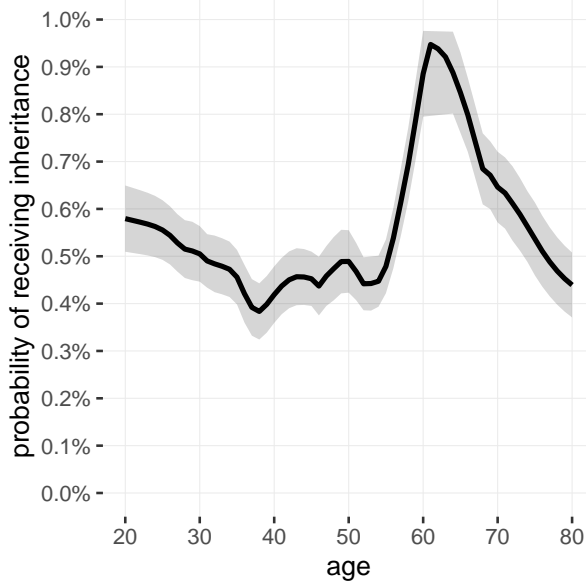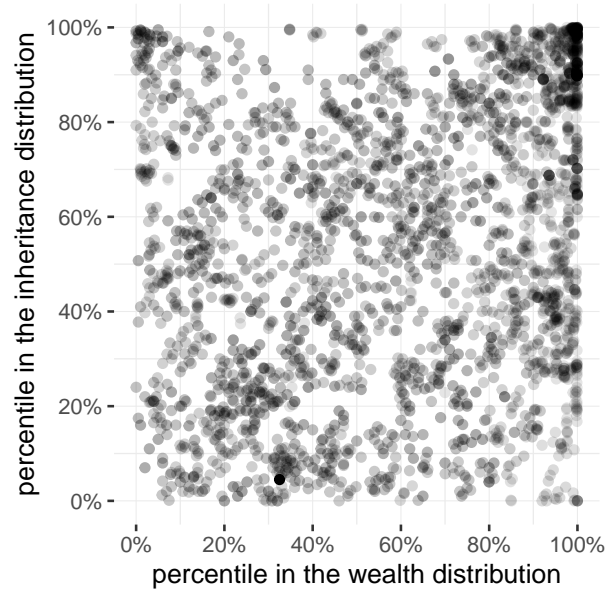
(a) Probability of Receiving Inheritance by Age
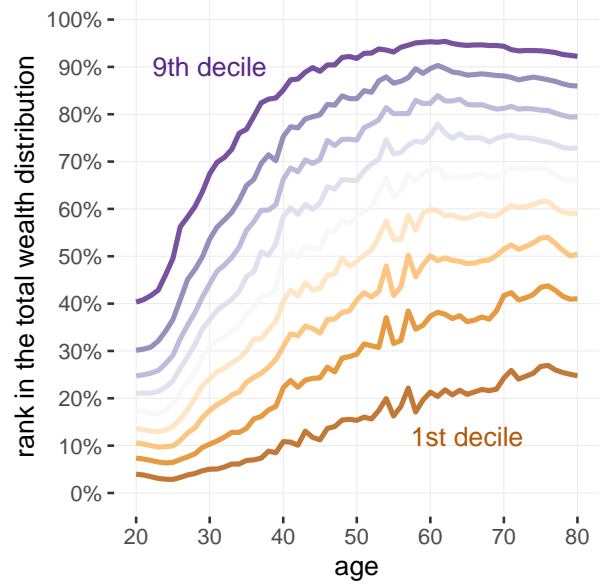
(b) Rank in the Wealth Distribution by Age

(c) Relative Probability of Inheritance
(Conditional on Age)

(d) Ranks in the Wealth and in the Inheritance
Distribution (Conditional on Age)

Source: Own computations using the Survey of Consumer Finances (SCF) (1989–2019). Gray ribbons correspond to the 95% confidence intervals. In figure B.2d, opacity is proportional to the weight of observations.

Figure B.2: Modeling of Inheritance

8

Extensive Margin   Let $D_i = 1$ if individual i receives inheritance, and $D_i = 0$ otherwise. Let A_i be their age and W_i their wealth. Assume that:

$$\mathbb{P}\{D_i = 1 | A_i = a, W_i = w\} = \mathbb{P}\{D_i = 1 | A_i = a\}\phi(F_{A_i=a}(w)) \tag{9}$$

where $F_{A_i=a}$ is the CDF of wealth conditional on age, and $\int_0^1 \phi(r)\,\mathrm{d}r = 1$. By construction, the expected value of the right-hand side of (9) conditional on age is equal to $\mathbb{P}\{D_i = 1 | A_i = a\}$ so that the specification makes probabilistic sense.[5] Note that $F_{A_i=a}(w)$ is the rank of w in the wealth distribution (conditional on age), which is how we can make the formula (9) consistent regardless of the shape of the wealth distribution.

The value of $\mathbb{P}\{D_i = 1 | A_i = a\}$ is determined by demography, so we only need to estimate ϕ. I start by calculating a rank in the wealth distribution conditional on age by running a nonparametric quantile regression of wealth on age for every percentile (see figure B.2b). I then regress the dummy D_i for having received inheritance on that rank, multiplied by $\mathbb{P}\{D_i = 1 | A_i = a\}$. I use ordinary least squares (OLS) and a cubic polynomial with coefficients constrained so that its integral over $[0, 1]$ equals one (see figure B.2c). As we can see, even after partialling out the effect of age, wealthier people still experience a higher probability of receiving an inheritance. I use that polynomial as my estimate of ϕ.

Intensive Margin   I account for the intensive margin by modeling the joint distribution of the ranks in the wealth distribution and the inheritance distribution (i.e., the copula), conditional on age and on having received an inheritance. I take the subsample of inheritance receivers and calculate their rank in the wealth and inheritance distribution using a nonparametric quantile regression as I did for the extensive margin.

The dependence between the two ranks is weak but significant (see figure B.2d): their Kendall's τ is equal to 8.3%. I represent this dependency parametrically using a bivariate copula. I select the most appropriate model out of a large family of 15 single-parameter copulas by finding the best fit according to the Akaike information criterion (AIC), which is the Joe copula.[67] I estimate its parameter so as to match the empirical value for Kendall's τ.

Estate Tax   I account for the federal estate tax using the complete estate tax schedule and exemption amount for each year. The top marginal estate tax rate has followed a clear inverted

---

[5]It is the direct result of a change of variable $r = F_{A_i=a}(w)$ and using the fact that $\frac{\partial}{\partial w}F_{A_i=a}(w) = f_{A_i=a}(w)$, so that $\int_{-\infty}^{+\infty} \phi(F_{A_i=a}(w))f_{A_i=a}(w)\,\mathrm{d}w = \int_0^1 \phi(r)\,\mathrm{d}r = 1$.

[6]The list of copulas includes the Gaussian copula, Student's t copula, the Clayton copula, the Gumbel copula, the Frank copula, the Joe copula, and rotated versions of these copulas.

[7]The Joe copula has the parametric form $C_\theta(u, v) = 1 - \left[(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right]^{1/\theta}$.
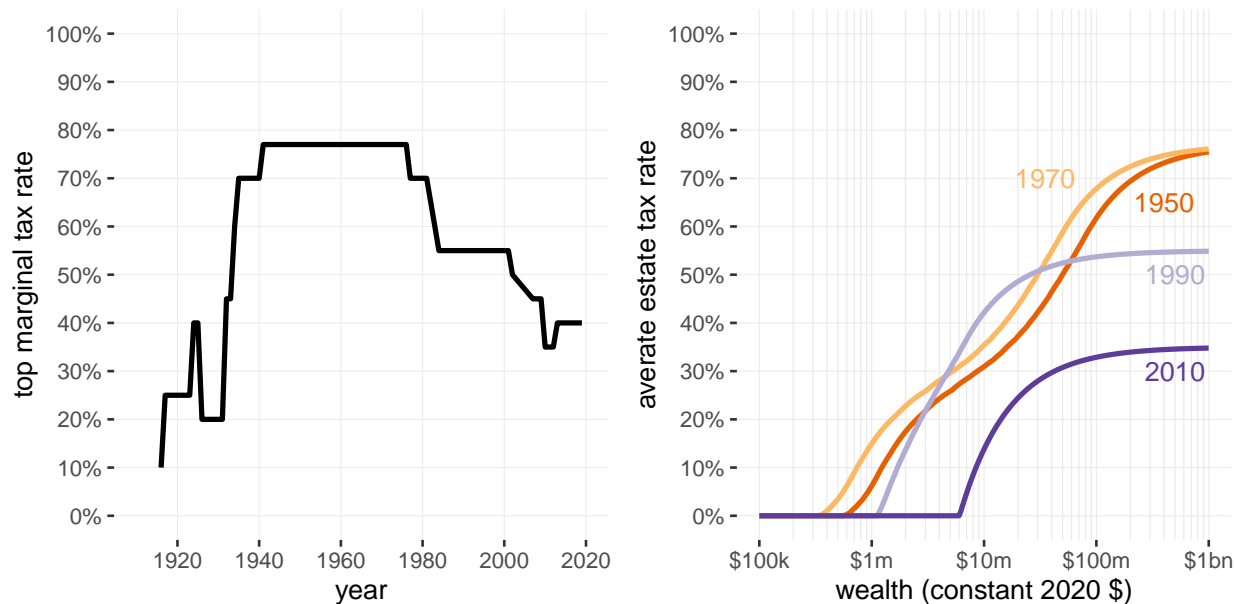
(a) Top Marginal Estate Tax since 1916   (b) Average Estate Tax Rate, by Wealth

Source: Own computation using the statutory schedules of the federal estate tax.

Figure B.3: Estate Tax

U-shaped pattern over the 20th century (figure B.3a), having been reduced by half since its mid-century peak. However, the changes to the overall progressivity of the estate tax are more ambiguous (figure B.3b). While the top marginal tax rate was very high in the 1950s, the top bracket did not kick in until extremely high levels of wealth. The 1980s reforms significantly reduced the top tax rate and increased the exemption amount so that by 1990, the very top and the upper middle of the wealth distribution were facing lower average tax rates. But $10m estates were actually facing slightly higher average tax rates. By now, however, the estate has been lowered so much that its profile is unambiguously less progressive than in the 1950s.

B.4   Marriages, Divorces and Assortative Mating

Aggregate Marriages and Divorce Rates by Age   I calculate population pyramids disaggregated by marital status using census microdata from the IPUMS USA database (Ruggles et al., 2022). I use this data to estimate the rate at which people get married and divorced by age and sex. This is possible assuming that mortality rates are the same regardless of marital status, and assuming that marriage rates are the same for people that are divorced are people that

have never been married. Under these conditions, we have for each sex:

$$(\text{marriage rate})_{a,t} = 1 - \frac{(\text{fraction never married})_{a+1,t+1}}{(\text{fraction never married})_{a,t}}$$

$$(\text{divorce rate})_{a,t} = \frac{(\text{fraction divorced})_{a+1,t+1} - [1 - (\text{marriage rate})_{a,t}] \times (\text{fraction divorced})_{a,t}}{(\text{fraction married})_{a,t}}$$

where a is age, and t is the year. This estimate being quite noisy, I winsorize the bottom 10% and the top 10% of values, and then apply a moving average with a ten-year age and year window. This gives the results of Figure B.4a. For the simulation, we anchor these numbers to the aggregate data on the crude rate of marriage and divorce from the NVSS (Center for Disease Control, 2022) (Figure B.4b).

Assortative Mating   I calculate the extent of assortative mating by estimating the copula between the rank of each spouse in the wealth distribution, conditional on age, at the time they get married. To that end, I use data from the SIPP, which collected data on wealth between 2013 and 2016. I select all people who just got married (i.e., people single in year $t-1$ but married in year t) and fit a parametric copula to the joint rank of each spouse in the wealth distribution, conditional on age (Figure B.4c). As in Section B.3, I select the best copula out of a large family of 15 single-parameter copulas according to the AIC, which is the Frank copula.[8] I estimate its parameter so as to match the empirical Kendall's τ, which is equal to 28%.

I also estimate the effect of assortative mating at divorce by looking at how wealth is divided between both spouses right before they separate. I also use the SIPP, select people who are married in year t but separated in year $t+1$, and extract the distribution of the share of the household's wealth that each spouse has.

C   Detailed Estimations and Results

Overall, the estimation procedure in this paper only involves fitting a series of lines. In practice, there are several methods for fitting lines, and this paper uses the approach that makes the most sense given the nature of the problem. This section describes the procedure used and discusses the sensitivity of estimates to the various hyperparameters.

---

[8]The formula of the Frank copula is $C(u,v) = -\frac{1}{\theta} \log\left[1 + \frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right]$.
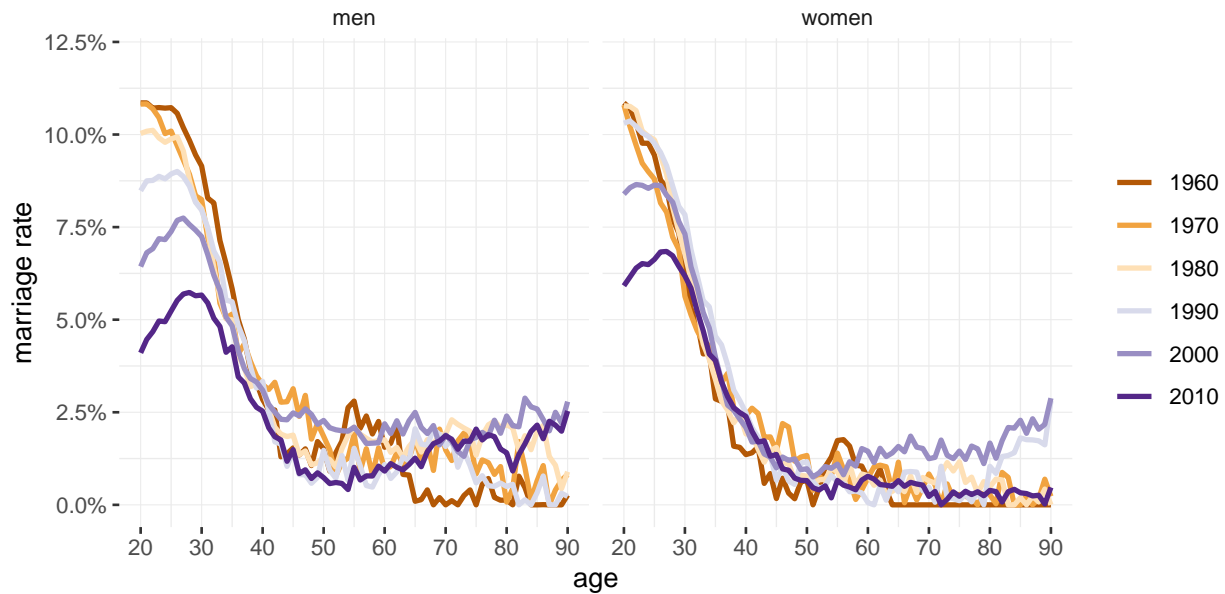
(a) Rates of Marriage by Age and Sex
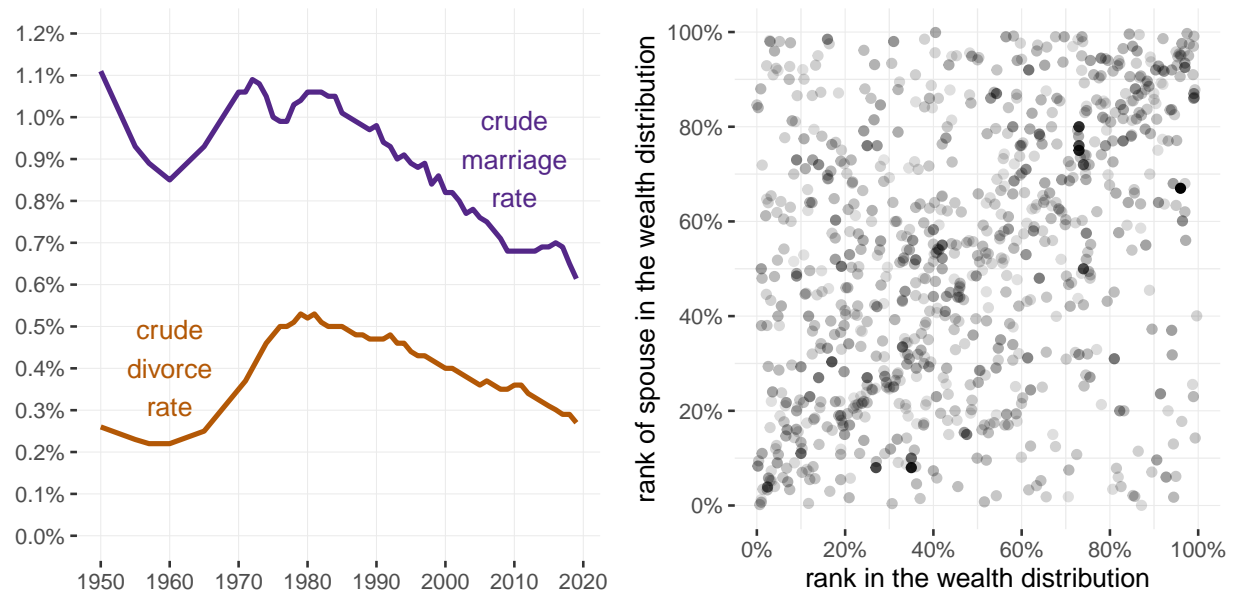
(b) Crude Aggregate Rates of Marriage and Divorce

(c) Ranks in the Wealth Distribution of Both Spouses at Marriage

Source: Own computations using the census microdata from IPUMS (Ruggles et al., 2022), National Vital Statistics System (NVSS) (Center for Disease Control, 2022), and Survey of Income and Program Participation (SIPP) (2013–2016). In figure B.4c, opacity is proportional to the weight of observations.

Figure B.4: Modeling of Marriage and Divorce

12

C.1 Estimation Procedure

Recall that the estimation procedure in this paper requires fitting a linear relationship of the form given by equation (5), which I will simplify as follows to emphasize the line-fitting aspect:

$$\forall t \in \{1, \ldots T\} \qquad y_{it} = a + b x_{it} \tag{10}$$

for each wealth bin $i \in \{1, \ldots, n\}$. The data are y_{it} and x_{it}. The parameters are a_i and b_i. While fitting this equation using OLS is an option, it is not a natural choice here for two reasons. First, that this equation does not hold perfectly in practice results from error terms that can be attributed simultaneously to y_{it} and x_{it}. In that sense, the setting is closer to an error-in-variable model, for which OLS are known to be biased. Second, and relatedly, there is an invariance in equation (10) that OLS breaks. Unlike usual econometric regressions, there is no obvious distinction between the outcome variable and the explanatory variable. The equation would make as much sense if the roles of x_{it} and y_{it} were reversed (and the meaning of the parameters a_i and b_i changed accordingly). But with OLS, regressing y on x yields a different result than regressing x on y.

The solution is to use a Deming (1943) regression, a generalization of orthogonal regression, which is a standard way of estimating error-in-variable models. The Deming (1943) estimator minimizes:

$$\min_{a,b,x_1^*,\ldots,x_T^*} \sum_{t=1}^{T} (y_t - a - b x_t^*)^2 + \delta (x_t - x_t^*)^2$$

where δ is an hyperparameter, defined separately, that captures the relative variance of the measurement error on y_{it} and x_{it}. When $\delta = 0$ the estimator collapses to OLS, and when $\delta = +\infty$ it collapses to OLS as well, but with the left-hand side and the right-hand side reversed. This estimator has an analytical solution, which is:

$$b_i = \frac{s_{yy} - \delta s_{xx} + \sqrt{(s_{yy} - \delta s_{xx})^2 + 4\delta s_{xy}^2}}{2 s_{xy}} \qquad\qquad a_i = \overline{y} - b\overline{x}$$

where:

$$\overline{x} = \tfrac{1}{T}\sum x_{it} \qquad\qquad \overline{y} = \tfrac{1}{T}\sum y_{it}$$

$$s_{xx} = \tfrac{1}{T}\sum_t (x_{it} - \overline{x})^2 \qquad s_{xy} = \tfrac{1}{T}\sum_t (x_{it} - \overline{x})(y_{it} - \overline{y}) \qquad s_{yy} = \tfrac{1}{T}\sum_t (y_{it} - \overline{y})^2$$

I estimate a variance of the measurement error on x_{it} and y_{it}, by taking a five-year moving average of these values (in each wealth bin), and calculate the variance of the difference

13

between the values and their moving average. This allows me to estimate δ. This is a rough estimate, so I check that results are robust to it in Section C.4. The choice of δ turns out to have limited impact, because the line actually fits the data closely: error terms all smalls, and therefore assuming that they are due to x_{it} or y_{it} makes little difference.

C.2   Estimation of Standard Errors

This paper's setting requires a nonstandard procedure for estimating standard errors for two reasons. First, because I use a Deming (1943) regression, which imply different standard errors than OLS. Second, because the error terms are not independent: there are correlated both across time and across wealth bins. To deal with these issues, I use a parametric bootstrap procedure that efficiently simulates error terms that reproduce the key features of the data.

First, I estimate an error term for each observation. Note that in geometrical terms, the Deming (1943) performs an oblique projection of observations onto the fitted line, according to a direction set by the hyperparameter δ. This projection is associated with the following scalar product in $\mathbb{R}^2$:

$$\langle (x_1, y_1), (x_2, y_2) \rangle_\delta \mapsto \delta x_1 x_2 + y_1 y_2$$

Given this behavior, the error terms that the Deming (1943) regression estimates for x_{it} and y_{it} are perfectly correlated, so it would not make sense to model them separately. Instead, I define a single error term e_{it} for every observation as the scalar product between (i) the unit vector orthogonal to the fitted line, and (ii) the residual vector $(x_{it} - x_{it}^*, y_{it} - y_{it}^*)$. The vector that is normal to the fitted line is $(b, -\delta)/\sqrt{\delta(b^2 + \delta)}$, where b is the slope of the fitted line, hence:

$$e_{it} = \frac{\langle (b, -\delta), (x_{it} - x_{it}^*, y_{it} - y_{it}^*) \rangle_\delta}{\sqrt{\delta(b^2 + \delta)}} = \sqrt{\frac{\delta}{b^2 + \delta}} [b(x_{it} - x_{it}^*) - (y_{it} - y_{it}^*)]$$

Then, for each wealth bin i, I estimate a coefficient ρ_i of autocorrelation across time as:

$$\rho_i = \left(\frac{1}{T} \sum_t e_{it}^2 \right)^{-1} \left(\frac{1}{T-1} \sum_t e_{it} e_{i,t-1} \right)$$

And I also estimate a correlation coefficient r of autocorrelation across wealth bins:

$$r = \frac{1}{T} \sum_t \left(\frac{1}{n} \sum_i e_{it}^2 \right)^{-1} \left(\frac{1}{n-1} \sum_i e_{it} e_{i-1,t} \right)$$

Unlike ρ_i, I do not vary the coefficient r across time. Indeed, there is no empirical evidence, and also no good *a priori* reasons, for it to vary as such. Instead, I average the value across all

14

periods, which gives a more robust estimate. I assume that autocorrelations across time and wealth bins follow each follow an $AR(1)$ process. For each wealth bin i, the correlation matrix across time is $\Omega_i^w = \text{Toeplitz}[1, \rho_i, \rho_i^2, \ldots, \rho_i^T]$. The correlation matrix across wealth bins is the same for all periods and is equal to $\Omega^t = \text{Toeplitz}[1, r, r^2, \ldots, r^n]$. I assume that the correlation matrix Ω between all error terms, across all periods and all wealth bins, has a two-way $AR(1)$ structure, which can be constructed as:

$$
\Omega = \begin{bmatrix} \Omega_1^w & & & \\ & \Omega_2^w & & \\ & & \ddots & \\ & & & \Omega_n^w \end{bmatrix}^{1/2} \underbrace{\begin{bmatrix} \Omega^t & & & \\ & \Omega^t & & \\ & & \ddots & \\ & & & \Omega^t \end{bmatrix}}_{T \text{ times}} \begin{bmatrix} \Omega_1^w & & & \\ & \Omega_2^w & & \\ & & \ddots & \\ & & & \Omega_n^w \end{bmatrix}^{1/2}
$$

where $(\cdot) \mapsto (\cdot)^{1/2}$ is the matrix square root for symmetric positive definite matrices. I estimate a standard error σ_i in every wealth bin i as:

$$
\sigma_i = \sqrt{\frac{1}{T} \sum_t e_{it}^2}
$$

and define the vector $\sigma = (\sigma_1, \ldots, \sigma_n)$. In line with the evidence, I assume that the error term is heteroscedastic over wealth bins but homoscedastic over time. The final, complete covariance matrix Σ for error terms is, therefore, equal to $\Sigma = A'\Omega A$, where

$$
A = \text{Diagonal}(\underbrace{\sigma, \ldots, \sigma}_{T \text{ times}})
$$

I simulate the error terms $\tilde{e}_{it}$ according to a multivariate normal distribution with mean zero and covariance Σ. The 2-dimensional residual is the product of this value with the unit vector normal to the fitted line, i.e. $(b, -\delta)/\sqrt{\delta(b^2 + \delta)}$. So I obtain bootstrap replications of the sample by adding the simulated error terms to the fitted values as follows:

$$
\tilde{x}_{it} = x_{it}^* + \tilde{e}_{it} b/\sqrt{\delta(b^2 + \delta)}
$$
$$
\tilde{y}_{it} = y_{it}^* - \tilde{e}_{it} \delta/\sqrt{\delta(b^2 + \delta)}
$$

I get bootstrap replications of the parameter values by running the Deming (1943) again regression on the simulated data.

C.3 Transformation of Parameters Due to the Inverse Hyperbolic Sine Transform

To facilitate the manipulation of the data and the estimations, I transform wealth using the inverse hyperbolic sine function, which we can define as:

$$\text{asinh}(x) \mapsto \log(x + \sqrt{x^2 + 1})$$

This function is useful because it is bijective, behaves logarithmically at large scales, but still tolerates zero or negative values, for which it behaves linearly. A helpful feature of continuous-time stochastic processes is that the dynamics of wealth and the dynamics of asinh(wealth) are directly related to one another through Itô's lemma. Concretely, assume that wealth follows a SDE with the following drift and diffusion:

$$\mu_t(w) = z_t(w) - c_t(w) \qquad \qquad \sigma_t(w) = \left(\psi_t^2(w) + \gamma_t^2(w)\right)^{1/2}$$

where:

$$z_t(w) \equiv y_t(w) + (r_t(w) - g_t)w \qquad \qquad \psi_t^2(w) \equiv v_t^2(w) + \phi_t^2(w)w^2$$

are, respectively, the drift and the diffusion induced by income. Then asinh(wealth) follows a SDE as well with the following drift and diffusion:

$$\tilde{\mu}_t(w) = \frac{\mu_t(w)}{\sqrt{1 + w^2}} - \frac{w}{2\sqrt{1 + w^2}}\left(\frac{\sigma_t(w)}{\sqrt{1 + w^2}}\right)^2$$

$$\tilde{\sigma}_t(w) = \frac{\sigma_t(w)}{\sqrt{1 + w^2}}$$

To perform the estimation, I must first transform the income parameters to conform to the asinh scale:

$$\tilde{z}_t(w) = \frac{z_t(w)}{\sqrt{1 + w^2}} - \frac{w}{2\sqrt{1 + w^2}}\left(\frac{\psi_t(w)}{\sqrt{1 + w^2}}\right)^2 \qquad \qquad \tilde{\psi}_t(w) = \frac{\psi_t(w)}{\sqrt{1 + w^2}}$$

Then I estimate $\tilde{c}(w)$ and $\tilde{\gamma}(w)$ using $\tilde{z}_t(w)$ as the income-induced drift, and $\tilde{\psi}_t(w)$ as the income-induced diffusion. I transform these parameters back into the linear scale:

$$c(w) = \tilde{c}_t(w)\sqrt{1 + w^2} - \frac{1}{2}\tilde{\gamma}(w)w \qquad \qquad \gamma(w) = \tilde{\gamma}(w)\sqrt{1 + w^2}$$

which gives estimates that can directly be interpreted as the mean and the variance of consumption.

C.4   Robustness Checks

The estimation involves several hyperparameters: they include δ in the Deming (1943) regression (see Section C.1), as well as a series of bandwidths that control the degree of smoothing involved when estimating specific densities and derivatives. In practice, the amount of noise in the data guides the choice of these hyperparameters. This section lists their default values and, more importantly, provides checks that confirm that the results are robust to variations of these parameters.

Hyperparameter	Role	Default value	Estimation method and kernel (if applicable)
δ	Determines the ratio of the variance of the error terms between both sides of the Deming (1943) regression.	Depends on the wealth bin (see Section C.1, as well as line "Bandwidth for estimating measurement error variance" in this table)	
Bandwidth for mean income over time	The average amount of income that is received by each wealth bin can sensibly vary over time with the business cycle. This value is smoothed over time to focus on the long-run dynamics instead.	12.5 years	locally constant regression with rectangular kernel
Bandwidth for income variance	The estimate of the variance of income conditional on wealth is smoothed across wealth, both to get a more stable estimate, and to be able to estimate the derivative with respect to wealth.	1 time average national income (asinh scale)	locally constant regression with rectangular kernel
Bandwidth for the derivative of the log density of wealth	This bandwidth parameter is used to calculate the derivative of the log density of wealth based on the histogram estimate of the density.	1.5 times average national income (asinh scale)	locally linear regression with rectangular kernel
Bandwidth for the inverse of the survival function of wealth	The inverse of the survival function of wealth $(F(w)/f(w))$ is computed as an intermediary step to obtain the left-hand side of equation (6). This estimate is smoothed over time using this bandwidth parameter.	2 years	locally constant regression with rectangular kernel
Bandwidth for auxiliary effects	The impact of auxiliary effects (demography, etc.) on the CDF of wealth are microsimulated for every year, and then are smoothed over time using this bandwidth.	10 years	locally constant regression with rectangular kernel
Bandwidth for estimating measurement error variance	To estimate the hyperparameter δ in the Deming (1943) regression, I estimate the variance of the measurement error for both variables in the equation, using the variance of the residual between observed values and smoothed version using this bandwidth parameter.	5 years	locally constant regression with rectangular kernel
Bandwidth for derivative of the diffusion	To adjust the value of the drift, I need to estimate the derivative of the diffusion with respect to wealth, using this bandwidth parameter.	0.5 time average national income (asinh scale)	locally linear regression with rectangular kernel

Table 3: List and Default Values for Hyperparameters

Table 3 describes the hyperparameters used in the estimation. Besides δ, all of them are bandwidth parameters used for smoothing out noise or short-term variations in the data. I apply

smoothing using locally constant regressions with a rectangular kernel (i.e., moving averages). When I need to estimate derivatives, I use the coefficient from a locally linear regression with a rectangular kernel as well.
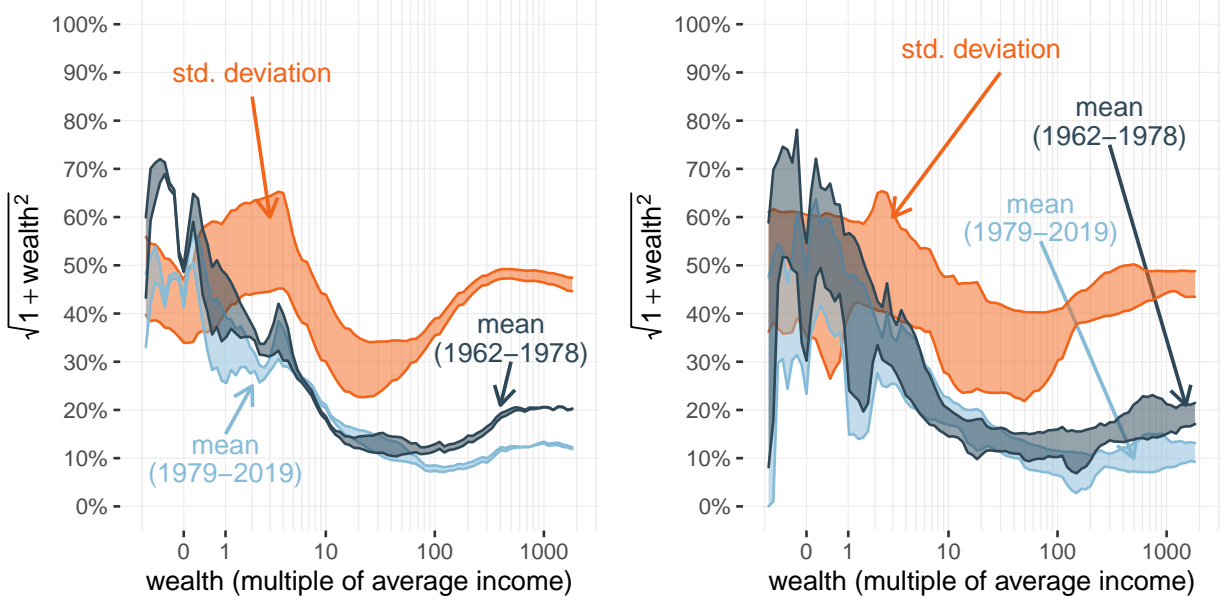
(a) δ Hyperparameter in Deming Regression
Note: Full range of estimates obtained by varying the hyperparameter δ in the Deming (1943) regression by factors $\alpha \in \{1/5, 1/2, 1, 2, 5\}$. See text for details.

(b) Bandwidth Hyperparameters
Note: 95% interval range of the estimates obtained by varying the bandwidth parameters separately at random by factors $\alpha \in \{1/5, 1/4, 1/3, 1/2, 1, 2, 3, 4, 5\}$. See text for details.

Figure C.5: Robustness Checks

In Figure C.5, I show how the results change when I vary the hyperparameters: I find that the key findings are robust. Figure C.5a focuses on δ, the hyperparameter used in the Deming (1943) regression. By default, this parameter is automatically determined using a rough estimate of the measurement error. But I can also modify this parameter directly and see how the results are affected. I adjust δ by a factor $\alpha \in \{1/5, 1/2, 1, 2, 5\}$, and then report the range of estimates we obtain using these different values. This range remains fairly narrow, especially for the drift, and at the top of the distribution. The limited impact of δ has a simple explanation. The parameter δ determines to what extent these residuals are attributed to the left-hand side or the right-hand side of equation (6). But the regression we run explains most of the data's variance, so the residuals remain small in all cases. As a result, where we attribute them has a limited impact.

Figure C.5b does a similar exercise, this time focusing on the bandwidth parameters (with δ now being automatically calculated, using these bandwidth parameters). I perform 100

estimations of the parameters, where I separately vary each bandwidth parameter at random by a coefficient $\alpha \in \{1/5, 1/4, 1/3, 1/2, 1, 2, 3, 4, 5\}$.[9] Then, I show the range of final estimates where 95% of estimates fall. These ranges are wider than the ones for the δ parameter alone, but the main results remain valid.

C.5 Wealth Distribution with a Wealth Tax and a Lump-sum Rebate
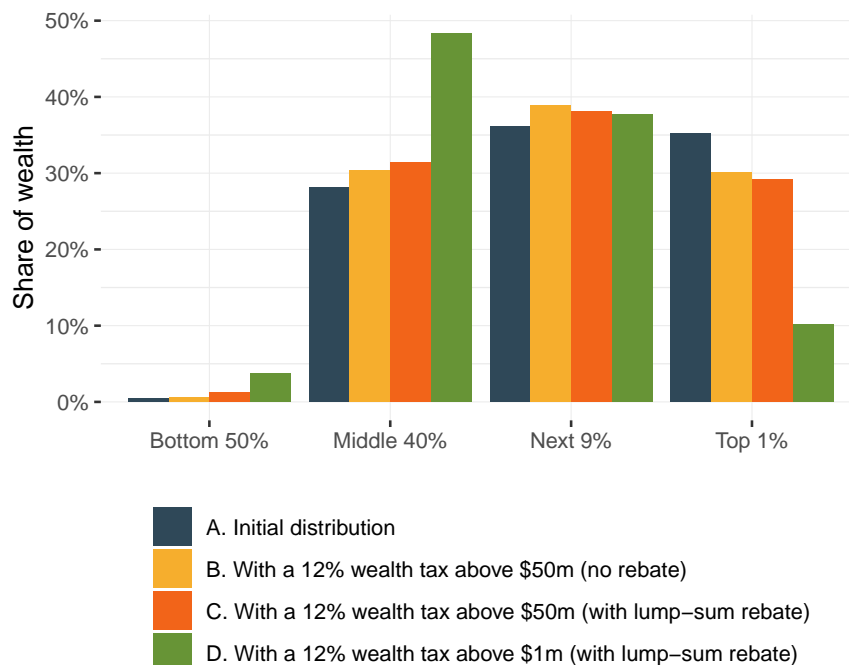
Note: This figure assumes benchmark parameters for behavioral responses (elasticity of tax avoidance $\epsilon = 1$ and elasticity of consumption $\eta = 1$.) For each bracket, the first bar shows the initial distribution of wealth, which corresponds to the year 2019 of the DINA data from Saez and Zucman (2020a). The second bar corresponds to the long-run distribution, with a wealth tax of 12% above $50m (which maximizes revenue in the benchmark calibration.) The third bar shows the same distribution when the government rebates the tax revenue lump sum every year. The fourth show the same effect for a tax with a much larger base (all wealth above $1m).

Figure C.6: Long-run Changes of the Wealth Distribution with a Wealth Tax and and Lump-sum Rebate

This section provides some details on how a wealth tax would impact the distribution of wealth in the long run. For this exercise, I focus on a revenue-maximizing linear tax above

---

[9]I make an exception for the bandwidth parameter used to calculate the derivative of the diffusion. I only vary that parameter by a factor $\alpha \in \{1, 2, 3, 4, 5\}$, because undersmoothing in that part of the estimation leads to unrealistically noisy estimates.

$50m, which in the benchmark calibration corresponds to a linear tax rate of 12%. The tax government rebates the tax lump sum every year, so the dynamics of wealth are now:

$$\mathrm{d}w_{it} = (\mu(w_{it}) - \tau(w_{it}) + \bar{\tau})\,\mathrm{d}t + \sigma_t(w_{it})\,\mathrm{d}B_{it}$$

where $\bar{\tau}$ is the average tax revenue. The steady-state wealth distribution is now characterized by:

$$f^*(w) \propto \theta(w)[\lambda(w)]^{\bar{\tau}} f(w) \qquad \text{where} \qquad \lambda(w) = \exp\left\{ 2 \int_{-\infty}^{w} \frac{1}{\sigma^2(s)}\,\mathrm{d}s \right\}$$

and where $\theta(w)$ is defined similarly as in the no-rebate case (and, in the current example, accounts for behavioral responses.) The value $\bar{\tau}$ must satisfy the equation:

$$\bar{\tau} - \frac{\int_{-\infty}^{+\infty} \tau(w)\theta(w)[\lambda(w)]^{\bar{\tau}} f(w)\,\mathrm{d}w}{\int_{-\infty}^{+\infty} \theta(w)[\lambda(w)]^{\bar{\tau}} f(w)\,\mathrm{d}w} = 0$$

which can be solved numerically. Figure C.6 shows the results from this exercise. It reports the wealth shares of four brackets (the bottom 50%, the middle 40%, the next 9%, and the top 1%) and four scenarios (no wealth tax, wealth tax without rebate, wealth tax with rebate, broader wealth tax with rebate). The wealth distribution remains highly unequal in all cases: the bottom 50% has virtually no wealth. Still, the wealth tax operates a significant amount of redistribution. The wealth tax alone primarily redistributes away from the top 1% to the benefit of the middle 40% and the next 9%. But it does not affect the bottom 50%. With the rebate, the bottom 50% more than doubles its share, but from a very low baseline (0.5%, to 1.3%). The tax with a broader base (all wealth above $1m) increases it a bit more, to 3.6%.

Appendix References

Alstadsæter, Annette et al. (2016). *Accounting for Business Income in Measuring Top Income Shares: Integrated Accrual Approach Using Individual and Firm Data from Norway*. Tech. rep. w22888. Cambridge, MA: National Bureau of Economic Research, w22888. DOI: 10.3386/w22888. URL: http://www.nber.org/papers/w22888.pdf (cit. on p. 5).

Center for Disease Control (2022). *National Vital Statistics System*. URL: https://www.cdc.gov/nchs/nvss/index.htm (cit. on pp. 11, 12).

Deming, W. Edwards (1943). *Statistical adjustment of data*. New York: Wiley (cit. on pp. 13–15, 17, 18).

Fritsch, F. N. and R. E. Carlson (1980). "Monotone Piecewise Cubic Interpolation". *SIAM Journal on Numerical Analysis* 17.2, pp. 238–246. DOI: 10.1137/0717021. URL: http://epubs.siam.org/doi/10.1137/0717021 (cit. on p. 6).

Gapminder (2019). *Children per women since 1800*. URL: https://www.gapminder.org/news/children-per-women-since-1800-in-gapminder-world/.

Gyöngy, I. (1986). "Mimicking the one-dimensional marginal distributions of processes having an ito differential". *Probability Theory and Related Fields* 71.4, pp. 501–516. DOI: 10.1007/BF00699039. URL: http://link.springer.com/10.1007/BF00699039 (cit. on p. 2).

Haines, Michael R. (1998). "Estimated Life Tables for the United States, 1850–1910". *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 31.4, pp. 149–169. DOI: 10.1080/01615449809601197. URL: http://www.tandfonline.com/doi/abs/10.1080/01615449809601197 (cit. on p. 6).

Human Fertility Collection (2019). Max Planck Institute for Demographic Research in Rostock, Germany and Vienna Institute of Demography in Vienna, Austria. URL: https://www.fertilitydata.org/.

Human Fertility Database (2019). Max Planck Institute for Demographic Research in Rostock, Germany and Vienna Institute of Demography in Vienna, Austria. URL: https://www.humanfertility.org/.

Human Life Table Database (2019). Max Planck Institute for Demographic Research in Rostock, Germany, Department of Demography at University of California at Berkeley, USA and Institut d'études démographiques (INED) in Paris, France. URL: https://www.lifetable.de/.

Human Mortality Database (2019). Max Planck Institute for Demographic Research in Rostock, Germany, Department of Demography at University of California at Berkeley, USA. URL: https://www.mortality.org.

Koenker, Roger (2019). *quantreg: Quantile Regression*. R package version 5.40. URL: https://CRAN.R-project.org/package=quantreg.

Menchik, Paul L. (1980). "Primogeniture, Equal Sharing, and the U.S. Distribution of Wealth". *The Quarterly Journal of Economics* 94.2, p. 299. DOI: 10.2307/1884542. URL: https://academic.oup.com/qje/article-lookup/doi/10.2307/1884542 (cit. on p. 7).

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (2018). "Distributional National Accounts: Methods and Estimates for the United States". *The Quarterly Journal of Economics* 133.2, pp. 553–609. DOI: 10.1093/qje/qjx043. URL: https://doi.org/10.1093/qje/qjx043 (cit. on pp. 3–5).

Robbins, Jacob A (2018). "Capital Gains and the Distribution of Income on the United States". URL: https://users.nber.org/~robbinsj/jr_inequ_jmp.pdf (cit. on p. 5).

Ruggles, Steven et al. (2022). *IPUMS USA: Version 12.0*. Version Number: 12.0 Type: dataset. DOI: 10.18128/D010.V12.0. URL: https://usa.ipums.org (cit. on pp. 6, 7, 10, 12).

Saez, Emmanuel and Gabriel Zucman (2016). "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data". *The Quarterly Journal of Economics* 131.2, pp. 519–578. DOI: 10.1093/qje/qjw004. URL: https://doi.org/10.1093/qje/qjw004 (cit. on p. 5).

— (2020a). "The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts". *Journal of Economic Perspectives* 34.4, pp. 3–26. DOI: 10.1257/jep.34.4.3. URL: https://pubs.aeaweb.org/doi/10.1257/jep.34.4.3 (cit. on pp. 3, 19).

— (2020b). *Trends in US Income and Wealth Inequality: Revising After the Revisionists*. Tech. rep. 27921. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w27921. URL: http://www.nber.org/papers/w27921.pdf (cit. on p. 3).

Schepsmeier, Ulf et al. (2018). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.1.8. URL: https://CRAN.R-project.org/package=VineCopula.

United Nations (2019). *World Population Prospects, 2019 Revision*. OCLC: 1142478963. URL: https://population.un.org/wpp/Publications/ (cit. on p. 6).

United States Census Bureau (2021). *National Intercensal Tables: 1900-1990*. URL: https://www.census.gov/content/census/en/data/tables/time-series/demo/popest/pre-1980-national.html (cit. on p. 6).