Ergodicity Economics

Ole Peters and Alexander Adamou

LONDON MATHEMATICAL LABORATORY

2018/06/26 at 19:59:34

Contents

1 Populations

The previous chapter developed a model of individual behaviour based on an assumed dynamic imposed on wealth. If we know the stochastic process that describes individual wealth, then we also know what happens at population level – each individual is represented by a realisation of the process, and we can compute the dynamics of wealth distributions. We answer questions about inequality and poverty in our model economy. It turns out that our decision criterion generates interesting emergent behaviour – cooperation, the sharing and pooling of resources, is often time-average growth optimal. This provides answers to the puzzles of why people cooperate, why there is an insurance market, and why we see socio-economic structure from the formation of firms to nation states with taxation and redistribution systems.

1.1 Every man for himself

{section:Every_man}

We have seen that risk aversion constitutes optimal behaviour under the assumption of multiplicative wealth growth and over time scales that are long enough for systematic trends to be significant. In this chapter we will continue to explore our null model, GBM. By "explore" we mean that we will let the model generate its world – if individual wealth was to follow GBM, what kind of features of an economy would emerge? We will see that cooperation and the formation of social structure also constitute optimal behaviour.

GBM is more than a random variable. It's a stochastic process, either a set of trajectories or a family of time-dependent random variables, depending on how we prefer to look at it. Both perspectives are informative in the context of economic modelling: from the set of trajectories we can judge what is likely to happen to an individual, *e.g.* by following a single trajectory for a long time; while the PDF of the random variable $x(t^*)$ at some fixed value of t^* tells us how wealth is distributed in our model.

We use the term wealth distribution to refer to the density function $\mathcal{P}_x(x)$ (not to the process of distributing wealth among people). This can be interpreted as follows: if I select a random individual (each individual with uniform probability $\frac{1}{N}$), the probability of the selected individual having wealth greater than x is given by the CDF $F_x(x) = \int_x^\infty \mathcal{P}_x(s)\,ds$. In a large population of N individuals, $\Delta x \mathcal{P}_x(x) N$ is the approximate number of individuals who have wealth between x and $x + \Delta x$. Thus, a broad wealth distribution with heavy tails indicates greater wealth inequality.

Examples:

- Under perfect equality everyone would have the same, meaning that the wealth distribution would be a Dirac delta function centred at the sample mean of x, that is
$$\mathcal{P}_x(x) = \delta(x - \langle x \rangle_N); \tag{1}$$
- Maximum inequality would mean that one individual owns everything and everyone else owns nothing, that is
$$\mathcal{P}_x(x) = \frac{N-1}{N}\delta(x-0) + \frac{1}{N}\delta(x - N\langle x \rangle_N). \tag{2}$$

1.1.1 Log-normal distribution

{section:Log-normal_wealt

At a given time, t, GBM produces a random variable, $x(t)$, with a log-normal distribution whose parameters depend on t. (A log-normally distributed random variable is one whose logarithm is a normally distributed random variable.) If each individual's wealth follows GBM,

$$dx = x(\mu dt + \sigma dW), \tag{3}$$ {eq:GBM}

with solution

$$x(t) = x(0)\exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W(t)\right], \tag{4}$$ {eq:GBM_sol}

then we will observe a log-normal distribution of wealth at each moment in time:

$$\ln x(t) \sim \mathcal{N}\left(\ln x(0) + \left(\mu - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right). \tag{5}$$ {eq:lognormal}

It will be convenient hereafter to assume the initial condition $x(0) = 1$ (and, therefore, $\ln x(0) = 0$) unless otherwise stated.

Note that the variance of $\ln x(t)$ increases linearly in time. We will develop an understanding of this shortly. As we will see, it indicates that any meaningful measure of inequality will grow over time in our simple model. To see what kind of a wealth distribution (Eq. 5) is, it is worth spelling out the lognormal PDF:

$$\mathcal{P}_x(x) = \frac{1}{x\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{[\ln x - (\mu - \frac{\sigma^2}{2})t]^2}{2\sigma^2 t}\right). \qquad (6)$$ {eq:PDFx}

This distribution is the subject of a wonderful book [2], sadly out-of-print now. We will find it useful to know a few of its basic properties. Of particular importance is the expected wealth under this distribution. This is

$$\langle x(t)\rangle = \exp(\mu t) \qquad (7)$$ {eq:exp_x}

or, equivalently, $\ln\langle x(t)\rangle = \mu t$. We could confirm this result by calculating $\langle x(t)\rangle = \int_0^\infty s\mathcal{P}_x(s)\,ds$, but this would be laborious. Instead we use a neat trick, courtesy of [18, Chapter 4.2], which will come in handy again in Sec. **??**. To compute moments, $\langle x^n\rangle$, of stochastic differential equations for x, like (Eq. 1.1.6), we find ordinary differential equations for the moments, which we know how to solve. For the first moment we do this simply by taking expectations of both sides of (Eq. 1.1.6). The noise term vanishes to turn the SDE for x into an ODE for $\langle x\rangle$:

$$\begin{aligned}
\langle dx\rangle &= \langle x(\mu dt + \sigma dW)\rangle & (8)\\
d\langle x\rangle &= \langle x\rangle \mu dt + \sigma \overbrace{\langle dW\rangle}^{=0} & (9)\\
&= \langle x\rangle \mu dt. & (10)
\end{aligned}$$

This is a very simple first-order linear differential equation for the expectation value of x. Its solution with initial condition $x(0) = 1$ is (Eq. 7).

For $\mu > 0$ the expected wealth grows exponentially over time, as do its population median and variance:

$$\begin{aligned}
\text{median}[x(t)] &= \exp[(\mu - \sigma^2/2)t]; & (11)\\
\text{var}[x(t)] &= \exp(2\mu t)[\exp(\sigma^2 t) - 1]. & (12)
\end{aligned}$$ {eq:median_x} {eq:var_x}

1.1.2 Two growth rates

{section:two_rates}

We will recap briefly here one of our key ideas, covered in detail in Sec. **??**, that the ensemble average of all possible trajectories of GBM grows at a different (faster) rate from that achieved by a single trajectory almost surely in the long-time limit. Understanding this difference was the key to developing a coherent theory of individual decision-making. We will see here that it is also crucial in understanding how wealth becomes distributed in a population of individuals whose wealths follow and, in particular, how we can measure the inequality in such a distribution. {eq:GBM}

From (Eq. **??**), we recall that the growth rate of the expected wealth is

$$g_{\langle\rangle} = \frac{d\ln\langle x\rangle}{dt} = \mu, \qquad (13)$$

while, from (Eq. **??**), the time-average growth rate of wealth is

$$\overline{g} = \frac{d \langle \ln x \rangle}{dt} = \mu - \frac{\sigma^2}{2}. \tag{14}$$

1.1.3 Measuring inequality

{section:Inequality_measu

In the case of GBM we have just seen how to compute the exact full wealth distribution $\mathcal{P}$. This is interesting but often we want only summary measures of the distribution. One such summary measure of particular interest to economists is inequality. How much inequality is there in a distribution like (Eq. 5)? And how does this quantity increase over time under GBM, as we have suggested?

Clearly, to answer these questions, we must quantify "inequality". In this section, and also in [1], we develop a natural way of measuring it, which makes use of the two growth rates we identified for the non-ergodic process. We will see that a particular inequality measure, known to economists as Theil's second index of inequality [38], is the difference between typical wealth (growing at the time-average growth rate) and average wealth (growing at the ensemble-average growth rate) in our model. Thus, the difference between the time average and ensemble average, the essence of ergodicity breaking, is the fundamental driver of the dynamics of inequality.

The two limits of inequality are easily identified: minimum inequality means that everyone has the same wealth, and maximum inequality means that one individual has all the wealth and everyone else has nothing. (This assumes that wealth cannot become negative.) Quantifying inequality in any other distribution is reminiscent of the gamble problem. Recall that for gambles we wanted make statements of the type "this gamble is more desirable than that gamble". We did this by collapsing a distribution to a scalar. Depending on the question that was being asked the appropriate way of collapsing the distribution and the resulting scalar can be different (the scalar relevant to an insurance company may not be relevant to an individual). In the case of inequality we also have a distribution – the wealth distribution – and we want to make statements of the type "this distribution is more unequal than that distribution". Again, this is done by collapsing the distribution to a scalar, and again many different choices of collapse and resulting scalar are possible. The Gini coefficient is a particularly well-known scalar of this type, the 80/20 ration is another, and many other measures exist.

In this context the expectation value is an important quantity. For instance, if everyone has the same wealth, everyone will own the average $\forall i, x_i = \langle x \rangle_N$, which converges to the expectation value for large N. Also, whatever the distribution of wealth, the total wealth is $N \langle x \rangle_N$ which converges to $N \langle x \rangle$ as N grows large. The growth rate of the expectation value, $g_{\langle\rangle}$, thus tells us how fast the average wealth and the total population wealth grow with probability one in a large ensemble. The time-average growth rate, $\overline{g}$, on the other hand, tells us how fast an individual's wealth grows with probability one in the long run. If the typical individual's wealth grows at a lower rate than the expectation value of wealth then there must be atypical individuals with very large wealths that account for the difference. This insight suggests the following measure of inequality.

<u>**Definition**</u> **Inequality, J, is the quantity whose growth rate is the**

difference between expectation-value and time-average growth rates,

$$\frac{dJ}{dt} = g_{\langle\rangle} - \overline{g}. \tag{15}$$

{eq:dJ}

Equation (15) defines the dynamic of inequality, and inequality itself is found by integrating over time:

$$J(t) = \int_0^t ds[g_{\langle\rangle}(s) - \overline{g}(s)]. \tag{16}$$

{eq:J}

This definition may be used for dynamics other than GBM. Whatever the wealth dynamic, typical minus average growth rates are informative of the dynamic of inequality. Within the GBM framework we can write the difference in growth rates as

$$\frac{dJ}{dt} = \frac{d\ln\langle x\rangle}{dt} - \frac{d\langle\ln x\rangle}{dt} \tag{17}$$

{eq:J_dyn}

and integrate over time to get

$$J(t) = \ln\langle x\rangle - \langle\ln x\rangle\,. \tag{18}$$

This quantity is known as the mean logarithmic deviation (MLD) or Theil's second index of inequality [38]. This is rather remarkable. Our general inequality measure, (Eq. 16), evaluated for the specific case of GBM, turns out to be a well-known measure of inequality that economists have identified independently, without considering non-ergodicity and ensemble average and time average growth rates. Merely by insisting on measuring inequality well, Theil used the GBM model without realising it!

Substituting the known values of the two growth rates into (Eq. 15) and integrating, we can evaluate the Theil inequality as a function of time:

$$J(t) = J(0) + \frac{\sigma^2}{2}t. \tag{19}$$

{eq:J_t}

Thus we see that, in GBM, our measure of inequality increases indefinitely.

1.1.4 Wealth condensation

{section:condensation}

The log-normal distribution generated by GBM broadens indefinitely, (Eq. 12). Likewise, the inequality present in the distribution – measured as the time-integrated difference between ensemble and time average growth rates – grows continually. A related property of GBM is the evolution towards wealth condensation. Wealth condensation means that a single individual will own a non-zero fraction of the total wealth in the population in the limit of large N, see *e.g.* [9]. In the present case an arbitrarily large share of total wealth will be owned by an arbitrarily small share of the population.

One simple way of seeing this is to calculate the fraction of the population whose wealths are less than the mean, *i.e.* $x(t) < \exp(\mu t)$. To do this, we define a new random variable, $z(t)$, whose distribution is the standard normal:

$$z(t) \equiv \frac{\ln x(t) - (\mu - \sigma^2/2)t}{\sigma t^{1/2}} \sim \mathcal{N}(0,1). \tag{20}$$

We want to know the mass of the distribution with $\ln x(t) < \mu t$ or, equivalently, $z < \sigma t^{1/2}/2$. This is

$$\Phi\left(\frac{\sigma t^{1/2}}{2}\right), \tag{21}$$

where Φ is the CDF of the standard normal distribution. This fraction tends to one as $t \to \infty$.

1.1.5 Rescaled wealth

{section:rescaled}

Over the decades, economists have arrived at many inequality measures, and have drawn up a list of conditions that particularly useful measures of inequality satisfy. Such measures are called "relative measures" [36, Appendix 4], and J is one of them. One of the conditions is that inequality measures should not change when x is divided by the same factor for everyone. Since we are primarily interested in inequality in this section it is useful to remove absolute wealth levels from the analysis and study an object called the rescaled wealth.

Definition **The rescaled wealth,**

$$y = \frac{x}{\langle x \rangle_N}, \tag{22}$$

{eq:rescaled}

is the proportion of the population-average wealth owned by an individual.

This quantity is useful, for instance because its numerical value does not depend on the currency used, it is a dimensionless number. Thus if my rescaled wealth, $y = 1/2$, this means that my wealth is half the average wealth, irrespective of whether I measure wealth in Kazakhstani Tenge or in Swiss Francs. For a large population, (Eq. 17) may be expressed in terms of y as $\frac{dJ}{dt} = -\frac{d\langle \ln y \rangle}{dt}$.

===AA: Insert result for lognormal distribution of rescaled wealth under self-averaging conditions $\langle x(t) \rangle_N \sim \langle x(t) \rangle$. ===

To make the model more tractable we consider the case $N \to \infty$, which replaces the finite-ensemble average by the expectation value, $\langle x \rangle_N \to \langle x \rangle$. The finite ensemble size has important effects but we will not discuss them here. $\langle x(t) \rangle = \exp(\mu t)$, just as for GBM without taxation, (Eq. 7). We are again interested in rescaled wealth, $y = \frac{x}{\langle x \rangle} = x e^{-\mu t}$ ((Eq. 22)), whose dynamic we derive using the chain rule

$$\begin{aligned} dy &= \frac{\partial y}{\partial t}\, dt + \frac{\partial y}{\partial x}\, dx + \frac{1}{2}\frac{\partial^2 y}{\partial x^2}\, dx^2 &(23)\\ &= -\mu y\, dt + \frac{1}{x}\, dx &(24)\\ &= y \sigma dW. &(25) \end{aligned}$$

{eq:ysde}

The first moment of y is trivially 1,

$$\langle y \rangle = \left\langle \frac{x}{\langle x \rangle} \right\rangle = 1. \tag{26}$$

1.1.6 v-normal distributions and Jensen's inequality {section:jensen}

So far we have confined our analysis to GBM, where wealths follow the dynamic specific by . However, as we discussed in the context of gambles, other wealth dynamics are possible. {eq:GBM}

===AA: Insert discussion of v-normal distributions and growth rates here. ===

1.1.7 Power law resemblance {section:power_law}

It is an established empirical observation [25] that the upper tails of real wealth distributions look more like a power law than a log-normal. Our trivial model does not strictly reproduce this feature, but it is instructive to compare the lognormal distribution to a power-law distribution. A power law PDF has the asymptotic form

$$\mathcal{P}_x(x) = x^{-\alpha}, \tag{27}$$ {eq:power_law}

for large arguments x. This implies that the logarithm of the PDF is proportional to the logarithm of its argument, $\ln \mathcal{P}_x(x) = -\alpha \ln x$. Plotting one against the other will yield a straight line, the slope being the exponent $-\alpha$.

Determining whether an empirical observation is consistent with such behaviour is difficult because the behaviour is to be observed in the tail (large x) where data are, by definition, sparse. A quick-and-dirty way of checking for possible power-law behaviour is to plot an empirical PDF against its argument on log-log scales, look for a straight line, and measure the slope. However, plotting any distribution on any type of scales results in some line. It may not be a straight line but it will have some slope everywhere. For a known distribution (power law or not) we can interpret this slope as a local apparent power-law exponent.

What is the local apparent power-law exponent of a log-normal wealth distribution near the expectation value $\langle x \rangle = \exp(\mu t)$, *i.e.* in the upper tail where approximate power law behaviour has been observed empirically? The logarithm of (Eq. 6) is

$$\begin{aligned}
\ln \mathcal{P}(x) = & \quad -\ln\left(x\sqrt{2\pi\sigma^2 t}\right) - \tfrac{[\ln x - (\mu - \frac{\sigma^2}{2})t]^2}{2\sigma^2 t} && (28)\\
= & \quad -\ln x - \tfrac{\ln(2\pi\sigma^2 t)}{2} - \tfrac{(\ln x)^2 - 2(\mu - \frac{\sigma^2}{2})t \ln x + (\mu - \frac{\sigma^2}{2})^2 t^2}{2\sigma^2 t}. && (29)
\end{aligned}$$

Collecting terms in powers of $\ln x$ we find

$$\ln \mathcal{P}(x) = -\frac{(\ln x)^2}{2\sigma^2 t} + \left(\frac{\mu}{\sigma^2} - \frac{3}{2}\right)\ln x - \frac{\ln(2\pi\sigma^2 t)}{2} - \frac{(\mu - \frac{\sigma^2}{2})^2 t}{2\sigma^2} \tag{30}$$

with local slope, *i.e.* apparent exponent,

$$\frac{d \ln \mathcal{P}(x)}{d \ln x} = -\frac{\ln x}{\sigma^2 t} + \frac{\mu}{\sigma^2} - \frac{3}{2}. \tag{31}$$

Near $\langle x \rangle$, $\ln x \sim \mu t$ so that the first two terms cancel approximately. Here the distribution will resemble a power-law with exponent $-3/2$ when plotted on doubly logarithmic scales. (The distribution will also look like a power-law where the first term is much smaller than the others, *e.g.* where $\ln x \ll \sigma^2 t$.)

We don't believe that such empirically observed power laws are merely a manifestation of this mathematical feature. Important real-world mechanisms that broaden real wealth distributions, *i.e.* concentrate wealth, are missing from the null model. However, it is interesting that the trivial model of GBM reproduces so many qualitative features of empirical observations.

1.2 Finite populations

{section:finite_populatio

So far we have considered the properties the random variable, $x(t)$, generated by GBM at a fixed time, t. Most of the mathematical objects we have discussed are, strictly speaking, relevant only in the limit $N \to \infty$, where N is the number of realisations of this random variable. For example, the expected wealth, $\langle x(t) \rangle$, is the limit of the sample mean wealth

$$\langle x(t) \rangle_N \equiv \frac{1}{N} \sum_{i=1}^{N} x_i(t), \tag{32}$$ {eq:sample}

as the sample size, N, grows large. In reality, human populations can be very large, say $N \sim 10^7$ for a nation state, but they are most certainly finite. Therefore, we need to be diligent and ask what the effects of this finiteness are. In particular, we will focus on the sample mean wealth under GBM. For what values of μ, σ, t, and N is this well approximated by the expectation value? And when it is not, what does it resemble?

1.2.1 Sums of lognormals

{section:sketch}

In [28] we studied the sample mean of GBM, which we termed the "partial ensemble average" (PEA). This is the average of N independent realisations the random variable $x(t)$, (Eq. 32). Here we sketch out some simple arguments about how this object depends on N and t.

Considering the two growth rates in Sec. 1.1.2, we anticipate the following tension:

A) for large N, the PEA should resemble the expectation value, $\exp(\mu t)$;

B) for long t, all trajectories in the sample – and, therefore, the sample mean – should grow like $\exp[(\mu - \sigma^2/2)t]$.

Situation A – when a sample mean resembles the corresponding expectation value – is known in statistical physics as "self-averaging." A simple strategy for estimating when this occurs is to look at the relative variance of the PEA,

$$R \equiv \frac{\mathrm{var}(\langle x(t) \rangle_N)}{\langle \langle x(t) \rangle_N \rangle^2}. \tag{33}$$

To be explicit, here the $\langle \cdot \rangle$ and $\mathrm{var}(\cdot)$ operators, without N as a subscript, refer to the mean and variance over all possible PEAs. The PEAs themselves, taken over finite samples of size N, are denoted $\langle \cdot \rangle_N$. Using standard results for the mean and variance of sums of independent random variables and inserting the results in (Eq. 7) and (Eq. 12), we get

$$R(N) = \frac{e^{\sigma^2 t} - 1}{N}. \tag{34}$$

If $R \ll 1$, then the PEA will likely be close to its own expectation value, which is equal to the expectation value of the GBM. Thus, in terms of N and t, $\langle x(t) \rangle_N \approx \langle x(t) \rangle$ when

$$t < \frac{\ln N}{\sigma^2}. \qquad (35)$$ {eq:short_t}

This hand-waving tells us roughly when the large-sample – or, as we see from (Eq. 35), short-time – self-averaging regime holds. A more careful estimate of the cross-over time in (Eq. 61) is a factor of 2 larger, but the scaling is identical.

For $t > \ln N/\sigma^2$, the growth rate of the PEA transitions from μ to its $t \to \infty$ limit of $\mu - \sigma^2/2$ (Situation B). Another way of viewing this is to think about what dominates the average. For early times in the process, all trajectories are close together and none dominate the PEA. However, as time goes by the distribution broadens exponentially. Since each trajectory contributes with the same weight to the PEA, after some time the PEA will be dominated by the maximum in the sample,

$$\langle x(t) \rangle_N \approx \frac{1}{N} \max_{i=1}^{N} \{x_i(t)\}, \qquad (36)$$

as illustrated in Fig. 1.

Self-averaging stops when even the "luckiest" trajectory is no longer close to the expectation value $\exp(\mu t)$. This is guaranteed to happen eventually because the probability for a trajectory to reach $\exp(\mu t)$ decreases towards zero as t grows. We know this from Sec. 1.1.4. Of course, this takes longer for larger samples, which have more chances to contain a lucky trajectory.

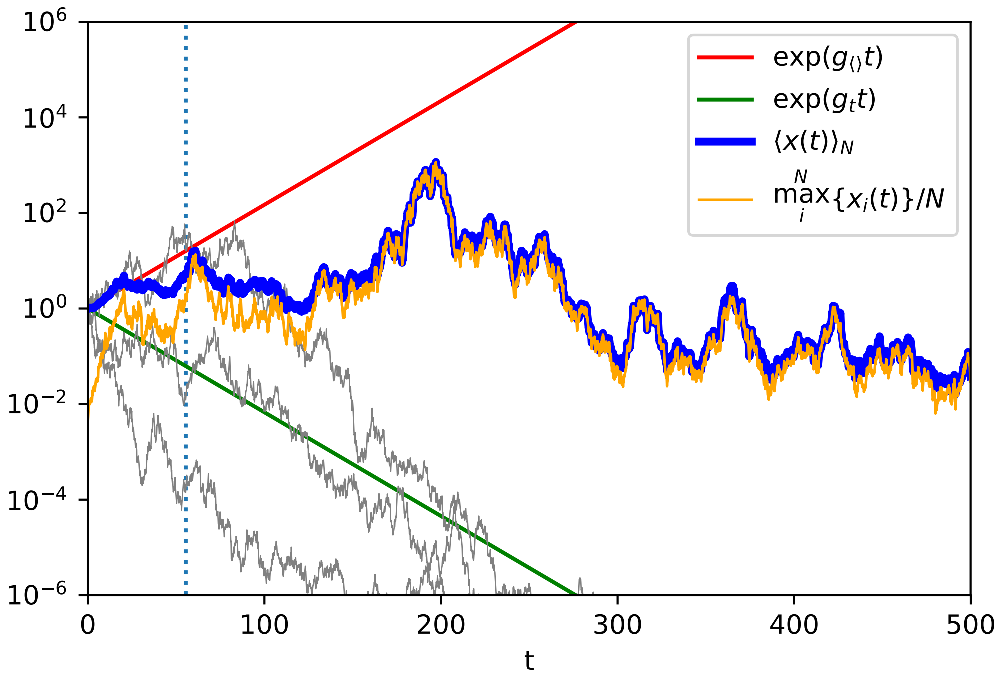

Figure 1: PEA and maximum in a finite ensemble of size $N = 256$. **Red line:** expectation value $\langle x(t)\rangle$. **Green line:** exponential growth at the time-average growth rate. In the $T \to \infty$ limit all trajectories grow at this rate. **Yellow line:** contribution of the maximum value of any trajectory at time t to the PEA. **Blue line:** PEA $\langle x(t)\rangle_N$. **Vertical line:** Crossover – for $t > t_c = \frac{2\ln N}{\sigma^2}$ the maximum begins to dominate the PEA (the yellow line approaches the blue line). **Grey lines:** randomly chosen trajectories – any typical trajectory soon grows at the time-average growth rate. **Parameters:** $N = 256$, $\mu = 0.05$, $\sigma = \sqrt{0.2}$. {fig:trajectories}

In [28] we analysed PEAs of GBM analytically and numerically. Using (Eq. 4) the PEA can be written as

$$\langle x \rangle_N = \frac{1}{N} \sum_{i=1}^{N} \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_i(t)\right], \tag{37}$$

{eq:PEA}

where $\{W_i(t)\}_{i=1\ldots N}$ are N independent realisations of the Wiener process. Taking the deterministic part out of the sum we re-write (Eq. 37) as

$$\langle x \rangle_N = \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t\right] \frac{1}{N} \sum_{i=1}^{N} \exp\left(t^{1/2} \sigma \xi_i\right), \tag{38}$$

{eq:PEA_2}

where $\{\xi_i\}_{i=1\ldots N}$ are N independent standard normal variates.

We found that typical trajectories of PEAs grow at $g_{\langle\rangle}$ up to a time t_c that is logarithmic in N, meaning $t_c \propto \ln N$. This is consistent with our analytical sketch. After this time, typical PEA trajectories begin to deviate from expectation-value behaviour, and eventually their growth rate converges to g_t. While the two limiting behaviours $N \to \infty$ and $t \to \infty$ can be computed exactly, what happens in between is less straightforward. The PEA is a random object outside these limits.

A quantity of crucial interest to us is the exponential growth rate experienced by the PEA,

$$g_{\text{est}}(t, N) \equiv \frac{\ln(\langle x(t) \rangle_N) - \ln(x(0))}{t - 0} = \frac{1}{t} \ln(\langle x(t) \rangle_N). \tag{39}$$

{eq:gest}

In [28] we proved that the $t \to \infty$ limit for any (finite) N is the same as for the case $N = 1$,

$$\lim_{t \to \infty} g_{\text{est}}(t, N) = \mu - \frac{\sigma^2}{2} \tag{40}$$

{eq:gest_2}

for all $N \geq 1$. Substituting (Eq. 38) in (Eq. 39) produces

$$\begin{aligned} g_{\text{est}}(t, N) &= \mu - \frac{\sigma^2}{2} + \frac{1}{t} \ln\left(\frac{1}{N} \sum_{i=1}^{N} \exp(t^{1/2} \sigma \xi_i)\right) & (41) \\ &= \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} + \frac{1}{t} \ln\left(\sum_{i=1}^{N} \exp(t^{1/2} \sigma \xi_i)\right). & (42) \end{aligned}$$

{eq:gest_4}

We didn't look in [28] at the expectation value of $g_{\text{est}}(t, N)$ for finite time and finite samples, but it's an interesting object that depends on N and t but is not stochastic. Note that this is not g_{est} of the expectation value, which would be the $N \to \infty$ limit of (Eq. 39). Instead it is the $S \to \infty$ limit,

$$\langle g_{\text{est}}(t, N) \rangle = \frac{1}{t} \langle \ln(\langle x(t) \rangle_N) \rangle = f(N, t), \tag{43}$$

{eq:gest_3}

where, as previously, $\langle \cdot \rangle$ without subscript refers to the average over all possible samples, *i.e.* $\lim_{S \to \infty} \langle \cdot \rangle_S$. The last two terms in (Eq. 42) suggest an exponential relationship between ensemble size and time. The final term is a tricky stochastic object on which the properties of the expectation value in (Eq. 43) will hinge. This term will be the focus of our attention: the sum of exponentials of normal random variates or, equivalently, log-normal variates.

1.2.2 The random energy model

Since the publication of [28] we have learned, thanks to discussions with J.-P. Bouchaud, that the key object in (Eq. 42) – the sum log-normal random variates – has been of interest to the mathematical physics community since the 1980s. The reason for this is Derrida's random energy model [13, 14].

It is defined as follows. Imagine a system whose energy levels are $2^K = N$ normally-distributed random numbers, ξ_i (corresponding to K spins). This is a very simple model of a disordered system, such as a spin glass, the idea being that the system is so complicated that we "give up" and simply model its energy levels as realisations of a random variable. (We denote the number of spins by K and the number of resulting energy levels by N, while Derrida uses N for the number of spins). In this model The partition function is then

$$Z = \sum_{i=1}^{N} \exp\left(\beta J \sqrt{\frac{K}{2}} \xi_i\right), \tag{44}$$ {eq:Z}

where the inverse temperature, β, is measured in appropriate units, and the scaling in K is chosen so as to ensure an extensive thermodynamic limit [13, p. 79]. J is a constant that will be determined below. The logarithm of the partition function gives the Helmholtz free energy,

$$F = -\frac{\ln Z}{\beta} \tag{45}$$

$$= -\frac{1}{\beta} \ln\left[\sum_{i=1}^{N} \exp\left(\beta J \sqrt{\frac{K}{2}} \xi_i\right)\right]. \tag{46}$$ {eq:F}

Like the growth rate estimator in (Eq. 39), this involves a sum of log-normal variates and, indeed, we can rewrite (Eq. 42) as

$$g_{\text{est}} = \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} - \frac{\beta F}{t}, \tag{47}$$ {eq:gest_5}

which is valid provided that

$$\beta J \sqrt{\frac{K}{2}} = \sigma t^{1/2}. \tag{48}$$ {eq:map}

Equation (48) does not give a unique mapping between the parameters of our GBM, (σ, t), and the parameters of the REM, (β, K, J). Equating (up to multiplication) the constant parameters, σ and J, in each model gives us a specific mapping:

$$\sigma = \frac{J}{\sqrt{2}} \quad \text{and} \quad t^{1/2} = \beta\sqrt{K}. \tag{49}$$ {eq:choice_1}

The expectation value of g_{est} is interesting. The only random object in (Eq. 47) is F. Knowing $\langle F \rangle$ thus amounts to knowing $\langle g_{\text{est}} \rangle$. In the statistical mechanics of the random energy model $\langle F \rangle$ is of key interest and so much about it is known. We can use this knowledge thanks to the mapping between the two problems.

Derrida identifies a critical temperature,

$$\frac{1}{\beta_c} \equiv \frac{J}{2\sqrt{\ln 2}}, \tag{50}$$ {eq:beta_c}

above and below which the expected free energy scales differently with K and β. This maps to a critical time scale in GBM,

$$t_c = \frac{2\ln N}{\sigma^2}, \tag{51}$$ {eq:t_c}

with high temperature ($1/\beta > 1/\beta_c$) corresponding to short time ($t < t_c$) and low temperature ($1/\beta < 1/\beta_c$) corresponding to long time ($t > t_c$). Note that t_c in (Eq. 61) scales identically with N and σ as the transition time, (Eq. 35), in our sketch.

In [13], $\langle F\rangle$ is computed in the high-temperature (short-time) regime as

$$\begin{aligned} \langle F\rangle &= E - S/\beta \qquad (52)\\ &= -\frac{K}{\beta}\ln 2 - \frac{\beta K J^2}{4}, \qquad (53) \end{aligned}$$ {eq:F_2}

and in the low-temperatures (long-time) regime as

$$\langle F\rangle = -KJ\sqrt{\ln 2}. \tag{54}$$ {eq:F_3}

Short time

We look at the short-time behavior first (high $1/\beta$, (Eq. 53)). The relevant computation of the entropy S in [13] involves replacing the number of energy levels $n(E)$ by its expectation value $\langle n(E)\rangle$. This is justified because the standard deviation of this number is $\sqrt{n}$ and relatively small when $\langle n(E)\rangle > 1$, which is the interesting regime in Derrida's case.

For spin glasses, the expectation value of F is interesting, supposedly, because the system may be self-averaging and can be thought of as an ensemble of many smaller sub-systems that are essentially independent. The macroscopic behavior is then given by the expectation value.

Taking expectation values and substituting from (Eq. 53) in (Eq. 47) we find

$$\langle g_{\text{est}}\rangle^{\text{short}} = \mu - \frac{\sigma^2}{2} + \frac{1}{t}\frac{KJ^2}{4T^2}. \tag{55}$$ {eq:gest_6}

From (Eq. 48) we know that $t = \frac{KJ^2}{2\sigma^2 T^2}$, which we substitute, to find

$$\langle g_{\text{est}}\rangle^{\text{short}} = \mu, \tag{56}$$ {eq:gest_7}

which is the correct behavior in the short-time regime.

Long time

Next, we turn to the expression for the long-time regime (low temperature, (Eq. 54)). Again taking expectation values and substituting, this time from (Eq. 54) in (Eq. 47), we find for long times

$$\langle g_{\text{est}}\rangle^{\text{long}} = \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} + \sqrt{\frac{2\ln N}{t}}\,\sigma, \tag{57}$$ {eq:gest_8}

which has the correct long-time asymptotic behavior. The form of the correction to the time-average growth rate in (Eq. 57) is consistent with [28] and [32], where it was found that approximately $N = \exp(t)$ systems are required for ensemble-average behavior to be observed for a time t, so that the parameter

$\ln N/t$ controls which regime dominates – if the parameter is small, then (Eq. 57) indicates that the long-time regime is relevant.

Figure 2 is a direct comparison between the results derived here, based on [13], and numerical results using the same parameter values as in [28], namely $\mu = 0.05, \sigma = \sqrt{0.2}, N = 256$ and $S = 10^5$.

Notice that $\langle g_{\text{est}} \rangle$ is not the (local) time derivative $\frac{\partial}{\partial t} \langle \ln(\langle x \rangle_N) \rangle$, but a time-average growth rate, $\left\langle \frac{1}{t} \ln \left(\frac{\langle x(t) \rangle_N}{\langle x(0) \rangle_N} \right) \right\rangle$. In [28] we used a notation that we've stopped using since then because it caused confusion – $\langle g \rangle$ there denotes the growth rate of the expectation value, which is not the expectation value of the growth rate.

It is remarkable that the expectation value $\langle g_{\text{est}}(N,t) \rangle$ so closely reflects the median, $q_{0.5}$, of $\langle x \rangle_N$, in the sense that

$$q_{0.5}(\langle x(t) \rangle_N) \approx \exp\left(\langle g_{\text{est}}(N,t) \rangle\, t\right). \tag{58}$$ {eq:quant_ave}

In [27] it was discussed in detail that $g_{\text{est}}(1,t)$ is an ergodic observable for (Eq. 1.1.6), in the sense that $\langle g_{\text{est}}(1,t) \rangle = \lim_{t \to \infty} g_{\text{est}}$. The relationship in (Eq. 58) is far more subtle. The typical behavior of GBM PEAs is complicated outside the limits $N \to \infty$ or $t \to \infty$, in the sense that growth rates are time dependent here. This complicated behaviour is well represented by an approximation that uses physical insights into spin glasses.

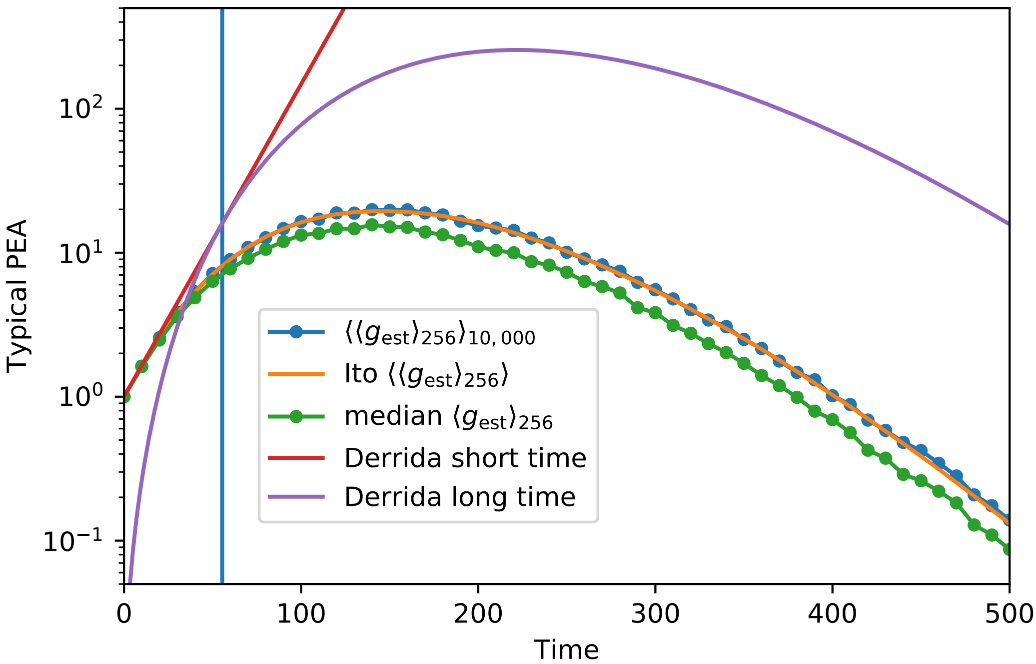

Figure 2: Lines are obtained by exponentiating the various exponential growth rates. **Blue line:** $\langle\langle g_{\text{est}}\rangle_{256}\rangle_{10,000}$ is the numerical mean (approximation of the expectation value) over a super-ensemble of $S = 10,000$ samples of g_{est} estimated in sub-ensembles of $N = 256$ GBMs each. **Green line:** median in a super-ensemble of S samples of g_{est}, each estimated in sub-ensembles of size N. **Yellow line:** (Eq. **??**) is an exact expression for $d\langle\ln\langle x\rangle_N\rangle$, derived using Itô calculus. We evaluate the expression by Monte Carlo, and integrate, $\langle\ln\langle x\rangle_N\rangle = \int_0^t d\langle\ln\langle x\rangle_N\rangle$. Exponentiation yields the yellow line. **Red line:** short-time behavior, based on the random energy model, (Eq. 56). **Purple line:** long-time behavior, based on the random energy model, (Eq. 57). **Vertical line:** Crossover between the regimes at $t_c = \frac{2\ln N}{\sigma^2}$, corresponding to $\beta_c = \frac{2(\ln 2)^{1/2}}{J}$. **Parameters:** $N = 256$, $S = 10,000$, $\mu = 0.05$, $\sigma = \sqrt{0.2}$. {fig:1}

2 Interactions

Insert abstract here.

2.1 Reallocation

{section:reallocation}

2.1.1 Introduction

{section:RGBM_intro}

In Sec. 1.1 we created a model world of independent trajectories of GBM. We studied how the distribution of the resulting random variables evolved over time. We saw that this is a world of broadening distributions, increasing inequality, and wealth condensation. We introduced cooperation to it in Sec. **??** and saw how this increases the time-average growth rate for those who pool and share all of their resources. In this section we study what happens if a large number of individuals pool and share only a fraction of their resources. This is reminiscent of the taxation and redistribution – which we shall call "reallocation" – carried out by populations in the real world.

We will find that, while full cooperation between two individuals increases their growth rates, sufficiently fast reallocation from richer to poorer in a large population has two related effects. Firstly, everyone's wealth grows in the long run at a rate close to that of the expectation value. Secondly, the distribution of rescaled wealth converges over time to a stable form. This means that, while wealth can still be distributed quite unequally, wealth condensation and the divergence of inequality no longer occur in our model. Of course, for this to be an interesting finding, we will have to quantify what we mean by "sufficiently fast reallocation."

We will also find that when reallocation is too slow or, in particular, when it goes from poorer to richer – which we will quantify as negative reallocation – no stable wealth distribution exists. In the latter case, the population splits into groups with positive and negative wealths, whose magnitudes grow exponentially.

Finally, having understood how our model behaves in each of these reallocation regimes, we will fit the model parameters to historical wealth data from the real world, specifically the United States. This will tell us which type of model behaviour best describes the dynamics of the US wealth distribution in both the recent and more distant past. You might find the results surprising – we certainly did!

2.1.2 The ergodic hypothesis in economics

{section:RGBM_EH}

Of course, we are not the first to study resource distributions and inequality in economics. This topic has a long history, going back at least as far as Vilfredo Pareto's work in the late 19$^{\text{th}}$ century [26] (in which he introduced the power-law distribution we discussed in Sec. 1.1.7). Economists studying such distributions usually assume that they converge in the long run to a unique and stable form, regardless of initial conditions. This allows them to study the stable distribution, for which many statistical techniques exist, and to ignore the transient phenomena preceding it, which are far harder to analyse. Paul Samuelson called this the "ergodic hypothesis" [35, pp. 11-12]. It's easy to see why: if this convergence happens, then the time average of the observable in question will equal its ensemble average over the stable distribution.[1]

Economics is often concerned with growth and a growing quantity cannot be ergodic in Samuelson's sense, because its distribution never stabilises. This

[1] Convergence to a unique and stable distribution is a sufficient but not necessary condition for an ergodic observable, as we have defined it.

suggests the simplifying ergodic hypothesis should never be made. Not so fast! Although rarely stated, a common strategy to salvage these techniques is to find a transformation of the non-ergodic process that produces a meaningful ergodic observable. If such an ergodic observable can be derived, then classical analytical techniques may still be used. We have already seen in the context of gambles that expected utility theory can be viewed as transformation of non-ergodic wealth increments into ergodic utility increments. Expectation values, which would otherwise be misleading, then quantify time-average growth of the decision-maker's wealth.

Studies of wealth distributions also employ this strategy. Individual wealth is modelled as a growing quantity. Dividing by the population average transforms this to a rescaled wealth, as in Sec. 1.1.5, which is hypothesised to be ergodic. For example, [4, p. 130] "impose assumptions ... that guarantee the existence and uniqueness of a limit stationary distribution." The idea is to take advantage of the simplicity with which the stable distribution can be analysed, *e.g.* to predict the effects of policies encoded in model parameters.

There is, however, an elephant in the room. To our knowledge, the validity of the ergodic hypothesis for rescaled wealth has never been tested empirically. It's certainly invalid for the GBM model world we studied previously because, as we saw in Sec. 1.1.5, rescaled wealth has an ever-broadening lognormal distribution. That doesn't seem to say much, as most reasonable people would consider our model world – containing a population of individuals whose wealths multiply noisily and who never interact – somewhat unrealistic! The model we are about to present will not only extend our understanding from this simple model world to one containing interactions, but also will allows us to test the hypothesis. This is because it has regimes, *i.e.* combinations of parameters, for which rescaled wealth is and isn't ergodic. This contrasts with models typically used by economists, which have the ergodic hypothesis "baked in."

If it is reasonable to assume a stable distribution exists, we must also consider how long convergence to would take after a change of parameters. It's no use if convergence in the model takes millennia, if we are using it to estimate the effect of a new tax policy over the next election cycle. Therefore, treating a stable model distribution as representative of the empirical wealth distribution implies an assumption of fast convergence. As the late Tony Atkinson pointed out, that "the speed of convergence makes a great deal of difference to the way in which we think about the model" [3]. We will also use our model to discuss this point. Without further ado, let us introduce it.

2.1.3 Reallocating GBM

{section:RGBM_model}

Our model, called Reallocating Geometric Brownian Motion (RGBM), is a system of N individuals whose wealths, $x_i(t)$, evolve according to the stochastic differential equation,

$$dx_i = x_i \left[(\mu - \tau)dt + \sigma dW_i\left(t\right)\right] + tau \left\langle x \right\rangle_N \; dt, \tag{59}$$

{eq:rgbm}

for all $i = 1 \dots N$. In effect, we have added to the GBM model a simple reallocation mechanism. Over a time step, dt, each individual pays a fixed proportion of its wealth, $\tau x_i dt$, into a central pot ("contributes to society") and gets back an equal share of the pot, $\tau \left\langle x \right\rangle_N dt$, ("benefits from society"). We can think of this as applying a wealth tax, say of 1% per year, to everyone's wealth and then redistributing the tax revenues equally. Note that the reallocation parameter, τ, is, like μ, a rate with dimensions per unit time. Note also that when $\tau = 0$, we recover our old friend, GBM, in which individuals grow their wealths without interacting.

RGBM is our null model of an exponentially growing economy with social structure. It is intended to capture only the most general features of the dynamics of wealth. A more complex model would treat the economy as a system of agents that interact with each other through a network of relationships. These relationships include trade in goods and services, employment, taxation, welfare payments, using public infrastructure (roads, schools, a legal system, social security, scientific research, and so on), insurance, wealth transfers through inheritance and gifts, and everything else that constitutes an economic network. It would be a hopeless task to list exhaustively all these interactions, let alone model them explicitly. Instead we introduce a single parameter – the reallocation rate, τ – to represent their net effect. If τ is positive, the direction of net reallocation is from richer to poorer. If negative, it is from poorer to richer.

We will see shortly that RGBM has both ergodic and non-ergodic regimes, characterised to a good approximation by the sign of τ. $\tau > 0$ produces an ergodic regime, in which wealths are positive, distributed with a Pareto tail, and confined around their mean value. $\tau < 0$ produces a non-ergodic regime, in which the population splits into two classes, characterised by positive and negative wealths which diverge away from the mean.

We offer a couple of health warnings. In RGBM, like in GBM, there are no additive changes akin to labour income and consumption. This is unproblematic for large wealths, where additive changes are dwarfed by capital gains. For small wealths, however, wages and consumption are significant and empirical distributions look rather different for low and high wealths [15]. We modelled earnings explicitly in [6] and found this didn't generate insights different from RGBM when fit to real wealth data. We note also, as [23, p. 41] put it, that our agents "do not marry or have children or die or even grow old." Therefore, the individual in our setup is best imagined as a household or a family, *i.e.* some long-lasting unit into which personal events are subsumed.

Having specified the model, we will use insights from Sec. 1.2 to understand how rescaled wealth is distributed in the ergodic and non-ergodic regimes. Then we will show briefly our results from fitting the model to historical wealth data from the United States. The full technical details of this fitting exercise are beyond the scope of these notes – if you are interested, you can find "chapter and verse" in [6]. Fitting τ to data will allow us to answer the important questions:

- What is the net reallocating effect of socio-economic structure on the wealth distribution?
- Are observations consistent with the ergodic hypothesis that the rescaled wealth distribution converges to a stable distribution?
- If so, how long does it take, after a change in conditions, for the rescaled wealth distribution to reach the stable distribution?

2.1.4 Model behaviour

{section:RGBM_behaviour}

It is instructive to write (Eq. 59) as

$$dx_i = \underbrace{x_i \left[\mu dt + \sigma dW_i\,(t)\right]}_{\text{Growth}} \; \underbrace{- \;\tau(x_i - \langle x \rangle_N) dt}_{\text{Reallocation}}. \tag{60}$$ {eq:rgbm_ou}

This resembles GBM with a mean-reverting term like that of [39] in physics and [40] in finance. It exposes the importance of the sign of τ. We discuss the two regimes in turn.

Positive τ

For $\tau > 0$, individual wealth, $x_i(t)$, reverts to the sample mean, $\langle x(t) \rangle_N$. We explored some of the properties of sample mean in Sec. 1.2 for wealths undergoing GBM. In particular, we saw that a short-time (or large-sample or low-volatility) self-averaging regime exists,

$$t < t_c \equiv \frac{2 \ln N}{\sigma^2}, \tag{61}$$ {eq:t_c}

where the sample mean is approximated well by the ensemble average,

$$\langle x(t) \rangle_N \sim \langle x(t) \rangle = \exp(\mu t). \tag{62}$$

(The final equality assumes, as previously, that $x_i(0) = 1$ for all i.) It turns out that the same self-averaging approximation can be made for wealths undergoing RGBM, (Eq. 59), when the reallocation rate, τ, is above some critical threshold:

$$\tau > \tau_c \equiv \frac{\sigma^2}{2 \ln N}. \tag{63}$$ {eq:tau_c}

Showing this is technically difficult [7] and we will confine ourselves to sketching the key ideas in Sec. 2.1.5 below. It won't have escaped your attention that $\tau_c = {t_c}^{-1}$ and, indeed, you will shortly have an intuition for why.

Fitting the model to data yields parameter values for which τ_c is extremely small. For example, typical parameters for US wealth data are $N = 10^8$ and $\sigma = 0.2$ year$^{-1/2}$, giving $\tau_c = 0.1\%$ year^{-1} (or $t_c = 900$ years). Accounting for the uncertainty in the fitted parameters makes this statistically indistinguishable from $\tau_c = 0$.

This means we can safely make the self-averaging approximation for the entire positive τ regime. That's great news, because it means we can rescale wealth by the ensemble average, $\langle x(t) \rangle = \exp(\mu t)$, as we did in Sec. 1.1.5 for GBM, and not have to worry about pesky finite N effects. Following the same procedure as there gives us a simple SDE in the rescaled wealth, $y_i(t) = x_i(t) \exp(-\mu t)$:

$$dy_i = y_i \sigma dW_i\,(t) - \tau(y_i - 1)dt. \tag{64}$$ {eq:rgbm_ou_re}

Note that the common growth rate, μ, has been scaled out as it was in Sec. 1.1.5.

The distribution of $y_i\,(t)$ can found by solving the corresponding Fokker-Planck equation, which we will do in Sec. 2.1.5. For now, we will just quote the result: a stable distribution exists with a power-law tail, to which the distribution of rescaled wealth converges over time. The distribution has a name – the Inverse Gamma Distribution – and a probability density function:

$$\mathcal{P}\,(y) = \frac{(\zeta - 1)^{\zeta}}{\Gamma\,(\zeta)} e^{-\frac{\zeta-1}{y}} y^{-(1+\zeta)}. \tag{65}$$ {eq:disti}

$\zeta = 1 + 2\tau/\sigma^2$ is the Pareto tail index (corresponding to $\alpha - 1$ in Sec. 1.1.7) and $\Gamma(\cdot)$ is the gamma function.

Example forms of the stationary distribution are shown in Figure 3. The usual stylised facts are recovered: the larger σ (more randomness in the returns) and the smaller τ (less social cohesion), the smaller the tail index ζ and the fatter the tail of the distribution. Fitted τ values give typical ζ values between 1 and 2 for the different datasets analysed, consistent with observed tail indices between 1.2 to 1.6 (see [6] for details). Not only does RGBM predict a realistic functional form for the distribution of rescaled wealth, but also it admits fitted parameter values which match observed tails. The inability to do this is a known weakness of earnings-based models (again, see [6] for discussion).

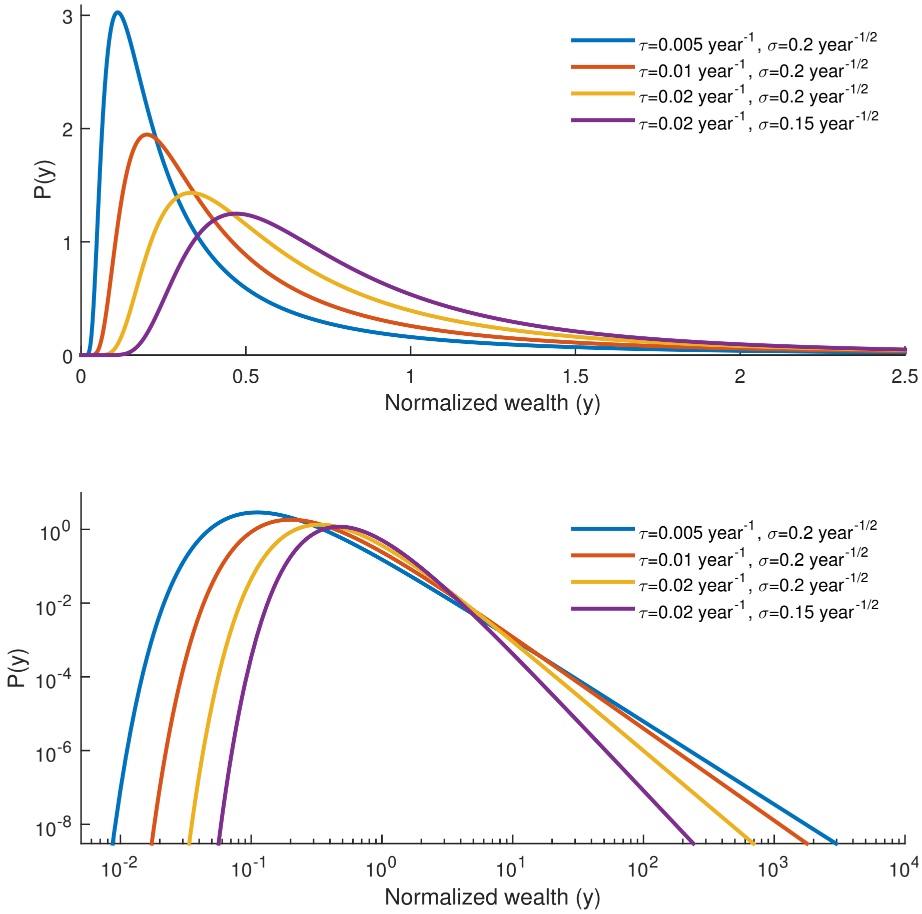

Figure 3: The stationary distribution for RGBM with positive τ. Top – linear scales; Bottom – logarithmic scales. {fig:dist}

For positive reallocation, (Eq. 64) and extensions of it have received much attention in statistical mechanics and econophysics [9, 8]. As a combination of GBM and a mean-reverting process it is a simple and analytically tractable stochastic process. [22] provide an overview of the literature and known results.

Negative τ

For $\tau < 0$ the model exhibits mean repulsion rather than reversion. The ergodic hypothesis is invalid and no stationary wealth distribution exists. The population splits into those above the mean and those below the mean. Whereas in RGBM with non-negative τ it is impossible for wealth to become negative, negative τ leads to negative wealth. No longer is total economic wealth a limit to the wealth of the richest individual because the poorest develop large negative wealth. The wealth of the rich in the population increases exponentially away from the mean, and the wealth of the poor becomes negative and exponentially large in magnitude, see Figure 4. Qualitatively, this echoes the findings that the rich are experiencing higher growth rates of their wealth than the poor [29, 42] and that the cumulative wealth of the poorest 50 percent of the American population was negative during 2008–2013 [33, 37].

Such splitting of the population is a common feature of non-ergodic processes. If rescaled wealth were an ergodic process, then individuals would, over long enough time, experience all parts of its distribution. People would spend 99 percent of their time as "the 99 percent" and 1 percent of their time as "the 1 percent". Therefore, the social mobility implicit in models that assume ergodicity might not exist in reality if that assumption is invalid. That inequality and immobility have been linked [12, 21, 5] may be unsurprising if both are viewed as consequences of non-ergodic wealth or income.

2.1.5 Derivation of the stable distribution

{section:RGBM_stable}

In transforming the differential equation for wealth, x, into a differential equation for rescaled wealth, y, we use the approximation $\langle x(t) \rangle_N = \langle x(0) \rangle_N e^{\mu t}$. In other words we assume that the population-average wealth grows like the expectation value of wealth.

It is known that this approximation is invalid for long times [28]. Specifically, over long times $\langle x(t) \rangle_N$ grows at the exponential rate $\mu - \sigma^2/2$, whereas the expectation value, $\langle x(t) \rangle$, grows at the exponential rate μ.

This raises the question for how long our approximation is valid. The answer depends on the sample size N, as it must because the expectation value is just the $N \to \infty$ limit of the population average. To assess whether the fluctuations in $\langle x(t) \rangle_N$ are important, we compare the variance of $\langle x(t) \rangle_N$ to the expectation value squared (equivalent to comparing the standard deviation to the expectation value). If the variance is smaller than the expectation value squared, then the approximation is acceptable. If this is not the case, then we cannot use this approximation.

The calculations that use the approximation relate to properties of the stationary distribution. This exists for τ above some positive threshold, *i.e.* with sufficiently strong reallocation. The coupling of wealth trajectories through reallocation lengthens the timescale over which the population average resembles the expected wealth. Therefore, we are on safe ground if we can show that the timescale on which the approximation is valid when $\tau = 0$ is longer than practically relevant timescales. This is a sufficient condition for the approximation to be valid when $\tau > 0$.

This means we work with a population of N independent GBMs, which

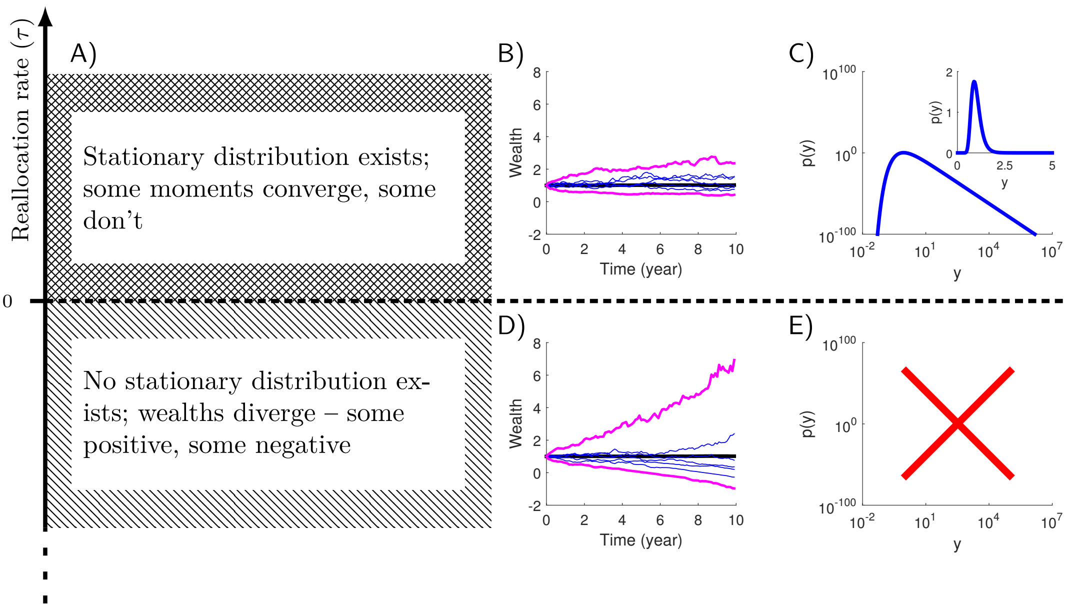

Figure 4: Regimes of RGBM. A) $\tau = 0$ separates the two regimes of RGBM. For $\tau > 0$, a stationary wealth distribution exists. For $\tau < 0$, no stationary wealth distribution exists and wealths diverge. B) Simulations of RGBM with $N = 1000$, $\mu = 0.021$ year^{-1} (presented after rescaling by $e^{\mu t}$), $\sigma = 0.14$ year$^{-1/2}$, $x_i(0) = 1$, $\tau = 0.15$ year^{-1}. Magenta lines: largest and smallest wealths, blue lines: five randomly chosen wealth trajectories, black line: sample mean. C) The stationary distribution to which the system in B) converges. Inset: same distribution on linear scales. D) Similar to B), with $\tau = -0.15$ year^{-1}. E) in the $\tau < 0$ regime, no stationary wealth distribution exists. {fig:regimes}

makes the computation of the variance easy. GBM is log-normally distributed,

$$\ln[x(t)] \sim \ln\mathcal{N}\ln\left(x(0) + \left[\mu - \frac{\sigma^2}{2}\right]t, \sigma^2 t\right). \tag{66}$$

From this, it follows that the expectation value of a single trajectory grows as

$$\langle x(t)\rangle = x(0)\,e^{\mu t}, \tag{67}$$

and the variance grows as

$$V[x(t)] = x(0)^2\,e^{2\mu t}\left[e^{\sigma^2 t} - 1\right]. \tag{68}$$

Because the wealth trajectories are independent, the variance of an average over N trajectories is one-N^{th} of the variance for the individual trajectory. We are now in a position to compare standard deviation and average as follows

$$\frac{V\left[\langle x(t)\rangle_N\right]}{\langle\langle x(t)\rangle_N\rangle^2} = \frac{e^{\sigma^2 t} - 1}{N}. \tag{69}$$

So long as this is less than one, $\langle x(t)\rangle$ is a good approximation for $\langle x(t)\rangle_N$. Rearranging and taking $N \gg 1$ gives an upper bound on the time for which this approximation holds:

$$t < t_c \equiv \frac{\ln N}{\sigma^2}. \tag{70}$$

For typical parameter values in our model, $N = 10^8$ and $\sigma = 0.16\,\text{year}^{-1/2}$, we find $t_c \approx 700$ years. It turns out that, strictly speaking, the minimum value of τ required for the stationary distribution to exist is not zero but proportional to the inverse of this timescale [7], *i.e.*

$$\tau > \tau_c \equiv \frac{\sigma^2}{2\ln N}. \tag{71}$$

{eq:tauc}

In essence, the inequality-increasing effects of multiplicative growth drive wealths apart on the timescale t_c, whereas the inequality-reducing effects of reallocation drive wealths back together on the timescale $1/\tau$. Thus, τ_c marks the point at which reallocation overcomes the forces that drive wealths apart, leading to a stationary distribution of rescaled wealth. In our case $\tau_c \approx 0.0007$ year^{-1}, which is in practice indistinguishable from zero.

The timescale t_c was derived assuming that everyone starts out equally, which is not generally the case. If the initial distribution of wealth is very unequal, then the variance of $\langle x \rangle_N$ will be dominated by the fluctuations experienced by the wealthiest individuals, and the approximation $\langle x \rangle_N \approx \langle x(0) \rangle_N e^{\mu t}$ becomes invalid more quickly. We confirmed numerically that the effect is negligible for our study: our results are indistiguishable whether we simulate dx (which requires an estimate for μ), or dy (where μ does not appear but the above-mentioned approximation is made).

We start again with the SDE for the rescaled wealth,

$$dy = \sigma\, y\, dW - \tau\,(y-1)\,dt. \tag{72}$$

This is an Itô equation with drift term $A = \tau(y-1)$ and diffusion term $B = y\sigma$.

Such equations imply ordinary second-order differential equations that describe the evolution of the pdf, called Fokker-Planck equations. The Fokker-Planck equation describes the change in probability density, at any point in (relative-wealth) space, due to the action of the drift term (like advection in a fluid) and due to the diffusion term (like heat spreading). In this case, we have

$$\frac{dp\,(y,t)}{dt} = \frac{\partial}{\partial y}\left[Ap\,(y,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left[B^2 p\,(y,t)\right]. \tag{73}$$

The steady-state Fokker-Planck equation for the pdf $p\,(y)$ is obtained by setting the time derivative to zero,

$$\frac{\sigma^2}{2}\left(y^2 p\right)_{yy} + \tau\left[(y-1)\,p\right]_y = 0. \tag{74}$$

{eq:fokker_planck}

Positive wealth subjected to continuous-time multiplicative dynamics with non-negative reallocation can never reach zero. Therefore, we solve Equation (74) with boundary condition $p\,(0) = 0$ to give

$$p\,(y) = C\,(\zeta)\, e^{-\frac{\zeta-1}{y}}\, y^{-(1+\zeta)}, \tag{75}$$

where

$$\zeta = 1 + \frac{2\tau}{\sigma^2} \tag{76}$$

and

$$C(\zeta) = \frac{(\zeta - 1)^{\zeta}}{\Gamma(\zeta)}, \tag{77}$$

with the gamma function $\Gamma(\zeta) = \int_0^\infty x^{\zeta-1} e^{-x}\, \mathrm{d}x$. The distribution has a power-law tail as $y \to \infty$, resembling Pareto's often confirmed observation that the frequency of large wealths tends to decay as a power law. The exponent of the power law, ζ, is called the Pareto parameter and is one measure of economic inequality.

2.1.6 Moments and convergence times

{RGBM_moments}

Our key finding is that under currently prevailing economic conditions it is not safe to assume the existence of stationary wealth distributions in models of wealth dynamics. Nevertheless, we present some results for the regime of our model where a stationary distribution exists. The full form of the distribution is derived in Appendix **??**. Because it has a power-law tail for large wealths, only the lower moments of the distribution exist, while higher moments diverge. Below, we derive a condition for the convergence of the variance and calculate its convergence time.

The variance of y is a combination of the first moment, $\langle y \rangle$ (the average), and the second moment, $\langle y^2 \rangle$:

$$V(y) = \langle y^2 \rangle - \langle y \rangle^2 \tag{78}$$

We thus need to find $\langle y \rangle$ and $\langle y^2 \rangle$ in order to determine the variance. The first moment of the rescaled wealth is, by definition, $\langle y \rangle = 1$.

To find the second moment, we start with the SDE for the rescaled wealth:

$$dy = \sigma y \, dW - \tau (y - 1)\, dt. \tag{79}$$

{eq:rescaledSDE}

This is an Itô process, which implies that an increment, df, in some (twice-differentiable) function $f(y, t)$ will also be an Itô process, and such increments can be found by Taylor-expanding to second order in dy as follows:

$$df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial y} dy + \frac{1}{2} \frac{\partial^2 f}{\partial y^2} dy^2 . \tag{80}$$

We insert $f(y, t) = y^2$ and obtain

$$d\left(y^2\right) = 2y dy + (dy)^2 . \tag{81}$$

{eq:diff2}

We substitute dy in Equation (81), which yields terms of order dW, dt, dW^2, dt^2, and $(dW\, dt)$. The scaling of Brownian motion allows us to replace dW^2 by dt, and we ignore $o(dt)$ terms. This yields

$$d\left(y^2\right) = 2\sigma y^2 \, dW - \left(2\tau - \sigma^2\right) y^2 \, dt + 2\tau y \, dt \tag{82}$$

Taking expectations on both sides, and using $\langle y \rangle = 1$, produces an ordinary differential equation for the second moment:

$$\frac{d\langle y^2 \rangle}{dt} = -\left(2\tau - \sigma^2\right) \langle y^2 \rangle + 2\tau \tag{83}$$

{eq:avediff2}

with solution

$$\langle y(t)^2 \rangle = \frac{2\tau}{2\tau - \sigma^2} + \left(\langle y(0)^2 \rangle - \frac{2\tau}{2\tau - \sigma^2} \right) e^{-\left(2\tau - \sigma^2\right)t} . \tag{84}$$ {eq:avediff3}

The variance $V(t) = \langle y(t)^2 \rangle - 1$ therefore follows

$$V(t) = V_\infty + (V_0 - V_\infty) e^{-\left(2\tau - \sigma^2\right)t} , \tag{85}$$ {eq:var1}

where V_0 is the initial variance and

$$V_\infty = \frac{2\tau}{2\tau - \sigma^2} . \tag{86}$$ {eq:varinf}

V converges in time to the asymptote, V_∞, provided the exponential in Equation (85) is decaying. This can be expressed as a condition on τ

$$\tau > \frac{\sigma^2}{2} . \tag{87}$$

Clearly, for negative values of τ the condition cannot be satisfied, and the variance (and inequality) of the wealth distribution will diverge. In the regime where the variance exists, $\tau > \sigma^2/2$, it also follows from Equation (85) that the convergence time of the variance is $1/\left(2\tau - \sigma^2\right)$.

As τ increases, increasingly high moments of the distribution become convergent to some finite value. The above procedure for finding the second moment (and thereby the variance) can be applied to the k^{th} moment, just by changing the second power y^2 to y^k, and any other cumulant can therefore be found as a combination of the relevant moments. For instance [22] also compute the third cumulant.

2.1.7 United States wealth data

{section:RGBM_data}

Wealth share data We analyze the wealth shares of the top quantiles of the US population, as estimated by three sources using different methods:

- The income tax method ("capitalization method") that uses information on capital income from individual income tax returns to estimate the underlying stock of wealth [34, 37]. "If we can observe capital income $k = rW$, where W is the underlying value of an asset and r is the known rate of return, then we can estimate wealth based on capital income and capitalization factor $1/r$ defined using the appropriate choice of rate of return" [19, p. 54]. Data availability: the wealth shares of the top 5, 0.5, 0.1 and 0.01 percent for 1917–2012 and of the top 10 and 1 percent for 1913–2014 (annually).

- The estate multiplier method that uses data from estate tax returns to estimate wealth for the upper tail of the wealth distribution [20]. "The basic idea is to think of decedents as a sample from the living population. The individual-specific mortality rate m_i becomes the sampling rate. If m_i is known, the distribution for the living population can be simply estimated by reweighting the data for decedents by inverse sampling weights $1/m_i$, which are called 'estate multipliers' " [19, p. 53]. Data availability: the wealth shares of the top 1, 0.5, 0.25, 0.1, 0.05 and 0.01 percent for 1916–2000 (annually, with several missing years).

- The survey-based method that uses data from the Survey of Consumer Finances (SCF) conducted by the Federal Reserve, plus defined-benefit pension wealth, plus the wealth of the members of the Forbes 400 [10]. Data availability: the wealth shares of the top 1 and 0.1 percent for 1989–2013 (for every three years).

These sources are based on different datasets and for different time periods. In the overlapping periods, they sometimes report markedly different wealth share estimates (see Figure 5).

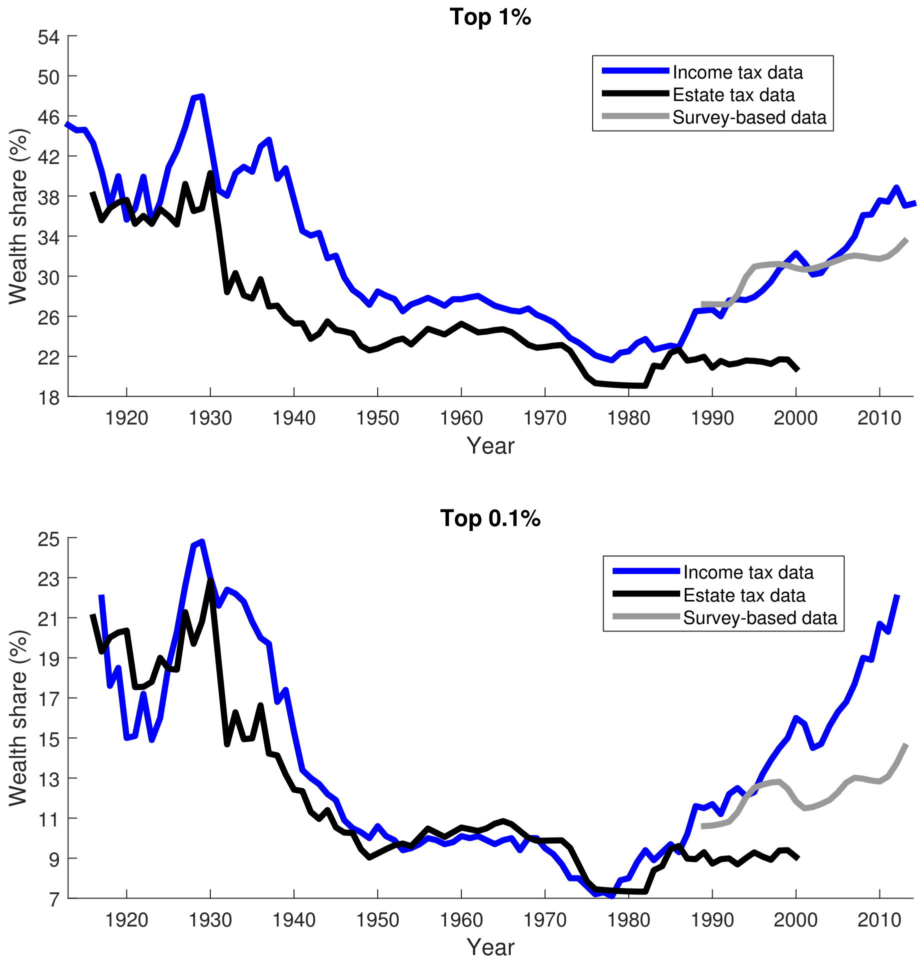

Figure 5: The top wealth shares in the US, 1913–2014. Sources – [34, 37] (blue); [20] (black); [10] (grey). {fig:data1}

[19] reviewed the advantages and disadvantages of the different methods (see also the comment by Kopczuk on [10]). He observed that "the survey-based and estate tax methods suggest that the share of wealth held by the top 1 percent has not increased much in recent decades, while the capitalization method suggests that it has" [19, p. 48].

Which method best reflects the recent trends in wealth inequality is a matter of ongoing debate. Each method suffers from bias. For example, the survey-based method suffers from some underrepresentation of families who belong to the top end of the distribution. The income tax method suffers from some practical difficulties – "not all categories of assets generate capital income that appears on tax returns. [...] Owner-occupied housing does not generate annual taxable capital income" [19, p. 54]. The estate tax method suffers from the need to accurately estimate mortality rates for the wealthy, known to be lower than those for the rest of the population. We refer the reader to [19, 10] for a thorough discussion. We analyze each data source separately.

Wealth Growth Rate We find numerically that the results of our analysis do not depend on μ. This is because wealth shares depend only on the distribution of rescaled wealth and, for $\tau > 0$, it is possible to scale out μ completely from the wealth dynamic to obtain Equation (64) for rescaled wealth. The fitted $\tau < 0$ values we find are not large or persistent enough to make our simulations significantly μ-dependent. However, formally, since we allow negative τ, we must simulate Equation (59) and not Equation (64). This requires us to specify a value of μ, which we estimate as $\mu = 0.021 \pm 0.001$ year^{-1} by a least-squares fit of historical per-capita private wealth in the US [30] to an exponential growth curve.

Volatility We must also specify the volatility parameter, σ, in Equation (59). In principle, this can vary with time. We have no access to real individual wealth trajectories, so we resort to estimating $\sigma(t)$ from other data. We find numerically that our results are not very sensitive to the details, so we need only a good "ballpark" estimate. We obtain that by assuming that the volatility in individual wealths tracks the volatility in the values of the companies that constitute the commercial and industrial base of the national economy. Therefore, for each year, we estimate $\sigma(t)$ as the standard deviation of daily logarithmic changes of the Dow Jones Industrial Average [31], which we annualise by multiplying by $(250/\text{year})^{1/2}$. The values usually lie between 0.1 and 0.2 year$^{-1/2}$, with an average of 0.16 year$^{-1/2}$. Running our empirical analysis with constant σ in this range had little effect on our results (see Appendix **??**) so, for simplicity, we present the analysis using $\sigma(t) = 0.16$ year$^{-1/2}$ for all t.

Fitting σ to stock market data means that we have only one model parameter – the effective reallocation rate, $\tau(t)$ – to fit to the historical wealth shares.

{sec:analysis}

Empirical Analysis The goal of the empirical analysis is to estimate $\tau(t)$ from the historical wealth data, using RGBM as our model. This estimation allows us to address two main questions:

1. Is it valid to assume ergodicity for the dynamics of relative wealth in the US? For the ergodic hypothesis to be valid, fitted values of $\tau(t)$ would have to be robustly positive.

2. If $\tau(t)$ is indeed positive, how long does it take for the distribution to converge to its asymptotic form?

We fit a time series, $\tau(t)$, that reproduces the annually observed wealth shares in the three datasets (see Sec. **??**): Income tax-based [34, 37], estate tax-based [20] and survey-based [10]. The wealth share, S_q, is defined as the proportion of total wealth, $\sum_i^N x_i$, owned by the richest fraction q of the population, *e.g.* $S_{10\%} = 80$ percent means that the richest 10 percent of the population own 80 percent of the total wealth.

For an empirical wealth share time series, $S_q^{\text{data}}(t)$, we proceed as follows.

– Step 1 Initialise N individual wealths, $\{x_i(t_0)\}$, as random variates of the inverse gamma distribution with parameters chosen to match $S_q^{\text{data}}(t_0)$.

– Step 2 Propagate $\{x_i(t)\}$ according to Equation (59) over Δt, using the value of τ that minimises the difference between the wealth share in the modelled population, $S_q^{\text{model}}(t + \Delta t, \tau)$, and $S_q^{\text{data}}(t + \Delta t)$. We use the Nelder-Mead algorithm [24].

– Step 3 Repeat Step 2 until the end of the time series.

We consider historical wealth shares of the richest $q = 10$, 5, 1, 0.5, 0.25, 0.1, 0.05 and 0.01 percent and obtain time series of fitted effective reallocation rates, $\tau_q(t)$, shown in Figure 6. For each value of q we perform a run of the simulation for $N = 10^8$. Since in practice dW is randomly chosen, each run of the simulation will result in slightly different $\tau_q(t)$ values. However, we found that the differences between such calculations are negligible.

Figure 6 (top) shows large annual fluctuations in $\tau_q(t)$. We are interested in longer-term changes in reallocation driven by structural economic and political changes. To elucidate these we smooth the data by taking a central 10-year moving average, $\widetilde{\tau}_q(t)$, where the window is truncated at the ends of the time series. To ensure the smoothing does not introduce artificial biases, we reverse the procedure and use $\widetilde{\tau}_q(t)$ to propagate the initially inverse gamma-distributed $\{x_i(t_0)\}$ and determine the wealth shares $S_q^{\text{model}}(t)$. The good agreement with $S_q^{\text{data}}(t)$ suggests that the smoothed $\widetilde{\tau}_q(t)$ is meaningful, see Figure 6 (bottom).

For the income tax method wealth shares [34], the effective reallocation rate, $\widetilde{\tau}(t)$, has been negative – *i.e.* from poorer to richer – since the mid-1980s. This holds for all of the inequality measures we derived from this dataset.

For the survey-based wealth shares [10], we observe briefer periods in which $\widetilde{\tau}(t) < 0$. The same is true for the estate tax data [20], see Figure 7. When $\tau(t)$ is positive, relevant convergence times are very long compared to the time scales of policy changes, namely at least several decades.

All three datasets indicate that making the ergodic hypothesis is an unwarranted restriction on models and analyses. The hypothesis makes it impossible to observe and reason about the most dramatic qualitative features of wealth dynamics, such as rising inequality, negative reallocation, negative wealth, and social immobility.

Convergence times In the ergodic regime it is possible to calculate how fast the wealth shares of different quantiles converge to their asymptotic value. We do this numerically. Starting with a population of equal wealths and assuming $\mu = 0.021$ year^{-1}, $\sigma = 0.16$ year$^{-1/2}$, and $\tau = 0.04$ year^{-1}, we let the system equilibrate for 3000 years, long enough for the distribution to reach its asymptotic form to numerical precision. We then create a "shock", by changing τ to a different "shock value", and allow the system to equilibrate again for 3000 years,

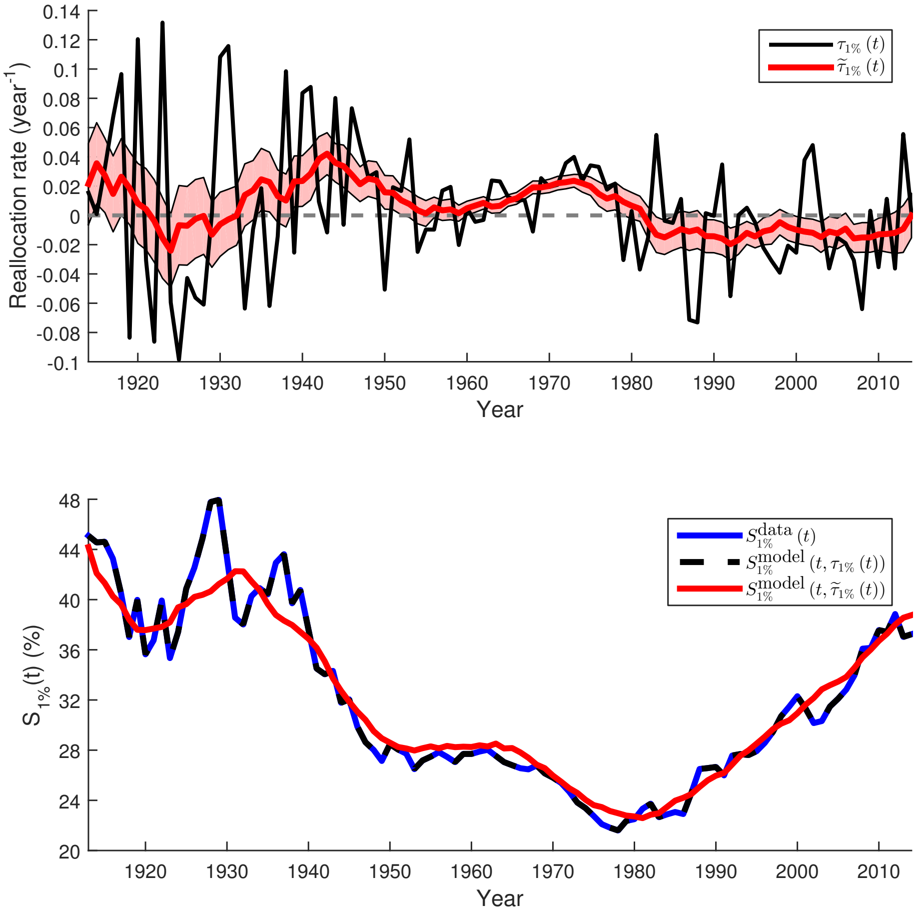

Figure 6: Fitted effective reallocation rates. Calculations done using $\mu = 0.021$ year^{-1} and $\sigma = 0.16$ year$^{-1/2}$. Top: $\tau_{1\%}(t)$ (black) and $\widetilde{\tau}_{1\%}(t)$ (red). Translucent envelopes indicate one standard error in the moving averages. Bottom: $S_{1\%}^{\text{data}}$ (blue), $S_{1\%}^{\text{model}}$ based on the annual $\tau_{1\%}(t)$ (dashed black), based on the 10-year moving average $\widetilde{\tau}_{1\%}(t)$ (red). {fig:tau}

see top panel of Figure 8. Following the shock, the wealth shares converge to their asymptotic values. We fit this convergence numerically with an exponential function and interpret the inverse of the exponential convergence rate as the convergence time. The bottom panel of Figure 8 shows the convergence times versus the shock value of τ.

In addition, it is possible to calculate the convergence time of the variance of the stationary distribution (and other cumulants and moments of interest). In the ergodic regime the stationary distribution has a finite variance only if $\tau > \sigma^2/2$ [22]. Convergence of the actual variance to the stationary variance occurs exponentially over a timescale $1/(2\tau - \sigma^2)$. Figure 9 shows the convergence times for different values of σ. See Appendix **??** for more details.

Convergence times for wealth shares and variance are long, ranging from a few decades to several centuries. This implies that empirical studies which as-

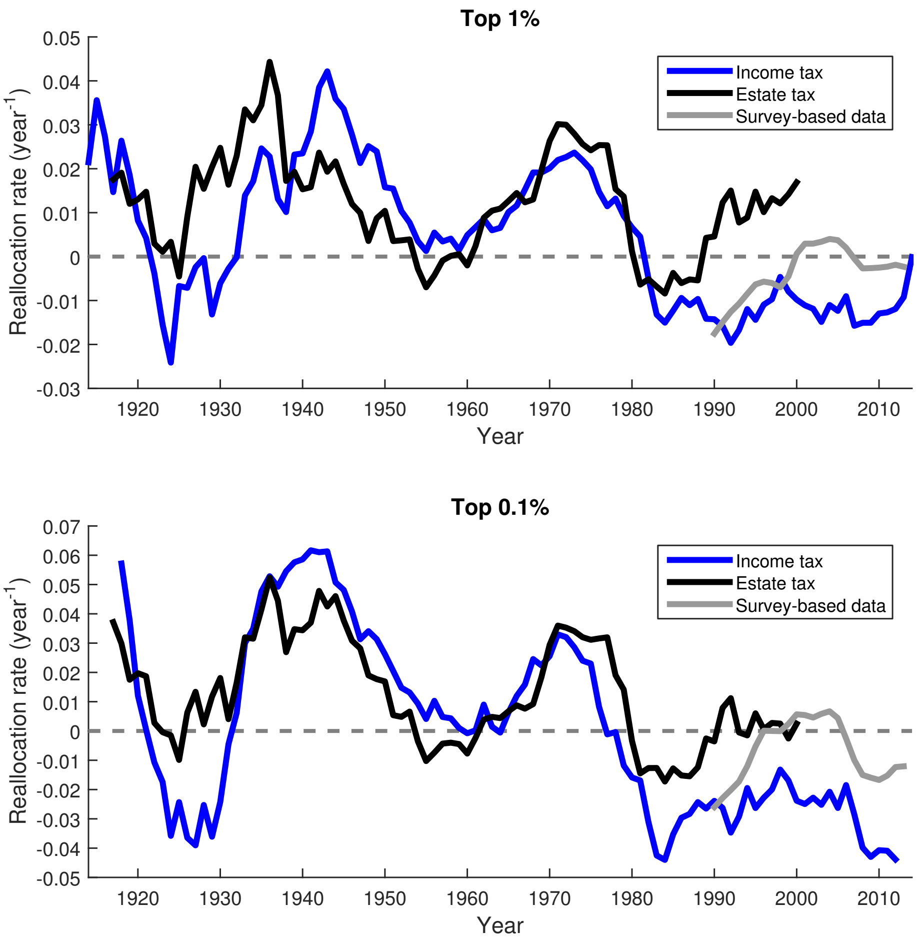

Figure 7: Effective reallocation rates for different datasets. {fig:shares_comp}

sume ergodicity and fast convergence will be inconsistent with the data. To test this, we simulate such a study by performing a different RGBM parameter fit. We find the reallocation rates, $\tau_q^{\text{eqm}}(t)$, that generate stationary distributions consistent with observed wealth shares. In other words, we assume instantaneous convergence.

Figure 10 contrasts $\tau_{1\%}^{\text{eqm}}(t)$ assuming ergodicity with $\widetilde{\tau}_{1\%}(t)$ without assuming ergodicity (using the income tax method dataset). If convergence were always possible and fast, then the two values would be identical within statistical uncertainties. They are not. In addition, the generally large discrepancies between the wealth inequality implied by $\tau_{1\%}^{\text{eqm}}(t)$ (bottom panel, Figure 10, green line) and as observed (bottom panel, Figure 10, blue line) indicate that the wealth distribution does not stay close to its asymptotic form. This means that the long convergence times we calculate are a practical methodological problem for conventional studies.

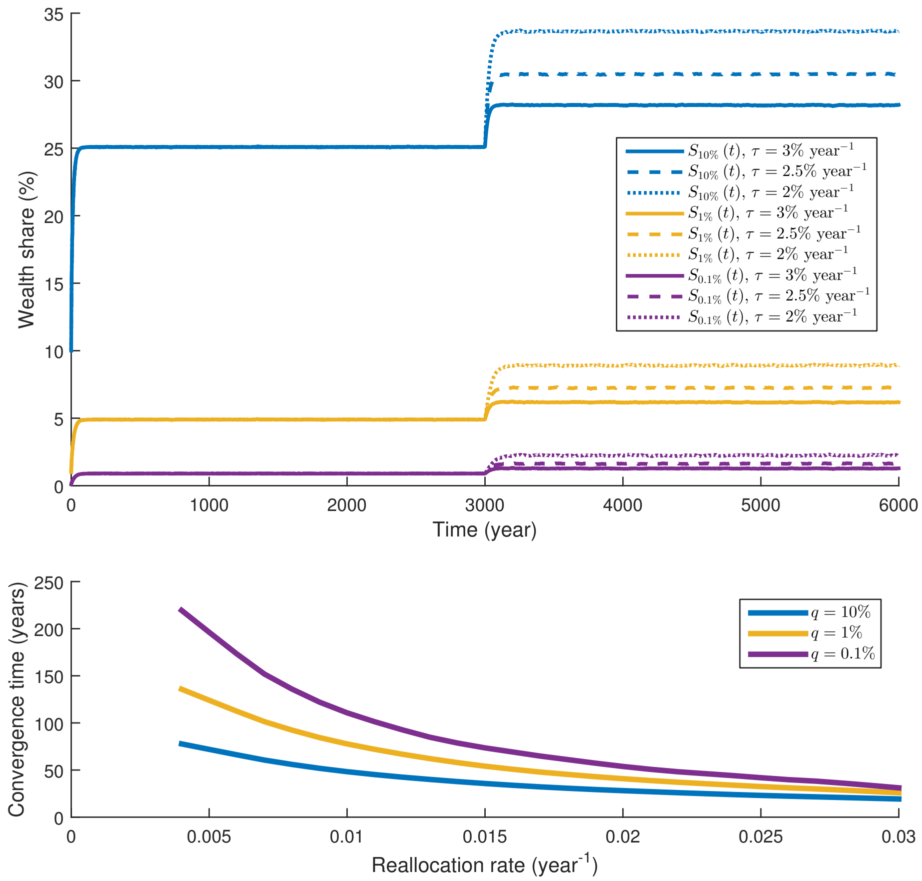

Figure 8: Wealth share convergence time. Top: The convergence of the wealth share for $q = 10$ percent (blue), $q = 1$ percent (yellow) and $q = 0.1$ percent (purple) following a change in the value of τ from 0.04 year^{-1} to 0.03 year^{-1} (solid), 0.025 year^{-1} (dashed) and 0.02 year^{-1} (dotted). Bottom: The wealth share exponential convergence time for $q = 10$ percent (blue), $q = 1$ percent (yellow) and $q = 0.1$ percent (purple) as a function of τ. {fig:conv}

Conclusions Studies of economic inequality often assume ergodicity of relative wealth. This assumption also goes under the headings of equilibrium, stationarity, or stability [1]. Specifically, it is assumed that:

1. the system can equilibrate, *i.e.* a stationary distribution exists to which the observed distribution converges in the long-time limit; and

2. the system equilibrates quickly, *i.e.* the observed distribution gets close to the stationary distribution after a time shorter than other relevant timescales, such as the time between policy changes.

Assumption 2 is often left unstated, but it is necessary for the stationary (model) distribution to resemble the observed (real) distribution. This matters because the stationary distribution is often a key object of study – model parameters are

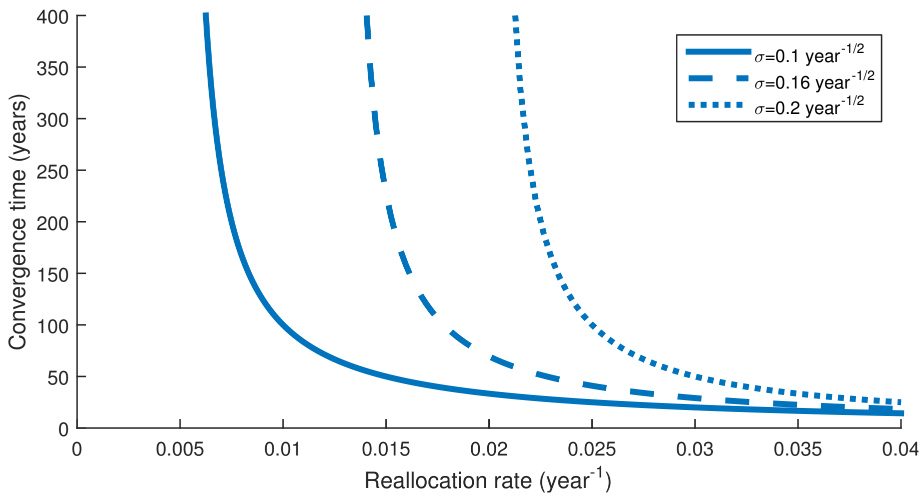

Figure 9: Variance convergence time {fig:convar}

found by fitting the stationary distribution to observed inequality, and effects of various model parameters on the stationary distribution are explored.

We do not assume ergodicity. Fitting τ in RGBM allows the data to speak without constraint as to whether the ergodic hypothesis is valid. We find it to be invalid because:

A. We observe negative τ values in all datasets analyzed, most notably using the income tax method, especially since about 1980. The wealth distribution is non-stationary and inequality increases for as long as these conditions prevail.

B. When we observe positive τ, the associated convergence times are mostly of the order of decades or centuries, see Figure 7 and Figure 8 (bottom). They are much longer than the periods over which economic conditions and policies change – they are the timescales of history rather than of politics.

The ergodic hypothesis precludes what we find. Item A above corresponds to reallocation that moves wealth from poorer to richer individuals, which is inconsistent with the ergodic hypothesis. In this sense the ergodic hypothesis is a set of blindfolds, hiding from view the most dramatic economic conditions. For the most recent data, the system is in a state best described by non-ergodic RGBM, $\tau < 0$ [34, 37] or $\tau \approx 0$ [10]. Therefore, each time we observe the wealth distribution, we see a snapshot of it either in the process of diverging or very far from its asymptotic form. It is much like a photograph of an explosion in space: it will show a fireball whose finite extent tells us nothing of the eventual distance between pieces of debris.

We also find that changes in the earnings distribution do not provide an adequate alternative explanation of the described dynamics of the wealth distribution. Although earnings have become more unequal over the recent decades in which wealth inequality has increased, their effect on the wealth distribution has been small and generally stabilizing rather than destabilizing. Treating

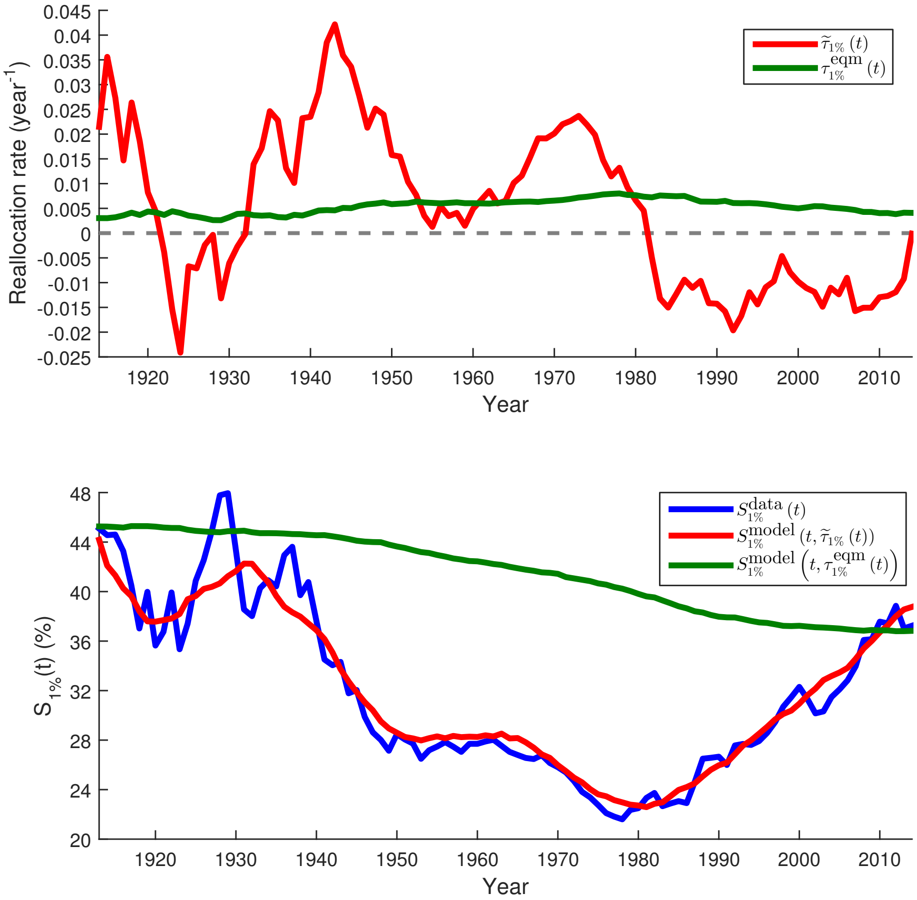

Figure 10: Comparison of dynamic and equilibrium reallocation rates. Top: $\widetilde{\tau}_{1\%}(t)$ (red, same as in the top of Figure 6). $\tau_{1\%}^{\mathrm{eqm}}(t)$ (green), defined such that $\lim_{t'\to\infty} S_{1\%}^{\mathrm{model}}\left(t', \tau_{1\%}^{\mathrm{eqm}}(t)\right) = S_{1\%}^{\mathrm{data}}(t)$. It is impossible by design for this value to be negative. The significant difference between the red and green lines demonstrates that the fast convergence assumption is invalid for the problem under consideration. Bottom: $S_{1\%}^{\mathrm{data}}$ (blue), $S_{1\%}^{\mathrm{model}}$ based on the 10-year moving average $\widetilde{\tau}_{1\%}(t)$ (red), based on $\tau_{1\%}^{\mathrm{eqm}}(t)$ (green). The reallocation rates found under the fast convergence assumption generate model wealth shares which bear little relation to reality. {fig:asymptau}

earnings explicitly in our model does not change fundamentally our conclusions.

The economic phenomena that trouble theorists most – such as diverging inequality, social immobility, and the emergence of negative wealth – are difficult to reproduce in a model that assumes ergodicity. In our simple model, this is easy to see: in the ergodic regime, $\tau > 0$, our model cannot reproduce these phenomena at all. One may be tempted to conclude that their existence is a sign of special conditions prevailing in the real world – collusion and conspiracies. But if we admit the possibility of non-ergodicity, $\tau \leq 0$, it becomes clear that these phenomena can easily emerge in an economy that does not actively guard against them.

List of Symbols

Δ Difference operator, for instance Δv is a difference of two values of v, for instance observed at two different times.

Δt A general time interval..

$g_{\mathbf{est}}$ Growth rate estimator for finite time and finite ensemble size.

$g_{\langle\rangle}$ Exponential growth rate of the expectation value.

$\overline{g}$ Time-average exponential growth rate.

$\mathcal{N}$ Normal distribution, $x \sim \mathcal{N}(\langle p \rangle, \mathrm{var}(p))$ means that the variable p is normally distributed with mean $\langle p \rangle$ and variance $\mathrm{var}(p)$..

p Probability, p_i is the probability of observing event i in a realization of a random variable.

$\mathcal{P}$ Probability density function.

t Time.

v Stationarity mapping function, so that $v(x)$ has stationary increments.

var Variance.

W Wiener process, $W(t) = \int_0^t dW$ is continuous and $W(t) \sim \mathcal{N}(0, \overline{g})$.

x Wealth.

References

[1] A. Adamou and O. Peters. Dynamics of inequality. *Significance*, 13(3):32–35, 2016.

[2] J. Aitchison and J. A. C. Brown. *The lognormal distribution*. Cambridge University Press, 1957.

[3] A. B. Atkinson. The timescale of economic models: How long is the long run? *The Review of Economic Studies*, 36(2):137–152, 1969.

[4] J. Benhabib, A. Bisin, and S. Zhu. The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica*, 79(1):123–157, 2011.

[5] Y. Berman. Understanding the mechanical relationship between inequality and intergenerational mobility. Available at: http://papers.ssrn.com/abstract=2796563, 2017.

[6] Y. Berman, O. Peters, and A. Adamou. An empirical test of the ergodic hypothesis: Wealth distributions in the United States. January 2017.

[7] J.-P. Bouchaud. Note on mean-field wealth models and the random energy model.

[8] J.-P. Bouchaud. On growth-optimal tax rates and the issue of wealth inequalities. *http://arXiv.org/abs/1508.00275*, August 2015.

[9] J.-P. Bouchaud and M. Mézard. Wealth condensation in a simple model of economy. *Physica A*, 282(4):536–545, 2000.

[10] J. Bricker, A. Henriques, J. Krimmel, and J. Sabelhaus. Measuring income and wealth at the top using administrative and survey data. *Brookings Papers on Economic Activity*, pages 261–331, 2016.

[11] M. Brzezinski. Do wealth distributions follow power laws? evidence from 'rich lists'. *Physica A: Statistical Mechanics and its Applications*, 406:155–162, 2014.

[12] M. Corak. Income inequality, equality of opportunity, and intergenerational mobility. *The Journal of Economic Perspectives*, 27(3):79–102, 2013.

[13] B. Derrida. Random-energy model: Limit of a family of disordered models. *Phys. Rev. Lett.*, 45(2):79–82, July 1980.

[14] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613–2626, September 1981.

[15] A. Drăgulescu and V. M. Yakovenko. Exponential and power-law probability distributions of wealth and income in the united kingdom and the united states. *Physica A: Statistical Mechanics and its Applications*, 299(1):213–221, 2001.

[16] X. Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1(1):255–294, 2009.

[17] O. S. Klass, O. Biham, M. Levy, O. Malcai, and S. Solomon. The forbes 400 and the pareto wealth distribution. *Economics Letters*, 90(2):290–295, 2006.

[18] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 1992.

[19] W. Kopczuk. What do we know about the evolution of top wealth shares in the united states? *The Journal of Economic Perspectives*, 29(1):47–66, 2015.

[20] W. Kopczuk and E. Saez. Top wealth shares in the united states, 1916–2000: Evidence from estate tax returns. *National Tax Journal*, 57(2):445–487, 2004.

[21] K. Liu, N. Lubbers, W. Klein, J. Tobochnik, B. Boghosian, and H. Gould. The effect of growth on equality in models of the economy. *arXiv*, 2013.

[22] Z. Liu and R. A. Serota. Correlation and relaxation times for a stochastic process with a fat-tailed steady-state distribution. *Physica A: Statistical Mechanics and its Applications*, 474:301–311, 2017.

[23] J. E. Meade. *Efficiency, Equality and The Ownership of Property*. Allen & Unwin, London, UK, 1964.

[24] J. A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

[25] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.*, 46(5):323–351, 2005.

[26] V. Pareto. *Cours d'économie Politique*. F. Rouge, Lausanne, Switzerland, 1897.

[27] O. Peters and M. Gell-Mann. Evaluating gambles using dynamics. *Chaos*, 26:23103, February 2016.

[28] O. Peters and W. Klein. Ergodicity breaking in geometric Brownian motion. *Phys. Rev. Lett.*, 110(10):100603, March 2013.

[29] T. Piketty. *Capital in the twenty-first century*. Harvard University Press, 2014.

[30] T. Piketty and G. Zucman. Capital is back: Wealth-income ratios in rich countries, 1700-2010. *The Quarterly Journal of Economics*, 129(3):1255–1310, 2014.

[31] Quandl. Dow Jones Industrial Average. http://www.quandl.com/data/BCB/UDJIAD1-Dow-Jones-Industrial-Average, 2016. Accessed: 04/19/2016.

[32] S. Redner. Random multiplicative processes: An elementary tutorial. *Am. J. Phys.*, 58(3):267–273, March 1990.

[33] J.-V. Rios-Rull and M. Kuhn. 2013 Update on the U.S. Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics. *Quarterly Review*, (April):1–75, 2016.

[34] E. Saez and G. Zucman. Wealth inequality in the united states since 1913: Evidence from capitalized income tax data. Technical report, National bureau of economic research, 2014.

[35] P. Samuelson. What classical and neoclassical monetary theory really was. *Canad. J. Econ.*, 1(1):1–15, February 1968.

[36] A. Sen. *On Economic Inequality*. Oxford: Clarendon Press, 1997.

[37] The World Wealth and Income Database. Usa top 10% and 1% and bottom 50% wealth shares, 1913–2014. http://wid.world/data/, 2016. Accessed: 12/26/2016.

[38] H. Theil. *Economics and information theory*. North-Holland Publishing Company, 1967.

[39] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Physical Review*, 36(5):823–841, Sep 1930.

[40] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

[41] P. Vermeulen. How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, 2017.

[42] E. N. Wolff. Household wealth trends in the united states, 1983–2010. *Oxford Review of Economic Policy*, 30(1):21–43, 2014.