

Ergodicity Economics

Ole Peters and Alexander Adamou



2018/06/08 at 16:59:17

Contents

1	Populations	3
1.1	Every man for himself	4
1.1.1	Log-normal wealth distribution	4
1.1.2	Inequality measure from two growth rates	6
1.2	Sums of lognormals and the random energy model	8
1.2.1	Introduction – why study GBM?	8
1.2.2	Non-ergodicity of GBM	9
1.2.3	PEAs in GBM – a sketch	10
1.2.4	Our previous work on PEAs	11
1.2.5	Mapping to the random energy model	13
1.2.6	Another route via Itô calculus	17
1.2.7	Discussion	18
1.3	Cooperation	18
1.4	Reallocation	22
1.4.1	Introduction	22
1.4.2	Model	24
1.4.3	Model positioning and definition	25
1.4.4	Model behaviour	27
2	Data	29
2.0.1	Wealth share data	29
2.0.2	Wealth Growth Rate	30
2.0.3	Volatility	31
2.0.4	Empirical Analysis	32
2.0.5	Convergence times	34
2.0.6	The effect of earnings	36
2.0.7	Conclusions	39
2.0.8	The derivation of self-averaging time t_c	41
2.0.9	The derivation of the stationary distribution	43
2.0.10	The effect of fixed versus time-varying σ	44
2.0.11	The derivation of the variance convergence time	46
2.1	Taxation	47
2.1.1	Wealth tax	48
2.1.2	Income tax	51
	List of Symbols	53
	References	54

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, *i.e.* a set of trajectories $x(t)$ or a family of random variables parameterized by t , 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 an individual trajectory for a long time, and the PDF of the random variable $x(t^*)$ at some fixed value of t^* is the wealth 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 $F_x(x) = \int_x^\infty ds \mathcal{P}_x(s)$. In a large population of N individuals, $\Delta x \mathcal{P}_x(x) N$ is the approximate number of individuals who have wealth near 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 centered at the mean of x , that is

$$\mathcal{P}_x(x) = \delta(x - \langle x \rangle_N); \quad (1)$$

- Maximum inequality would mean that one individual owns everything and everyone else owns nothing, that is

$$\mathcal{P}_x(x) = \frac{N-1}{N} \delta(x-0) + \frac{1}{N} \delta(x - N \langle x \rangle_N). \quad (2)$$

1.1.1 Log-normal wealth distribution

{section:Log-normal_wealt}

GBM is log-normally distributed. If each individual’s wealth follows GBM,

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

with solution

$$x(t) = x_0 \exp \left[\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W(t) \right], \quad (4)$$

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

$$\ln x(t) \sim \mathcal{N} \left(\ln x_0 + \left(\mu - \frac{\sigma^2}{2} \right) t, \sigma^2 t \right). \quad (5) \quad \{\text{eq:lognormal}\}$$

We notice that the variance of $\ln x$ increases linearly in time – we will develop an understanding of this fact shortly. As we will see (though we cannot conclude

this yet), 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

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

INSERT PROPERTIES OF GBM DISTRIBUTION HERE (note that log-normal is an insufficient description of the distribution of a GBM - the time-dependence of the random variable must be included)

It is a well-established empirical observation [37] that the upper tails of real wealth distributions tend to 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

$$\mathcal{P}_x(x) = x^{-\alpha}, \quad (7) \quad \{\text{eq:power_law}\}$$

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 behavior 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 $\langle x \rangle$, *i.e.* in the upper tail where approximate power law behavior has been observed empirically? The logarithm of (Eq. 6) is

$$\ln \mathcal{P}(x) = -\ln x \sqrt{2\pi\sigma^2 t} - \frac{[\ln(x) - (\mu - \frac{\sigma^2}{2})t]^2}{2\sigma^2 t} \quad (8)$$

$$= -\ln x - \frac{\ln(2\pi\sigma^2 t)}{2} - \frac{(\ln(x))^2 + [(\mu - \frac{\sigma^2}{2})t]^2 - 2(\mu - \frac{\sigma^2}{2})t \ln x}{2\sigma^2 t} \quad (9)$$

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

$$\ln \mathcal{P}(x) = [\ln x]^2 \times \frac{-1}{2\sigma^2 t} + \ln x \times \left(-\frac{3}{2} + \frac{\mu}{\sigma^2}\right) - \frac{\ln(2\pi\sigma^2 t)}{2} - \frac{[(\mu - \frac{\sigma^2}{2})t]^2}{2\sigma^2 t} \quad (10)$$

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

$$\frac{d \ln \mathcal{P}(x)}{d \ln x} = -\frac{\ln x}{\sigma^2 t} - \frac{3}{2} + \frac{\mu}{\sigma^2}. \quad (11)$$

If the first term is small compared to the others, this distribution will look like a power law when plotted on double-logarithmic scales. We don't believe that the

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 chimes with so many qualitative features of empirical observations.

1.1.2 Inequality measure from two growth rates

{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 we often only want summary measures of the distribution. A distributional property of particular interest to economists is inequality. How much inequality is there in a distribution like (Eq. 5), and in what sense does this quantity increase in time under GBM as we have pointed out? Clearly, we should quantify “inequality”. In this section 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 [57], is the difference between typical wealth (that grows at the time-average growth rate) and average wealth (that grows at the ensemble-average growth rate) in our model. Thus, the difference between the time average and ensemble average, the essence of ergodicity breaking, fundamentally drives 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 has more desirability 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 has more inequality 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 $N \langle x \rangle$. The growth rate of the expectation value, $g_{\langle x \rangle}$, 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, \bar{g} , 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 wealth that account for the difference. This suggests the following measure of inequality.

Definition inequality $J(t)$, is the quantity that grows in time at

the rate of the difference between expectation-value and time-average growth rates,

$$\frac{dJ}{dt} = g_{\langle} - \bar{g}. \quad (12) \quad \{\text{eq:dJ}\}$$

Equation (12) defines the dynamic of inequality, and inequality itself is found by integrating over time

$$J(t) = \int_0^t ds [g_{\langle}(s) - \bar{g}(s)]. \quad (13) \quad \{\text{eq:J}\}$$

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 down the two growth rates explicitly and find

$$\frac{dJ}{dt} = \frac{d \ln \langle x \rangle}{dt} - \frac{d \langle \ln x \rangle}{dt}. \quad (14) \quad \{\text{eq:J_dyn}\}$$

Integrating over time

$$J(t) = \ln \langle x \rangle - \langle \ln x \rangle, \quad (15)$$

which is known as the mean logarithmic deviation (MLD) or Theil's second index of inequality [57]. This is rather remarkable. Our general inequality measure, (Eq. 13), evaluated for the specific case of GBM, turns out to be a well-known measure of inequality that economists identified independently, without considering non-ergodicity and ensemble average and time average growth rates. Merely by insisting of measuring inequality well, Theil implicitly used the GBM model!

The expectation value $\langle \ln x \rangle = \ln x_0 + \left(\mu - \frac{\sigma^2}{2}\right)t$ is given by (Eq. 5). It differs from the logarithm of the expectation value, $\langle \ln x \rangle \neq \ln \langle x \rangle$, which we now compute. In order to do this we introduce a useful trick that will come in handy again in Sec. 2.1.1 (the general procedure is described in [29, Chapter 4.2]): to compute moments, $\langle x^n \rangle$, of stochastic differential equations for x , like (Eq. 83), we find solvable ordinary differential equations for the moments. For the first moment we do this simply by taking expectations of both sides of (Eq. 83). The noise term disappears, and we turn the SDE for x into an ODE for $\langle x \rangle$

$$\langle dx \rangle = \langle x(\mu dt + \sigma dW) \rangle \quad (16)$$

$$d \langle x \rangle = \langle x \rangle \mu dt + \sigma \overbrace{\langle dW \rangle}^{=0} \quad (17)$$

$$= \langle x \rangle \mu dt. \quad (18)$$

This is a (very simple) first order linear differential equation for the first moment (*i.e.* the expectation value) of x . Its solution is

$$\langle x \rangle = \langle x_0 \rangle \exp(\mu t) \quad (19) \quad \{\text{eq:exp_x}\}$$

so that $\ln \langle x \rangle = \ln \langle x_0 \rangle + \mu t$. Now we can carry out the differentiations on the RHS of (Eq. 14) and find the Theil inequality as a function of time

$$J(t) = J(0) + \frac{\sigma^2}{2}t. \quad (20)$$

Inequality increases indefinitely. This result implies an 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.* [13]. In the present case an arbitrarily large share of total wealth will be owned by an arbitrarily small share of the population.

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” [54, 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,

$$y = \frac{x}{\langle x \rangle_N}, \quad (21) \quad \{\text{eq:rescaled}\}$$

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. Equation (14), may be expressed in terms of y as $\frac{dJ}{dt} = -\frac{d(\ln y)}{dt}$.

1.2 Sums of lognormals and the random energy model

{section:REM}

1.2.1 Introduction – why study GBM?

The basic phenomenon of self-reproduction occurs from the smallest chemical compound capable of building copies of itself to arguably the most complex known social system, namely the global economy. Not only is the economy made by humans who self-reproduce and whose cells self-reproduce but capitalism itself taps into the power of self-reproduction: capital means resources which can be deployed in order to generate resources, which can be deployed in order. . . . Realism usually requires models that include noise, to account for the multitude of effects not explicitly modelled. GBM is a simple, intuitive, and analytically tractable model of noisy multiplicative growth. A deep understanding of this basic model of self-reproduction is crucial and only in the last few decades have we achieved this.

A quantity, x , of self-reproducing resources (such as biomass or capital) follows GBM if it evolves over time, t , according to the Itô stochastic differential equation,

$$dx = x(\mu dt + \sigma dW_t). \quad (22) \quad \{\text{eq:GBM}\}$$

dt denotes the infinitesimal time increment and dW_t the infinitesimal increment in a Wiener process, which is a normal variate with $\langle dW_t \rangle = 0$ and $\langle dW_t dW_s \rangle = \delta(t - s)dt$. μ and σ are constant parameters called the drift and volatility. Put simply, relative changes in resources, dx/x , are assumed to be

drawn independently from a stationary normal distribution at each time step. (Eq. 83) represents the continuous-time limit of this process. Its solution,

$$x(t) = x(0) \exp \left[\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W(t) \right], \quad (23) \quad \{\text{eq:GBM_sol}\}$$

yields exponential growth at a noisy rate. Hereafter, we shall assume $x(0) = 1$ and neglect it, except where retaining it is illustrative. The distribution of $x(t)$ is a time-dependent log-normal,

$$\ln(x(t)) \sim \mathcal{N} \left(\left(\mu - \frac{\sigma^2}{2} \right) t, \sigma^2 t \right), \quad (24)$$

whose mean, median, and variance grow (or decay) exponentially in time:

$$\langle x(t) \rangle = \exp(\mu t); \quad (25) \quad \{\text{eq:mean_x}\}$$

$$\text{median}(x(t)) = \exp[(\mu - \sigma^2/2)t]; \quad (26) \quad \{\text{eq:median_x}\}$$

$$\text{var}(x(t)) = \exp(2\mu t)[\exp(\sigma^2 t) - 1]. \quad (27) \quad \{\text{eq:var_x}\}$$

GBM is a standard model in finance for self-reproducing quantities such as stock prices. Since relative changes are modelled as normal variates, the central limit theorem means that GBM is an attractor process for a large class of multiplicative dynamics. Any quantity whose relative changes are random variables with finite mean and variance will behave like a GBM after a sufficiently long time. We work with GBM because it is standard and general. It exemplifies important qualitative and universal features of multiplicative growth.

Specifically, we are interested in GBM as a model of economic resources, owned by some entity. We will think of x as measured in currency units, such as dollars.

1.2.2 Non-ergodicity of GBM

The non-ergodicity of this growth process manifests itself in an intriguing way as a difference between the growth of the expectation value of x and the growth of x over time. Imagine a world where people's wealth follows (Eq. 83). In such a world each person's wealth grows exponentially at rate

$$g_t = \mu - \sigma^2/2 \quad (28) \quad \{\text{eq:gt}\}$$

with probability 1 if we observe the person's wealth for a long time. The expectation value of each person's wealth grows exponentially at

$$g_{\langle \rangle} = \mu. \quad (29) \quad \{\text{eq:gave}\}$$

The expectation value, by definition, is the average over an ensemble of N realizations of x in the limit $N \rightarrow \infty$. Important insights follow: the aggregate wealth in our model economy does not grow at the same rate as an individual's wealth [1] (meaning that GDP may be a flawed reflection of national economic well-being); inequality grows indefinitely [13, 8] (even without interactions between individuals); and pooling and sharing resources accelerates growth [42, 12].

1.2.3 PEAs in GBM – a sketch

{section:sketch}

The following question often emerges in applications of these results. Equation (28) and (Eq. 29) are limiting cases ($t \rightarrow \infty$ and $N \rightarrow \infty$, respectively): what happens when time and population size are finite? In [44] we studied the “partial ensemble average” (PEA) of GBM, details of which are in Sec. 1.2.4. The PEA is the sample mean of N independent realisations of GBM:

$$\langle x(t) \rangle_N \equiv \frac{1}{N} \sum_{i=1}^N x_i(t). \quad (30)$$

Here we sketch out some simple arguments about how this object depends on N and t .

Considering (Eq. 28) and (Eq. 29) together, we expect the following tension:

- A) for large N , the PEA should resemble the expectation value, $\exp(\mu t)$;
- B) for long t , all trajectories 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,

$$R \equiv \frac{\text{var}(\langle x(t) \rangle_N)}{\langle \langle x(t) \rangle_N \rangle^2}. \quad (31)$$

To be explicit, here the $\langle \cdot \rangle$ and $\text{var}(\cdot)$ 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. 25) and (Eq. 27), we get

$$R(N) = \frac{e^{\sigma^2 t} - 1}{N}. \quad (32)$$

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

$$t < \frac{\ln N}{\sigma^2}. \quad (33) \quad \{\text{eq:short_t}\}$$

This hand-waving tells us roughly when the large-sample (or, as we see from (Eq. 33), short-time) self-averaging regime holds. A more careful estimate of the cross-over time in (Eq. 49) 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 $t \rightarrow \infty$ 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, but as time goes by the distribution broadens exponentially. Because each trajectory contributes with the same weight to the PEA, after some time the PEA will be dominated by the maximum in the sample,

$$\langle x(t) \rangle_N \approx \frac{1}{N} \max_{i=1}^N \{x_i(t)\}, \quad (34)$$

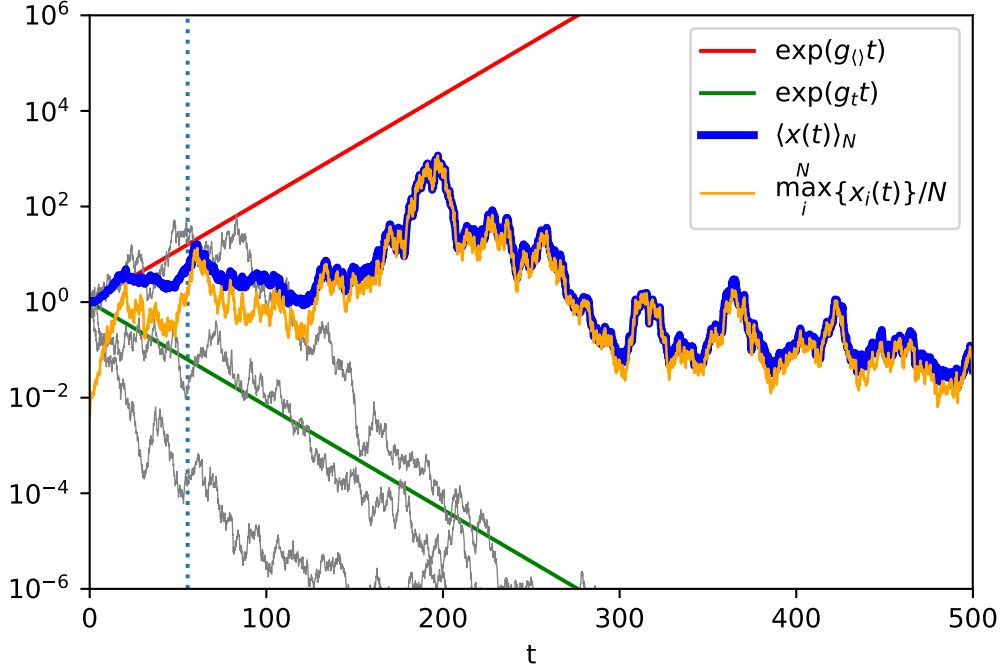


Figure 1: PEA and maximum in a finite ensemble of size $N = 256$. **Red line:** expectation value $\langle x(t) \rangle$. **Green line:** exponential growth at the time-average growth rate. In the $T \rightarrow \infty$ limit all trajectories grow at this rate. **Yellow line:** contribution of the maximum 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}

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. Of course, this takes longer for larger samples, which have more chances to contain a lucky trajectory.

1.2.4 Our previous work on PEAs

{section:prevwork}

In [44] we analysed PEAs of GBM analytically and numerically. Using (Eq. 23) the PEA can be written as

$$\langle x \rangle_N = \frac{1}{N} \sum_{i=1}^N \exp \left[\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_i(t) \right], \quad (35) \quad \{\text{eq:PEA}\}$$

where $\{W_i(t)\}_{i=1\dots N}$ are N independent realisations of the Wiener process. Taking the deterministic part out of the sum we re-write (Eq. 35) as

$$\langle x \rangle_N = \exp \left[\left(\mu - \frac{\sigma^2}{2} \right) t \right] \frac{1}{N} \sum_{i=1}^N \exp \left(t^{1/2} \sigma \xi_i \right), \quad (36) \quad \{\text{eq:PEA_2}\}$$

where $\{\xi_i\}_{i=1\dots N}$ 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 sketch in Sec. 1.2.3. After this time, typical PEA-trajectories begin to deviate from expectation-value behavior, and eventually their growth rate converges to g_t . While the two limiting behaviours $N \rightarrow \infty$ and $t \rightarrow \infty$ can be computed exactly, what happens in between is less straight-forward. The PEA is a random object outside these limits. In [44] we dealt with this issue numerically by creating a super-sample of S samples, each consisting of N trajectories. In this way we were able to study the median of $\langle x \rangle_N(t)$, representing the behavior of typical trajectories.

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

$$g_{\text{est}}(t, N) \equiv \frac{\ln(\langle x(t) \rangle_N) - \ln(x(0))}{t - 0} = \frac{1}{t} \ln(\langle x(t) \rangle_N). \quad (37) \quad \{\text{eq:gest}\}$$

In [44] we proved that the $t \rightarrow \infty$ limit for any (finite) N is the same as for the case $N = 1$,

$$\lim_{t \rightarrow \infty} g_{\text{est}}(t, N) = \mu - \frac{\sigma^2}{2} \quad (38) \quad \{\text{eq:gest_2}\}$$

for all $N \geq 1$. Substituting (Eq. 36) in (Eq. 37) produces

$$g_{\text{est}}(t, N) = \mu - \frac{\sigma^2}{2} + \frac{1}{t} \ln \left(\frac{1}{N} \sum_{i=1}^N \exp(t^{1/2} \sigma \xi_i) \right) \quad (39)$$

$$= \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} + \frac{1}{t} \ln \left(\sum_{i=1}^N \exp(t^{1/2} \sigma \xi_i) \right). \quad (40) \quad \{\text{eq:gest_4}\}$$

We didn't look in [44] 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 $N \rightarrow \infty$ limit of (Eq. 37). Instead it is the $S \rightarrow \infty$ limit,

$$\langle g_{\text{est}}(t, N) \rangle = \frac{1}{t} \langle \ln(\langle x(t) \rangle_N) \rangle = f(N, t), \quad (41) \quad \{\text{eq:gest_3}\}$$

where, as in Sec. 1.2.3, $\langle \cdot \rangle$ without subscript refers to the average over all possible samples, *i.e.* $\lim_{S \rightarrow \infty} \langle \cdot \rangle_S$. The last two terms in (Eq. 40) 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. 41) will hinge. This term will be the focus of our attention: the sum of exponentials of normal random variates or, equivalently, log-normal variates.

1.2.5 Mapping to the random energy model

Since the publication of [44] we have learned, thanks to discussions with J.-P. Bouchaud, that the key object in (Eq. 40) – the sum of exponentials of 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 [20, 21]. It is defined as follows. Imagine a system whose energy levels are $2^K = N$ random numbers ξ_i (corresponding to $K = \ln N / \ln 2$ 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 just model its energy levels as realizations of a random variable. (We denote the number of spins by K and the number of resulting energy levels by N , whereas Derrida uses N for the number of spins). The partition function is then

$$Z = \sum_{i=1}^N \exp \left(\beta J \sqrt{\frac{K}{2}} \xi_i \right), \quad (42) \quad \{\text{eq:Z}\}$$

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 [20, p. 79]. J is a constant that will be determined below. The logarithm of the partition function gives the Helmholtz free energy,

$$F = -\frac{\ln Z}{\beta} \quad (43)$$

$$= -\frac{1}{\beta} \ln \left[\sum_{i=1}^N \exp \left(\beta J \sqrt{\frac{K}{2}} \xi_i \right) \right]. \quad (44) \quad \{\text{eq:F}\}$$

Like the growth rate estimator in (Eq. 37), this involves a sum of log-normal variates and, indeed, we can rewrite (Eq. 40) as

$$g_{\text{est}} = \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} - \frac{\beta F}{t}, \quad (45) \quad \{\text{eq:g est}_5\}$$

which is valid provided that

$$\beta J \sqrt{\frac{K}{2}} = \sigma t^{1/2}. \quad (46) \quad \{\text{eq:map}\}$$

Equation (46) 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:

$$\sigma = \frac{J}{\sqrt{2}} \quad \text{and} \quad t^{1/2} = \beta \sqrt{K}. \quad (47) \quad \{\text{eq:choice}_1\}$$

The expectation value of g_{est} is interesting. The only random object in (Eq. 45) 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,

$$\frac{1}{\beta_c} \equiv \frac{J}{2\sqrt{\ln 2}}, \quad (48) \quad \{\text{eq:beta_c}\}$$

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

$$t_c = \frac{2 \ln N}{\sigma^2}, \quad (49) \quad \{\text{eq:t_c}\}$$

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. 49) scales identically with N and σ as the transition time, (Eq. 33), in our sketch.

In [20], $\langle F \rangle$ is computed in the high-temperature (short-time) regime as

$$\langle F \rangle = E - S/\beta \quad (50)$$

$$= -\frac{K}{\beta} \ln 2 - \frac{\beta K J^2}{4}, \quad (51) \quad \{\text{eq:F_2}\}$$

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

$$\langle F \rangle = -KJ\sqrt{\ln 2}. \quad (52) \quad \{\text{eq:F_3}\}$$

Short time

We look at the short-time behavior first (high $1/\beta$, (Eq. 51)). The relevant computation of the entropy S in [20] 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. 51) in (Eq. 45) we find

$$\langle g_{\text{est}} \rangle^{\text{short}} = \mu - \frac{\sigma^2}{2} + \frac{1}{t} \frac{K J^2}{4T^2}. \quad (53) \quad \{\text{eq:gest_6}\}$$

From (Eq. 46) we know that $t = \frac{K J^2}{2\sigma^2 T^2}$, which we substitute, to find

$$\langle g_{\text{est}} \rangle^{\text{short}} = \mu, \quad (54) \quad \{\text{eq:gest_7}\}$$

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. 52)). Again taking expectation values and substituting, this time from (Eq. 52) in (Eq. 45), we find for long times

$$\langle g_{\text{est}} \rangle^{\text{long}} = \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} + \sqrt{\frac{2 \ln N}{t}} \sigma, \quad (55) \quad \{\text{eq:gest_8}\}$$

which has the correct long-time asymptotic behavior. The form of the correction to the time-average growth rate in (Eq. 55) is consistent with [44] and [50], 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. 55) indicates that the long-time regime is relevant.

Figure 2 is a direct comparison between the results derived here, based on [20], and numerical results using the same parameter values as in [44], 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 [44] 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

$$q_{0.5}(\langle x(t) \rangle_N) \approx \exp(\langle g_{\text{est}}(N, t) \rangle). \quad (56) \quad \{\text{eq:quant_ave}\}$$

In [43] it was discussed in detail that $g_{\text{est}}(1, t)$ is an ergodic observable for (Eq. 83), in the sense that $\langle g_{\text{est}}(1, t) \rangle = \lim_{t \rightarrow \infty} g_{\text{est}}$. The relationship in (Eq. 56) is far more subtle. The typical behavior of GBM PEAs is complicated outside the limits $N \rightarrow \infty$ or $t \rightarrow \infty$, in the sense that growth rates are time dependent here. This complicated behavior is well represented by an approximation that uses physical insights into spin glasses. Beautiful!

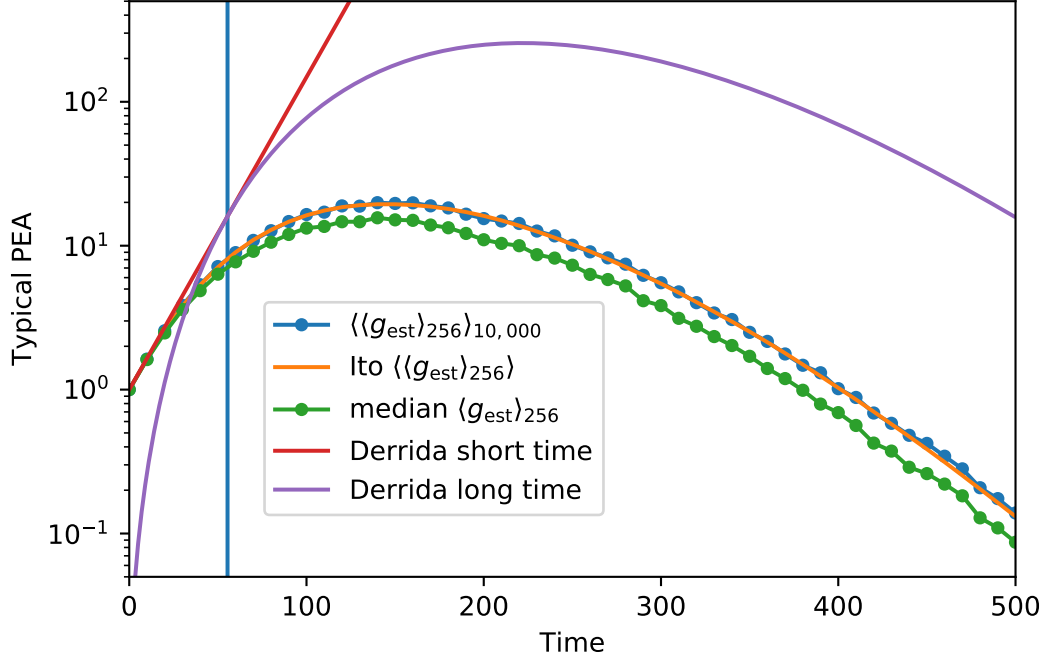


Figure 2: Lines are obtained by exponentiating the various exponential growth rates. **Blue line:** $\langle\langle g_{\text{est}} \rangle_{256} \rangle_{10,000}$ is the numerical mean (approximation of the expectation value) over a super-ensemble of $S = 10,000$ samples of g_{est} estimated in sub-ensembles 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. 66) 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. 54). **Purple line:** long-time behavior, based on the random energy model, (Eq. 55). **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}

1.2.6 Another route via Itô calculus

Another way to find the expectation value of PEA growth rates, (Eq. 41), is to compute $\langle d \ln \langle x \rangle_N \rangle$ using Itô calculus. We compute this directly, without invoking the random energy model. To apply Itô calculus, we will need the first two partial derivatives of $d \ln \langle x \rangle_N$ with respect to x_i .

$$\frac{\partial \ln \langle x \rangle_N}{\partial x_i} = \frac{1}{N \langle x \rangle_N} \quad (57)$$

and

$$\frac{\partial^2 \ln \langle x \rangle_N}{\partial x_i^2} = -\frac{1}{N^2 \langle x \rangle_N^2} \quad (58)$$

Now, Taylor-expanding $d \ln \langle x \rangle_N$ we find

$$d \ln \langle x \rangle_N = \sum_i \frac{\partial \ln \langle x \rangle_N}{\partial x_i} dx_i + \frac{1}{2} \sum_i \sum_j \frac{\partial^2 \ln \langle x \rangle_N}{\partial x_i \partial x_j} dx_i dx_j + \dots \quad (59)$$

$$\approx \frac{1}{N \langle x \rangle_N} \sum_i dx_i - \frac{1}{2N^2 \langle x \rangle_N^2} \sum_i \sum_j dx_i dx_j. \quad (60)$$

The double-sum can be split into ‘diagonal’ ($i = j$) terms and cross-terms as

$$\frac{1}{N \langle x \rangle_N} \sum_i x_i (\mu dt + \sigma dW_i) - \frac{1}{2N^2 \langle x \rangle_N^2} \left(\sum_i dx_i^2 + \sum_j \sum_{i \neq j} dx_i dx_j \right) \quad (61)$$

Parts of the cross-terms are negligible because they are of order dt^2 and the rest vanishes when taking the expectation value, as we see by writing out one cross term

$$dx_i dx_j = x_i x_j (\mu^2 dt^2 + \mu \sigma dt dW_i + \mu \sigma dt dW_j + \sigma^2 dW_i dW_j). \quad (62)$$

We therefore drop these terms now, as we take the expectation value, using the Wiener identity $\langle dW_i^2 \rangle = dt$ for the final term

$$\langle d \ln \langle x \rangle_N \rangle = \left\langle \mu dt - \frac{1}{2N^2 \langle x \rangle_N^2} \sum_i x_i^2 (\mu^2 dt^2 + \sigma^2 dt) \right\rangle \quad (63)$$

$$= \mu dt - \left\langle \frac{1}{2} \sigma^2 \frac{\langle x^2 \rangle_N}{N \langle x \rangle_N^2} \right\rangle dt + O(dt^2). \quad (64)$$

Discarding terms of higher-than-first order in dt and re-writing the last term as a fraction of sums, we thus have

$$\frac{\langle d \ln \langle x \rangle_N \rangle}{dt} = \mu - \frac{1}{2} \sigma^2 \left\langle \frac{\frac{1}{N} \sum_i x_i^2}{N \left(\frac{1}{N} \sum_i x_i \right)^2} \right\rangle \quad (65)$$

$$= \mu - \frac{1}{2} \sigma^2 \left\langle \frac{\sum_i x_i^2}{(\sum_i x_i)^2} \right\rangle \quad (66) \quad \{\text{eq:Ito_sums}\}$$

This expression has the correct behaviour: for short times, all x_i are essentially identical, and the second term is $-\frac{1}{2} \sigma^2 \frac{1}{N}$, which is negligible if N is large. So, for short times we see expectation-value behaviour. For long times, the largest x_i will dominate both the numerator and the denominator, and we have $-\frac{1}{2} \sigma^2$: the full Itô correction is felt for long times.

1.2.7 Discussion

The Itô result is exact. A Monte-Carlo estimate of (Eq. 66) (which is easy to obtain) is shown in Fig. 2 (yellow line). This agrees well with numerical observations. The approximations from the random energy model have the right shape and asymptotic behavior, though they're not on the same scale as the median PEA. This is, of course, not surprising because these estimates are not designed to coincide with the median PEA. Quantitatively they are closer to a higher quantile of the distribution of PEAs. An intriguing question is this: is our computation using Itô calculus helpful to compute the expected free energy of the random energy model?

1.3 Cooperation

{section:Cooperation}

Under multiplicative growth, fluctuations are undesirable because they reduce time-average growth rates. In the long run, wealth $x_1(t)$ with noise term σ_1 will outperform wealth $x_2(t)$ with a larger noise term $\sigma_2 > \sigma_1$, in the sense that

$$\bar{g}(x_1) > \bar{g}(x_2) \quad (67)$$

with probability 1.

For this reason it is desirable to reduce fluctuations. One protocol that achieves this is resource pooling and sharing. In Sec. 1.1 we explored the world created by the model of independent GBMs. This is a world where everyone experiences the same long-term growth rate. We want to explore the effect of the invention of cooperation. As it turns out cooperation increases growth rates, and this is a crucial insight.

Suppose two individuals, $x_1(t)$ and $x_2(t)$ decide to meet up every Monday, put all their wealth on a table, divide it in two equal amounts, and go back to their business, *i.e.* they submit their wealth to our toy dynamic (Eq. 83). How would this operation affect the dynamic of the wealth of these two individuals?

Consider a discretized version of (Eq. 83), such as would be used in a numerical simulation. The non-cooperators grow according to

$$\delta x_i(t) = x_i(t) \left[\mu \delta t + \sigma \sqrt{\delta t} \xi_i \right], \quad (68) \quad \{\text{eq:discrete_nonc_grow}\}$$

$$x_i(t + \delta t) = x_i(t) + \delta x_i(t), \quad (69) \quad \{\text{eq:discrete_nonc_coop}\}$$

where ξ_i are standard normal random variates, $\xi_i \sim \mathcal{N}(0, 1)$.

We imagine that the two previously non-cooperating entities, with resources $x_1(t)$ and $x_2(t)$, cooperate to produce two entities, whose resources we label $x_1^c(t)$ and $x_2^c(t)$ to distinguish them from the non-cooperating case. We envisage equal sharing of resources, $x_1^c = x_2^c$, and introduce a cooperation operator, \oplus , such that

$$x_1 \oplus x_2 = x_1^c + x_2^c. \quad (70)$$

In the discrete-time picture, each time step involves a two-phase process. First there is a growth phase, analogous to (Eq. 83), in which each cooperator increases its resources by

$$\delta x_i^c(t) = x_i^c(t) \left[\mu \delta t + \sigma \sqrt{\delta t} \xi_i \right]. \quad (71) \quad \{\text{eq:discrete_coop_grow}\}$$

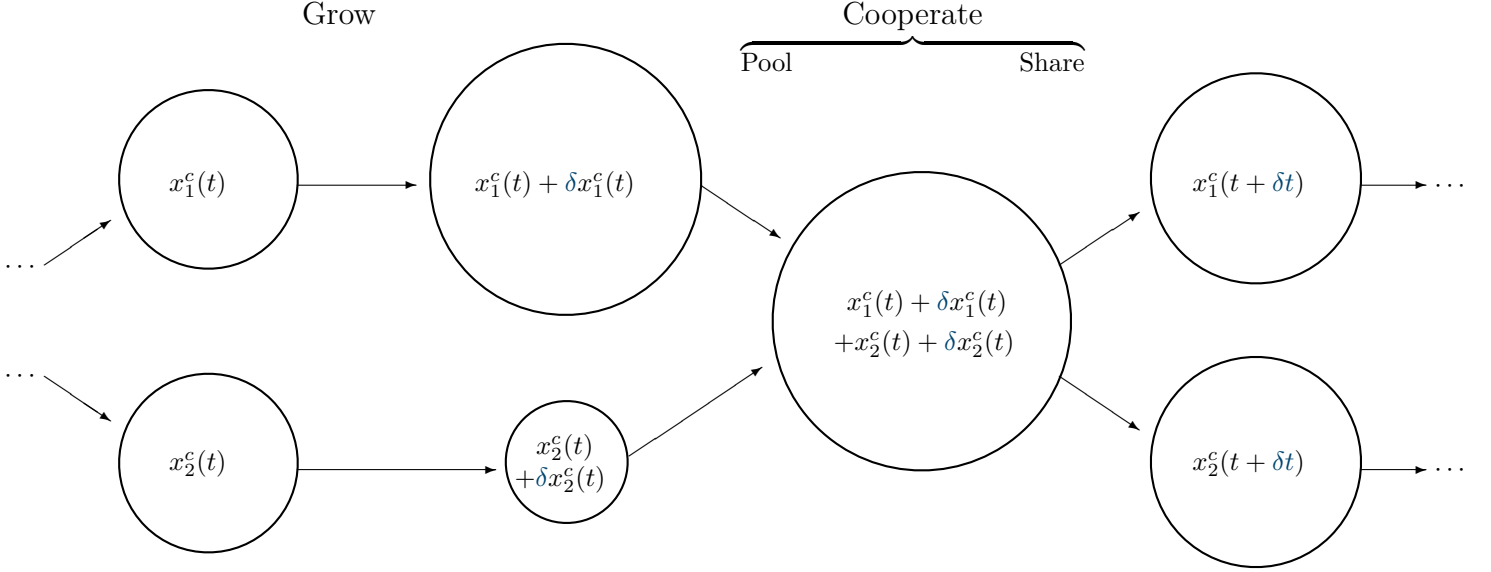


Figure 3: Cooperation dynamics. Cooperators start each time step with equal resources, then they *grow* independently according to (Eq. 71), then they *co-operate* by *pooling* resources and *sharing* them equally, then the next time step begins.

{fig:dynamics}

This is followed by a cooperation phase, replacing (Eq. 69), in which resources are pooled and shared equally among the cooperators:

$$x_i^c(t + \delta t) = \frac{x_1^c(t) + \delta x_1^c(t) + x_2^c(t) + \delta x_2^c(t)}{2}. \quad (72) \quad \{\text{eq:discrete_coop_coop}\}$$

With this prescription both cooperators and their sum experience the following dynamic:

$$(x_1 \oplus x_2)(t + \delta t) = (x_1 \oplus x_2)(t) \left[1 + \left(\mu \delta t + \sigma \sqrt{\delta t} \frac{\xi_1 + \xi_2}{2} \right) \right]. \quad (73) \quad \{\text{eq:discrete_cooperate}\}$$

For ease of notation we define

$$\xi_{1 \oplus 2} = \frac{\xi_1 + \xi_2}{\sqrt{2}}, \quad (74)$$

which is another standard Gaussian, $\xi_{1 \oplus 2} \sim \mathcal{N}(0, 1)$. Letting the time increment $\delta t \rightarrow 0$ we recover an equation of the same form as (Eq. 83) but with a different fluctuation amplitude,

$$d(x_1 \oplus x_2) = (x_1 \oplus x_2) \left(\mu dt + \frac{\sigma}{\sqrt{2}} dW_{1 \oplus 2} \right). \quad (75)$$

The expectation values of a non-cooperator, $\langle x_1(t) \rangle$, and a corresponding cooperator, $\langle x_1^c(t) \rangle$, are identical. Based on expectation values, we thus cannot see any benefit of cooperation. Worse still, immediately after the growth phase,

the better-off entity of a cooperating pair, $x_1^c(t_0) > x_2^c(t_0)$, say, would increase its expectation value from $\frac{x_1^c(t_0) + x_2^c(t_0)}{2} \exp(\mu(t - t_0))$ to $x_1^c(t_0) \exp(\mu(t - t_0))$ by breaking the cooperation. But it would be foolish to act on the basis of this analysis – the short-term gain from breaking cooperation is a one-off, and is dwarfed by the long-term multiplicative advantage of continued cooperation. An analysis based on expectation values finds that there is no reason for cooperation to arise, and that if it does arise there are good reasons for it to end, *i.e.* it will be fragile. Because expectation values are inappropriately used to evaluate future prospects, the observation of widespread cooperation constitutes a conundrum.

The solution of the conundrum comes from considering the time-average growth rate. The non-cooperating entities grow at $g_t(x_i) = \mu - \frac{\sigma^2}{2}$, whereas the cooperating unit benefits from a reduction of the amplitude of relative fluctuations and grows at $g_t(x_1 \oplus x_2) = \mu - \frac{\sigma^2}{4}$, and we have

$$g_t(x_1 \oplus x_2) > g_t(x_i) \quad (76)$$

for any non-zero noise amplitude. Imagine a world where cooperation does not exist, just like in Sec. 1.1. Now introduce into this world two individuals who have invented cooperation – very quickly this pair of individuals will be vastly more wealthy than anyone else. To keep up, others will have to start cooperating. The effect is illustrated in Fig. 4 by direct simulation of (Eq. 68)–(Eq. 69) and (Eq. 73).

Imagine again the pair of cooperators outperforming all of their peers. Other entities will have to form pairs to keep up, and the obvious next step is for larger cooperating units to form – groups of 3 may form, pairs of pairs, cooperation clusters of n individuals, and the larger the cooperating group the closer the time-average growth rate will get to the expectation value. For n cooperators, $x_1 \oplus x_2 \dots \oplus x_n$ the spurious drift term is $-\frac{\sigma^2}{2n}$, so that the time-average growth approaches expectation-value growth for large n . The approach to this upper bound as the number of cooperators increases favours the formation of social structure.

We may generalise to different drift terms, μ_i , and noise amplitudes, σ_i , for different individual entities. Whether cooperation is beneficial in the long run for any given entity depends on these parameters as follows. Entity 1 will benefit from cooperation with entity 2 if

$$\mu_1 - \frac{\sigma_1^2}{2} < \frac{\mu_1 + \mu_2}{2} - \frac{\sigma_1^2 + \sigma_2^2}{8}. \quad (77)$$

We emphasize that this inequality may be satisfied also if the expectation value of entity 1 grows faster than the expectation value of entity 2, *i.e.* if $\mu_1 > \mu_2$. An analysis of expectation values, again, is utterly misleading: the benefit conferred on entity 1 due to the fluctuation-reducing effect of cooperation may outweigh the cost of having to cooperate with an entity with smaller expectation value.

Notice the nature of the Monte-Carlo simulation in Fig. 4. No ensemble is constructed. Only individual trajectories are simulated and run for a time that is long enough for statistically significant features to rise above the noise. This method teases out of the dynamics what happens over time. The significance of any observed structure – its epistemological meaning – is immediately clear: this is what happens over time for an individual system (a cell, a person’s wealth, *etc.*). Simulating an ensemble and averaging over members to remove noise does

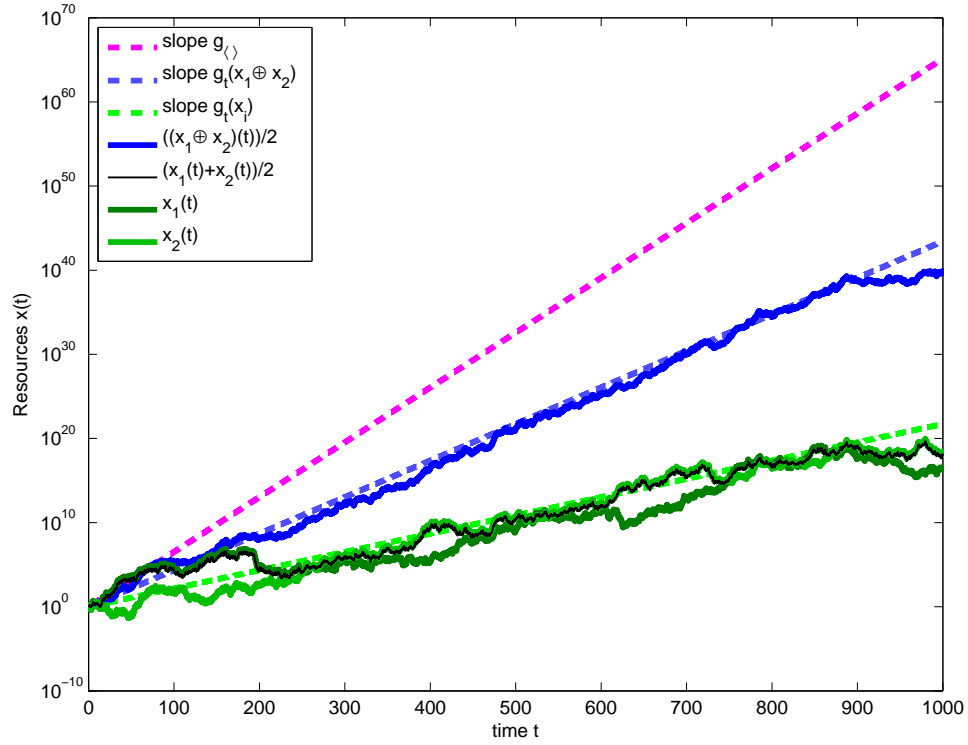


Figure 4: Typical trajectories for two non-cooperating (green) entities and for the corresponding cooperating unit (blue). Over time, the noise reduction for the cooperator leads to faster growth. Even without effects of specialisation or the emergence of new function, cooperation pays in the long run. The black thin line shows the average of the non-cooperating entities. While in the logarithmic vertical scale the average traces the more successful trajectory, it is far inferior to the cooperating unit. In a very literal mathematical sense the whole, $(x_1 \oplus x_2)(t)$, is more than the sum of its parts, $x_1(t) + x_2(t)$. The algebra of cooperation is not merely that of summation.

{fig:cooperate}

not tell the same story. The resulting features may not emerge over time. They are what happens on average in an ensemble, but – at least for GBM – this is not what happens to the individual with probability 1. For instance the pink dashed line in Fig. 4 is the ensemble average of $x_1(t)$, $x_2(t)$, and $(x_1 \oplus x_2)(t)/2$, and it has nothing to do with what happens in the individual trajectories over time.

When judged on expectation values, the apparent futility of cooperation is unsurprising because expectation values are the result for infinitely many cooperators, and adding further cooperators cannot improve on this.

In our model the advantage of cooperation, and hence the emergence of social structure in the broadest sense – is purely a non-linear effect of fluctuations – cooperation reduces the magnitude of fluctuations, and over time (though not in expectation) this implies faster growth.

Another generalisation is partial cooperation – entities may share only a proportion of their resources, resembling taxation and redistribution. We discuss this in the next section.

1.4 Reallocation

1.4.1 Introduction

{reallocation}

{sec:introduction}

Paul Samuelson identified the “ergodic hypothesis” in the mindset of classical economic theorists, defining it as “a belief in unique long-run equilibrium independent of initial conditions” [53, pp. 11-12]. Classically, the equilibrium is studied and the transient phenomena preceding it are ignored. This approach simplifies analysis greatly and is widespread. Here we present an empirical test of the validity of the ergodic hypothesis as applied to the distribution of wealth. We find it inconsistent with historical data from the United States (US), raising doubts about the conclusions of studies in which it is made.

Economics is often concerned with growth. A growing quantity cannot be ergodic in Samuelson’s sense because it does not tend to an equilibrium value. Naïvely, then, the ergodic hypothesis seems an inappropriate modelling choice in this context, ruling out a swathe of analytical techniques. Although this is rarely stated explicitly, a common strategy to salvage these techniques is the following: find a transformation of the non-ergodic process that produces a meaningful ergodic observable. If such an ergodic observable can be derived, then classical techniques may still be useful. For instance, [43] show that utility theory can be viewed as an attempt to transform non-ergodic growing wealth into ergodic growth rates. 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, which is hypothesised to be ergodic. Formally, the distribution of the stochastic process for the rescaled wealth is assumed to converge to a unique and time-independent distribution. For example, [5] “impose assumptions [...] that guarantee the existence and uniqueness of a limit stationary distribution” (p. 130). Similar formulations appear in [55, 10, 46, 18, 19, 26]. These studies take advantage of the simplicity with which the stationary distribution can then be analyzed, *e.g.* to test the effects of policies encoded in model parameters.

Treating the stationary model distribution as representative of the empirical wealth distribution implies an assumption of fast convergence, *i.e.* that the actual distribution approaches its asymptotic form on a timescale shorter than that of relevant changes in economic conditions. For example, the top rate of income tax in the US was reduced from 91 percent to 70 percent between 1963 and 1965. The next change occurred about 20 years later, in 1982, when it was reduced to 50 percent. In this example, convergence would be fast if the model distribution were close to its asymptotic form after, say, 5 years. Convergence would be slow if this took, say, 100 years. [2, p. 137] argued that “the speed of convergence makes a great deal of difference to the way in which we think about the model”. It determines the practical relevance of the stationary distribution [2, 17].

Recent research has marked a shift away from studying the stationary distribution in isolation. [24, 7, 27, 9] study the dynamics of wealth and income, rather than how they are distributed asymptotically. Some of these studies also consider convergence times [24, 9].

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. Here we present such a test. For the test to be meaningful, a model is required that does not assume ergodicity from the outset. It must have regions of parameter space (“regimes”) in which the ergodic hypothesis for rescaled wealth does and does not hold.

Our model satisfies this condition. Its basis is that individual wealth undergoes noisy exponential growth. Social structure is represented by a wealth reallocation mechanism in which a fraction of everyone’s wealth is pooled and shared. The sign of the reallocation rate parameter determines whether a stationary distribution exists: if positive (corresponding to reallocation from richer to poorer) then it does; if negative (from poorer to richer) then it doesn’t.

We estimate model parameters using three different datasets of historical wealth shares in the US population. These estimated parameters – and not an *a priori* assumption – tell us whether a stationary distribution exists. When it does, it has a Pareto tail, consistent with data and standard models [38, 22]. We evaluate convergence times, when convergence is possible, to assess whether the ergodic hypothesis is acceptable in practice.

Our findings invalidate the ergodic hypothesis. The fitted reallocation rate is not robustly positive for any dataset we analyze. Indeed, for one dataset we find it to be consistently negative for the last thirty years or so. We cannot overstate our surprise at this finding. Most theorists would consider a model in which individual wealths grow independently, *i.e.* with no reallocation, an extreme and unrealistic model of an advanced Western economy with socio-political institutions and infrastructure. We would expect to infer consistent positive reallocation from data for such an economy. We find the opposite: that, from the 1980s to the present, the US economy is best described in our model as one in which wealth is systematically reallocated from poorer to richer.

Other datasets yield reallocation rates for which a stationary distribution does exist, but with convergence times of decades or centuries. Whichever data are used, our analysis does not provide the unequivocal endorsement of the ergodic hypothesis that would justify its ubiquitous use in the field. Policy recommendations based on models which assume the existence of a stationary distribution and fast convergence may be ineffective in practice. Worse still, such

inappropriately-constrained models may paint a misleading picture of reality, for example that with taxation and public spending our economies are positively redistributive. This could lead to policy prescriptions which run counter to policy goals.

We also extend our model to treat explicitly additive changes in wealth, akin to labor income and consumption. This allows us to answer the question whether rescaled wealth is inherently unstable, or whether it is inherently stable with increases in its inequality driven by increasingly unequal earnings. We find that earnings had only a small effect on the dynamics of the wealth distribution over the last century and that changes in their distribution do not explain adequately the observed instability in wealth.

Our contribution is threefold. Firstly, we develop a theoretical model which describes the dynamics of wealth and allows us to test empirically the validity of the ergodic hypothesis. Naturally, this cannot be done using models in which the ergodic hypothesis is implicit.

Secondly, we use the model to assess how fast the wealth distribution converges to its asymptotic form, when this exists. We find convergence times greater than the typical intervals between changes in policy and other determinants of the distribution. This casts doubt over the ability of models with fast convergence to interpret observed changes in inequality and yield effective policy.

Thirdly, our study extends a body of theory of economic systems without assuming ergodicity. [43] show that a sound decision theory can be developed by finding ergodic growth rates for non-ergodic wealth processes. In the present context, [1] derive a fundamental measure of wealth inequality from such growth rates. Elsewhere, [40] resolves the St. Petersburg paradox by maximizing individual performance over time and [39] applies the same reasoning to derive an optimal leverage for stock-market investments. [41] extend this work to derive constraints on price fluctuations in freely-traded assets, resolving the equity premium puzzle of [35].

Indiscriminate use of the ergodic hypothesis can mean that non-ergodic processes are analyzed with methods appropriate only for ergodic processes. This is a severe methodological problem, which we argue is behind many open problems in economics. Our work demonstrates the importance of establishing empirically and analytically which observables may be treated legitimately as ergodic.

1.4.2 Model

{sec:model}

We call our model Reallocating Geometric Brownian Motion (RGBM). Individual wealth undergoes random multiplicative growth, modelled as Geometric Brownian Motion (GBM), and is reallocated among individuals by a simple pooling and sharing mechanism. Thus everyone's wealth is coupled to the total wealth in the economy. We view RGBM as a 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.

RGBM has both ergodic and non-ergodic regimes, characterised by the sign of the reallocation rate parameter. Reallocation from richer to poorer produces an ergodic regime, in which wealths are positive, distributed with a Pareto tail, and confined around their mean value. Reallocation from poorer to richer 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. If the reallocation rate is zero, RGBM reduces to GBM, in which individual wealths grow independently and no social structure is represented.

1.4.3 Model positioning and definition

We use the literature review of [19] to place our model in the context of existing work. They introduce “a simple accounting framework, originally due to [34]” (p. 9):

“At birth, individuals are endowed with a, possibly individual-specific, fraction of the contemporaneous, aggregate capital stock. The only source of income in the economy is an idiosyncratic rate of return on individual wealth.

In a given period, individual wealth, normalised by the average capital stock, accrues at the exponential rate

$$r_t^i - g + s_t^i, \quad (78) \quad \{\text{eq:meade}\}$$

where r_t^i is the realised rate of return for individual i of current age t and s_t^i is the ratio between the individual’s flow of (dis)saving and her wealth at the beginning of the period.”

Our model is similar. We denote by $x_i(t)$ the wealth at time t of the i^{th} member of a population of N individuals. We define

$$y_i(t) \equiv \frac{x_i(t)}{\langle x(t) \rangle_N} \quad (79) \quad \{\text{eq:yi}\}$$

as the same individual’s wealth rescaled by the population average,

$$\langle x(t) \rangle_N \equiv \frac{1}{N} \sum_{i=1}^N x_i(t). \quad (80)$$

Equation (78) implies the following differential equation for the rescaled wealth:

$$dy_i = y_i [r_i(t) - g + s_i(t)] dt. \quad (81) \quad \{\text{eq:rescaled}\}$$

[34] assumes that the population-average wealth grows exponentially at rate g . Under this assumption, individual wealth obeys

$$dx_i = x_i [r_i(t) + s_i(t)] dt. \quad (82) \quad \{\text{eq:unscaled}\}$$

The sum $r_i(t) + s_i(t)$ is the time-varying exponential growth rate of individual wealth. [34] interprets this in terms of idiosyncratic rates of return and saving. We offer a simpler interpretation by decomposing $[r_i(t) + s_i(t)] dt$ into:

- μdt – a common, constant part, representing economic growth; and
- $\sigma dW_i(t)$ – an idiosyncratic, random part, representing individual growth.

Taking $dW_i(t)$ to be the increment in a Wiener process (normally distributed with zero mean and variance dt) we arrive at GBM,

$$dx_i = x_i [\mu dt + \sigma dW_i(t)], \quad (83) \quad \{\text{eq:GBM}\}$$

with drift μ and volatility σ . This is our basic model for the evolution of individual wealth in the absence of social structure. Unsurprisingly, Equation (83) is also the most influential model of asset price dynamics in financial economics.

Wealth evolves multiplicatively under GBM. There are no additive changes akin to labor 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. Indeed, empirical distributions exhibit different regularities for low and high wealths [22]. There are other multiplicative effects at play. For instance, investing in one's health, housing, and education may be closer to a multiplicative investment than to wealth-depleting consumption. Nonetheless, we leave a red flag here to indicate that our model's realism is questionable for small wealths. We treat additive earnings explicitly in Sec. ?? and find that including them in a less parsimonious model of wealth accumulation does not alter fundamentally our conclusions.

We note that t in our framework denotes time, rather than the age of an agent. As [34] puts it, our agents “do not marry or have children or die or even grow old” (p. 41). 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.

Finally, we add to the system a term that is absent from many other models in the literature. This term represents redistributive social structure. We imagine, as the simplest social interaction, that each individual pays a fixed proportion of its wealth, $x_i \tau dt$, into a central pot (“contributes to society”) and gets back an equal share of the pot, $\langle x \rangle_N \tau dt$, (“benefits from society”):

$$dx_i = x_i [(\mu - \tau)dt + \sigma dW_i(t)] + \langle x \rangle_N \tau dt. \quad (84) \quad \{\text{eq:rgbm}\}$$

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, paying taxes, using centrally-organised infrastructure (roads, schools, a legal system, social security, scientific research, and so on), insurance schemes, wealth transfers through inheritance and gifts, and everything else that constitutes an economic network. It would be a hopeless task to produce an exhaustive list of all these interactions, let alone include them as model components. 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.

Fitting τ to data will allow us to answer questions such as:

- 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 stationary distribution?
- if so, how long does it take, after a change in conditions, for the rescaled wealth distribution to reach the stationary distribution?

To be clear, we do not want to separate different processes that affect the wealth distribution, as would be useful if we wanted to understand the effect of a specific process, say income tax, on economic inequality. We aim for the opposite – a model that summarises in as few parameters as possible everything that affects the wealth distribution, crucially without assuming ergodicity. For example, if effects of death and inheritance were separated out explicitly with additional parameters, rather than being included in τ , the fitted τ values would no longer tell us if the system as a whole is in an ergodic regime.

1.4.4 Model behaviour

{sec:model_behavior}

Equation (84) is our model for the evolution of wealth with social structure and the basis for the empirical study that follows. It is instructive to write it as

$$dx_i = \underbrace{x_i [\mu dt + \sigma dW_i(t)]}_{\text{Growth}} - \underbrace{\tau(x_i - \langle x \rangle_N) dt}_{\text{Reallocation}}. \quad (85) \quad \{\text{eq:rgbm_ou}\}$$

This can be thought of as GBM with a mean-reverting term like that of [58] in physics and [59] in finance. This representation exposes the importance of the sign of τ . We discuss the two regimes in turn.

Positive τ For $\tau > 0$, wealth, x_i , reverts to the population average, $\langle x \rangle_N$. The large-sample approximation, $\langle x(t) \rangle_N \propto e^{\mu t}$, is valid¹ and yields a simple differential equation for the rescaled wealth,

$$dy_i = y_i \sigma dW_i(t) - \tau(y_i - 1) dt, \quad (86) \quad \{\text{eq:rgbm_ou_re}\}$$

in which the common growth rate, μ , has been scaled out. The distribution of $y_i(t)$ can found by solving the corresponding Fokker-Planck equation (also known as the Kolmogorov forward equation). A stationary distribution exists with a Pareto tail, see Appendix 2.0.9. It is known as the inverse gamma distribution and has probability density function,

$$\mathcal{P}(y) = \frac{(\zeta - 1)^\zeta}{\Gamma(\zeta)} e^{-\frac{\zeta-1}{y}} y^{-(1+\zeta)}, \quad (87) \quad \{\text{eq:disti}\}$$

where $\zeta = 1 + (2\tau/\sigma^2)$ is the Pareto tail index, $\Gamma(\cdot)$ is the gamma function, and the index i has been dropped. Example forms of the stationary distribution are shown in Figure 5. 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 and the fatter the tail of the distribution. Moreover, the fitted τ values we obtain in Sec. ?? give typical ζ values between 1 and 2 for the different datasets analyzed, consistent with observed tail indices between 1.2 to 1.6 [28, 23, 15, 60]. Thus, not only does RGBM predict a realistic functional form for the distribution of rescaled wealth, but also it admits fitted parameters which match observed tail thicknesses. The inability to do the latter is a known shortcoming in models of earnings-based wealth accumulation, see Sec. ??.

¹Strictly speaking, the large-sample approximation and resulting rescaled-wealth process, Equation (86), hold only for $\tau > \tau_c$. However, $\tau_c \approx 0$ for realistic model parameters and fits to data do not allow us to distinguish it from zero. Nonetheless, the derivation of τ_c is instructive, see Appendix 2.0.8.

Equation (86) and extensions of it have received much attention in statistical mechanics and econophysics [13, 12]. As a combination of GBM and an Ornstein-Uhlenbeck process, it is a simple and analytically tractable stochastic process. [33] provide an overview of the literature and known results.

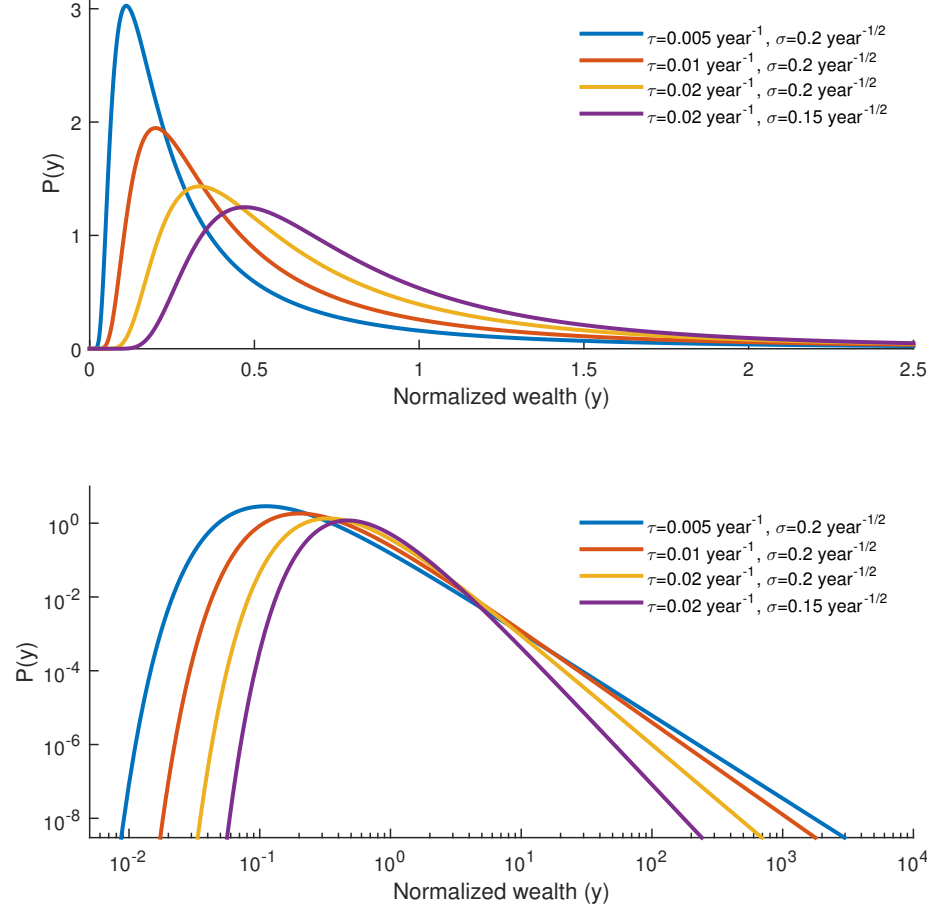


Figure 5: The stationary distribution for RGBM with positive τ . Top – linear scales; Bottom – logarithmic scales.

{fig:dist}

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 6. Qualitatively, this echoes the findings that the rich are experiencing higher growth rates of their wealth than the poor [45, 61] and that the cumulative wealth of the poorest 50 percent

of the American population was negative during 2008–2013 [51, 56].

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”. The social mobility that is, therefore, implicit in models that assume ergodicity might not exist in reality if that assumption is invalid. That inequality and immobility have been linked [16, 32, 6] may be unsurprising when both are viewed as consequences of non-ergodic wealth or income.

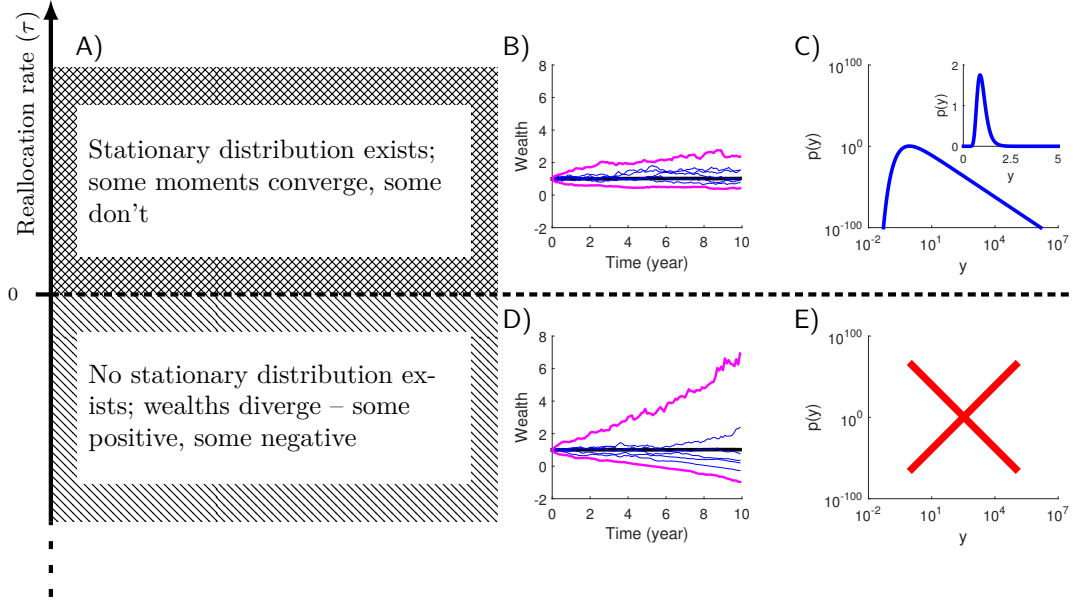


Figure 6: Regimes of RGBM. A) $\tau = 0$ separates the two regimes of RGBM. For $\tau > 0$, a stationary wealth distribution exists. For $\tau < 0$, no stationary wealth distribution exists and wealths diverge. B) Simulations of RGBM with $N = 1000$, $\mu = 0.021 \text{ year}^{-1}$ (presented after rescaling by $e^{\mu t}$), $\sigma = 0.14 \text{ year}^{-1/2}$, $x_i(0) = 1$, $\tau = 0.15 \text{ year}^{-1}$. 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 $\tau = -0.15 \text{ year}^{-1}$. E) in the $\tau < 0$ regime, no stationary wealth distribution exists.

{fig:regimes}

2 Data

{sec:data}

2.0.1 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 [52, 56]. “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” [30, 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 [31]. “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’ ” [30, 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 [14]. 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 7).

[30] reviewed the advantages and disadvantages of the different methods (see also the comment by Kopczuk on [14]). 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” [30, 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” [30, 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 [30, 14] for a thorough discussion. We analyze each data source separately.

2.0.2 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 (86) 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 (84)

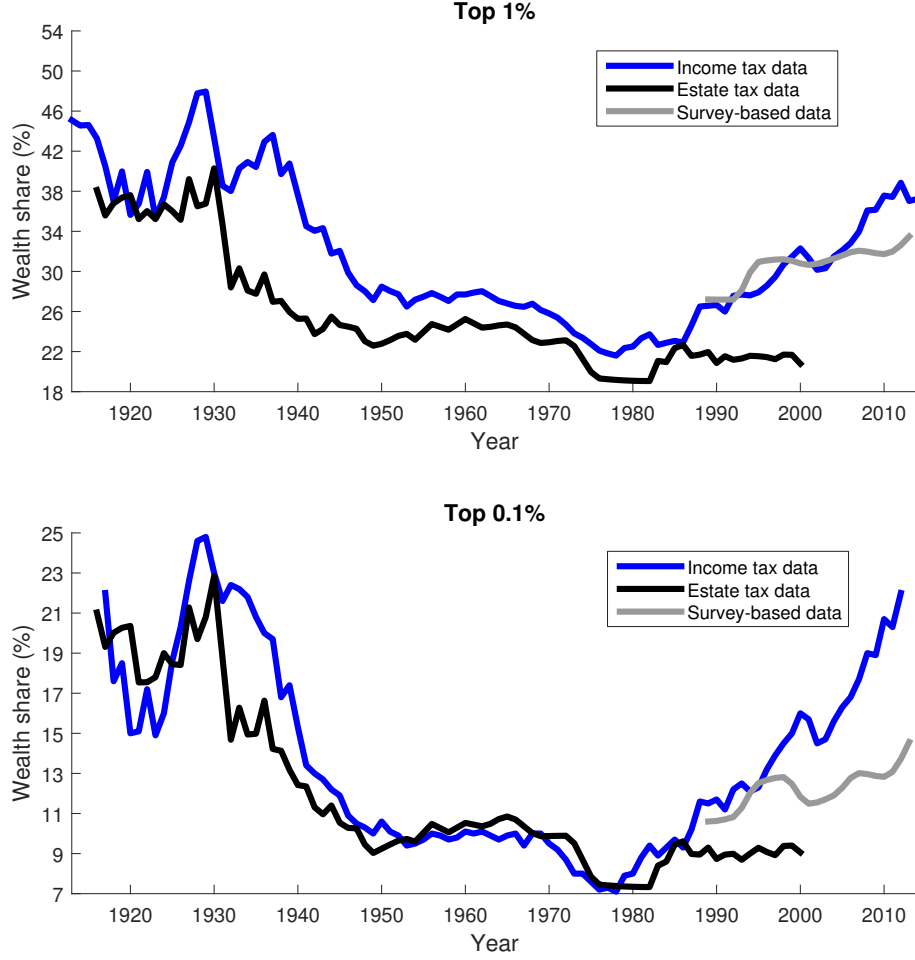


Figure 7: The top wealth shares in the US, 1913–2014. Sources – [52, 56] (blue); [31] (black); [14] (grey).

{fig:data1}

and not Equation (86). This requires us to specify a value of μ , which we estimate as $\mu = 0.021 \pm 0.001 \text{ year}^{-1}$ by a least-squares fit of historical per-capita private wealth in the US [47] to an exponential growth curve.

2.0.3 Volatility

We must also specify the volatility parameter, σ , in Equation (84). 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 “ball-park” 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 [49], which we annualise by multiplying

by $(250/\text{year})^{1/2}$. The values usually lie between 0.1 and 0.2 $\text{year}^{-1/2}$, with an average of 0.16 $\text{year}^{-1/2}$. Running our empirical analysis with constant σ in this range had little effect on our results (see Appendix 2.0.10) so, for simplicity, we present the analysis using $\sigma(t) = 0.16 \text{ year}^{-1/2}$ 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.

2.0.4 Empirical Analysis

{sec: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 [52, 56], estate tax-based [31] and survey-based [14]. 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 (84) 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 [36].
- 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 8. 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 8 (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, $\tilde{\tau}_q(t)$, 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 $\tilde{\tau}_q(t)$ 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 $\tilde{\tau}_q(t)$ is meaningful, see Figure 8 (bottom).

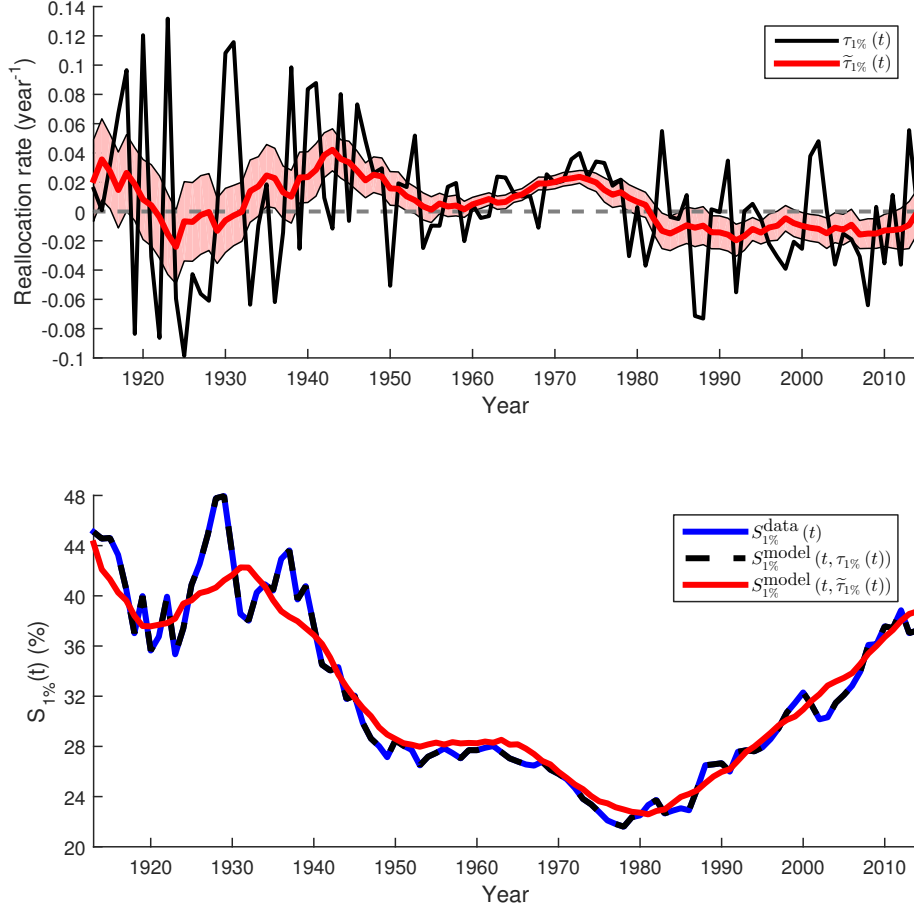


Figure 8: Fitted effective reallocation rates. Calculations done using $\mu = 0.021 \text{ year}^{-1}$ and $\sigma = 0.16 \text{ year}^{-1/2}$. Top: $\tau_{1\%}(t)$ (black) and $\tilde{\tau}_{1\%}(t)$ (red). Translucent envelopes indicate one standard error in the moving averages. Bottom: $S_{1\%}^{\text{data}}(t)$ (blue), $S_{1\%}^{\text{model}}(t, \tau_{1\%}(t))$ (dashed black), based on the 10-year moving average $\tilde{\tau}_{1\%}(t)$ (red).

{fig:tau}

For the income tax method wealth shares [52], the effective reallocation rate, $\tilde{\tau}(t)$, 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 [14], we observe briefer periods in which $\tilde{\tau}(t) < 0$. The same is true for the estate tax data [31], see Figure 9. 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.

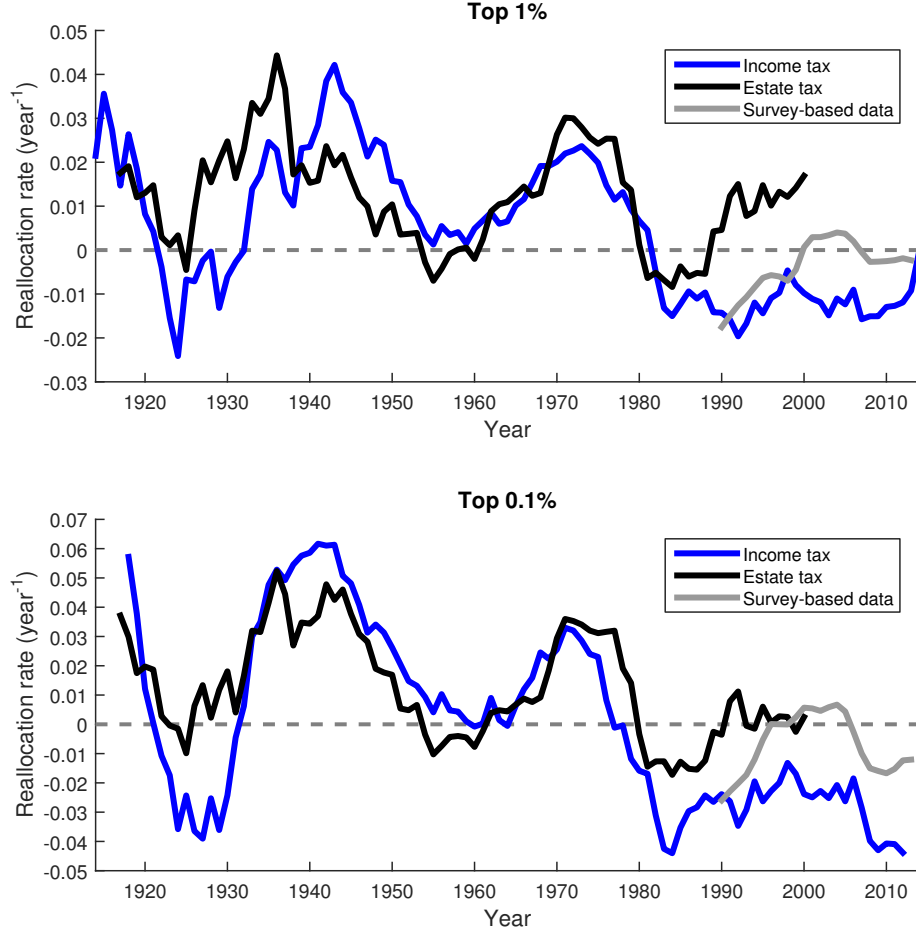


Figure 9: Effective reallocation rates for different datasets.

{fig:shares_comp}

2.0.5 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 $\mu = 0.021 \text{ year}^{-1}$, $\sigma = 0.16 \text{ year}^{-1/2}$, and $\tau = 0.04 \text{ year}^{-1}$, 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, see top panel of Figure 10. 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 10 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 $\tau >$

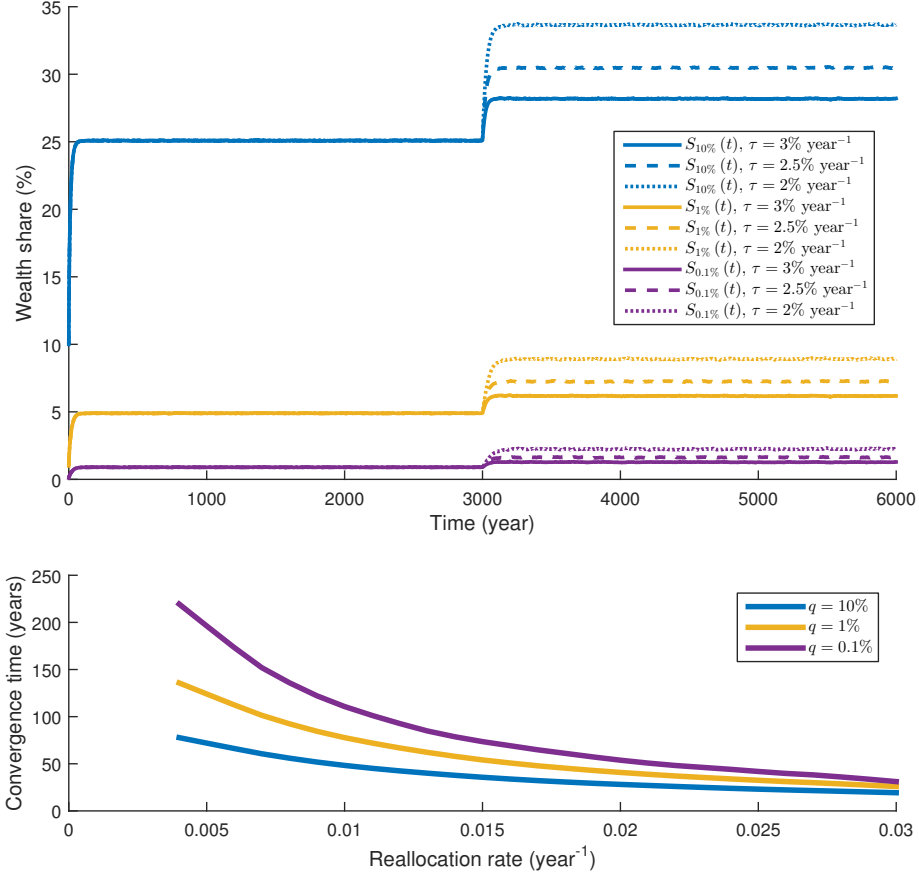


Figure 10: Wealth share convergence time. Top: The convergence of the wealth share for $q = 10$ percent (blue), $q = 1$ percent (yellow) and $q = 0.1$ percent (purple) following a change in the value of τ from 0.04 year^{-1} to 0.03 year^{-1} (solid), 0.025 year^{-1} (dashed) and 0.02 year^{-1} (dotted).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

$\sigma^2/2$ [33]. Convergence of the actual variance to the stationary variance occurs exponentially over a timescale $1/(2\tau - \sigma^2)$. Figure 11 shows the convergence times for different values of σ . See Appendix 2.0.11 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 assume 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 12 contrasts $\tau_{1\%}^{\text{eqm}}(t)$ assuming ergodicity with $\tilde{\tau}_{1\%}(t)$ without assuming ergodicity (using the income tax method dataset). If convergence were always possible and fast, then the two values would be identical within statis-

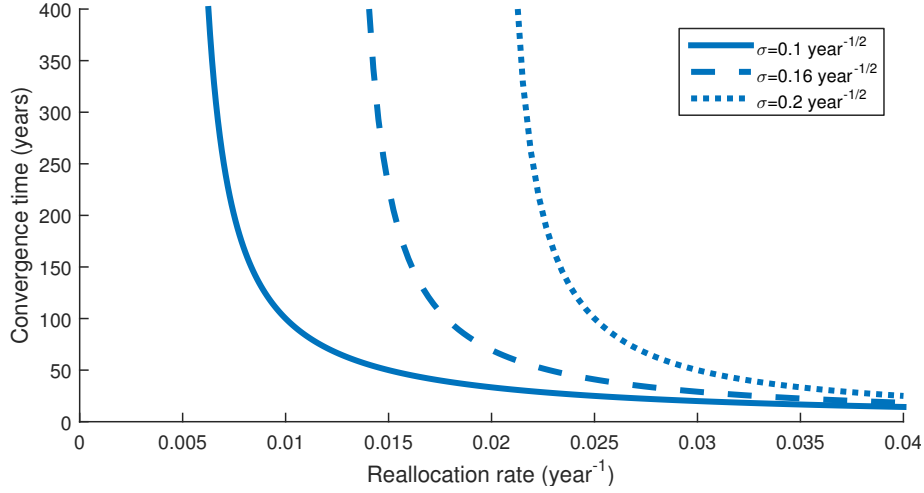


Figure 11: Variance convergence time

{fig:convar}

tical uncertainties. They are not. In addition, the generally large discrepancies between the wealth inequality implied by $\tau_{1\%}^{\text{eqm}}(t)$ (bottom panel, Figure 12, green line) and as observed (bottom panel, Figure 12, 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.

2.0.6 The effect of earnings

{sec:earnings}

In a recent review, [4, p. 593] remark that “the literature has largely emphasised the role of earnings inequality in explaining wealth inequality” and point to empirical failures of this approach. The most popular models, in which agents accumulate wealth through stochastic earnings and precautionary savings, predict a positive correlation between earnings and wealth inequality, which is not observed in cross-country data, and struggle “to reproduce the thick right tail of the wealth distribution observed in the data” [4, p. 593]. The latter failure is also noted by [25]. [4, p. 595] conclude that “other factors, like stochastic idiosyncratic returns on wealth” must be at play.

Consistent with these observations, RGBM models wealth accumulation as a multiplicative process. Changes in individual wealth are composed of terms proportional to either individual wealth or average wealth, in effect generating stochastic idiosyncratic returns. Additive changes akin to labor income and consumption are not treated explicitly. Instead their effects are wrapped into the reallocation rate, τ . While this parsimony makes the model tractable – in that we can write down the rescaled wealth distribution, Equation (87), for positive τ – it makes it impossible to disentangle the various real-world drivers of the observed increase in wealth inequality. It is, therefore, legitimate to ask whether wealth is inherently unstable (because it is reallocated negatively) or whether the increase in wealth inequality is really due to changes in the earnings distribution. If the latter, then the negative reallocation found in RGBM would be an artefact of an underspecified model and an unreliable indicator that rescaled wealth is

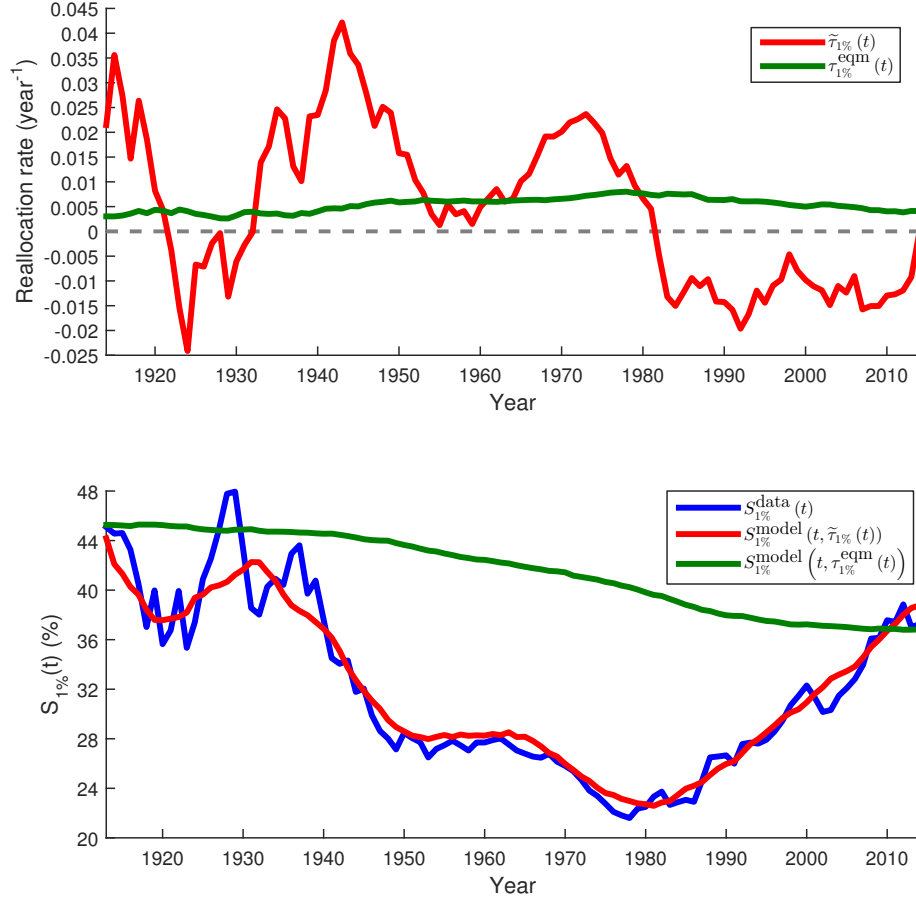


Figure 12: Comparison of dynamic and equilibrium reallocation rates. Top: $\tilde{\tau}_{1\%}(t)$ (red, same as in the top of Figure 8). $\tau_{1\%}^{eqm}(t)$ (green), defined such that $\lim_{t' \rightarrow \infty} S_{1\%}^{model}(t', \tau_{1\%}^{eqm}(t)) = S_{1\%}^{data}(t)$. It is impossible by design 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: $S_{1\%}^{data}(t)$ (blue), $S_{1\%}^{model}$ based on the 10-year moving average $\tilde{\tau}_{1\%}(t)$ (red), based on $\tau_{1\%}^{eqm}(t)$ (green). The reallocation rates found under the fast convergence assumption generate model wealth shares which bear little relation to reality.

{fig:asymptau}

non-ergodic.

To ensure we are not fooling ourselves, we check this by adding to Equation (84) a term representing earnings, for which data are available. We find that earnings have had a small effect on the dynamics of the wealth distribution over the last century. Since about 2000 this may have contributed to the increase in wealth inequality, but in general the effect has been stabilizing. This implies that the purely multiplicative dynamics of wealth, *i.e.* excluding additive earnings, have exhibited greater negative reallocation than fits to RGBM might suggest.

We separate earnings out as follows, in a model we call Earnings Geometric Brownian Motion (EGBM):

$$dx_i = x_i \left[(\mu(t) - \tau^{\text{EGBM}}(t)) dt + \sigma(t) dW_i(t) \right] + \langle x \rangle_N \tau^{\text{EGBM}}(t) dt + \epsilon_i(t) dt. \quad (88) \quad \{\text{eq:egbm}\}$$

The notation is similar to RGBM, with the addition of the individual earnings term $\epsilon_i(t)$. We consider earnings after spending and after tax, as they directly change wealth. This changes the role of the reallocation rate, as reflected in the notation: τ^{EGBM} reflects changes in wealth for reasons other than additive income and consumption.

Using the same techniques as for the RGBM fits, we fit a time series, $\tau^{\text{EGBM}}(t)$, that reproduces annually observed wealth shares in the US (for the income tax method dataset). This requires us to find individual earnings in Equation (88), which we do as follows. We assume that the earnings are log-normally distributed (see for example [48]) and then fit the two parameters of the distribution every year. This is done using the mean earnings (calculated as total disposable national income multiplied by the personal savings rate and divided by the population size, as reported by [47]) and the top 1% income share data (as reported in [56]).

Since the joint wealth-earnings distribution is not known fully, we analyze two bounding scenarios. First, a “correlated” scenario in which earnings are perfectly correlated with wealth. Every year, the richest individual has the highest earnings, the second richest individual has the second highest earnings, and so on. This is the least stabilizing possibility. Second, an “anti-correlated” scenario in which the poorest individual has the highest earnings, the second poorest individual has the second highest earnings, and so on. This is the most stabilizing possibility. We fit $\tau^{\text{EGBM}}(t)$ for each scenario, and compare the results to the fitted $\tau(t)$ values for which the effects of earnings have not been disentangled from other effects. See Figure 13.

In general, disentangling earnings has a small effect on the fitted reallocation rates. For the correlated scenario – which is the more realistic of the two because the correlation between wealth and earnings is positive [51] – the fitted reallocation rates $\tau^{\text{EGBM}}(t)$ are within the statistical error of $\tau(t)$, into which the effects of earnings have been subsumed.

When earnings have a stabilizing effect, the fitted reallocation rate is lower under EGBM than under RGBM. Even in the correlated scenario, which is the least stabilizing, earnings have a stabilizing effect for years 1930–2000. During 2000–2010, we find that earnings destabilise the wealth distribution in the correlated scenario. However, since the difference between the fitted values is small, we can conclude that the contribution of earnings to the reported increase in wealth inequality from 1985 to 2010 is negligible. Moreover, the earnings contribution cannot explain the invalidity of the ergodic hypothesis that constitutes our main finding.

We suspect the main reason for the small effect of earnings is the low ratio of earnings to wealth. Figure 14 shows the ratio of average earnings (after spending) to average wealth in the US during the period under study. Typically this ratio is around 1%, at which level earnings play only a secondary role in the dynamics of wealth. This echoes the observations of [45, 47, 9]. Under such conditions, it is unsurprising that models reliant on earnings as the primary mechanism of wealth accumulation fail to resemble reality [25, 4]. By including

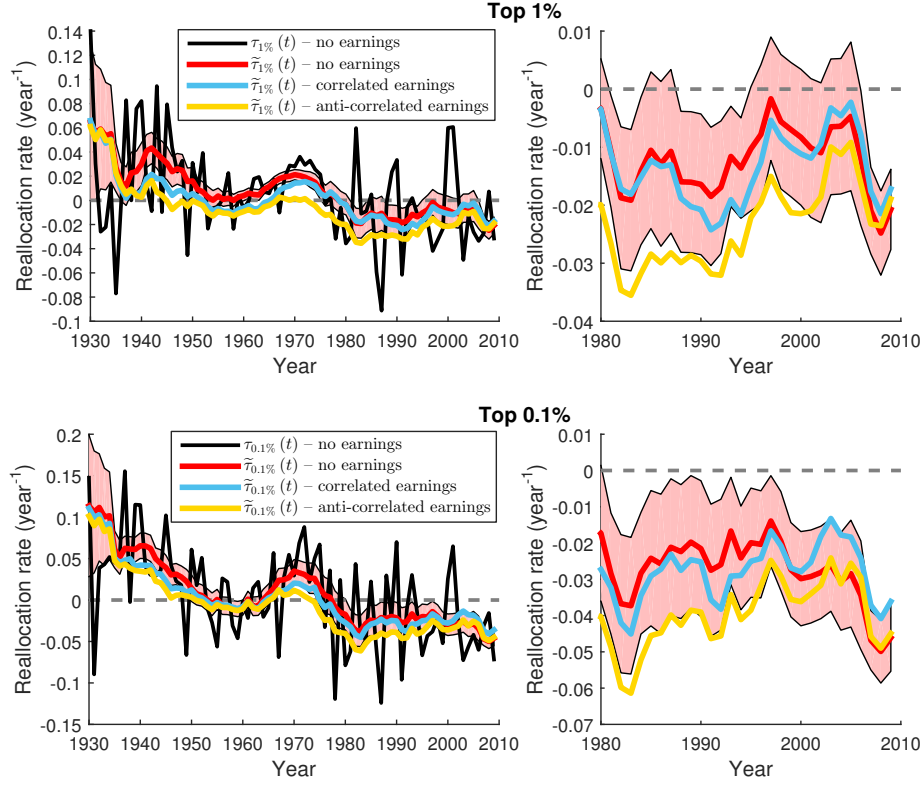


Figure 13: The effective reallocation rates for the RGBM and EGBM models. The red translucent envelopes indicate one standard error in the moving averages of the RGBM fitted reallocation rates.

{fig:egbm}

a multiplicative growth mechanism for wealth, RGBM avoids these pitfalls. Not only does it predict the Pareto tail of the stationary wealth distribution, when it exists, but also it allows for fitted tail exponents that reproduce the observed tail thickness (see Sec. ??). Retreating to the simplicity of RGBM is, therefore, justified.

2.0.7 Conclusions

{sec: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

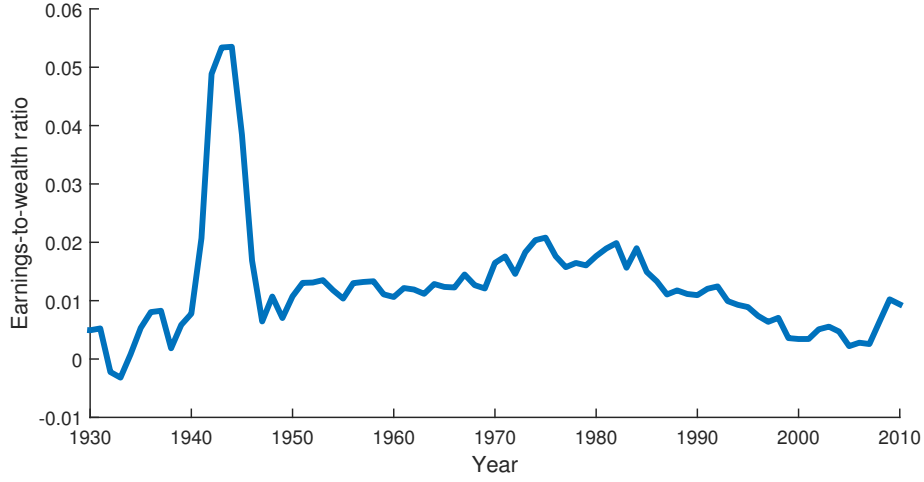


Figure 14: Time series of the ratio of average earnings after spending to average private wealth in the US. Data from [47].

{fig:ratio}

the stationary distribution is often a key object of study – model parameters are 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 9 and Figure 10 (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, $\tau < 0$ [52, 56] or $\tau \approx 0$ [14]. 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 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.

2.0.8 The derivation of self-averaging time t_c

{app:tauc}

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 [44]. 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 $N \rightarrow \infty$ 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 makes the computation of the variance easy. GBM is log-normally distributed,

$$\ln[x(t)] \sim \ln \mathcal{N} \ln \left(x(0) + \left[\mu - \frac{\sigma^2}{2} \right] t, \sigma^2 t \right). \quad (89)$$

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

$$\langle x(t) \rangle = x(0) e^{\mu t}, \quad (90)$$

and the variance grows as

$$V[x(t)] = x(0)^2 e^{2\mu t} \left[e^{\sigma^2 t} - 1 \right]. \quad (91)$$

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

$$\frac{V[\langle x(t) \rangle_N]}{\langle \langle x(t) \rangle_N \rangle^2} = \frac{e^{\sigma^2 t} - 1}{N}. \quad (92)$$

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:

$$t < t_c \equiv \frac{\ln N}{\sigma^2}. \quad (93)$$

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 [11], *i.e.*

$$\tau > \tau_c \equiv \frac{\sigma^2}{2 \ln N}. \quad (94) \quad \{\text{eq:tauc}\}$$

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 $\tau_c \approx 0.0007 \text{ year}^{-1}$, 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 indistinguishable whether we simulate dx (which requires an estimate for μ), or dy (where μ does not appear but the above-mentioned approximation is made).

2.0.9 The derivation of the stationary distribution

{app:stat}

We start again with the SDE for the rescaled wealth,

$$dy = \sigma y dW - \tau (y - 1) dt. \quad (95)$$

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

$$\frac{dp(y, t)}{dt} = \frac{\partial}{\partial y} [Ap(y, t)] + \frac{1}{2} \frac{\partial^2}{\partial y^2} [B^2 p(y, t)]. \quad (96)$$

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

$$\frac{\sigma^2}{2} (y^2 p)_{yy} + \tau [(y - 1)p]_y = 0. \quad (97) \quad \{\text{eq:fokker_planck}\}$$

Positive wealth subjected to continuous-time multiplicative dynamics with non-negative reallocation can never reach zero. Therefore, we solve Equation (97) with boundary condition $p(0) = 0$ to give

$$p(y) = C(\zeta) e^{-\frac{\zeta-1}{y}} y^{-(1+\zeta)}, \quad (98)$$

where

$$\zeta = 1 + \frac{2\tau}{\sigma^2} \quad (99)$$

and

$$C(\zeta) = \frac{(\zeta - 1)^\zeta}{\Gamma(\zeta)}, \quad (100)$$

with the gamma function $\Gamma(\zeta) = \int_0^\infty x^{\zeta-1} e^{-x} dx$. The distribution has a power-law tail as $y \rightarrow \infty$, 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.

2.0.10 The effect of fixed versus time-varying σ

{app:fixed_sig}

Our analysis requires us to set the volatility parameter, σ , in the RGBM model. We estimate this as a time-varying quantity, $\sigma(t)$, using data from the American stock market. For each year in the analysis, we define $\sigma(t)$ as the standard deviation of the daily logarithmic changes in the Dow Jones Industrial Average (DJIA) for that year [49], annualised by multiplying by $(250/\text{year})^{1/2}$. We present this in Figure 15.

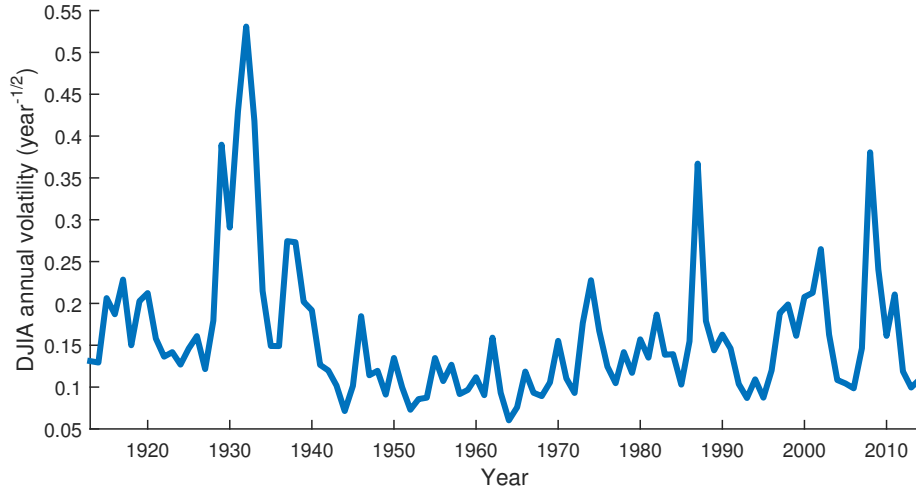


Figure 15: Annualised $\sigma(t)$ estimated from daily changes in the DJIA. Data taken from [49].

{fig:signat}

In the results presented in Sec. ??, we used a fixed value – $\sigma = 0.16$, the average of the time-varying $\sigma(t)$ – for simplicity. In order to show this has little effect on the estimated reallocation rates, a comparison is presented in Figure 16.

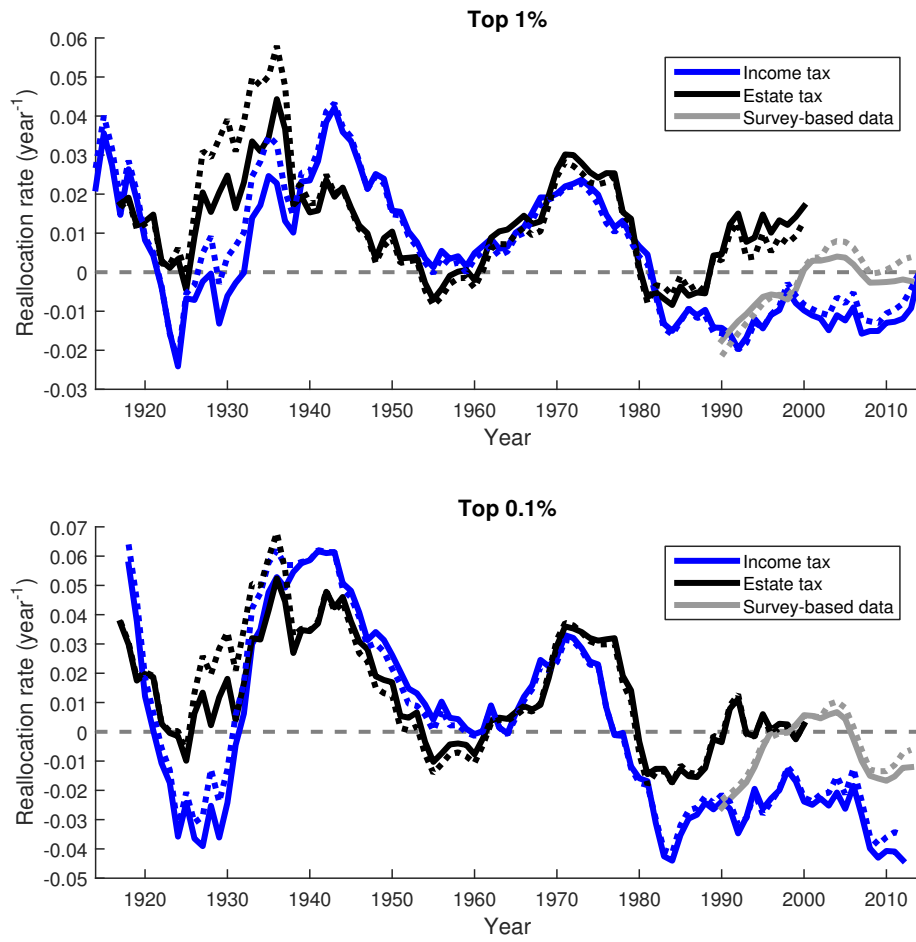


Figure 16: Smoothed effective reallocation rates with fixed and time-varying σ . Solid lines – fixed σ ; Dotted lines – time-varying σ .

{fig:sigmat_tau}

2.0.11 The derivation of the variance convergence time

{app:var_conv}

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 2.0.9. 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, $\langle y^2 \rangle$:

$$V(y) = \langle y^2 \rangle - \langle y \rangle^2 \quad (101)$$

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:

$$dy = \sigma y dW - \tau (y - 1) dt. \quad (102) \quad \{\text{eq:rescaledSDE}\}$$

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:

$$df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial y} dy + \frac{1}{2} \frac{\partial^2 f}{\partial y^2} dy^2. \quad (103)$$

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

$$d(y^2) = 2y dy + (dy)^2. \quad (104) \quad \{\text{eq:diff2}\}$$

We substitute dy in Equation (104), 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

$$d(y^2) = 2\sigma y^2 dW - (2\tau - \sigma^2) y^2 dt + 2\tau y dt \quad (105)$$

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

$$\frac{d\langle y^2 \rangle}{dt} = - (2\tau - \sigma^2) \langle y^2 \rangle + 2\tau \quad (106) \quad \{\text{eq:avediff2}\}$$

with solution

$$\langle y(t)^2 \rangle = \frac{2\tau}{2\tau - \sigma^2} + \left(\langle y(0)^2 \rangle - \frac{2\tau}{2\tau - \sigma^2} \right) e^{-(2\tau - \sigma^2)t}. \quad (107) \quad \{\text{eq:avediff3}\}$$

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

$$V(t) = V_\infty + (V_0 - V_\infty) e^{-(2\tau - \sigma^2)t}, \quad (108) \quad \{\text{eq:var1}\}$$

where V_0 is the initial variance and

$$V_\infty = \frac{2\tau}{2\tau - \sigma^2}. \quad (109) \quad \{\text{eq:varinf}\}$$

V converges in time to the asymptote, V_∞ , provided the exponential in Equation (108) is decaying. This can be expressed as a condition on τ

$$\tau > \frac{\sigma^2}{2}. \quad (110)$$

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 (108) that the convergence time of the variance is $1/(2\tau - \sigma^2)$.

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 [33] also compute the third cumulant.

2.1 Taxation

{section:Taxation}

In Sec. 1.1 we created a world of independent GBMs; in Sec. 1.3 we introduced to this world the invention of cooperation and saw that it increases long-time growth for those who participate in resource-pooling and sharing. In this section we study what happens if a large number of individuals pool and share a small fraction of their resources, which is reminiscent of taxation and redistribution carried out in a large population. We will find that while cooperation for 2 individuals increases their growth rates, sufficient cooperation in a large population has two related effects. Firstly, everyone's wealth grows asymptotically at a rate close to that of the expectation value. Secondly, wealth condensation and the divergence of inequality no longer occur.

We introduce a model that applies a flat wealth tax rate and every individual, irrespective of his wealth, receives the same benefit from the collected tax, in absolute terms. This mimicks the actions of a central agency that collects each year from everyone 1% of his wealth and pays $1/N^{\text{th}}$ of the total collected amount to each individual. A similar model will be used for income tax, see (Eq. 130) in Sec. 2.1.2.

Of course this isn't how taxation works in reality – wealth taxes are usually only collected in the form of inheritance tax and sometimes property or land tax; often progressive rates are applied, and how tax takings are actually redistributed is very unclear. Who benefits from government activity? Infrastructure is built, benefits payments made, healthcare and education provided, a legal system is maintained of courts that can enforce contracts and enable corporate structures, police and an army may provide security. Individuals will benefit from these different aspects to very different degrees. Our model ignores this and lets everyone benefit equally.

Despite the simplicity of the setup the following important feature emerges: there is a critical tax rate. This qualitative result applies both to income tax and to wealth tax.

Definition Critical tax rate

Below the critical tax rate the variance of rescaled wealth increases indefinitely. Above the critical tax rate it stabilizes to an asymptotic value in the limit $t \rightarrow \infty$.

Section 1.3 was concerned with growth, here we are concerned with inequality. We will therefore work with the rescaled wealth, y , introduced in (Eq. 81). Equation (83) defines the dynamic of x . From it we can find the dynamic for $f(x) = y$ using Itô calculus

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial t} dt + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} dx^2 \quad (111)$$

$$= dy = -\mu y dt + \frac{y}{x} dx \quad (112)$$

$$= y \sigma dW. \quad (113)$$

2.1.1 Wealth tax

{section:Wealth_tax}

We investigate the situation where each individual's wealth is taxed at a rate of $0 \leq \tau \leq 1$ per unit time, and the total tax thus raised is redistributed equally among the population. This is modelled by the stochastic wealth process,

$$dx = x[(\mu - \tau) dt + \sigma dW] + \tau \langle x \rangle_N dt, \quad (114) \quad \{\text{eq:wsde}\}$$

which is a modified version of (Eq. 83) – the term $-\tau x dt$ was added to represent tax collection, and the term $+\tau \langle x \rangle_N dt$ to represent redistribution of collected tax. To make the model more tractable we consider the case $N \rightarrow \infty$, which replaces the finite-ensemble average by the expectation value, $\langle x \rangle_N \rightarrow \langle x \rangle$. The finite ensemble size has important effects but we will not discuss them here. Total wealth is conserved by the taxation and redistribution process in this model, and the expectation value is unaffected, $\langle x(t) \rangle = \langle x_0 \rangle e^{\mu t}$, just as for GBM without taxation, (Eq. 19). We are again interested in rescaled wealth, $y = \frac{x}{\langle x \rangle} = x e^{-\mu t}$ ((Eq. 81)), whose dynamic we derive using the chain rule

$$dy = \frac{\partial y}{\partial t} dt + \frac{\partial y}{\partial x} dx + \frac{1}{2} \frac{\partial^2 y}{\partial x^2} dx^2 \quad (115)$$

$$= -\mu y dt + \frac{1}{x} dx \quad (116) \quad \{\text{eq:ysde}\}$$

$$= y(-\tau dt + \sigma dW) + \tau dt. \quad (117)$$

The first moment of y is trivially 1,

$$\langle y \rangle = \left\langle \frac{x}{\langle x \rangle} \right\rangle = 1. \quad (118)$$

We compute the dynamic of the second moment of y , to first order in dt , using the chain rule again,

$$d(y^2) = \frac{\partial(y^2)}{\partial t} dt + \frac{\partial(y^2)}{\partial y} dy + \frac{1}{2} \frac{\partial^2(y^2)}{\partial y^2} (dy)^2 \quad (119)$$

$$= 2y dy + (dy)^2 \quad (120)$$

$$= 2y^2(-\tau dt + \sigma dW) + 2y\tau dt + y^2\sigma^2 dt. \quad (121)$$

Taking expectation values yields

$$\langle dy^2 \rangle = \langle -2y^2(-\tau dt + \sigma dW) + 2y\tau dt - y^2\sigma^2 dt \rangle \quad (122)$$

$$= d\langle y^2 \rangle = (\sigma^2 - 2\tau)\langle y^2 \rangle dt + 2\tau dt. \quad (123)$$

This equation is an inhomogeneous first-order ordinary differential equation for the second moment. Perhaps it's more recognizable when written in standard form as

$$\left(\frac{d}{dt} - (\sigma^2 - 2\tau) \right) \langle y^2 \rangle = 2\tau. \quad (124) \quad \{\text{eq:ODE2}\}$$

Such equations are solvable using the method of integrating factors, see *e.g.* [3, Chpater 1.5]. The solution of the dynamic (Eq. 124) is the second moment of the distribution of rescaled wealth as a function of time, namely

$$\langle y^2 \rangle = \frac{2\tau}{2\tau - \sigma^2} + \left(\langle y_0^2 \rangle - \frac{2\tau}{2\tau - \sigma^2} \right) e^{-(2\tau - \sigma^2)t}. \quad (125)$$

This can be rewritten in terms of the variance of rescaled wealth, $V = \langle y^2 \rangle - 1$, as

$$V(t) = V_\infty + (V_0 - V_\infty)e^{-(2\tau - \sigma^2)t}, \quad (126) \quad \{\text{eq:wvar}\}$$

where V_0 is the initial variance and

$$V_\infty \equiv \frac{\sigma^2}{2\tau - \sigma^2}. \quad (127)$$

V converges in time to the asymptote, V_∞ , provided the exponent in (Eq. 126) is decaying. This can be expressed as a condition on τ ,

$$\boxed{\tau > \tau_c \equiv \frac{\sigma^2}{2}}, \quad (128) \quad \{\text{eq:wstab}\}$$

which defines the critical tax rate, τ_c . Above this critical tax rate, $\tau > \tau_c$, the variance of the rescaled-wealth distribution stabilises. Below it, the variance grows beyond all bounds. We believe that the divergence or convergence of the variance signals an important change in systemic behavior, but we hasten to point out the following caveat: a finite second moment does not guarantee finiteness of higher moments. A deeper analysis of ODEs of the type of (Eq. 124), which we don't reproduce here, reveals that any finite wealth tax rate implies that all moments of order $n > \frac{2\tau}{\sigma^2} + 1$ diverge. Under the flat wealth tax investigated here, the wealth distribution never fully stabilizes. In the language often used by economists in this debate, an ergodic wealth distribution does not exist for our model.

Caveats aside, (Eq. 126) also allows us to identify a characteristic timescale over which the variance stabilises for supercritical taxation,

$$T_s = \frac{1}{2\tau - \sigma^2}. \quad (129)$$

τ_c may be viewed as the tax rate at which T_s diverges.

Numerical simulations confirm that the above analytical results are informative for finite ensembles. Fig. 17 compares the evolution of the empirical

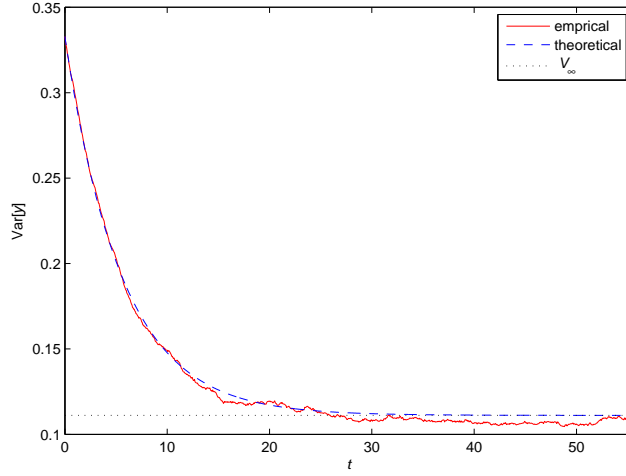


Figure 17: Wealth tax. The empirical variance of the rescaled wealths of $N = 10^4$ realisations of (Eq. 114) with uniformly-distributed initial wealths (red); the theoretical variance for the infinite ensemble, $V(t)$ (blue dashed); and the asymptotic theoretical variance, V_∞ (black dotted). Parameter values are $\mu = 0.05$, $\sigma^2 = 0.02$, and $\tau = 0.1$ per unit time.

{fig:var_wealth}

variance of the rescaled wealths of $N = 10^4$ realisations of the stochastic wealth process in (Eq. 114) with the theoretical result for the infinite ensemble in (Eq. 126). Parameter values were $\mu = 0.05$, $\sigma^2 = 0.02$, and $\tau = 0.1$ per unit time, of which the first two are realistic for a time unit of one year [41] (assuming individual wealth processes share parameters with stock market indices). The differences are finite-sample effects. Fig. 18 shows the initial distribution of rescaled wealths, which was chosen to be uniform, and the final distribution at the end of the period shown in Fig. 17.

The simulated parameter values give a critical tax rate of 1% pa. This is broadly in line with genuine annual wealth and property taxes in the few countries in which they are levied. Under the simulated tax rate of 10% pa, the stabilisation time is $T_s \approx 6$ years. It is hard to imagine a wealth tax of this magnitude being politically feasible in the real world. In our simple model, the tax rate could be set either to achieve convergence of inequality to a desired level, reflected by V_∞ , or over a desired timescale, represented by T_s .

It is interesting to connect this with the most widely levied wealth tax: the inheritance tax. In the UK this is levied at 40% of the value of an individual's estate (above a certain threshold) upon death. We can surmise that an individual will typically hold most of his wealth for the human generation time of around 30 years, this being a sensible estimate of the time between inheriting or otherwise accumulating his wealth and passing it on. Using our plausible parameter values, an inheritance tax of 40% corresponds to an annually compounded wealth tax of $1 - (0.6)^{1/30} \approx 1.7\%$ pa and a stabilisation time of around 70 years. The former is close to and, notably, above the critical rate of 1% pa, suggesting that variance stabilisation may be an influential criterion in the determination of our

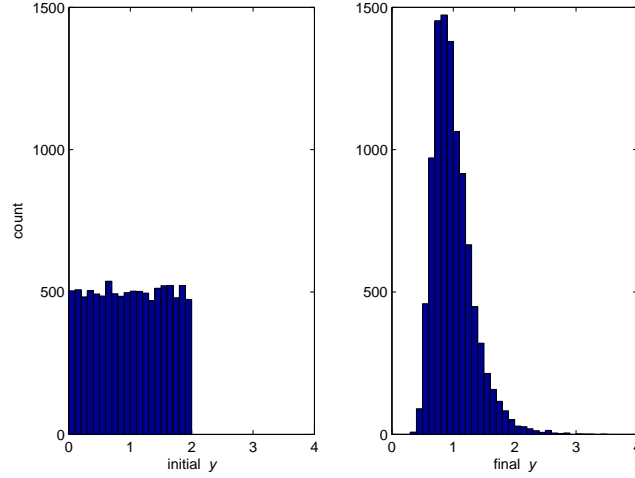


Figure 18: Histograms of the initial (left) and final (right) empirical distributions of the rescaled wealth for the same realisations of (Eq. 114) used in Fig. 17.

{fig:hist_wealth}

taxes.

2.1.2 Income tax

{section:Income_tax}

In our very simple model, we have seen that a flat wealth tax can stabilize the variance of the rescaled-wealth distribution. In this section we show that in a similarly simple model an income tax can achieve the same result. We introduce a model of income tax under which a fraction, $0 \leq \tau \leq 1$, of each individual's deterministic wealth increment, $\mu x dt$, is deducted and the total tax raised is redistributed equally. This is modelled by the stochastic wealth process,

$$dx = x[\mu(1 - \tau) dt + \sigma dW] + \mu\tau \langle x \rangle_N dt. \quad (130) \quad \{\text{eq:isde}\}$$

Again, we consider the large-population limit $N \rightarrow \infty$, corresponding to the replacement $\langle x \rangle_N \rightarrow \langle x \rangle$. For positive drift, $\mu > 0$, the deterministic increment, $\mu x dt$, is guaranteed to be positive. It can be thought of as the income derived from that individual's activities, such as employment, on which governments typically levy taxes. Note that τ in (Eq. 130) is a dimensionless number, whereas it is a rate of dimension “per unit time” in (Eq. 114). The form of (Eq. 130) is identical to (Eq. 114) with the parameter transformation $\tau \rightarrow \mu\tau$. Thus we can immediately deduce the dynamic for the rescaled wealth as

$$dy = y(-\mu\tau dt + \sigma dW) + \mu\tau dt. \quad (131) \quad \{\text{eq:itax}\}$$

The variance stabilisation condition analogous to (Eq. 128) becomes

$$\tau > \tau_c \equiv \frac{\sigma^2}{2\mu}. \quad (132) \quad \{\text{eq:icrit}\}$$

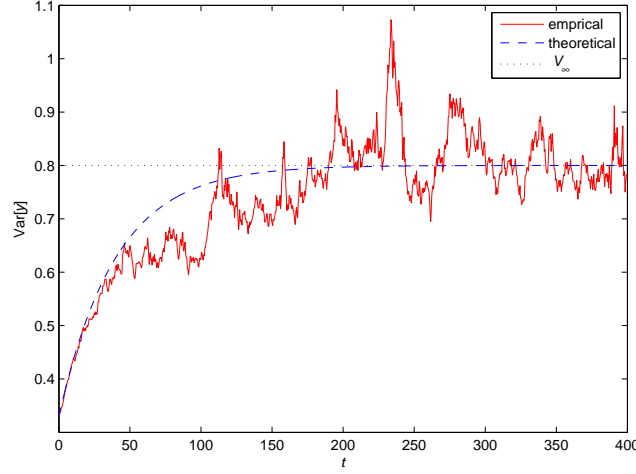


Figure 19: Income tax. The empirical variance of the rescaled wealths of 10^4 realisations of (Eq. 130) with uniformly-distributed initial wealths (red); the theoretical variance for the infinite ensemble, $V(t)$ (blue dashed); and the asymptotic theoretical variance, V_∞ (black dotted). Parameter values are $\mu = 0.05$ and $\sigma^2 = 0.02$ per unit time, and $\tau = 0.45$.

{fig:var_income}

This defines the critical income tax, τ_c , above which the variance converges to its asymptotic value,

$$V_\infty = \frac{\sigma^2}{2\mu\tau - \sigma^2}, \quad (133)$$

according to

$$V(t) = V_\infty + (V_0 - V_\infty)e^{-(2\mu\tau - \sigma^2)t}. \quad (134) \quad \{\text{eq:ivar}\}$$

Finally, the stabilisation time is

$$T_s = \frac{1}{2\mu\tau - \sigma^2}. \quad (135)$$

Fig. 19 compares the evolution of the empirical variance of the rescaled wealths of 10^4 realisations of the stochastic wealth process in (Eq. 130) with the theoretical result for the infinite ensemble. Parameter values were $\mu = 0.05$ and $\sigma^2 = 0.02$ per unit time, and $\tau = 0.45$. The latter is the UK's limiting income tax rate for large incomes, which will be the determining tax rate for variance stabilisation.

The finite-sample deviations from the infinite-ensemble result are larger in Fig. 19 than in Fig. 17. This is due entirely to the simulated parameter values: (Eq. 114) and (Eq. 130) can be made equivalent by choosing different parameters.

Fig. 20 shows the initial distribution of rescaled wealths, which was chosen to be uniform, and the final distribution at the end of the period shown in Fig. 19. The distribution of wealths under income tax has an appreciably longer tail than under wealth tax. As before this is a function of the parameter choices.

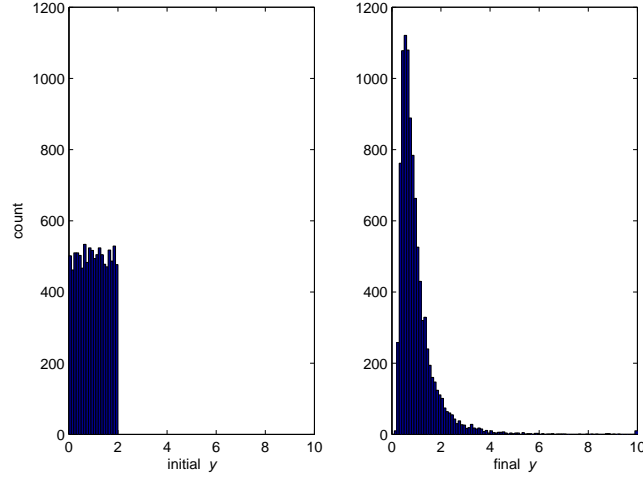


Figure 20: Histograms of the initial (left) and final (right) empirical distributions of the rescaled wealth for the same realisations of (Eq. 131) used in Fig. 19.

{fig:hist_income}

The simulated parameter values have a critical income tax rate of $\tau_c = 0.2$ and a stabilisation time of $T_s = 40$ years. Thus the UK sets its income tax at a level which, at least in this simple framework, has a variance stabilising effect.

List of Symbols

δ Most frequently used to express a difference, for instance δx is a difference between two wealths x . It can be the Kronecker delta function, a function of two arguments with properties $\delta(i, j) = 1$ if $i = j$ and $\delta(i, j) = 0$ otherwise. It can also be the Dirac delta function of one argument, $\int f(x)\delta(x - x_0)dx = f(x_0)$.

δt A time interval corresponding to the duration of one round of a gamble or, mathematically, the period over which a single realisation of the constituent random variable of a discrete-time stochastic process is generated..

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_{est} Growth rate estimator for finite time and finite ensemble size.

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

\bar{g} Time-average exponential growth rate.

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

- 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 $W(t) \sim \mathcal{N}(0, \bar{g})$.
- x Wealth.

References

- [1] A. Adamou and O. Peters. Dynamics of inequality. *Significance*, 13(3):32–35, 2016.
- [2] A. B. Atkinson. The timescale of economic models: How long is the long run? *The Review of Economic Studies*, 36(2):137–152, 1969.
- [3] C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*. Springer, New York, 1978.
- [4] J. Benhabib, A. Bisin, and M. Luo. Earnings inequality and other determinants of wealth inequality. *American Economic Review*, 107(5):593–597, 2017.
- [5] 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.
- [6] Y. Berman. Understanding the mechanical relationship between inequality and intergenerational mobility. Available at: <http://papers.ssrn.com/abstract=2796563>, 2017.
- [7] Y. Berman, E. Ben-Jacob, and Y. Shapira. The dynamics of wealth inequality and the effect of income distribution. *PloS one*, 11(6):e0199999, 2016.
- [8] Y. Berman, O. Peters, and A. Adamou. An empirical test of the ergodic hypothesis: Wealth distributions in the United States. January 2017.
- [9] Y. Berman and Y. Shapira. Revisiting $r > g$ —the asymptotic dynamics of wealth inequality. *Physica A: Statistical Mechanics and its Applications*, 467:562–572, 2017.
- [10] T. Bewley. The permanent income hypothesis: A theoretical formulation. *Journal of Economic Theory*, 16(2):252–292, 1977.
- [11] J.-P. Bouchaud. Note on mean-field wealth models and the random energy model.

- [12] J.-P. Bouchaud. On growth-optimal tax rates and the issue of wealth inequalities. <http://arXiv.org/abs/1508.00275>, August 2015.
- [13] J.-P. Bouchaud and M. Mézard. Wealth condensation in a simple model of economy. *Physica A*, 282(4):536–545, 2000.
- [14] 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.
- [15] M. Brzezinski. Do wealth distributions follow power laws? evidence from ‘rich lists’. *Physica A: Statistical Mechanics and its Applications*, 406:155–162, 2014.
- [16] M. Corak. Income inequality, equality of opportunity, and intergenerational mobility. *The Journal of Economic Perspectives*, 27(3):79–102, 2013.
- [17] F. A. Cowell. Piketty in the long run. *The British Journal of Sociology*, 65(4):708–720, 2014.
- [18] M. De Nardi. Quantitative models of wealth inequality: A survey. 2015. Working Paper 21106, NBER.
- [19] M. De Nardi, G. Fella, and F. Yang. Piketty’s book and macro models of wealth inequality. 2015. Working Paper 21730, NBER.
- [20] B. Derrida. Random-energy model: Limit of a family of disordered models. *Phys. Rev. Lett.*, 45(2):79–82, July 1980.
- [21] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613–2626, September 1981.
- [22] 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.
- [23] X. Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1(1):255–294, 2009.
- [24] X. Gabaix, J.-M. Lasry, P.-L. Lions, and B. Moll. The dynamics of inequality. *Econometrica: Journal of the Econometric Society*, 84(6):2071–2111, 2016.
- [25] J. Hubmer, P. Krusell, and A. A. Smith Jr. The historical evolution of the wealth distribution: A quantitative-theoretic investigation. 2016. Working Paper 23011, NBER.
- [26] C. I. Jones. Pareto and piketty: The macroeconomics of top income and wealth inequality. *The Journal of Economic Perspectives*, 29(1):29–46, 2015.
- [27] B. Kaymak and M. Poschke. The evolution of wealth inequality over half a century: The role of taxes, transfers and technology. *Journal of Monetary Economics*, 77:1–25, 2016.

- [28] 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.
- [29] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Science & Business Media, 1992.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] J. E. Meade. *Efficiency, Equality and The Ownership of Property*. Allen & Unwin, London, UK, 1964.
- [35] R. Mehra and E. C. Prescott. The equity premium puzzle. *J. Monetary Econ.*, 15:145–161, 1985.
- [36] J. A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- [37] M. E. J. Newman. Power laws, Pareto distributions and Zipf’s law. *Contemp. Phys.*, 46(5):323–351, 2005.
- [38] V. Pareto. *Cours d’économie Politique*. F. Rouge, Lausanne, Switzerland, 1897.
- [39] O. Peters. Optimal leverage from non-ergodicity. *Quant. Fin.*, 11(11):1593–1602, November 2011.
- [40] O. Peters. The time resolution of the St Petersburg paradox. *Phil. Trans. R. Soc. A*, 369(1956):4913–4931, December 2011.
- [41] O. Peters and A. Adamou. Stochastic market efficiency. *SFI working paper 13-06-022*, 2013.
- [42] O. Peters and A. Adamou. The evolutionary advantage of cooperation. *arXiv:1506.03414*, June 2015.
- [43] O. Peters and M. Gell-Mann. Evaluating gambles using dynamics. *Chaos*, 26:23103, February 2016.
- [44] O. Peters and W. Klein. Ergodicity breaking in geometric Brownian motion. *Phys. Rev. Lett.*, 110(10):100603, March 2013.

- [45] T. Piketty. *Capital in the twenty-first century*. Harvard University Press, 2014.
- [46] T. Piketty and E. Saez. A theory of optimal inheritance taxation. *Econometrica: Journal of the Econometric Society*, 81(5):1851–1886, 2013.
- [47] 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.
- [48] M. Pinkovskiy and X. Sala-i-Martin. Parametric estimations of the world distribution of income. 2009. Working Paper 15433, NBER.
- [49] Quandl. Dow Jones Industrial Average. <http://www.quandl.com/data/BCB/UDJIAD1-Dow-Jones-Industrial-Average>, 2016. Accessed: 04/19/2016.
- [50] S. Redner. Random multiplicative processes: An elementary tutorial. *Am. J. Phys.*, 58(3):267–273, March 1990.
- [51] 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.
- [52] 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.
- [53] P. Samuelson. What classical and neoclassical monetary theory really was. *Canad. J. Econ.*, 1(1):1–15, February 1968.
- [54] A. Sen. *On Economic Inequality*. Oxford: Clarendon Press, 1997.
- [55] J. E. Stiglitz. Distribution of income and wealth among individuals. *Econometrica: Journal of the Econometric Society*, 37(3):382–397, 1969.
- [56] 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.
- [57] H. Theil. *Economics and information theory*. North-Holland Publishing Company, 1967.
- [58] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Physical Review*, 36(5):823–841, Sep 1930.
- [59] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.
- [60] P. Vermeulen. How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, 2017.
- [61] E. N. Wolff. Household wealth trends in the united states, 1983–2010. *Oxford Review of Economic Policy*, 30(1):21–43, 2014.