Ergodicity Economics

Ole Peters and Alexander Adamou

2018/06/27 at 10:26:24

Contents

1 Populations

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

1.1 Every man for himself

{section:Every_man}

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

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

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

Examples:

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

1.1.1 Log-normal distribution

{section:Log-normal_wealt

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

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

with solution

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

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

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

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

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

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

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

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

or, equivalently, $\ln \langle x(t)\rangle = \mu t$. We could confirm this result by calculating $\langle x(t)\rangle = \int_0^\infty s\mathcal{P}_x(s)\, ds$, but this would be laborious. Instead we use a neat trick, courtesy of [?, Chapter 4.2], which will come in handy again in Sec. ??. To compute moments, $\langle x^n\rangle$, of stochastic differential equations for x, like (Eq. 1.1.2), we find ordinary differential equations for the moments, which we know how to solve. For the first moment we do this simply by taking expectations of both sides of (Eq. 1.1.2). The noise term vanishes to turn the SDE for x into an ODE for $\langle x\rangle$:

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

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

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

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

1.1.2 Two growth rates

{section:two_rates}

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

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

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

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

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

1.1.3 Measuring inequality

{section:Inequality_measu

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

Clearly, to answer these questions, we must quantify "inequality". In this section, and also in [**?**], 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 [**?**], is the difference between typical wealth (growing at the time-average growth rate) and average wealth (growing at the ensemble-average growth rate) in our model. Thus, the difference between the time average and ensemble average, the essence of ergodicity breaking, is the fundamental driver of the dynamics of inequality.

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

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

Definition **Inequality, J, is the quantity whose growth rate is the**

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

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

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

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

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

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

and integrate over time to get

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

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

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

$$J(t) = \frac{\sigma^2}{2} t. \tag{19}$$ {eq:J_t}

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

1.1.4 Wealth condensation

{section:condensation}

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

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

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

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

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

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

1.1.5 Rescaled wealth

{section:rescaled}

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" [**?**, 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, we can remove absolute wealth levels from the analysis and study an object called the rescaled wealth.

Definition **The rescaled wealth,**

$$y_i(t) = \frac{x(t)}{\langle x(t)\rangle_N}, \tag{22}$$ {eq:rescaled}

is the proportion of the sample mean wealth – *i.e.* **the wealth averaged over the finite population – owned by an individual.**

This quantity is useful because its numerical value does not depend on the currency used: it is a dimensionless number. Thus if my rescaled wealth is $y_i(t) = 1/2$, this means that my wealth is half the average wealth, irrespective of whether I measure it in Kazakhstani Tenge or in Swiss Francs. The sample mean rescaled wealth is easily calculated:

$$\langle y_i(t)\rangle_N = \left\langle \frac{x(t)}{\langle x(t)\rangle_N} \right\rangle_N = 1. \tag{23}$$

If the population size, N, is large enough, then we might expect the sample mean wealth, $\langle x(t)\rangle_N$, to be close to the ensemble average, $\langle x(t)\rangle$, which is simply its $N \to \infty$ limit. We will discuss more carefully when this approximation holds for wealths following GBM in Sec. **??**. Let's assume for now that it does hold. The rescaled wealth is then well approximated as

$$y_i(t) = \frac{x_i(t)}{\langle x(t)\rangle} = x_i(t)\exp(-\mu t). \tag{24}$$

Now that we have an expression for y in terms of x and t, we can derive the dynamic for rescaled wealth using Itô's formula (just as we did to find the wealth dynamic for a general utility function in Sec. **??**). We start with

$$dy = \frac{\partial y}{\partial t}\,dt + \frac{\partial y}{\partial x}\,dx + \frac{1}{2}\frac{\partial^2 y}{\partial x^2}\,dx^2 \tag{25}$$

$$= -\mu y\,dt + \frac{y}{x}\,dx, \tag{26}$$ {eq:ysde}

and then substitute (Eq. 1.1.2) for dx to get

$$dy = y\sigma dW. \tag{27}$$ {eq:GBM_y}

Thus $y(t)$ follows a very simple GBM with zero drift and volatility σ. This means that rescaled wealth, like wealth, has an ever-broadening lognormal distribution:

$$\ln y(t) \sim \mathcal{N}\left(-\frac{\sigma^2}{2}t, \sigma^2 t\right). \quad (28)$$ {eq:lognormal_y}

Finally, noting that $\langle \ln y \rangle = \langle \ln x \rangle - \ln \langle x \rangle$ gives us a simple expression for our inequality measure in (Eq. 18) in terms of the rescaled wealth:

$$J(t) = -\langle \ln y \rangle. \quad (29)$$

1.1.6 v-normal distributions and Jensen's inequality

{section:jensen}

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

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

1.1.7 Power law resemblance

{section:power_law}

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

$$\mathcal{P}_x(x) = x^{-\alpha}, \quad (30)$$ {eq:power_law}

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

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

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

$$\ln \mathcal{P}(x) = -\ln\left(x\sqrt{2\pi\sigma^2 t}\right) - \frac{[\ln x - (\mu - \frac{\sigma^2}{2})t]^2}{2\sigma^2 t} \quad (31)$$

$$= -\ln x - \frac{\ln(2\pi\sigma^2 t)}{2} - \frac{(\ln x)^2 - 2(\mu - \frac{\sigma^2}{2})t \ln x + (\mu - \frac{\sigma^2}{2})^2 t^2}{2\sigma^2 t}. \quad (32)$$

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

$$\ln \mathcal{P}(x) = -\frac{(\ln x)^2}{2\sigma^2 t} + \left(\frac{\mu}{\sigma^2} - \frac{3}{2}\right)\ln x - \frac{\ln(2\pi\sigma^2 t)}{2} - \frac{(\mu - \frac{\sigma^2}{2})^2 t}{2\sigma^2} \quad (33)$$

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

$$\frac{d \ln \mathcal{P}(x)}{d \ln x} = -\frac{\ln x}{\sigma^2 t} + \frac{\mu}{\sigma^2} - \frac{3}{2}. \tag{34}$$

Near $\langle x \rangle$, $\ln x \sim \mu t$ so that the first two terms cancel approximately. Here the distribution will resemble a power-law with exponent $-3/2$ when plotted on doubly logarithmic scales. (The distribution will also look like a power-law where the first term is much smaller than the others, *e.g.* where $\ln x \ll \sigma^2 t$.) We don't believe that such empirically observed power laws are merely a manifestation of this mathematical feature. Important real-world mechanisms that broaden real wealth distributions, *i.e.* concentrate wealth, are missing from the null model. However, it is interesting that the trivial model of GBM reproduces so many qualitative features of empirical observations.

1.2 Finite populations

{section:finite_populatio

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

$$\langle x(t) \rangle_N \equiv \frac{1}{N} \sum_{i=1}^{N} x_i(t), \tag{35}$$ {eq:sample}

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

1.2.1 Sums of lognormals

{section:sketch}

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

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

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

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

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

$$R \equiv \frac{\operatorname{var}(\langle x(t) \rangle_N)}{\langle \langle x(t) \rangle_N \rangle^2}. \tag{36}$$

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

$$R(N) = \frac{e^{\sigma^2 t} - 1}{N}. \tag{37}$$

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

$$t < \frac{\ln N}{\sigma^2}. \tag{38}$$ {eq:short_t}

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

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

$$\langle x(t)\rangle_N \approx \frac{1}{N} \max_{i=1}^{N}\{x_i(t)\}, \tag{39}$$

as illustrated in Fig. **??**.

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

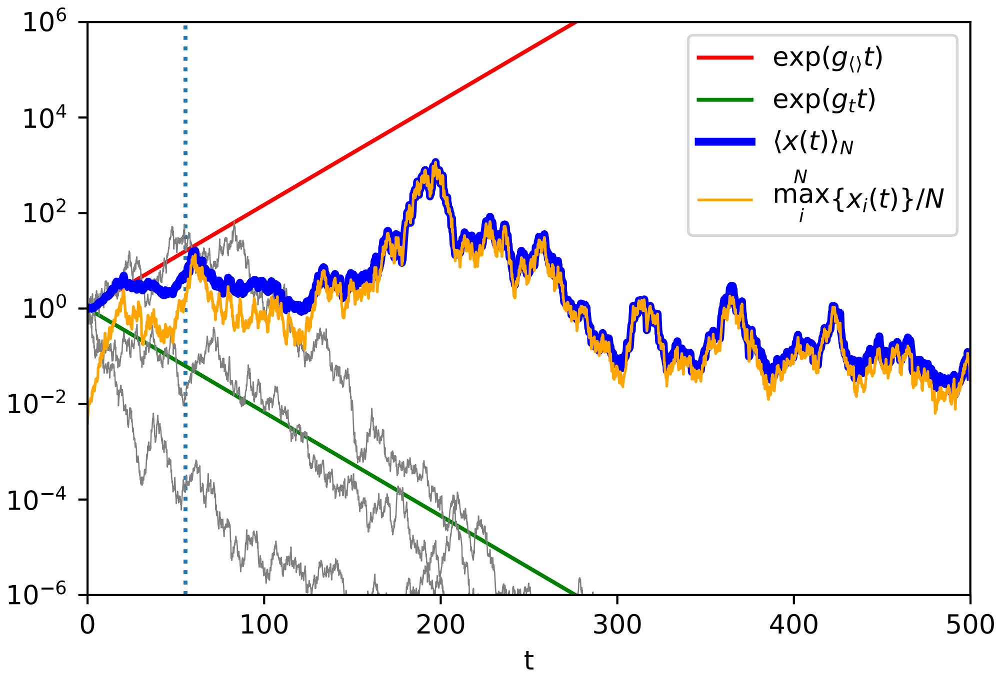

Figure 1: PEA and maximum in a finite ensemble of size $N = 256$. **Red line:** expectation value $\langle x(t)\rangle$. **Green line:** exponential growth at the time-average growth rate. In the $T \to \infty$ limit all trajectories grow at this rate. **Yellow line:** contribution of the maximum value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

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

$$\langle x \rangle_N = \frac{1}{N}\sum_{i=1}^{N} \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_i(t)\right], \tag{40}$$ {eq:PEA}

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

$$\langle x \rangle_N = \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t\right] \frac{1}{N}\sum_{i=1}^{N} \exp\left(t^{1/2}\sigma\xi_i\right), \tag{41}$$ {eq:PEA_2}

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

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

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

$$g_{\text{est}}(t, N) \equiv \frac{\ln(\langle x(t) \rangle_N) - \ln(x(0))}{t - 0} = \frac{1}{t}\ln(\langle x(t) \rangle_N). \tag{42}$$ {eq:gest}

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

$$\lim_{t\to\infty} g_{\text{est}}(t, N) = \mu - \frac{\sigma^2}{2} \tag{43}$$ {eq:gest_2}

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

$$\begin{aligned} g_{\text{est}}(t, N) &= \mu - \frac{\sigma^2}{2} + \frac{1}{t}\ln\left(\frac{1}{N}\sum_{i=1}^{N}\exp(t^{1/2}\sigma\xi_i)\right) & (44) \\ &= \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} + \frac{1}{t}\ln\left(\sum_{i=1}^{N}\exp(t^{1/2}\sigma\xi_i)\right). & (45) \end{aligned}$$ {eq:gest_4}

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

$$\langle g_{\text{est}}(t, N) \rangle = \frac{1}{t}\langle \ln(\langle x(t) \rangle_N) \rangle = f(N, t), \tag{46}$$ {eq:gest_3}

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

1.2.2 The random energy model

{section:REM}

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

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

$$Z = \sum_{i=1}^{N} \exp\left(\beta J \sqrt{\frac{K}{2}} \xi_i\right), \tag{47}$$ {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 [?, p. 79]. J is a constant that will be determined below. The logarithm of the partition function gives the Helmholtz free energy,

$$\begin{aligned} F &= -\frac{\ln Z}{\beta} && (48) \\ &= -\frac{1}{\beta} \ln\left[\sum_{i=1}^{N} \exp\left(\beta J \sqrt{\frac{K}{2}} \xi_i\right)\right]. && (49) \end{aligned}$$ {eq:F}

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

$$g_{\text{est}} = \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} - \frac{\beta F}{t}, \tag{50}$$ {eq:gest_5}

which is valid provided that

$$\beta J \sqrt{\frac{K}{2}} = \sigma t^{1/2}. \tag{51}$$ {eq:map}

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

$$\sigma = \frac{J}{\sqrt{2}} \quad \text{and} \quad t^{1/2} = \beta\sqrt{K}. \tag{52}$$ {eq:choice_1}

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

Derrida identifies a critical temperature,

$$\frac{1}{\beta_c} \equiv \frac{J}{2\sqrt{\ln 2}}, \tag{53}$$ {eq:beta_c}

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

$$t_c = \frac{2 \ln N}{\sigma^2}, \tag{54}$$ {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. **??**) scales identically with N and σ as the transition time, (Eq. **??**), in our sketch.

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

$$\begin{aligned} \langle F \rangle &= E - S/\beta & (55)\\ &= -\frac{K}{\beta}\ln 2 - \frac{\beta K J^2}{4}, & (56) \end{aligned}$$ {eq:F_2}

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

$$\langle F \rangle = -KJ\sqrt{\ln 2}. \tag{57}$$ {eq:F_3}

Short time

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

$$\langle g_{\text{est}} \rangle^{\text{short}} = \mu - \frac{\sigma^2}{2} + \frac{1}{t}\frac{KJ^2}{4T^2}. \tag{58}$$ {eq:gest_6}

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

$$\langle g_{\text{est}} \rangle^{\text{short}} = \mu, \tag{59}$$ {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. **??**)). Again taking expectation values and substituting, this time from (Eq. **??**) in (Eq. **??**), we find for long times

$$\langle g_{\text{est}} \rangle^{\text{long}} = \mu - \frac{\sigma^2}{2} - \frac{\ln N}{t} + \sqrt{\frac{2 \ln N}{t}}\,\sigma, \tag{60}$$ {eq:gest_8}

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

Figure **??** is a direct comparison between the results derived here, based on [**?**], and numerical results using the same parameter values as in [**?**], 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 [**?**] we used a notation that we've stopped using since then because it caused confusion – $\langle g \rangle$ there denotes the growth rate of the expectation value, which is not the expectation value of the growth rate.

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

$$q_{0.5}(\langle x(t) \rangle_N) \approx \exp\left(\langle g_{\text{est}}(N,t) \rangle t\right). \tag{61}$$ {eq:quant_ave}

In [**?**] it was discussed in detail that $g_{\text{est}}(1,t)$ is an ergodic observable for (Eq. 1.1.2), in the sense that $\langle g_{\text{est}}(1,t) \rangle = \lim_{t\to\infty} g_{\text{est}}$. The relationship in (Eq. **??**) is far more subtle. The typical behavior of GBM PEAs is complicated outside the limits $N \to \infty$ or $t \to \infty$, in the sense that growth rates are time dependent here. This complicated behaviour is well represented by an approximation that uses physical insights into spin glasses.

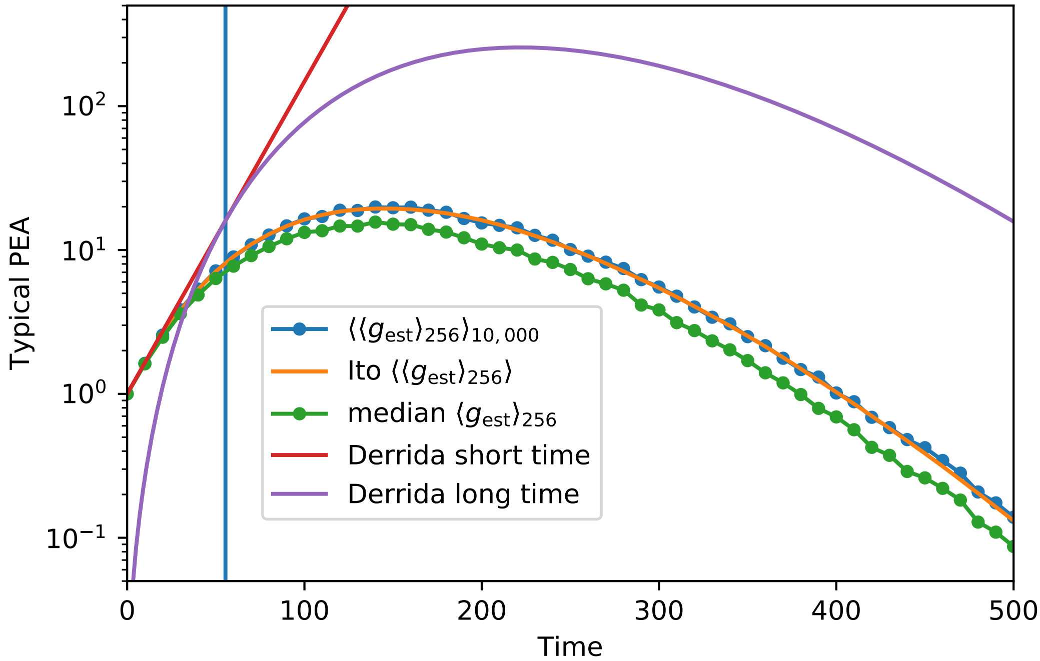

Figure 2: Lines are obtained by exponentiating the various exponential growth rates. **Blue line:** $\langle\langle g_{\mathrm{est}}\rangle_{256}\rangle_{10{,}000}$ is the numerical mean (approximation of the expectation value) over a super-ensemble of $S = 10,000$ samples of g_{est} estimated in sub-ensembles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

2 Interactions

Insert abstract here.

2.1 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

$$\overline{g}(x_1) > \overline{g}(x_2) \tag{62}$$

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. 1.1.2). How would this operation affect the dynamic of the wealth of these two individuals?

Consider a discretized version of (Eq. 1.1.2), such as would be used in a numerical simulation. The non-cooperators grow according to

$$\delta x_i(t) = x_i(t)\left[\mu\delta t + \sigma\sqrt{\delta t}\,\xi_i\right], \tag{63}$$ {eq:discrete_nonc_grow}

$$x_i(t+\delta t) = x_i(t) + \delta x_i(t), \tag{64}$$ {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

$$x_1 \oplus x_2 = x_1^c + x_2^c. \tag{65}$$

In the discrete-time picture, each time step involves a two-phase process. First there is a growth phase, analogous to (Eq. 1.1.2), in which each cooperator increases its resources by

$$\delta x_i^c(t) = x_i^c(t)\left[\mu\delta t + \sigma\sqrt{\delta t}\,\xi_i\right]. \tag{66}$$ {eq:discrete_coop_grow}

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

$$x_i^c(t+\delta t) = \frac{x_1^c(t) + \delta x_1^c(t) + x_2^c(t) + \delta x_2^c(t)}{2}. \tag{67}$$ {eq:discrete_coop_coop}

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

$$(x_1 \oplus x_2)(t+\delta t) = (x_1 \oplus x_2)(t)\left[1 + \left(\mu\delta t + \sigma\sqrt{\delta t}\,\frac{\xi_1+\xi_2}{2}\right)\right]. \tag{68}$$ {eq:discrete_cooperate}

For ease of notation we define

$$\xi_{1\oplus 2} = \frac{\xi_1+\xi_2}{\sqrt{2}}, \tag{69}$$

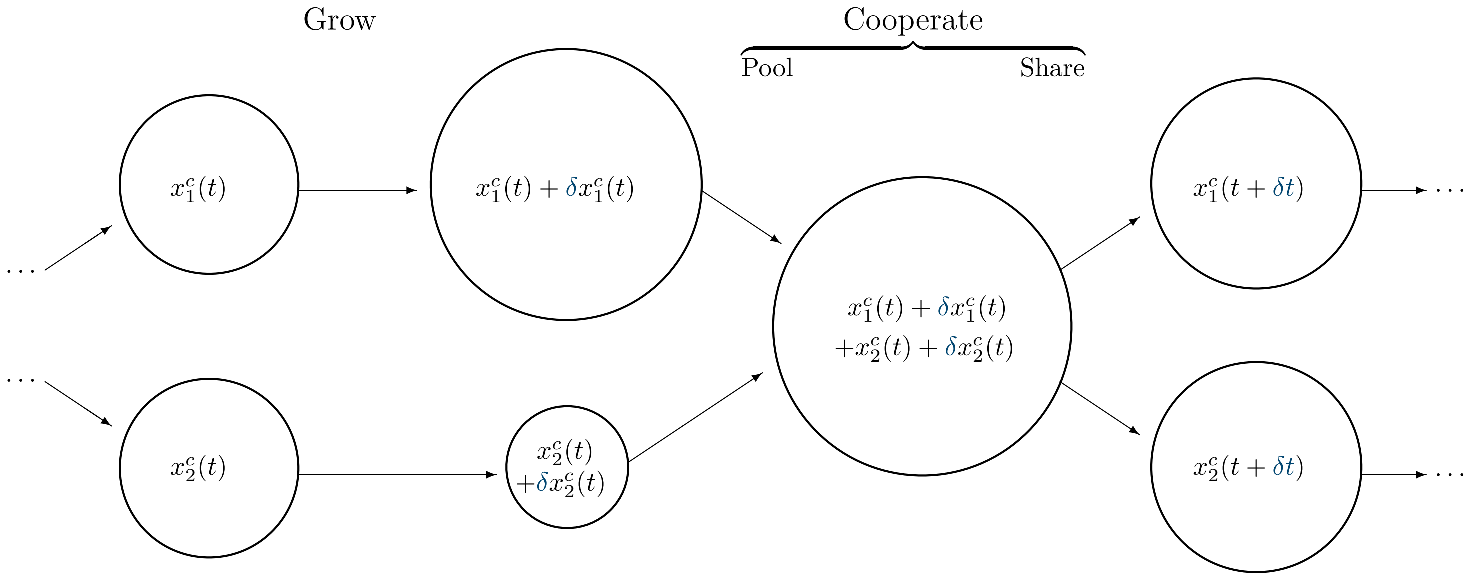

Figure 3: Cooperation dynamics. Cooperators start each time step with equal resources, then they *grow* independently according to (Eq. **??**), then they *cooperate* by *pooling* resources and *sharing* them equally, then the next time step begins. {fig:dynamics}

which is another standard Gaussian, $\xi_{1\oplus 2} \sim \mathcal{N}(0,1)$. Letting the time increment $\delta t \to 0$ we recover an equation of the same form as (Eq. 1.1.2) but with a different fluctuation amplitude,

$$d(x_1 \oplus x_2) = (x_1 \oplus x_2)\left(\mu dt + \frac{\sigma}{\sqrt{2}} dW_{1\oplus 2}\right). \tag{70}$$

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

$$g_t(x_1 \oplus x_2) > g_t(x_i) \tag{71}$$

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. **??** by direct simulation of (Eq. **??**)–(Eq. **??**) and (Eq. **??**).

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, $x_1 \oplus x_2 ... \oplus x_n$ 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

$$\mu_1 - \frac{\sigma_1^2}{2} < \frac{\mu_1 + \mu_2}{2} - \frac{\sigma_1^2 + \sigma_2^2}{8}. \tag{72}$$

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. **??**. 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 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. **??** is the ensemble average of $x_1(t)$, $x_1(t)$, 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.

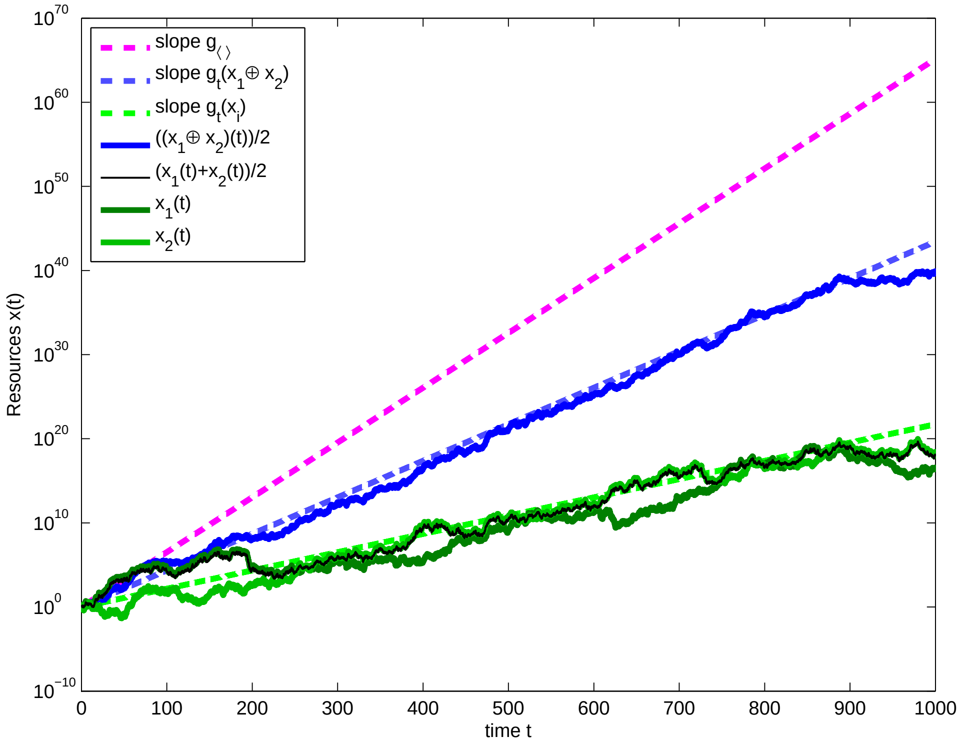

Figure 4: Typical trajectories for two non-cooperating (green) entities and for the corresponding cooperating unit (blue). Over time, the noise reduction for the cooperator leads to faster growth. Even without effects of specialisation or the emergence of new function, cooperation pays in the long run. 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}

2.2 Reallocation

{section:reallocation}

2.2.1 Introduction

{section:RGBM_intro}

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

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

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

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

2.2.2 The ergodic hypothesis in economics

{section:RGBM_EH}

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

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

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

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

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

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

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

2.2.3 Reallocating GBM

{section:RGBM_model}

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

$$dx_i = x_i\left[(\mu - \tau)dt + \sigma dW_i\left(t\right)\right] + tau\left\langle x\right\rangle_N\ dt, \tag{73}$$ {eq:rgbm}

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

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

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

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

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

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

2.2.4 Model behaviour

{section:RGBM_behaviour}

It is instructive to write (Eq. ??) as

$$dx_i = \underbrace{x_i \left[\mu dt + \sigma dW_i\left(t\right)\right]}_{\text{Growth}} \ - \ \underbrace{\tau(x_i - \langle x \rangle_N)dt}_{\text{Reallocation}} . \quad (74)$$ {eq:rgbm_ou}

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

Positive τ

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

$$t < t_c \equiv \frac{2 \ln N}{\sigma^2}, \quad (75)$$ {eq:t_c}

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

$$\langle x(t) \rangle_N \sim \langle x(t) \rangle = \exp(\mu t). \quad (76)$$ {eq:rgbm_self}

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

$$\tau > \tau_c \equiv \frac{\sigma^2}{2 \ln N}. \quad (77)$$ {eq:tau_c}

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

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

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

$$dy_i = y_i \sigma dW_i\left(t\right) - \tau(y_i - 1)dt. \quad (78)$$ {eq:rgbm_ou_re}

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

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

$$\mathcal{P}\left(y\right) = \frac{\left(\zeta - 1\right)^{\zeta}}{\Gamma\left(\zeta\right)} e^{-\frac{\zeta - 1}{y}} y^{-(1+\zeta)}. \quad (79)$$ {eq:disti}

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

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

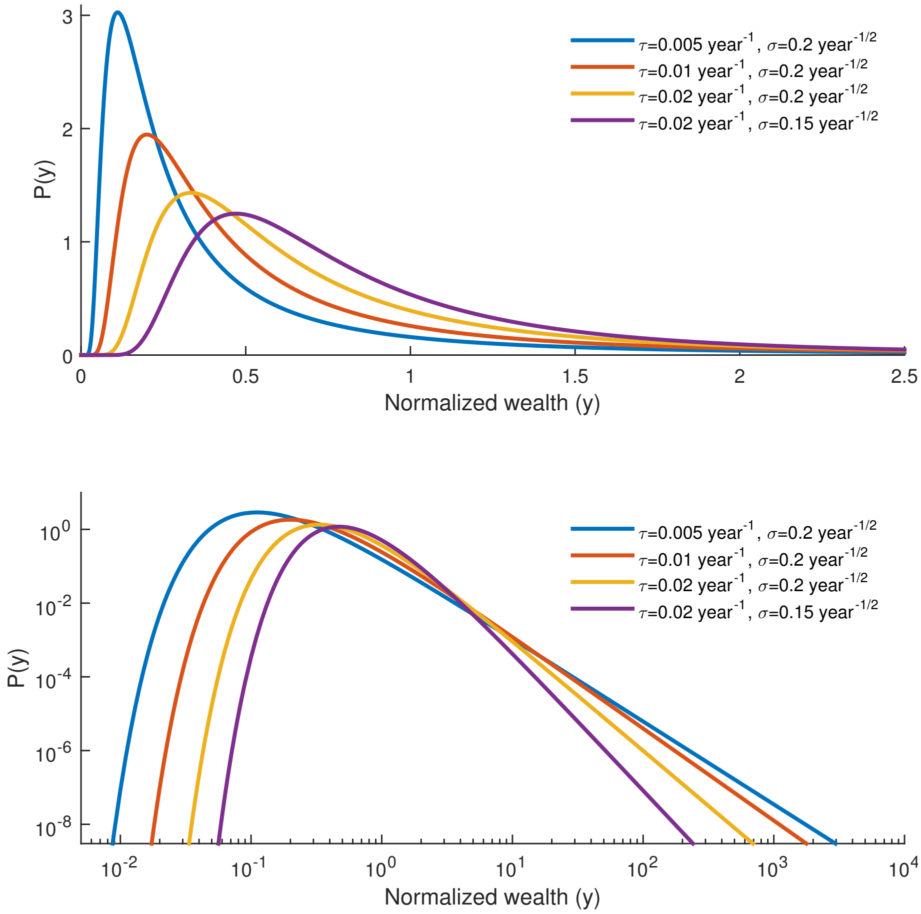

Figure 5: The stationary distribution for RGBM with positive τ. Top – linear scales; Bottom – logarithmic scales. {fig:dist}

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

Negative τ

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

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

2.2.5 Derivation of the stable distribution

{section:RGBM_stable}

In this section we will sketch the argument for why we can make the self-averaging approximation, (Eq. **??**), in RGBM with sufficiently fast positive reallocation, (Eq. **??**). This is shown rigorously in [**?**]. Then we will solve the Fokker-Planck equation for the rescaled wealth and derive the inverse gamma distribution, (Eq. **??**). If you are happy to believe the quoted results in Sec. **??**, then you can skip the Fokker-Planck bit safely.

We presented arguments in Sec. **??** for why wealth in GBM is self-averaging, $\langle x(t)\rangle_N \sim \langle x(t)\rangle = \exp(\mu t)$ for short time. By mapping from GBM to the random energy model in Sec. **??**, we showed that "short time" means $t < t_c$, where $t_c = 2\ln N/\sigma^2$. We can think of this as follows: t_c is the timescale over which the inequality-increasing effects of noisy multiplicative growth drive wealths apart, such that a finite sample of wealths stops self-averaging and becomes dominated by a few trajectories.

Let's now think about what happens when we add reallocation to GBM, creating RGBM. τ is the reallocation rate, so τ^{-1} is reallocation timescale, *i.e.* the timescale over which the inequality-reducing effects of reallocation pull wealths together. If $\tau^{-1} > t_c$, then reallocation happens too slowly to prevent the expiry of self-averaging. However, if $\tau^{-1} < t_c$, then reallocation pulls wealths together more quickly than they get driven apart, continually "resetting" the sample and allowing self-averaging to be maintained indefinitely. Converting this condition into a reallocation rate, we get $\tau > {t_c}^{-1}$, as in (Eq. **??**). As mentioned in Sec. **??**, this becomes indistinguishable from $\tau > 0$ for realistic parameters, so the self-averaging approximation can be made safely for all positive τ.

We can now approximate the rescaled wealth, $y_i(t) = x_i(t)/\langle x(t)\rangle_N$, as

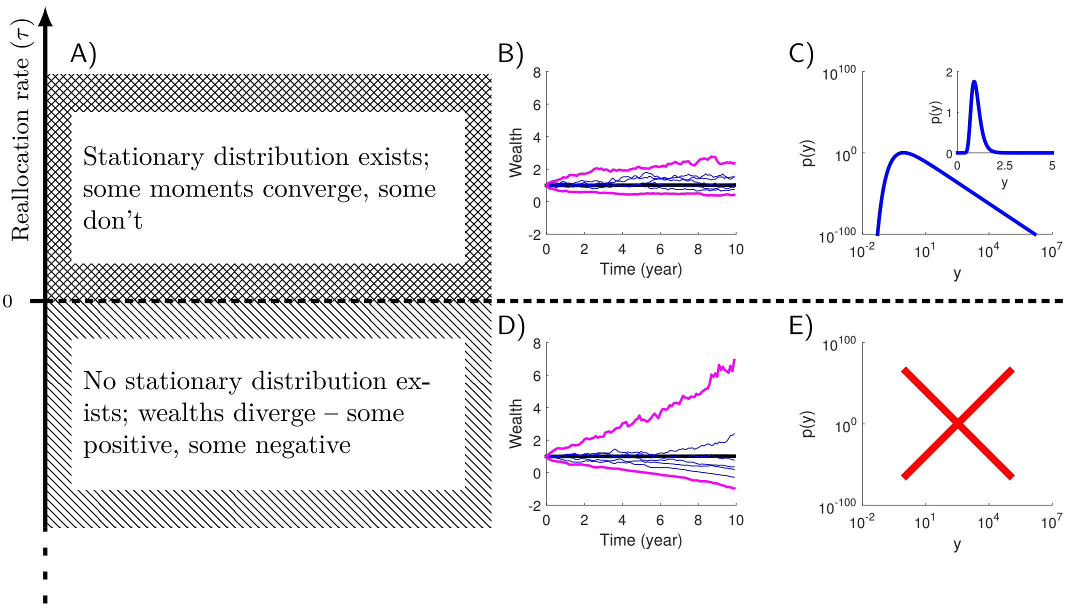

Figure 6: Regimes of RGBM. A) $\tau = 0$ separates the two regimes of RGBM. For $\tau > 0$, a stationary wealth distribution exists. For $\tau < 0$, no stationary wealth distribution exists and wealths diverge. B) Simulations of RGBM with $N = 1000$, $\mu = 0.021$ year^{-1} (presented after rescaling by $\exp(\mu t)$), $\sigma = 0.14$ year$^{-1/2}$, $x_i(0) = 1$, $\tau = 0.15$ year^{-1}. Magenta lines: largest and smallest wealths, blue lines: five randomly chosen wealth trajectories, black line: sample mean. C) The stationary distribution to which the system in B) converges. Inset: same distribution on linear scales. D) Similar to B), with $\tau = -0.15$ year^{-1}. E) in the $\tau < 0$ regime, no stationary wealth distribution exists. {fig:regimes}

$y_i(t) = x_i(t)\exp(-\mu t)$, which follows the SDE:

$$dy = \sigma\, y\, dW - \tau\,(y-1)\,dt. \tag{80}$$

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

$$\frac{d\mathcal{P}\,(y,t)}{dt} = \frac{\partial}{\partial y}\left[A\mathcal{P}\,(y,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left[B^2\mathcal{P}\,(y,t)\right]. \tag{81}$$

The steady-state Fokker-Planck equation for the PDF, $\mathcal{P}\,(y)$, is obtained by setting the time derivative to zero,

$$\frac{\sigma^2}{2}\left(y^2\mathcal{P}\right)_{yy} + \tau\left[(y-1)\,\mathcal{P}\right]_y = 0. \tag{82}$$ {eq:fokker_planck}

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

$$\mathcal{P}(y) = C(\zeta)\, e^{-\frac{\zeta-1}{y}} y^{-(1+\zeta)}\,, \tag{83}$$

where

$$\zeta = 1 + \frac{2\tau}{\sigma^2} \tag{84}$$

and

$$C(\zeta) = \frac{(\zeta-1)^{\zeta}}{\Gamma(\zeta)}\,, \tag{85}$$

with the gamma function $\Gamma(\zeta) = \int_0^\infty x^{\zeta-1} e^{-x}\, \mathrm{d}x$. The distribution has a power-law tail as $y \to \infty$, resembling Pareto's oft-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.

2.2.6 Moments and convergence times

{RGBM_moments}

The inverse gamma distribution, (Eq. **??**), has a power-law tail. This means that, for positive reallocation, while some of the lower moments of the stable rescaled wealth distribution may exist, higher moments will not. Specifically, the k^{th} moment diverges if $k > \zeta$.

If we find parameters consistent with positive reallocation when we fit our model to data, we will be interested in whether certain statistics – such as the variance – exist. We will also want to know how long it takes the distribution to converge sufficiently to its stable form for them to be meaningful. Here we derive a condition for the convergence of the variance and calculate its convergence time, noting also the general procedure for other statistics.

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

$$V(y) = \langle y^2 \rangle - \langle y \rangle^2 \tag{86}$$

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

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

$$dy = \sigma y\, dW - \tau (y-1)\, dt. \tag{87}$$ {eq:rescaledSDE}

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

$$df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial y} dy + \frac{1}{2} \frac{\partial^2 f}{\partial y^2} dy^2\,. \tag{88}$$

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

$$d\left(y^2\right) = 2y dy + (dy)^2\,. \tag{89}$$ {eq:diff2}

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

$$d\left(y^2\right) = 2\sigma y^2\,dW - \left(2\tau - \sigma^2\right) y^2\,dt + 2\tau y\,dt \tag{90}$$

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

$$\frac{d\langle y^2\rangle}{dt} = -\left(2\tau - \sigma^2\right)\langle y^2\rangle + 2\tau \tag{91}$$

{eq:avediff2}

with solution

$$\langle y\,(t)^2\rangle = \frac{2\tau}{2\tau - \sigma^2} + \left(\langle y\,(0)^2\rangle - \frac{2\tau}{2\tau - \sigma^2}\right) e^{-\left(2\tau - \sigma^2\right)t}\,. \tag{92}$$

{eq:avediff3}

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

$$V\,(t) = V_\infty + \left(V_0 - V_\infty\right) e^{-\left(2\tau - \sigma^2\right)t}\,, \tag{93}$$

{eq:var1}

where V_0 is the initial variance and

$$V_\infty = \frac{2\tau}{2\tau - \sigma^2}\,. \tag{94}$$

{eq:varinf}

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

$$\tau > \frac{\sigma^2}{2}, \tag{95}$$

which is the same as the condition we noted previously for the 2^{nd} moment to exist: $\zeta > 2$.

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 (**??**) that the convergence time of the variance is $1/\left(2\tau - \sigma^2\right)$.

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

List of Symbols

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

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

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

$\mathcal{P}$ Probability density function.

var Variance.

References

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

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

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

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

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

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

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

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

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

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

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

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

[13] 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.

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

[15] 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.

[16] 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.

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

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

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

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

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

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

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

[24] 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.

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

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

[27] 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.

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

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

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

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