Abstract

Contents

1 Populations

1.1 Every man for himself

{section:Every_man}

We have seen that risk aversion constitutes optimal behaviour under the assumptions of multiplicative wealth growth, and 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
$$\mathcal{P}_x(x) = \delta(x - \langle x \rangle_N) \tag{1}$$
- 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 wealth distribution

{section:Log-normal_wealt

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

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

with solution

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

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

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

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

{eq:PDFx}

It is a well-established empirical observation [4] 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 aymptotic form

$$\mathcal{P}_x(x) = x^{-\alpha}, \tag{7}$$

{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

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

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

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

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

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

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

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 intersting, 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 [7], 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, $g_{\langle\rangle}$, 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, g_t, 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 a-typical 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,**

$$\frac{dJ}{dt} = g_{\langle\rangle} - g_t. \tag{12}$$

{eq:dJ}

Equation (12) defines the dynamic of inequality, and inequality itself

is found by integrating over time

$$J(t) = \int_0^t ds[g_{\langle\rangle}(s) - g_t(s)]. \tag{13}$$ {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

$$\frac{dJ}{dt} = \frac{d\ln\langle x\rangle}{dt} - \frac{d\langle\ln x\rangle}{dt}. \tag{14}$$ {eq:J_dyn}

Integrating over time

$$J(t) = \ln\langle x\rangle - \langle\ln x\rangle\,, \tag{15}$$

which is known as the mean logarithmic deviation (MLD) or Theil's second index of inequality [7]. 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. 1.3.1 (the general procedure is described in [3, Chapter 4.2]): to compute moments, $\langle x^n\rangle$, of stochastic differential equations for x, like (Eq. 3), 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. 3). The noise term disappears, and we turn the SDE for x into an ODE for $\langle x\rangle$

$$\begin{aligned}
\langle dx\rangle &= \langle x(\mu dt + \sigma dW)\rangle && (16)\\
d\langle x\rangle &= \langle x\rangle\,\mu dt + \sigma\,\overbrace{\langle dW\rangle}^{=0} && (17)\\
&= \langle x\rangle\,\mu dt. && (18)
\end{aligned}$$

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

$$\langle x\rangle = \langle x_0\rangle\exp(\mu t) \tag{19}$$ {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

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

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.* [2]. 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" [6, Appendix 4], and J is one of them. One of the conditions is that inequality measures should not change when x is divided by the same factor for everyone. Since we are primarily interested in inequality in this section it is useful to remove absolute wealth levels from the analysis and study an object called the rescaled wealth.

Definition **The rescaled wealth,**

$$y = \frac{x}{\langle x \rangle_N}, \tag{21}$$

{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 $\frac{dJ}{dt} = -\frac{d\langle \ln y \rangle}{dt}$.

1.2 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

$$g_t(x_1) > g_t(x_2) \tag{22}$$

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

Consider a discretized version of (Eq. 3), such as would be used in a numerical simulation. The non-cooperators grow according to

$$\begin{aligned} \Delta x_i(t) &= x_i(t)\left[\mu \Delta t + \sigma \sqrt{\Delta t}\, \xi_i\right], & (23) \\ x_i(t+\Delta t) &= x_i(t) + \Delta x_i(t), & (24) \end{aligned}$$

{eq:discrete_nonc_grow}
{eq:discrete_nonc_coop}

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

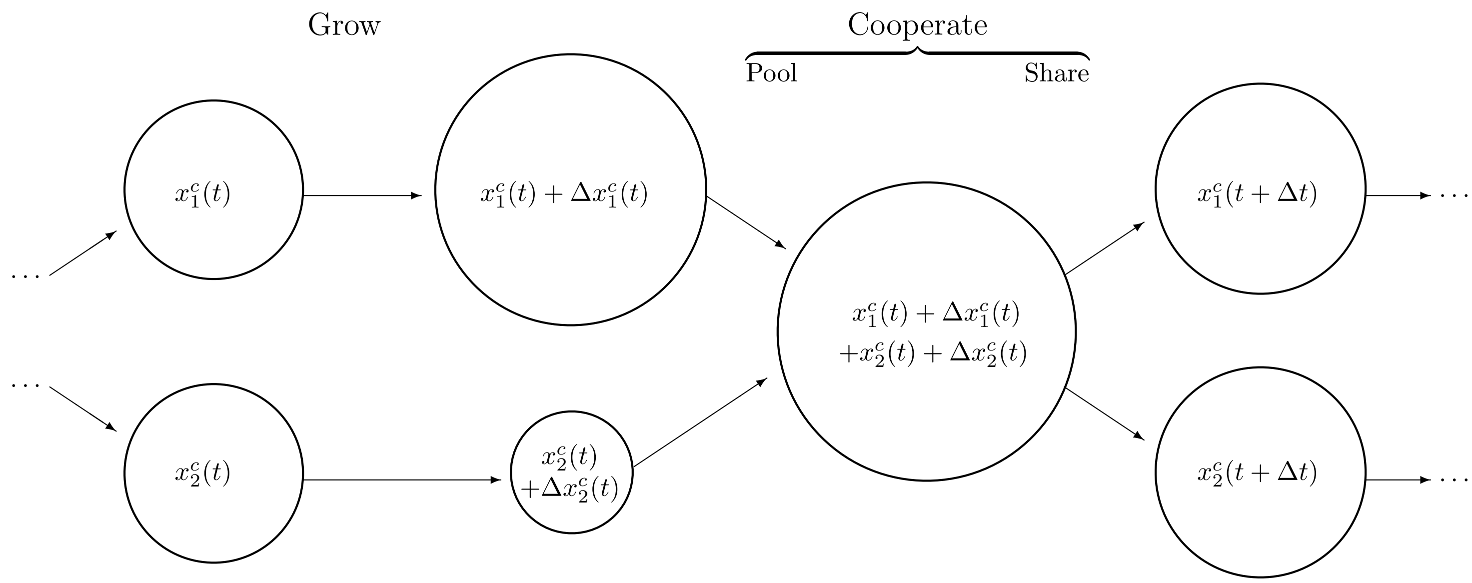

Figure 1: Cooperation dynamics. Cooperators start each time step with equal resources, then they *grow* independently according to (Eq. 26), then they *cooperate* by *pooling* resources and *sharing* them equally, then the next time step begins. {fig:dynamics}

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

$$x_1 \oplus x_2 = x_1^c + x_2^c. \tag{25}$$

In the discrete-time picture, each time step involves a two-phase process. First there is a growth phase, analogous to (Eq. 3), in which each cooperator increases its resources by

$$\Delta x_i^c(t) = x_i^c(t) \left[\mu \Delta t + \sigma \sqrt{\Delta t}\, \xi_i\right]. \tag{26}$$ {eq:discrete_coop_grow}

This is followed by a cooperation phase, replacing (Eq. 24), in which resources are pooled and shared equally among the cooperators:

$$x_i^c(t + \Delta t) = \frac{x_1^c(t) + \Delta x_1^c(t) + x_2^c(t) + \Delta x_2^c(t)}{2}. \tag{27}$$ {eq:discrete_coop_coop}

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

$$(x_1 \oplus x_2)(t + \Delta t) = (x_1 \oplus x_2)(t) \left[1 + \left(\mu \Delta t + \sigma \sqrt{\Delta t}\, \frac{\xi_1 + \xi_2}{2}\right)\right]. \tag{28}$$ {eq:discrete_cooperate}

For ease of notation we define

$$\xi_{1 \oplus 2} = \frac{\xi_1 + \xi_2}{\sqrt{2}}, \tag{29}$$

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. 3) but with a different fluctuation amplitude,

$$d(x_1 \oplus x_2) = (x_1 \oplus x_2)\left(\mu dt + \frac{\sigma}{\sqrt{2}} dW_{1\oplus 2}\right). \tag{30}$$

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

$$g_t(x_1 \oplus x_2) > g_t(x_i) \tag{31}$$

for any non-zero noise amplitude. Imagine a world where cooperation does not exist, just like in Sec. **??**. 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. 2 by direct simulation of (Eq. 23)–(Eq. 24) and (Eq. 28).

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

$$\mu_1 - \frac{\sigma_1^2}{2} < \frac{\mu_1 + \mu_2}{2} - \frac{\sigma_1^2 + \sigma_2^2}{8}. \tag{32}$$

We emphasize that this inequality may be satisfied also if the expectation value of entity 1 grows faster than the expectation values of entity 2, *i.e.* if $\mu_1 > \mu_2$. An

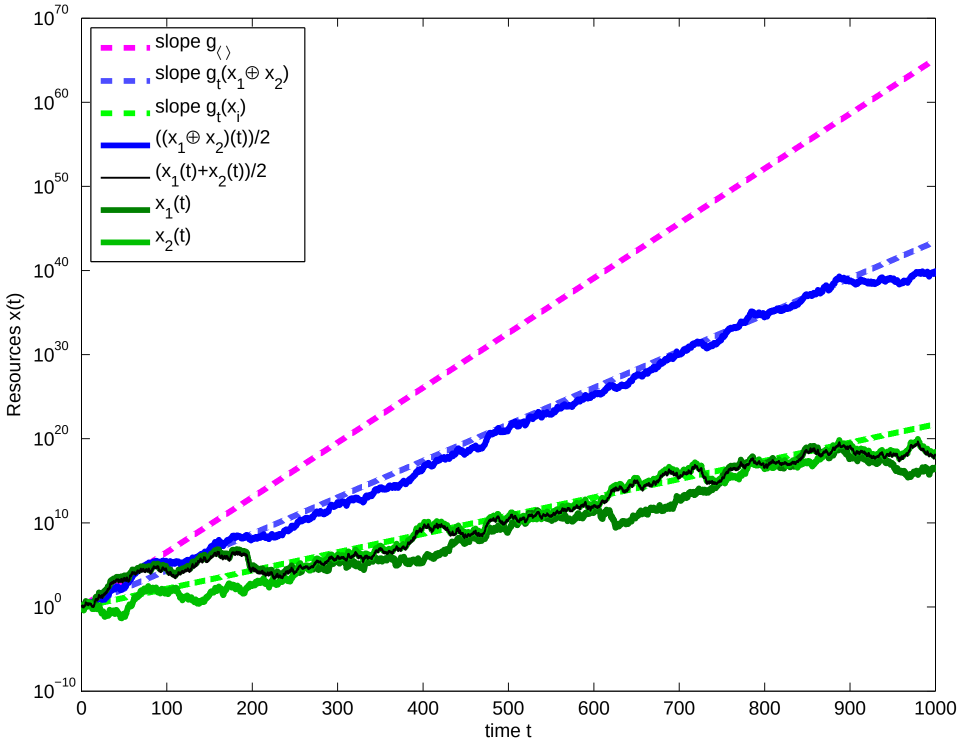

Figure 2: Typical trajectories for two non-cooperating (green) entities and for the corresponding cooperating unit (blue). Over time, the noise reduction for the cooperator leads to faster growth. Even without effects of specialisation or the emergence of new function, cooperation pays in the long run. 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}

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. 2. 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. 2 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.

1.3 Taxation

{section:Taxation}

In Sec. 1.1 we created a world of independent GBMs; in Sec. 1.2 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^{th} of the total collected amount to each individual. A similar model will be used for income tax, see (Eq. 52) in Sec. 1.3.2.

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.

<u>**Definition**</u> **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 $t \to \infty$.

Section 1.2 was concerned with growth, here we are concerned with inequality. We will therefore work with the rescaled wealth, y, introduced in (Eq. 21). Equation (3) defines the dynamic of x. From it we can find the dynamic for $f(x) = y$ using Itô calculus

$$\begin{aligned} df &= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}dx^2 && (33)\\ = dy &= -\mu y dt + \frac{y}{x}dx && (34)\\ &= y\sigma dW && (35)\end{aligned}$$

1.3.1 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,

$$dx = x[(\mu - \tau)\, dt + \sigma\, dW] + \tau \left\langle x \right\rangle_N dt, \qquad (36)$$ {eq:wsde}

which is a modified version of (Eq. 3) – the term $-\tau x dt$ was added to represent tax collection, and the term $+\tau \left\langle x \right\rangle_N dt$ to represent redistribution of collected tax. To make the model more tractable we consider the case $N \to \infty$, which replaces the finite-ensemble average by the expectation value, $\left\langle x \right\rangle_N \to \left\langle x \right\rangle$. 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, $\left\langle x(t) \right\rangle = \left\langle x_0 \right\rangle e^{\mu t}$, 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. 21)), whose dynamic we derive using the chain rule

$$\begin{aligned} dy &= \frac{\partial y}{\partial t}\, dt + \frac{\partial y}{\partial x}\, dx + \frac{1}{2}\frac{\partial^2 y}{\partial x^2}\, dx^2 && (37)\\ &= -\mu y\, dt + \frac{1}{x}\, dx && (38)\\ &= y(-\tau dt + \sigma dW) + \tau dt. && (39)\end{aligned}$$

{eq:ysde} (38)

The first moment of y is trivially 1,

$$\left\langle y \right\rangle = \left\langle \frac{x}{\left\langle x \right\rangle} \right\rangle = 1. \qquad (40)$$

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

$$\begin{aligned} d(y^2) &= \frac{\partial(y^2)}{\partial t}dt + \frac{\partial(y^2)}{\partial y}dy + \frac{1}{2}\frac{\partial^2(y^2)}{\partial y^2}(dy)^2 & (41)\\ &= 2ydy + (dy)^2 & (42)\\ &= 2y^2(-\tau dt + \sigma dW) + 2y\tau dt + y^2\sigma^2 dt. & (43)\end{aligned}$$

Taking expectation values yields

$$\begin{aligned} \langle dy^2 \rangle &= \langle -2y^2(-\tau dt + \sigma dW) + 2y\tau dt - y^2\sigma^2 dt \rangle & (44)\\ = d\langle y^2\rangle &= (\sigma^2 - 2\tau)\langle y^2 \rangle dt + 2\tau dt. & (45)\end{aligned}$$

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

$$\left(\frac{d}{dt} - (\sigma^2 - 2\tau)\right)\langle y^2\rangle = 2\tau. \quad (46) \quad \texttt{\{eq:ODE2\}}$$

Such equations are solvable using the method of integrating factors, see *e.g.* [1, Chpater 1.5]. The solution of the dynamic (Eq. 46) is the second moment of the distribution of rescaled wealth as a function of time, namely

$$\langle y^2\rangle = \frac{2\tau}{2\tau - \sigma^2} + \left(\langle y_0^2\rangle - \frac{2\tau}{2\tau-\sigma^2}\right)e^{-(2\tau-\sigma^2)t}. \quad (47)$$

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

$$V(t) = V_\infty + (V_0 - V_\infty)e^{-(2\tau-\sigma^2)t}, \quad (48) \quad \texttt{\{eq:wvar\}}$$

where V_0 is the initial variance and

$$V_\infty \equiv \frac{\sigma^2}{2\tau - \sigma^2}. \quad (49)$$

V converges in time to the asymptote, V_∞, provided the exponential in (Eq. 48) is decaying. This can be expressed as a condition on τ,

$$\boxed{\tau > \tau_c \equiv \frac{\sigma^2}{2},} \quad (50) \quad \texttt{\{eq:wstab\}}$$

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. 46), 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.

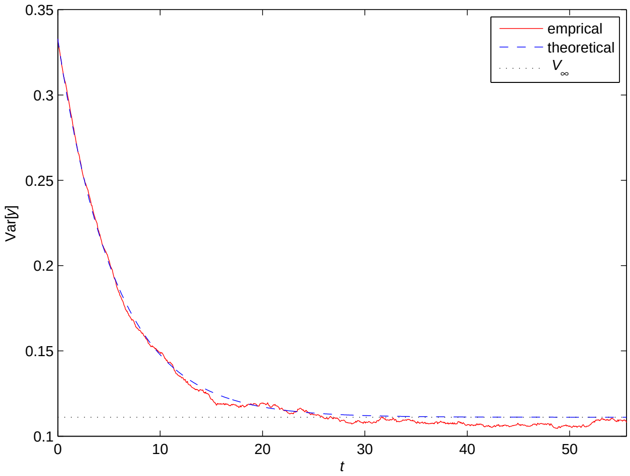

Figure 3: Wealth tax. The empirical variance of the rescaled wealths of $N = 10^4$ realisations of (Eq. 36) with uniformly-distributed initial wealths (red); the theoretical variance for the infinite ensemble, $V(t)$ (blue dashed); and the asymptotic theoretical variance, V_∞ (black dotted). Parameter values are $\mu = 0.05$, $\sigma^2 = 0.02$, and $\tau = 0.1$ per unit time. {fig:var_wealth}

Caveats aside, (Eq. 48) also allows us to identify a characteristic timescale over which the variance stabilises for supercritical taxation,

$$T_s = \frac{1}{2\tau - \sigma^2}. \tag{51}$$

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. 3 compares the evolution of the empirical variance of the rescaled wealths of $N = 10^4$ realisations of the stochastic wealth process in (Eq. 36) with the theoretical result for the infinite ensemble in (Eq. 48). 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 [5] (assuming individual wealth processes share parameters with stock market indices). The differences are finite-sample effects. Fig. 4 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. 3.

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 es-

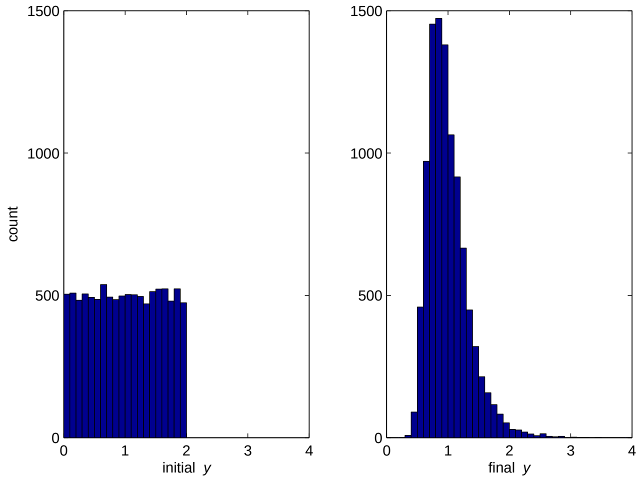

Figure 4: Histograms of the initial (left) and final (right) empirical distributions of the rescaled wealth for the same realisations of (Eq. 36) used in Fig. 3. {fig:hist_wealth}

tate (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% corresonds 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 taxes.

1.3.2 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 determinsitic wealth increment, $\mu x\, dt$, is deducted and the total tax raised is redistributed equally. This is modelled by the stochastic wealth process,

$$dx = x[\mu(1-\tau)\, dt + \sigma\, dW] + \mu\tau \left\langle x \right\rangle_N dt. \qquad (52)$$ {eq:isde}

Again, we consider the large-population limit $N \to \infty$, corresponding to the replacement $\left\langle x \right\rangle_N \to \left\langle x \right\rangle$. 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. 52) is a dimensionless number, whereas it is a rate of dimension "per unit time" in (Eq. 36). The form of (Eq. 52) is identical to (Eq. 36) with the parameter transformation $\tau \to \mu\tau$. Thus we can immediately deduce the dynamic for the rescaled wealth as

$$dy = y(-\mu\tau\, dt + \sigma\, dW) + \mu\tau\, dt. \tag{53}$$ {eq:itax}

The variance stabilisation condition analogous to (Eq. 50) becomes

$$\boxed{\tau > \tau_c \equiv \frac{\sigma^2}{2\mu}.} \tag{54}$$ {eq:icrit}

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

$$V_\infty = \frac{\sigma^2}{2\mu\tau - \sigma^2}, \tag{55}$$

according to

$$V(t) = V_\infty + (V_0 - V_\infty)e^{-(2\mu\tau-\sigma^2)t}. \tag{56}$$ {eq:ivar}

Finally, the stabilisation time is

$$T_s = \frac{1}{2\mu\tau - \sigma^2}. \tag{57}$$

Fig. 5 compares the evolution of the empirical variance of the rescaled wealths of 10^4 realisations of the stochastic wealth process in (Eq. 52) 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. 5 than in Fig. 3. This is due entirely to the simulated parameter values: (Eq. 36) and (Eq. 52) can be made equivalent by choosing different parameters.

Fig. 6 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. 5. 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. 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.

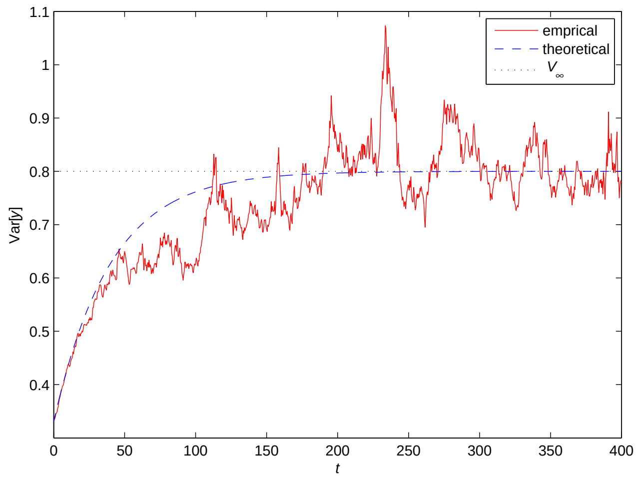

Figure 5: Income tax. The empirical variance of the rescaled wealths of 10^4 realisations of (Eq. 52) with uniformly-distributed initial wealths (red); the theoretical variance for the infinite ensemble, $V(t)$ (blue dashed); and the asymptotic theoretical variance, V_∞ (black dotted). Parameter values are $\mu = 0.05$ and $\sigma^2 = 0.02$ per unit time, and $\tau = 0.45$. {fig:var_income}

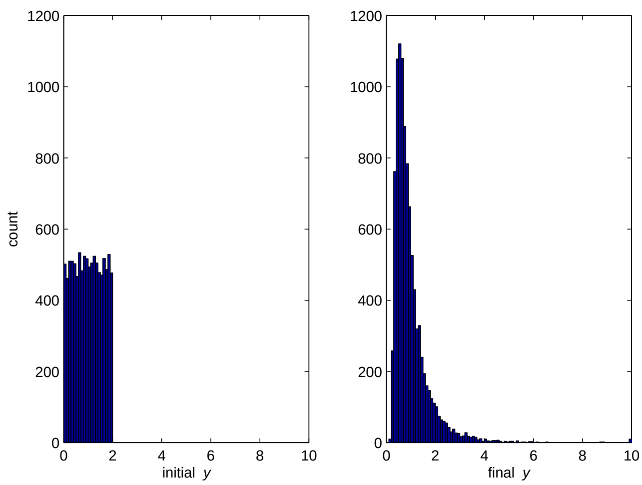

Figure 6: Histograms of the initial (left) and final (right) empirical distributions of the rescaled wealth for the same realisations of (Eq. 53) used in Fig. 5. {fig:hist_income}

References

[1] C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*. Springer, New York, 1978.

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

[3] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag, 1999.

[4] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.*, 46(5):323–352, 2005.

[5] O. Peters and A. Adamou. Stochastic market efficiency. *SFI working paper 13-06-022*, 2013.

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

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