

Mobility, Mixing and Ergodicity: A Physically-Motivated Measure for Economic Mobility

Viktor Stojkoski ^{*} Alexander Adamou[†] Yonatan Berman[‡] Colm Connaughton [§]
Ole Peters[¶] ^{||}

August 2, 2020

Abstract

Keywords: mobility, inequality, ergodicity economics

^{*}Macedonian Academy of Sciences and Arts, vstojkoski@manu.edu.mk

[†]London Mathematical Laboratory, a.adamou@lml.org.uk

[‡]London Mathematical Laboratory, y.berman@lml.org.uk

[§]London Mathematical Laboratory and University of Warwick, c.p.connaughton@warwick.ac.uk

[¶]London Mathematical Laboratory and Santa Fe Institute, o.peters@lml.org.uk

^{||}We thank...

1 Introduction

Economic mobility describes “dynamic aspects of inequality” (?). It quantifies how wealth (or income¹) ranks of individuals evolve over time. Intuitively, when mobility is high, ranks evolve quickly, and the chances of an individual to change her position in the wealth distribution over a given time period are high. When mobility is low, individuals are unlikely to change their rank in the distribution over time, or that it changes slowly.

? described several required properties of statistical measures of mobility and set the standard for such measures. Mobility measures are assumed to be derived from a transition matrix, a bistochastic matrix describing the conditional probabilities of individuals to move between different ranks over some time period. An example for such a measure, used extensively in the mobility literature, is the rank correlation, the correlation between individual wealth ranks in two points in time. Another canonical measure of mobility, used most typically in studies of intergenerational income mobility, is the intergenerational earnings elasticity (IGE), defined as the regression coefficient between log-incomes of parents and children.²

The standard statistical measures of mobility have several limitations. First, they are generally incomparable. For example, a rank correlation of 0.3 would have a different meaning if it corresponds to a period of one year or ten years. Second, for a given time period, it is not possible to tell whether some value of rank correlation is high or low, as this is a dimensionless quantity. Thus, rank correlation over some time period can only be high or low in comparison to other economies over a similar time period, or to other time periods of the same length. In addition, the interpretation of the rank correlation depends on the underlying wealth distribution and its dynamics. The same rank correlation cannot be interpreted similarly when the underlying wealth distribution remains unchanged, and when it becomes more and more unequal. The IGE has similar limitations, but also others. Most notably, the IGE is sensitive by design to the level of inequality and not only to the transition matrix, as discussed in detail in the mobility literature (*e.g.* ?).

This paper introduces mixing time, a property of stochastic processes, as a measure of mobility. When wealth is an ergodic observable (?), and assuming the wealth distribution approaches a steady state, if the wealths of an arbitrary group of individuals is followed over time, the distribution of wealth within this group will gradually become similar to the steady-state wealth distribution. The characteristic time of this process is the mixing time. Put simply, it is the time scale over which individuals mix into the wealth distribution. When mixing is rapid, *i.e.* the mixing time is short relative to the window of observation, we could interpret that as high wealth mobility. Slow mixing is interpreted as low mobility.

We then consider Reallocating Geometric Brownian Motion (RGBM (???)) as a model for wealth dynamics and study mixing in this model. In RGBM, individual wealth undergoes random mul-

¹We focus on wealth in this paper, but it applies also to income

²In fact, the rank correlation and the IGE are both measures of immobility, and to consider them as measures of mobility one has to consider their complement.

multiplicative growth, modeled as Geometric Brownian Motion (GBM), and is reallocated among individuals by a simple pooling and sharing mechanism. RGBM is a null model of an exponentially growing economy with social structure. It has three parameters representing economic growth, random shocks to individual wealth, and economic interaction among agents, quantified by the reallocation rate. This model is known to reproduce several important stylized facts. In particular, when the reallocation rate is positive, the wealth distribution converges to a stationary distribution with a Pareto tail. The model has both ergodic and non-ergodic regimes, characterized by the sign of the reallocation rate parameter (?).

We find that in RGBM the mixing time scales with the inverse of the reallocation rate. As the reallocation rate becomes higher, *i.e.* when a larger share of each individual’s wealth is pooled and then shared per unit time, mixing time becomes shorter proportionally, and mobility increases. As the reallocation rate approaches zero, mixing times get longer, and mobility lower. Since decreasing reallocation rates also lead to increasing inequality, this result is in line with the empirical observation that as inequality increases mobility decreases, and vice versa (?).

YB: Here we need to describe how mixing time and rank correlation or other standard measures are related in RGBM YB

In practice, many economic systems are best modeled as non-ergodic (?). In particular, ? argue that the US economy is best described in RGBM as one in which wealth is systematically reallocated from poorer to richer, *i.e.* the reallocation rate is negative. In such a case there is no mixing, so the mixing time is infinite. Thus, measuring mobility using standard measures under this regime may be misleading. The thorough study of RGBM in this regime is outside of the scope of this paper and left for future work.

The paper is organized as follows. Section 2 discusses the concept of mixing time and how it provides a physically-motivated measure for mobility. Section 4 studies mobility using mixing times in reallocating geometric Brownian motion as a model for wealth. We conclude in Section 6.

2 Standard measures of economic mobility

We begin with an overview of standard measures of economic mobility. These measures describe properties of the bivariate joint wealth distribution in two points in time. To characterize these measures we utilize two widely used measures of economic mobility: Spearman’s rank correlation and the intragenerational earnings elasticity. We also describe the wealth transition matrix, or copula, from which various mobility measures can be derived.

Spearman's rank correlation: Spearman's rank correlation is defined on a joint distribution of wealth at two points in time, t_m and t_n ($t_m < t_n$). It is defined as

$$r_{t_m, t_n} = 1 - \frac{6 \sum_i [rg(x_i(t_m)) - rg(x_i(t_n))]^2}{N(N^2 - 1)}, \quad (2.1)$$

where $rg(x)$ is the rank transformation of x , $x_i(t)$ is the wealth of individual i in period t and N is the population size. This measure is bounded between -1 and 1 . $r_{t_m, t_n} = 1$ suggests perfect immobility, a state in which there is no change in wealth ranks between the two points in time. Lower values suggest greater economic mobility.

Intragenerational earnings elasticity: The intragenerational earnings elasticity is defined as the slope b_{t_m, t_n} of the regression

$$\log(x_i(t_n)) = b_0 + b_{t_m, t_n} \log(x_i(t_m)) + u_i, \quad (2.2)$$

where b_0 is the intercept and u_i is the error term. This is a simple linear regression and therefore,

$$b_{t_m, t_n} = \text{corr}(\log(x(t_n)), \log(x(t_m))) \frac{\text{var}(\log(x(t_n)))}{\text{var}(\log(x(t_m)))}, \quad (2.3)$$

where $\text{corr}(x, y)$ is the correlation between the variables x and y and $\text{var}(x)$ is the variance of x . As with the rank correlation, lower intragenerational earnings elasticity also indicates greater mobility. However, this measure is unbounded and may take on any real values.

Wealth transition matrix: The wealth transition matrix disaggregates wealth rankings and summarizes economic mobility in a transition matrix \mathbf{A} in which the elements A_{kl} quantify the probability that an individual in wealth quantile k in period t_m is found in wealth quantile l in period t_n . In a perfectly mobile economy, the entries of the transition matrix are all equal to each other. This would correspond to 0 rank correlation. In an immobile economy, on the other hand, the largest values are concentrated in the diagonal entries. A perfectly immobile case, of rank correlation 1, would correspond to the identity transition matrix.

2.1 Limitations

To visualize the limitations of the standard measures, in Fig. 1 we construct a simplified example in economy with 9 individuals, during 3 time periods. At the left panel we show the dynamics of the wealth ranks. Besides this, we also highlight the amount of wealth owned by a person with a colored circle, with darker colored and larger circle implying a wealthier person.

In our example only the four richest individuals changed their wealth rank. In particular, in the second period, t_2 , the second richest person becomes the richest, while the richest person in the

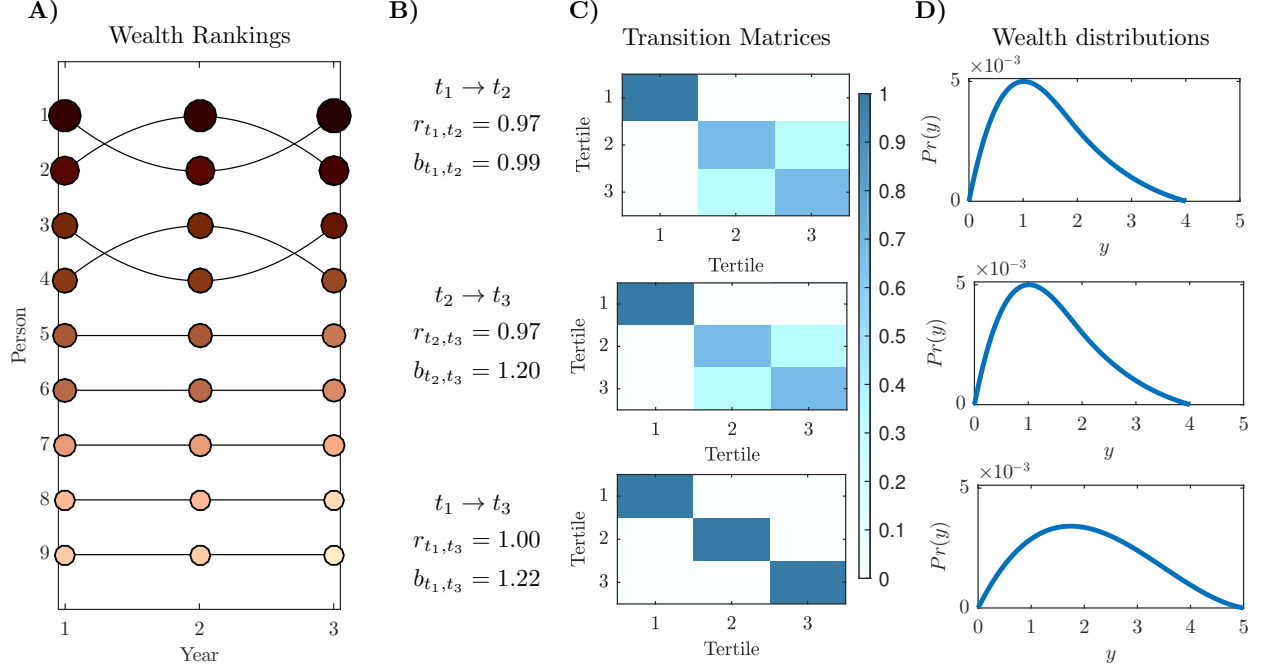


Figure 1: **Mobility measures in a simplified example.** **A)** Dynamics of the wealth rankings in a simplified economy consisting of 9 individuals, during 3 time periods. **B)** Estimated Spearman's rank correlation and intragenerational earnings elasticity during the studied periods. **C)** The corresponding mobility transition matrices. **D)** The distribution of wealth for the 3 time periods.

first period t_1 becomes second richest. In addition, the third and fourth richest individuals also switch their positions. In period t_3 , the dynamics are reversed, thus ending up with the same wealth rankings as in t_1 . The only difference is that in this period, the distribution of wealth has changed, with the richest individual being richer than in the previous periods. If we were to look at the rank correlation, it shows that mobility was the same between t_1 and t_2 , and between t_2 and t_3 . However, overall, between t_1 and t_3 it would show no mobility ($r_{t_1, t_3} = 1$). The intragenerational earnings elasticity gives us similar results. In fact, since inequality increased in the last period, it further suggests that the economy became less mobile.

The same conclusions hold when looking at the transition matrices estimated by dividing the wealth rankings in tertiles. The transition matrices between t_1 and t_2 and between t_2 and t_3 suggest that there is a $p = 1/3$ probability for an individual in the second tertile to climb up to the richest tertile and vice versa. Yet, we structured our economy in a way that allows movement only between the individuals at the edge of the tertile, so the transition matrices fail to adequately represent the movement of the typical person in the tertile.

This simple example shows that standard measures of economic mobility represent aggregate values of the changes in the wealth rankings of the individuals which constitute the population between two time periods. Therefore, they rely on a relevant time period for which the wealth rankings are compared. A rank correlation of 0.3 would have a different meaning if it corresponds to a period

of one year or ten years. Also, for a given time period, it is not possible to tell whether some value of rank correlation is high or low, as this is a dimensionless quantity. Thus, rank correlation over some time period can only be high or low in comparison to other economies over a similar time period, or to other time periods of the same length. The standard measures do not quantify the mobility of the typical representative of the population. In our simplified economy the bottom tertile is perfectly immobile. Regardless of the rank correlation value, it is not representative of the dynamics of wealth of individuals belonging to this tertile. In addition, quantifying mobility using the transition matrix (and deriving various functionals of it as mobility indicators), implicitly assumes the Markov property – that if transition matrix \mathbf{A} represents the period between t_1 and t_2 and \mathbf{B} the period between t_2 and t_3 , then \mathbf{AB} would represent the transition matrix between t_1 and t_3 . This is not inline with empirical findings (??).

3 Mixing time as a measure of mobility

The concept of mixing time comes from probability theory and statistical physics. It allows overcoming the limitations discussed above. A mixing time would provide a characteristic time scale over which individuals mix into the wealth distribution. Figuratively, we can think of the economy as a cup of coffee and of some person’s wealth as milk poured in the coffee. The mixing time quantifies the time required for the milk to blend with the coffee. This enables the measure to be used for appropriate comparison between different time periods and economies. As we will see in the following sections, there is a relationship between the mixing time and standard mobility measures, whenever mixing time is a finite quantity. However, the standard measures may still indicate that there is some level of mobility even when mixing does not occur (*i.e.* the mixing time is infinite). This is because mixing time depends on the existence of mobility between *every* quantile in the wealth distribution. In other words, finite mixing times require wealth ranks to be ergodic, and we will argue that in the absence of ergodicity, mobility cannot be properly defined.³

3.1 Measuring mixing times

In physical terms, mixing describes the property of a dynamical system of being strongly intertwined. That is, for any set of particles moving according to the laws of some dynamical system, regardless of their initial condition, and after a sufficient period of time, the mixing time, the fraction of the particles found within a particular region in the phase space (the space of the variable x characterizing the particles, which can be their wealth in our case), is proportional to the volume of that region in the phase space. Mixing is strongly related to the concept of ergodicity. However, the latter is a broader concept: An observable x is said to be ergodic if its time-average is to its ensemble averages at any given time. Put differently, ergodicity implies that every trajectory spreads

³This idea was already described by ?.

around the phase space according to the steady state distribution of x . Hence, every dynamical system that is mixing is also ergodic, but the opposite is not necessarily true.

In economics terms, mixing implies that there always exists a stationary distribution to which some rescaled transformation of a person's wealth converges. This, subsequently, indicates that mobility between every possible quantile in the population will exist. In this sense, mixing time gives an estimate for the relevant time-scale for this kind of mobility to be feasible. It overcomes the limitations of the standard measures in several ways. First, it is measured in temporal units, thus allowing for deriving conclusions regarding the extent of mobility in an economy without requiring comparisons to other economies and/or to other time periods. Second, mixing time will be a finite quantity only when each individual is able to move across the whole distribution of wealth. Therefore, it does not represent an aggregate measure of the mobility in the economy, and instead evaluates the feasibility of the typical representative of the population to change its wealth status. Lastly, mixing time does not follow the mathematical assumptions of the standard measures of mobility, thus allowing for distinct statistical properties of the stationary wealth distribution over time.

The described properties of a mixing economy can be utilized to develop an estimation procedure for mixing time. In particular, the first step in estimating the mixing time is setting a target distribution to which the rescaled transformation of wealth converges. In reality, it is difficult to recognize the exact form of this distribution as it depends on the wealth dynamics which may be highly random. As a crude estimate, one should use a stationary distribution to which an approximate form of the wealth dynamics converges. Once a stationary distribution is chosen, it is necessary to select a subsample of the population which represent the typical representatives in the population. For example, these can be the persons whose wealth is closest to either the mean/median/mode wealth, depending on the properties of the stationary distribution.

The next step involves tracking the wealth dynamics of each person in the subsample for a sufficient amount of time, and, at certain time points, quantifying the differences between the subsample and the stationary wealth distribution. This can be done by utilizing usual statistical distribution distance measures such as the Kullback-Leibler divergence, the Hellinger distance or the Kolmogorov-Smirnov statistic. In a mixing economy the statistic will exhibit two states. First, there will be a mixing time state during which the two distributions will slowly converge towards each other, and the statistic will slowly decrease. After this state, there will be a stable state with no significant differences between the stationary and subsample wealth distributions.

The last step is to utilize the data only for the mixing time state and estimate the slope of the regression in which the dependent variable is the log of the statistic and the independent is time. Mixing time is the additive inverse of the reciprocal value of the slope ($-1/slope$).

Fig. 2 summarizes the procedure. In it, with the blue line we plot the log of an arbitrary distance statistic as a function of time, whereas the black dashed line gives the regression line for the relationship between the statistic and time during the three different states. The inset plots provide

snapshots for the wealth distribution of the subsample at different time points (red dashed lines). For comparison, the snapshots also include the form of the selected stationary distribution (black line).

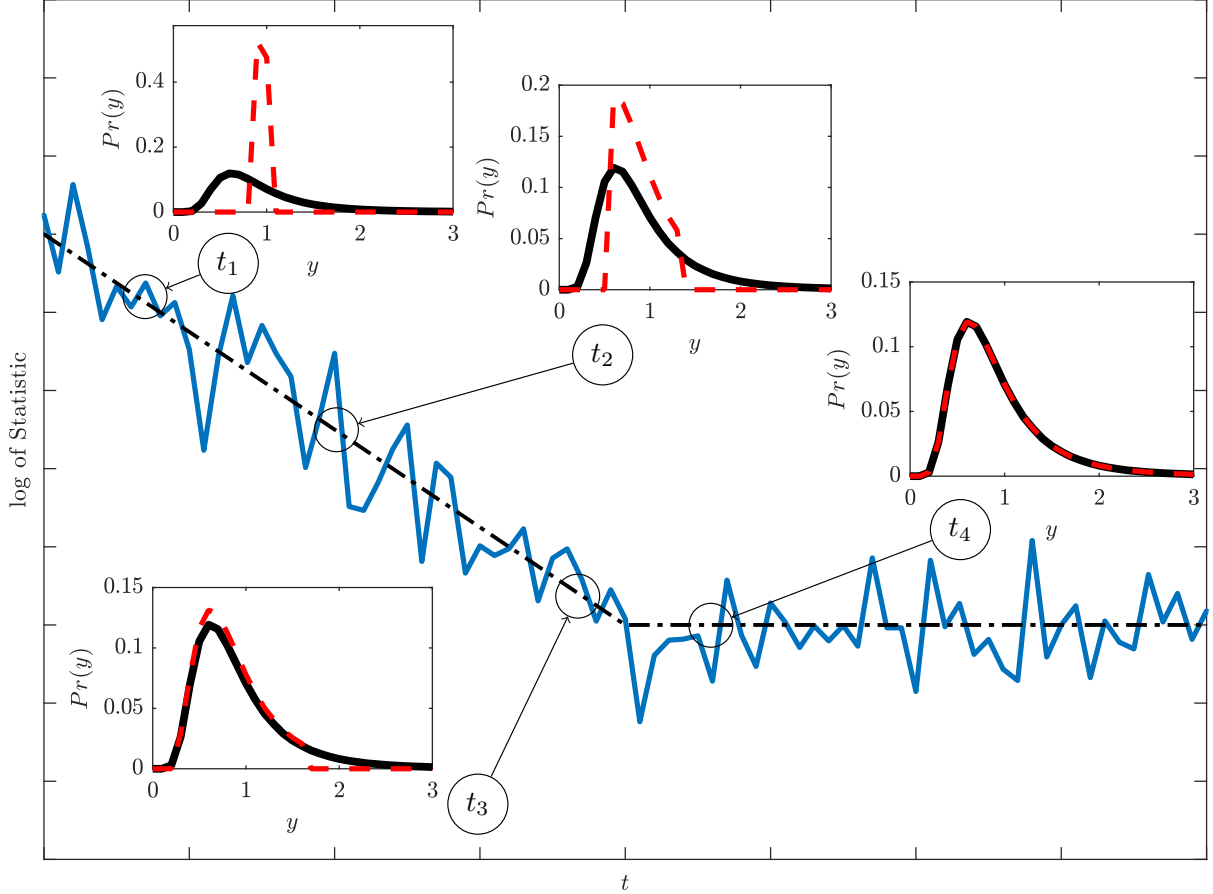


Figure 2: **Estimating mixing time.** Blue line: The log of the statistical distance between the subsample and population wealth distributions. **Black dash-dotted line:** Estimated slope of the regression of the log of the statistic on time. Inset plots: Snapshots of the empirical population wealth distribution (black line) and the subsample wealth distribution (red dashed line) at time periods t_1 , t_2 , t_3 and t_4 .

4 Mixing time in a simple model of an economy

4.1 Reallocating geometric Brownian motion

To illustrate the application of mixing time in economic systems, we use Reallocating geometric Brownian motion (RGBM) as a simple model for wealth dynamics (?). Under RGBM, the dynamics

of the wealth of person i is specified as

$$dx_i = x_i (\mu dt + \sigma dW_i) - \tau (x_i - \langle x \rangle_N) dt, \quad (4.1)$$

with $\mu > 0$ being the drift term, $\sigma > 0$ the noise amplitude, and dW_i is an independent Wiener increment, $W_i(t) = \int_0^t dW_i$. τ is a parameter that quantifies the reallocation of wealth. It implies that every dt , everyone in the economy contributes a fraction τdt of their wealth to a central pool. The pool is then shared equally across the population ($\langle x \rangle_N$ is per-capita wealth). The parameter aggregates a multitude of effects, *e.g.* collective investment in infrastructure, education, social programs, taxation, rents paid, or private profits. The dynamic behavior of RGBM is strictly dependent on the relation between τ and σ , and can be both ergodic and non-ergodic. Since the existence of mixing is predicated on ergodic dynamics, we focus on the ergodic regime. **In RGBM, for large enough population, the transition to this regime approximately occurs when $\tau > \frac{\sigma^2}{2}$.**

Then, the model exhibits mean-reversion as each x_i reverts to the population average $\langle x \rangle_N$. The large population approximation $\langle x(t) \rangle_N = \exp[\mu t]$ is valid, and it can be used to write a mean field equation for the dynamics of the rescaled wealth $y_i = x_i / \langle x \rangle_N$ as

$$dy = y \sigma dW - \tau (y - 1) dt. \quad (4.2)$$

Notice that the mean field representation leads to decoupling of the dynamics of each individual i , and hence we can neglect the subscript i in equation Eq. (4.2). This is a convenient form of RGBM which allows for deriving powerful analytical results regarding the properties of the model. For instance, we can derive the stationary distribution for the rescaled wealth, which reads

$$p(y) = \frac{(\theta - 1)^\theta}{\Gamma(\theta)} \exp\left(-\frac{\theta - 1}{y}\right) y^{-(1+\theta)}, \quad (4.3)$$

where $\theta = 1 + \frac{2\tau}{\sigma^2}$ and $\Gamma(\cdot)$ is the Gamma function. The distribution has a power-law tail as $y \rightarrow \infty$, resembling Pareto's often confirmed observation that the frequency of large wealths tends to decay as a power law. The exponent of the power law, θ , is called the Pareto tail parameter, and can be used as a measure of economic inequality (?). More importantly, the usual stylized facts are recovered: the larger σ (more randomness in the dynamics) and the smaller τ (less reallocation), the smaller the tail index and the fatter the tail of the distribution (higher inequality).

4.2 Mixing time in RGBM

In RGBM, mixing time can be analytically estimated from its correlation function,

$$\text{corr}(y(t), y(t + \delta)) = \exp[-\tau \delta]. \quad (4.4)$$

This function measures the statistical relation between observations of rescaled wealth at two different time points t and $t - \delta$. In a mixing economy the subsample distribution will eventually converge to the same distribution as the one of the whole population. Therefore, we can reduce our analysis to simply studying the convergence time of the subsample distribution to the stationary distribution. Moreover, since in RGBM the dynamics of rescaled wealth of each individual are decoupled, we can further reduce the analysis to examining the convergence time of the process described with Eq. (4.2) to the stationary distribution (Eq. (4.3)). This can be done by defining the time period t^* as the transition point at which the distribution of rescaled wealth converges to the stationary distribution. Then, we can rewrite correlation function terms of this period as,

$$\text{corr}(y(t^*), y(t)) = \exp[-\tau(t^* - t)] , \quad (4.5)$$

for any time period t that belongs to the mixing time state. Obviously, the magnitude of any statistical measure $S(t)$ will be proportional to the value of this correlation, and its log will be of the form

$$\log(S(t)) \propto -\tau(t^* - t) , \quad (4.6)$$

which implies that the mixing time in RGBM is $1/\tau$. The interpretation behind this result is fairly intuitive – in an economy in which reallocation from the rich to the poor is larger, there will be also more mixing. On the other hand, as the reallocation rate approaches zero, mixing times get longer, and mobility lower. Since decreasing reallocation rates also lead to increasing inequality, this result is in line with the empirical observations which suggests that as inequality increases mobility decreases, and vice versa (?). Figure 3 compares our analytical expression with numerical simulations.

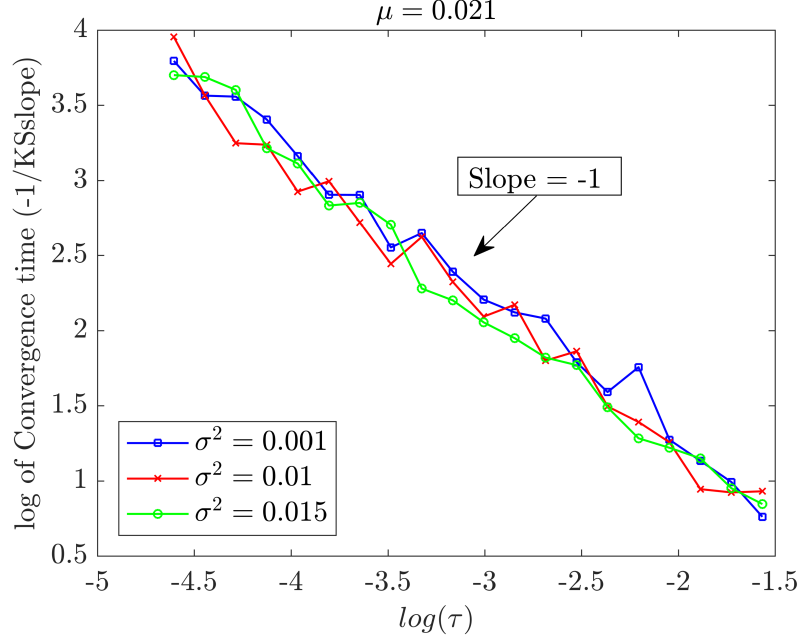


Figure 3: Mixing time in RGBM. **Describe the figure.**

4.3 Mixing time and standard measures of mobility

Unlike the mixing time, the standard measures mobility in RGBM depend on the noise amplitude σ , and not only on τ . This is a consequence of the randomness playing a significant role in the wealth dynamics when we consider time frames that are shorter than the mixing time. In what follows we evaluate the relationship between mixing time and the standard measures of mobility in RGBM.

Spearman's rank correlation: In the ergodic regime of RGBM the Spearman's rank correlation is inversely related to the mixing time. However, the rank correlation is also dependent on the noise amplitude σ and the temporal difference δ between the two periods that are being compared. Larger values for both parameters lead to a greater economic mobility. This can be seen in Fig. 4A where we plot the additive inverse of log of the rank correlation divided by δ as a function of τ for various noise amplitudes.

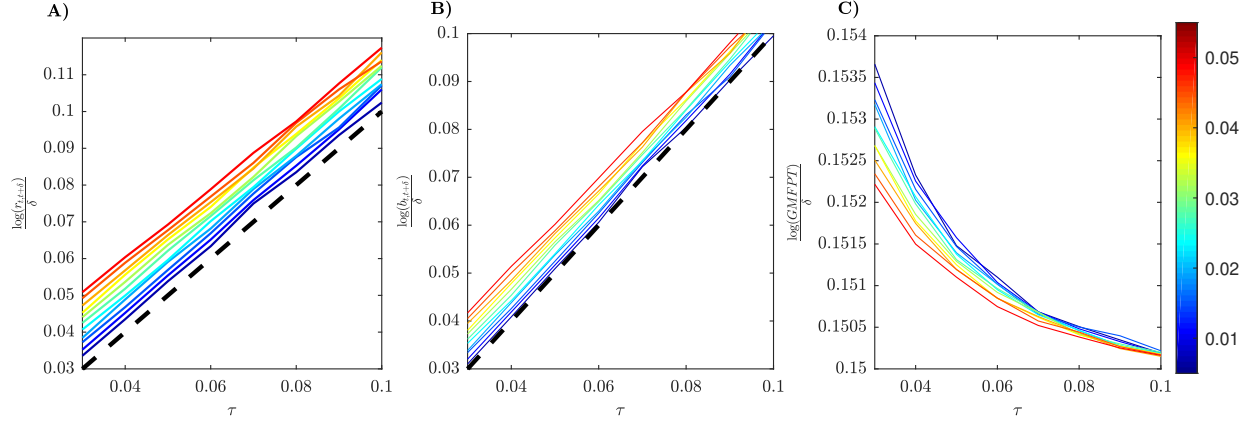


Figure 4: **Mixing time and standard measures of economic mobility.** **A)** Log of the Spearman's rank correlation divided by the temporal difference as a function of τ . **B)** Same as **A)**, only on the y axis is the log of the Intragenerational earnings elasticity. **C)** Same as **A)**, only on the y axis is the log of the GMFPT. **A-C)** The line colors correspond to the magnitude of the noise amplitude, as highlighted by the colormap. In the simulation, $\delta = 20$ and $N = 10^4$.

Intragenerational earnings elasticity: Identically to the properties of the Spearman Rank correlation and as evidenced in Fig. 4B, the Intragenerational earnings elasticity is also dependent on both the noise amplitude and the magnitude of reallocation. **The approximate mathematical relationship between the intragenerational earnings elasticity and mixing time is studied in the Appendix.**

Transition matrices: As the transition matrix divides the wealth distribution into Q quantiles, it gives us a total of Q^2 measures for mobility, each quantifying the relationship between pairs of quantiles. Instead of studying each of those measures, we will focus on one aggregate quantity called the Global Mean First Passage Time (GMFPT). Formally, the mean first passage time (MFPT) between quantiles k and l gives an estimate of the time required for an individual whose wealth belongs to quantile k to move to quantile l , and the GMFPT is the average across all possible MFPT pairs. By construction, this quantity is measured in the same units as mixing time, and as such it can give us a direct relation between the properties of the standard measures and mixing time in RGBM. As inferred from Fig. 4C, the behavior of the transition matrices is similar to the one of the previous two measures. In other words, it is dependent on both the noise amplitude and the temporal differences between the periods for which it is estimated. However, differently from the previous measures it appears that there is no linear transformation for the relationship between GMFPT and the mixing time.

4.4 The Great Gatsby curve in RGBM

We now turn our attention to applying mixing times in studying the relationship between mobility and inequality. A convenient illustration of this relationship is the Great Gatsby curve, which describes the association between inequality and the intergenerational earnings elasticity across countries (??). The curve represents a simple two-dimensional visualization where on the x-axis is given the static inequality and on the y-axis the economic mobility. It was first introduced in a talk by Krueger ?, who showed empirically using country level data that economic *immobility* and static inequality are positively and linearly related. Formally, static inequality is defined as the extent of concentration in the distribution of wealth among the population. Hence, various measures which quantify the dispersion of the stationary distribution Eq. (4.3) can be used to evaluate this concept in RGBM. Here, we use the most widely used measure, the Gini coefficient. Fig. 5 displays mixing time as a function of the Gini index in RGBM. RGBM appropriately reproduces empirical observations by suggesting that inequality and immobility are positively related (?). However, in RGBM the relationship is nonlinear and it looks like an inverted sigmoid curve.

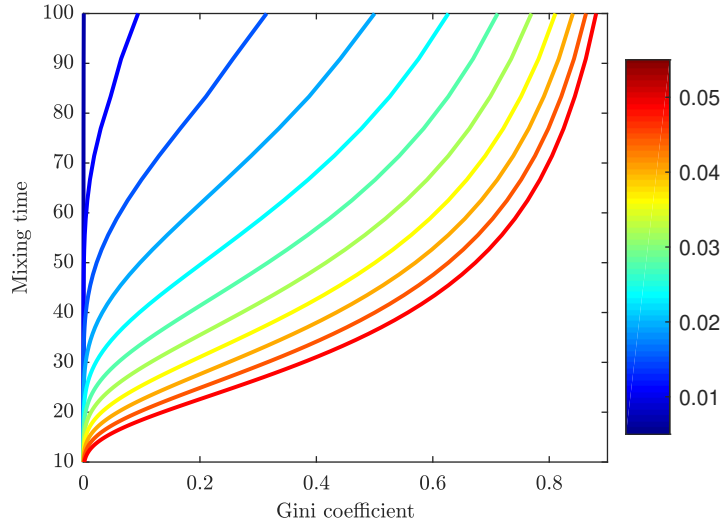


Figure 5: **The Great Gatsby curve in RGBM.** Mixing time as a function of the Gini Coefficient in RGBM. The line colors correspond to the magnitude of the noise amplitude, as highlighted by the colormap.

5 Discussion

- comment on non-ergodicity
- comment on real data?

In a non-ergodic economy mixing time will exist, but under a suitable transformation of the wealth

variable.

6 Conclusion