The Fractal Nature of Inequality in a Fast Growing World*

Guido Cozzi† Fabio Privileggi‡

June 12, 2009

Abstract

In this paper we investigate wealth inequality/polarization properties related to the support of the limit distribution of wealth in innovative economies characterized by uninsurable individual risk. We work out two simple successive generation examples, one with stochastic human capital accumulation and one with R&D, and prove that intense technological progress makes the support of the wealth distribution converge to a fractal Cantor-like set. Such limit distribution implies the disappearance of the middle class, with a "gap" between two wealth clusters that widens as the growth rate becomes higher. Hence, we claim that in a highly meritocratic world in which the payoff of the successful individuals is high enough, and in which social mobility is strong, societies tend to become unequal and polarized. We also show that a redistribution scheme financed by proportional taxation does not help cure society's inequality/polarization – on the contrary, it might increase it – whereas random taxation may well succeed in filling the gap by giving rise to an artificial middle class, but it hardly makes such class sizeable enough. Finally, we investigate how disconnection, a typical feature of Cantor-like sets, is related to inequality in the long run.

JEL Classification Numbers: C61, D31, O33

Keywords: Wealth Inequality, Growth, Fractal, Polarization**,** Pulverization.

*We wish to thank Karl Shell, Neil Wallace, Manuel Santos, Salvador Ortigueira and all the participants of the *Cornell-PSU Macro Workshop* at Pennsylvania State University (March 2002) for insightful and constructive comments.We thank also Philippe Aghion, Jess Benhabib, Michele Boldrin and all the participants of the Summer School in Economic Theory on *Intertemporal Macroeconomics: Growth, Fluctuations, and the Open Economy* held in Venice (Italy, July 2002) for their valuable comments and suggestions. We are also grateful to Victor Rios-Rull, John Knowles, Dirk Krueger and Raouf Boucekkine for fruitful discussions. We are especially indebted to Luigi Montrucchio for inspiration and encouragement. All the typical caveats apply.

†Dept. of Economics, Adam Smith Building, University of Glasgow, G12 8RT, Glasgow (UK); tel. +44(0)141-3306603; e-mail G.Cozzi@lbss.gla.ac.uk.

‡Dept. of Public Policy and Public Choice, University of Eastern Piedmont; e-mail: fabio.privileggi@sp.unipmn.it.

1 Introduction

How do we predict a fast growing and unequal society's wealth distribution to look like? In a global highly competitive and technologically turbulent economy individual success or failure may substantially alter one's position in the social scale. We argue that societies in a twin peak world would tend to look polarized with a complex (fractal) structure.

This is proved by constructing a simple competitive economy with successive generations and uninsurable individual risk to show how easily the support of their limit distribution of individual relative wealth levels can look like a peculiar geometric object called Cantor set, provided that the exogenous growth rate is high enough. A Cantor set is a fractal on the real line, that is, a totally disconnected set with self-similar structure with an evident characteristic: it exhibits a "hole" in the middle. Our definition of (extreme) inequality is based on such hole, which may obviously be interpreted as the lack of a middle class, which, in turn, is often identified with the term 'polarization' by the mainstream literature on inequality.

Emerging phenomena of income or wealth inequality and polarization has been lately observed in many economies. D'Ambrosio and Wolff (2008) document an overall increase in US wealth polarization in the 1983-2004; Wolff (2007) analyzes the dramatic debt-related squeeze in the middle-class share of total wealth during the early 2000s; Drew-Becker and Gordon (2007) show convincing evidence that 80-90% of the wage distribution fail to grow at the productivity growth rate, whereas only the top quintile captures the increase in the productivity growth. Therefore the gap between top incomes and lower incomes widens over time. They have shown that the 50-80% quantile of the income distribution in the US have been steadily declining from 1966 to 2001.

Our main prediction is a positive relationship between polarization and growth, which has proved empirically significant and robust in the multi-country regressions by Roine, Vlachos and Waldenström (2007). In their panel data analysis they showed that throughout the 20^{th} century growth in the developed world has been "pro-rich". According to their findings, "high income groups in society have a larger share of their income tied to the actual development of the economy", while those who fail to tie their income to the growth of productivity fail to the lower tail of the distribution. US evidence suggests that computerization leads to displace 'middle skilled' workers, and "[d]isplacing this 'middle' generates polarization" (Autor, Katz and Kearney, 2006). In a large panel of countries Lundberg and Squire (2003) find that growth has a positive effect on income inequality. Perloff and Wu (2005) show that during the fast growing period 1985-2001 in China income inequality increased dramatically both nation-wide and in urban areas. From the theoretical point of view, the literature on income inequality/polarization appears to be already rich enough, both from the perspective of the possible consequences that inequality/polarization may have on growth rates[1] and from the per-

[1] See, *e.g.*, Loury (1981), Banerjee and Newmann (1993), Galor and Zeira (1993), Alesina and Rodrik (1994), Persson and Tabellini (1994), Benhabib and Rustichini (1996), Benabou (1996a), Aghion and Bolton (1997), Piketty (1997), Aghion, Caroli and Garcia-Penalosa (1999), Benabou (2000), Alesina and Angeletos (2005a, 2005b). For related empirical contributions see, *e.g.*, Alesina and Rodrik (1994), Perotti (1996), Benabou (1996b), Benhabib and Spiegel (1997), Barro (1999, 2000), Forbes (2000), Alesina and La Ferrara (2005).

spective of analyzing what aspects of innovation and growth may generate inequality/polarization.[2] However, this is the first paper that enquires on the fractal properties of the support of the equilibrium wealth distribution generated by the individual adoption of fast evolving technological change.

In this paper, economies with (possibly) polarized wealth distribution in the long run are analyzed by means of Iterated Function Systems (IFS) to describe their dynamics and their limit distribution. Even if the IFS approach seems to be capable of unveiling new aspects of economic dynamics, the application of such methodology to economic models seems to be at an early stage: up to our knowledge, very few works appeared along this line, and none of them with the aim to explain wealth or income inequality. Some examples are Bhattacharya and Majumdar (1999a, 1999b, 2001 and, for an excellent survey, 2007), who dealt mainly with IFS with random monotone maps, Montrucchio and Privileggi (1999), Mitra, Montrucchio and Privileggi (2004), and Mitra and Privileggi (2004, 2006 and 2009), who studied stochastic optimal growth models converging to invariant probabilities supported on Cantor sets. From a different perspective, Bucci, Kunze and La Torre (2008) applied the Collage Theorem to the inverse problem of finding a suitable IFS in order to estimate some key-parameters of a Lucas-Uzawa type model (see also Kunze, La Torre and Vrscay, 2007a, 2007b, 2009, and La Torre and Mendivil, 2008).

Our analysis is characterized by markets with equal opportunities for all individuals; such equal opportunities fuel a strong mobility engine that, if associated to high growth rates, may generate inequality. Mobility is introduced through stochastic labor income heterogeneity, which represents the ability of the individuals to adopt better and better technologies. If better technologies entail some adoption uncertainty at the individual level and if such risk is uninsurable, due to the unobservable or unverifiable individual commitment into a learning effort, income heterogeneity becomes a natural consequence of aggregate growth, and the faster the aggregate growth the relatively larger the magnitude of the uncertain part of the individual resources.

A faster growing environment implies stronger family mobility prospects, because a successful individual from a poor family can more easily overtake the unsuccessful individuals of a richer family, but it means a tendency for the middle class to disappear as well.[3] Hence a "hole" in the middle of the support of the wealth distribution is more likely to appear the faster the pace of technological growth: the wealth distribution becomes polarized into a high and a low wealth classes. However, the random dynamical system that governs the individual assignment across the ever expanding social wealth distribution is not only polarizing the wealth distribution, but will mirror the central hole everywhere through the wealth distribution itself: the absence of a middle class at the social level implies the absence of "middle subclasses" at all levels, due to the diversity of the destinies of the different individuals who travel stochastically through the society's wealth distribution. It follows

[2]It is worth mentioning the literature on skill-biased technological change as well as directed technological change, such as Acemoglu (1999, 2002 and 2007), Card and Di Nardo (2002), and Autor, Katz and Kearney (2006). For an excellent survey, see Acemoglu (2008).

[3]As Mookherjee and Ray (2005, p. 13) notice: "if growth (from neutral technical progress) causes wages to grow at a uniform rate, then fast growing countries are more likely to display wide spans, since higher growth in wages across generations will dull the level of desired bequests". Similarly in Mookherjee and Ray (2002). Unlike these models and that in Mookherjee and Ray (2003), in our paper uncertainty plays a major role.

that the very same process that generates a wealth distribution which is disconnected in the middle multiplies such disconnection at infinity (in all its subintervals), thus generating a totally disconnected support of the wealth distribution. Therefore we reach what we can call a "pulverized" society. Such a "fractal society" is an intriguing mix of polarization and pulverization.

This kind of "polarization/pulverization" of the aggregate wealth distribution differs from the traditional idea of "polarization": though if we photograph the wealth distribution at each point in time we get a highly "polarized" picture, when we track the processes for the successive wealth levels of any single individual we observe a strong mobility. Dynamically, such societies are not polarized in "durable classes", but they show a tremendous impact of mobility. Indeed, it is the amplitude of such mobility that generates polarization: the very fact that the gains of a lucky poor can make her richer than an unlucky rich is at the same time an important mobility aspect and the cause of polarization.

We will obtain "fractalized" wealth distributions from two versions of a simple macroeconomic model with no aggregate uncertainty and individual idiosyncratic income risk. Our specifications generate enough linearity in the random dynamical system which immediately translates into well known properties of the Barnsley IFS used to generate the Cantor set. The choice of such a simple (textbook-like) model allows us to examine in depth the most direct relationship between growth rate and wealth inequality in a dynamic framework.

An important consequence of our main result regards the effect of a fiscal policy aimed at eliminating polarization/pulverization through income taxation of those who are successful and redistribution to the unlucky individuals. Intuitively, since such policy directly attacks the mechanism responsible for the "fractalization" of society, one would expect that this would easily reach its target. We show that this is not the case. In fact, simple redistribution schemes can never eliminate polarization/pulverization of society. What's more, even if the free workings of the private economy itself did not imply socioeconomic disconnection, a direct taxation of wealth of all individuals may be able to induce polarization/pulverization of society. Also the adoption of a random taxation scheme, which has in principle the potential of creating an artificial middle class in a polarized economy, proves essentially ineffective whenever the incentive compatibility constraint is sufficiently tight.

A closer look at how inequality is being affected by the interplay between pulverization and polarization – two apparently contradictory aspects related to the same phenomenon that generates a Cantor support for the limit distribution – in the long run is given by calculating the limit of the Gini coefficient of the marginal distributions as time tends to infinity: we find that inequality remains positive for the invariant wealth distribution.

The main assumption underlying the (stochastic) dynamics in both models under study is that there are only two states of nature: 'failure' or 'success'. Such framework allows the best outcome under the low realization to be worse than the worst outcome under the high realization whenever the growth rate is large enough, as we shall prove in our main result. The choice of such an assumption, if on one hand plays a key role in establishing a direct relationship between growth and wealth polarization, on the other hand may appear extreme and unrealistic. At the end of the paper we shall show, by means of a heuristic but robust argument, that the main idea developed in the 'two shocks setting' actually generalizes to i.i.d. stochastic processes defined by a density – *i.e.*, quite the opposite scenario of having the "highly discrete" process of only two states – provided that such density is bimodal, in the sense that it concentrates most of the weight on the boundaries of its state space.

The paper is organized as follows. In Section 2 the two macroeconomic models of technological change are introduced. Section 3 is devoted to a self-contained review of the basic mathematical methods we use to analyze the possibly fractal support of the limit distribution for a random dynamical system. In Section 4 we provide sufficient conditions for the limit wealth distribution to have a Cantor support, which we interpret as a polarized/pulverized distribution; such conditions are expressed in terms of (exogenous) growth rate and degree of intergenerational altruism of the population. In Section 5 the main implications of the analysis of Section 3 on the inefficacy of inequality-eliminating policies are reported in detail. In Section 6 we focus on a closer examination of the interplay between inequality and what we have somewhat tentatively called "pulverization". Finally, Section 7 shows the robustness of our approach by proving that smooth perturbations of our discrete stochastic process do not affect the main result. Section 8 concludes with some comments, while the Appendix A contains the proof of the main result of Section 6 and Appendix B explains the formula for the approximation in Section 7.

2 Technology and Growth

In this section we introduce two simple macroeconomic models with exogenously evolving technology. In the first one, we assume a sequence of successive generations of altruistic individuals who take consumption and bequest decisions on their wealth accumulated out of a stochastic income acquired at the utility cost of learning a technology that is new at every generation. The second model hinges on the same framework of the first one, but allows for exploitation of new discoveries by means of patents which expire after one generation. Both models are characterized by a strong mobility engine (equal opportunities for all individuals) and uninsurable individual risk. Unlike the mainstream literature, no imperfections on credit markets or barriers to access education are assumed. On the other hand, uncertainty is modeled in a standard fashion, similar to that adopted in Aghion and Bolton (1997): there are only two states of nature describing achievements of economic agents, either 'success', with probability $0 < p < 1$, or 'failure', with probability $1 - p$.

2.1 Adoption of New Technologies

Consider an infinite horizon discrete time economy with a continuum of infinitely lived families that will be indexed by i. With no loss of generality we shall normalize population over the unit interval, *i.e.*, $i \in [0, 1]$. Each family is formed by a one-period lived altruistic individuals whose preferences are represented by the following "warm glow" (see Andreoni, 1989) utility function

$$u(c, b, e) = c^{1-\beta} b^{\beta} - e$$

where $c > 0$ denotes end-of-life consumption, $b > 0$ the bequest left to the unique heir, $e \geq 0$ a learning effort,[4] and $0 < \beta < 1$ the degree of intergenerational altruism. As, for example, in Banerjee and Newman (1993), Galor and Zeira (1993), or Piketty (1997), such Cobb-Douglas altruistic preferences imply that a fraction β of each individual's end of life wealth will be passed over to her child. Hence, the indirect utility of end-of-life wealth W is linear (risk neutral preferences) and equal to

$$U(W) = (1-\beta)^{1-\beta} \beta^{\beta} W - e.$$

The end-of-life wealth W of each family is uncertain at the beginning of each generation: it depends on the wealth level inherited from the past, that is on the bequest left by the ancestor, and on individual success in learning the technology that becomes available during her lifetime.

Individuals of generation t are endowed with one unit of labor time which they will inelastically use to produce a perishable consumption good at the common productivity level $A_t > 0$. At the beginning of period t, a new General Purpose Technology (see Helpman, 1998) appears exogenously and every individual has to learn it in order to successfully enter production. Learning technology A_t requires an effort that entails a certain utility cost $e_t > 0$. Whether an individual exerts the required effort for learning such technology is something that cannot be observed by anybody but the individual.[5] Moreover "success" in the adoption of the technology is not sure, but it occurs to each individual with probability $0 < p < 1$ constant through time, independently of all other individuals. In other words, all individuals of the same generation face the same opportunity of success. Since the (exertion of) learning effort is unobservable, borrower-creditor interaction lasts one period only and individual's offspring cannot be sanctioned; accordingly, no idiosyncratic risk can be insured.

Technology is assumed to evolve exogenously: $A_t = \gamma A_{t-1}$, where $\gamma > 1$. Consistently, we will assume that $e_t = \gamma e_{t-1}$, that is, learning a more advanced technology requires more effort.

Provided that individual $i \in [0,1]$ alive in period t undertakes the learning effort e_t at the beginning of her life, her end-of-period *income* Y_t will be:

$$Y_t^i = \begin{cases} 0 & \text{with probability } 1-p \\ A_t & \text{with probability } p \end{cases} \tag{1}$$

Notice that in this model income derives from the "ability" in the use of current technologies and entails no utility loss. Failure to gain an effective education might be the outcome of cognitive and non-cognitive skills, responsible of school drop-outs.[6] Hence the economy of this section is characterized by skill-biased technological change (see Acemoglu, 2008).

[4] As will become clear later, each agent chooses to exert effort e between two values: zero and a strictly positive fixed amount which depends on time.

[5] Specifically, it is not the amount of learning effort which is not observable, but whether an individual undertakes such effort at all.

[6] Early interventions in favor of disadvantaged-children – such as Perry Preschool and other programs discussed by Cuhna and Heckman (2006, 2007, and 2009) – may modify the probabilities of success at different wealth levels. This is very important, but it will not alter the main results of this paper, as we focus our attention to the support of the wealth distribution.

The evolution of technology yields $A_t = \gamma^t A_0$ and that of effort $e_t = \gamma^t e_0$, with both A_0 and e_0 strictly positive. Individual i wealth at the beginning of her life in period t is given by the bequest inherited from period $t-1$:

$$b_t^i = \beta W_{t-1}^i,$$

where W_{t-1}^i represents the wealth accumulated by her ancestor at the end of time $t-1$. Provided that individual i will perform effort e_t in order to learn technology A_t, her expected indirect utility conditional to the past wealth and the performed effort is given by

$$\begin{aligned}\mathbb{E}\left[U\left(W_t^i\right) \mid \left(W_{t-1}^i, e_t\right)\right] &= (1-\beta)^{1-\beta}\beta^\beta \mathbb{E}\left[W_t^i | W_{t-1}^i\right] - e_t \\ &= (1-\beta)^{1-\beta}\beta^\beta \left[p\left(\beta W_{t-1}^i + A_t\right) + (1-p)\beta W_{t-1}^i\right] - e_t \\ &= (1-\beta)^{1-\beta}\beta^\beta \left(\beta W_{t-1}^i + pA_t\right) - e_t\end{aligned} \tag{2}$$

where the probability of success p in adopting technology A_t does not depend on time.

We shall assume the following.

Assumption 1

$$0 < e_0 < (1-\beta)^{1-\beta}\beta^\beta p A_0.$$

Assumption 1 implies that the expected indirect utility obtained by exerting effort e_t is greater than the certain effort for all $t \geq 0$, thus rational individuals will always put the required effort into learning the new technology. It follows that the intergenerational motion of the wealth of family $i \in [0,1]$ is described by

$$W_t^i = \begin{cases} \beta W_{t-1}^i & \text{with probability } 1-p \\ \beta W_{t-1}^i + A_t & \text{with probability } p. \end{cases} \tag{3}$$

Let $b_0^i \geq 0$ denote the "original" bequest available at the beginning of period $t=0$ to family i, then

$$W_0^i = \begin{cases} b_0^i & \text{with probability } 1-p \\ b_0^i + A_0 & \text{with probability } p \end{cases}$$

Since A_t grows exogenously through time, the random dynamical system (3) described by the two maps $f_1(W,t) = \beta W$ and $f_2(W,t) = \beta W + A_t$ evolves along increasing sets of possible wealths. In particular, at the end of period t generation i will be endowed with some wealth W_t^i in the interval

$$\left[\beta b_0^i, \beta b_0^i + \left(\frac{1-(\beta/\gamma)^{t+1}}{\gamma-\beta}\right)\gamma^{t+1}A_0\right] \tag{4}$$

which, since $\gamma > 1$, $0 < \beta < 1$ and $A_0 > 0$, diverges to $[0, +\infty)$ as $t \to +\infty$.

However, notice that, since $0 < \beta < 1$, both f_1 and f_2 in (3) are contractions in the variable W, that is, wealth grows only thanks to technological parameter A_t as time elapses. Hence, a better highlighting of the features of this dynamics can be obtained by transforming system (3) into an equivalent law of motion adjusted by the productivity level A_t, which turns out to be a contractive process eventually remaining bounded inside a compact set, which we shall call *trapping region*.

Dividing (3) by A_t we get the equivalent system in terms of $w_t^i = W_t^i / A_t$:

$$w_t^i = \begin{cases} (\beta/\gamma)\, w_{t-1}^i & \text{with probability } 1-p \\ (\beta/\gamma)\, w_{t-1}^i + 1 & \text{with probability } p \end{cases} \tag{5}$$

whose trapping region, as can be easily shown, is the interval $\left[0, (1-(\beta/\gamma))^{-1}\right]$. Let

$$\alpha = \frac{\beta}{\gamma}, \tag{6}$$

which implies $0 < \alpha < 1$, and consider the linear transformation $y_t^i = (1-\alpha)\, w_t^i$ of (5). With this change of variable we obtain the following productivity-adjusted dynamic:

$$y_t^i = \begin{cases} \alpha y_{t-1}^i & \text{with probability } 1-p \\ \alpha y_{t-1}^i + (1-\alpha) & \text{with probability } p, \end{cases} \tag{7}$$

which, as we shall see more in detail in Section 3.3, has the unit interval $[0,1]$ as trapping region.

The stochastic dynamic (7) defines two possible levels of (productivity adjusted) wealth at time t of individual i, y_t^i, provided that her wealth at time $t-1$ is y_{t-1}^i. The lower level is reached with probability $1-p$ while the upper level is reached with probability of success p.

System (7) belongs to an important family of random dynamical systems known in the literature as *(Hyperbolic) Iterated Function Systems* (IFS). Before studying thoroughly IFS (7), which is the topic of Section 3, we turn our attention to a second, slightly more sophisticated, model, mainly to show that dynamics of the form expressed in (7) can be easily replicated.

2.2 Schumpeterian Growth with Patents

While keeping the same framework of Section 2.1, let us now assume that every individual of generation t at the beginning of her economic life has the same probability $0 < p < 1$ of discovering a better production method that allows the productivity of a number $\theta \geq 1$ of individuals to jump to the new technological frontier $A_t = \gamma A_{t-1}$, provided she undertook an indivisible innovation effort $e_t = \gamma e_{t-1}$.

To render growth endogenous we will assume that productivity growth rate γ is an increasing and bounded[7] function of the aggregate innovative effort $\int_0^1 e_0^i di$, where 1 is the constant (normalized) population size.[8] Inventions are immediately patented and the patents expire after one generation.

[7]With this simple assumption – that may be motivated by some kinds of congestion effects – we eliminate Jones (1995) scale effects.

[8]It would not be difficult to allow for population growth. Interestingly, as will become clearer throughout the paper, offspring's division of bequest would reinforce inequality in this model and/or even generate it.

We will assume that each individual can run only one research project during her life. Hence we are building a simple Schumpeterian model in which the entrepreneurs are new people (Schumpeter, 1934, 1939) who try to adapt the ever-evolving society knowledge frontier to their sphere of production, as in Aghion and Howitt (1998) and Howitt (1999). The parallel with Aghion and Howitt (1998, Chapter 3) and Howitt (1999) cross-sector spillover is in our assumption that A_t evolves as an increasing function of social R&D adoption effort. This adds a zero growth equilibrium due to R&D coordination failure: if each individual expects nobody to exert effort she will be better off not exerting it. In the rest of the analysis we will concentrate only on the positive growth equilibrium.

Unlike usual Schumpeterian models we are here assuming a limited productive capacity per firms and/or a limited number of patent licensees. In fact we will assume that in order to implement each successful innovation the cooperation of θ workers (including the innovator) is necessary. Hence, by the law of large numbers, in the steady state there will be a fraction p of innovators, and a fraction $p\theta$ of individuals employed in all innovative productive processes. Since we keep the whole population normalized to 1, in order to let all innovators carry on their activity, the fraction $p\theta$ of employed individuals cannot exceed 1, that is, the number of workers for each activity must be bounded by

$$1 \leq \theta \leq \frac{1}{p}. \tag{8}$$

If the RHS of (8) holds with equality, the society is perfectly divided in a fraction p of entrepreneurs/ innovators and a fraction $1-p$ of workers. If the RHS of (8) holds with strict inequality, then there will be a fraction of people who will be treated as self-employed in production processes that use the technology A_{t-1} available from the last period. Since patents expire after one period, the technology A_{t-1}, available only for the innovators at time $t-1$, becomes of public domain at time t. Alternatively, if $\theta = 1/p$, equilibrium unemployment would result in this simple economy.

Therefore we shall assume that, at each period t, both employed workers in the innovative sectors and self-employed workers in the old sectors perceive salaries equal to their productivity under the old technology A_{t-1}. In this last scenario there will be a fraction $0 < p\theta < 1$ of individuals employed in the A_t technology sector and a fraction $1-p\theta$ of individuals employed in the A_{t-1} technology sector. Of these families, only a fraction p is able to reap the benefits of the innovative technology A_t (each by employing $\theta-1$ workers) by means of patents, while the other fraction $1-p$, being they employed in the innovative sector or self-employed in the old sector, is remunerated by the productivity of the A_{t-1} technology.[9]

The innovations of this model can alternatively be interpreted as the discovery of an "entrepreneurial talent" that allows the innovator to found a firm that permits a more efficient use of θ workers by making them use the best productive practices available in her firm. As in Cuhna and Heckman (2006, 2007, and 2009), and Heckman (2008), non-cognitive abilities matter as well as cognitive abilities. In this sense, the model of this section can be viewed as an education model of the firm:

[9] If $\theta > 1/p$, the innovators would not be able to implement their discoveries, and in a competitive equilibrium all profits would be zero, leading to a society with a unique wealth group without inequality.

in the particular case $\theta = 1$ the individual is only able to privately accumulate the "state of the art" human capital. Unlike the previous example, the technology learned by generation t will be observed by everybody when it is operated, and, afterwards, every family will become able to use it at no additional educational cost. With $\theta = 1$ this model depicts an economy similar to that of the previous example, except for a perfect educational spillover which allows the wealth of the children of the unlucky generation to instantaneously reach the level of the lucky members of the previous cohort.

Let us turn our attention to the evolution of wealth through time in this model. In every period, p "innovators" will appear and $p\theta \leq 1$ skilled workers will be producing with the cutting-edge technology, paying their extra productivity to each successful innovator. The innovator – as a patent holder or as an entrepreneur – is able to extract the complete productivity increment for one period, thereby rendering the appropriable technology of every non-innovator equal to the same value A_{t-1}. In other words, besides directly benefitting from the new technology A_t, each single innovator in period t can appropriate the productivity gains of the non-innovators workers employed in her firm. Her end-of-period income is thus equal to

$$A_t + (A_t - A_{t-1})(\theta - 1) = [1 + \theta(\gamma - 1)] A_{t-1}. \tag{9}$$

Hence, the wealth of individual i at the end of period t, provided she undertook the indivisible innovation effort e_t at the beginning of the period, will be

$$W_t^i = \begin{cases} \beta W_{t-1}^i + A_{t-1} & \text{with probability } 1-p \\ \beta W_{t-1}^i + [1 + \theta(\gamma - 1)] A_{t-1} & \text{with probability } p. \end{cases} \tag{10}$$

The unlucky will get only the one-period lagged productivity A_{t-1} wealth, being her self-employed or employed by some patent holder firm; in the latter case she must pay the full monopolistic rent to the successful patent holder who employs her, though she can choose between different patent holders.

Once again, we need to make sure that all families find it convenient to undertake the indivisible innovation effort e_t at the beginning of each period t. The individual i expected utility gain conditional on effort e_t is given by

$$\begin{aligned} \mathbb{E}\left[U\left(Y_t^i\right)|e_t\right] &= p\rho[1 + \theta(\gamma - 1)] A_{t-1} + (1-p)\rho A_{t-1} - e_t \\ &= \rho[1 + p\theta(\gamma - 1)] A_{t-1} - e_t, \end{aligned}$$

where $\rho = (1-\beta)^{1-\beta}\beta^{\beta}$, while the individual i certain utility gain obtained by exerting zero effort is given by

$$U(A_{t-1}) = \rho A_{t-1}.$$

To achieve our goal,

$$\mathbb{E}\left[U\left(Y_t^i\right)|e_t\right] > U(A_{t-1})$$

must hold, which easily translates into the next assumption.

Assumption 2

$$0 < e_0 < \rho\theta\left(1 - 1/\gamma\right) pA_0.$$

where $\rho = \left(1-\beta\right)^{1-\beta}\beta^{\beta}$.

Notice that in this case nobody ends up with a zero wealth, but instead even the "poorest" segment of the population improves its standards of living at the same steady rate $\gamma - 1$ as the richest. In particular, at the end of period t each individual i will have some wealth W_t^i laying in the interval

$$\left[\beta b_0^i + \left(\frac{1-\left(\beta/\gamma\right)^{t+1}}{\gamma-\beta}\right)\gamma^t A_0, \beta b_0^i + \left(\frac{1-\left(\beta/\gamma\right)^{t+1}}{\gamma-\beta}\right)\gamma^t\left[1+\theta\left(\gamma-1\right)\right]A_0\right].$$

This is a consequence of the temporary nature of patents that allows the inventors to "exploit" the unlucky only for a limited lapse of time and, upon expiry, makes that innovation available for everybody to be freely used.

Following the same technique as in Section 2.1, divide both equations in (10) by A_t to get the productivity-adjusted dynamic

$$w_t^i = \begin{cases} \alpha w_{t-1}^i + 1/\gamma & \text{with probability } 1-p \\ \alpha w_{t-1}^i + \left[1+\theta\left(\gamma-1\right)\right]/\gamma & \text{with probability } p, \end{cases} \tag{11}$$

where $\alpha = \beta/\gamma$. Through the affine transformation $y_t^i = \left[\theta\left(\gamma-1\right)\right]^{-1}\left[\gamma\left(1-\alpha\right)w_t^i - 1\right]$ of (11), it is immediately seen that we obtain the same IFS as in (7), taking the relevant values on the interval $[0,1]$.

3 Iterated Function Systems and their Attractor

In this section we provide a self-contained description of the mathematical toolkit necessary to handle IFS of the kind defined in (7). We shall confine our attention to IFS constituted by maps which are contractions, since we heavily rely on a basic result on convergence of IFS requiring this property. Then, we shall generalize the idea of normalizing linear dynamics over a compact interval (specifically, $[0,1]$) already used in the previous sections, and we shall carefully study the geometric properties of the the fixed point – the attractor – of such normalized IFS. On these geometric properties is based the definition of wealth polarization/pulverization that will be used in subsequent sections.

3.1 A Well Known Result on IFS

There is a huge literature available on IFS, which has grown very fast since, a few decades ago, it proved useful in techniques for generating approximated images of fractals on computer screens. Exhaustive treatment can be found, among others, in Hutchinson (1981), Barnsley and Demko (1985), Edgar (1990), Vrscay (1991), Stark and Bressloff (1993), Lasota and Mackey (1994), and Falconer

(1997, 2003). For a simplified exposition, focused on discussing an optimal growth model exhibiting the same dynamics as in (7), see also Mitra, Montrucchio and Privileggi (2004).

Let X be some compact subset of $\mathbb{R}$ and consider a pair of maps, $f_1 : X \to X$, $f_2 : X \to X$, such that $f_1 < f_2$ with some constant $0 < \alpha_j < 1$ such that $|f_j(x) - f_j(y)| \leq \alpha_j |x - y|$ for all $x, y \in X$ and $j = 1, 2$. Given a fixed probability $0 < p < 1$, the triple $\{f_1, f_2, p\}$ defines the (contractive) IFS

$$x_t = \begin{cases} f_1(x_{t-1}) & \text{with probability } 1 - p \\ f_2(x_{t-1}) & \text{with probability } p. \end{cases} \tag{12}$$

on the compact set X. System (12) induces an operator T on $\mathbb{R}$, called *Barnsley operator*, defined by

$$T(B) = f_1(B) \cup f_2(B), \qquad B \subseteq X, \tag{13}$$

where $f_j(B)$ denotes the image of B through f_j, $j = 1, 2$. Successive iterations of T transform B into a sequence of sets $B_t = T[T^{t-1}(B)]$ through time. We are interested in properties of the limiting set, if it exists, to which the sequence B_t might eventually converge. A set $A \subseteq X$ is called an *invariant set* or *attractor* for (12) if it is compact and satisfies $T(A) = A$. It is a set such that, once entered by the IFS, successive iterations of T keep the system inside it.

Since (12) describes a stochastic dynamical system, another important aspect of the IFS is the evolution through time of marginal probability distributions. Given any initial distribution ν_0 over X, it is interesting to study how this probability evolves according to (12). Let $\mathcal{B}$ be the σ-algebra of Borel measurable subsets of X and $\mathcal{P}$ the space of probability measures on $(X, \mathcal{B})$. Define the Markov operator $M : \mathcal{P} \to \mathcal{P}$ as

$$M\nu(B) = (1 - p)\,\nu\left[f_1^{-1}(B)\right] + p\nu\left[f_2^{-1}(B)\right], \qquad \text{for all } B \in \mathcal{B} \tag{14}$$

where $\nu \in \mathcal{P}$ and $f_j^{-1}(B)$ denotes the preimage set $\{x \in X : f_j(x) \in B\}$, $j = 1, 2$. Operator M is often called *Foias operator*. As we did for operator T, we want to study successive iterations of M starting from some initial probability ν_0, $\nu_t(B) = M[M^{t-1}\nu_0(B)]$, which yields the evolution of marginal probabilities of the system as time elapses. A probability distribution $\nu^* \in \mathcal{P}$ is said to be invariant with respect to M if

$$\nu^* = M\nu^*. \tag{15}$$

An invariant probability distribution is usually interpreted in economics as the stochastic steady state to which the economy might eventually converge starting from some initial distribution ν_0 (see for example Stokey and Lucas, 1989, and Montrucchio and Privileggi, 1999).

Below we recall an important result available for the fixed point of our IFS. Recall that the *support* of a probability distribution ν is the smallest closed set $S \subseteq X$ such that $\nu(S) = 1$, and that a sequence ν_t of probabilities *converges weakly* to ν^* if $\lim_{t\to\infty} \int f d\nu_t = \int f d\nu^*$ for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$.

Theorem 1 *Consider the IFS described by* $\{f_1, f_2, p\}$.

i) *There is a unique attractor for the IFS; that is, a unique compact set* $A \subseteq X$, *such that* $f_1(A) \cup f_2(A) = A$.

ii) *There is a unique probability distribution* ν^* *on* $(X, \mathcal{B})$ *satisfying the functional equation (15), that is,*

$$\nu^*(B) = (1-p)\,\nu^*\left[f_1^{-1}(B)\right] + p\nu^*\left[f_2^{-1}(B)\right] \qquad \text{for all } B \in \mathcal{B}. \tag{16}$$

iii) A *is the support of* ν^* *and, for any probability*[10] ν_0 *on* $(X, \mathcal{B})$, *the sequence* $\nu_t = M^t \nu_0$ *for* $t = 0, 1, 2, \ldots$, *converges weakly to* ν^*.

The original proof relies on a contraction mapping argument and dates back to Hutchinson (1981). See also Lasota and Mackey (1994) and Falconer (2003) for further discussion.

3.2 Scaling Maps

Consider the IFS (12) and assume that the maps f_1, f_2 are increasing. Let a and b be their fixed points respectively, that is, $f_1(a) = a$ and $f_2(b) = b$, as in figure 1. Since the maps f_1, f_2 are both contractions, it is readily seen that, as time elapses, values x_t that are admissible eventually must lay inside the interval $[a, b]$, that is, $[a, b]$ is the trapping region of (12). In other words, the portion of the maps f_1, f_2 which is relevant in the long run is included in the square T in figure 1 (where the plots of f_1 and f_2 are in bold). Hence, with no loss of generality, we may let $X = [a, b]$.

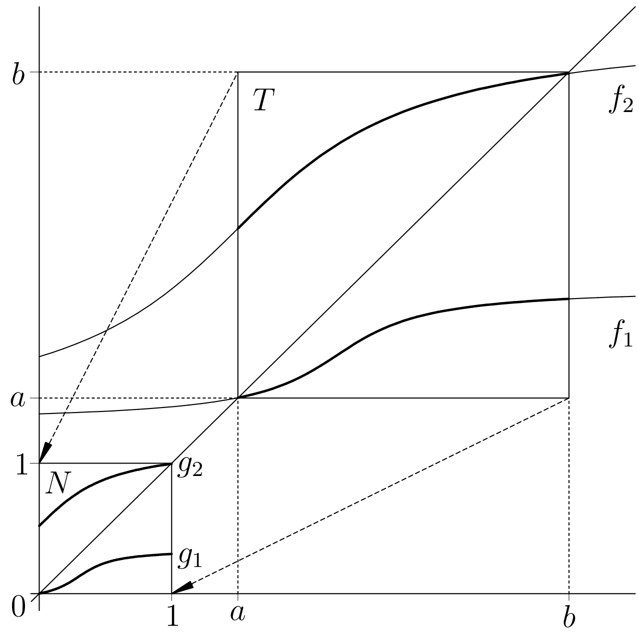


FIGURE 1: normalization of two contractive maps f_1, f_2 over the unit square.

[10]To be precise, weak convergence holds for any initial probability ν such that $\int |x-a|\, d\nu < \infty$ for some constant a. See Section 2.1.2 in Mitra, Montrucchio and Privileggi (2004) for more details.

For any increasing contractive maps $f_1,\ f_2$, such relevant region can be "normalized" over the interval $[0,1]$ (that is, the square T can be transformed into the square N in figure 1) by the following two transformations:

1. by a *rigid translation* towards the origin, so that the fixed point a becomes the origin itself, and

2. by *scaling* the whole system *by a factor* $k = b - a$.

The outcome of such transformation is a new IFS

$$y_t = \begin{cases} g_1\left(y_{t-1}\right) & \text{with probability } 1-p \\ g_2\left(y_{t-1}\right) & \text{with probability } p \end{cases} \tag{17}$$

where the maps g_j are given by

$$g_j\left(y\right) = k^{-1}\left[f_j\left(ky+a\right)-a\right], \qquad j = 1, 2, \tag{18}$$

with $k = b - a$, as can be easily checked. Figure 1 illustrates this translation/scaling procedure that transforms the original relevant region T into the new "normalized" relevant region N, which is the unit square. Such normalization can be generalized to maps $f_1 < f_2$ that are not necessarily monotone,[11] see Cozzi and Privileggi (2002) for details.

Transformations that are translations and scaling are called similarities (see Falconer, 2003, pp. 7 and 8). A *similarity* has the property of transforming sets into geometrically similar ones, in the sense that it preserves relative distances between points of the original set; formally, it is a transformation $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $\left|S\left(x\right)-S\left(y\right)\right| = k\left|x-y\right|$ for all $x, y \in \mathbb{R}^n$ and some constant *ratio* or *scale* $k > 0$. Therefore, by construction, the IFS (17) obtained through (18), has graph similar to the graph of the original IFS (12); this can be easily checked by noting that the graphs inside the squares T and N in figure 1 are themselves similar. With a slight abuse of terminology, we shall say that *the IFS* (12) and (17) *are similar*.

An important consequence of the normalization procedure described above is that the invariant sets of both (12) and (17) have the same geometric properties, as they are generated by similar systems. Thus, similar IFS have similar attractors, and studying the geometric features of the attractor of the normalized IFS (17) is equivalent to studying the geometry of (12).

3.3 Normalized Linear IFS

If the maps f_j are linear and with same slope $0 < \alpha < 1$, that is, of the form

$$x_t = \begin{cases} \alpha x_{t-1} + z_1 & \text{with probability } 1-p \\ \alpha x_{t-1} + z_2 & \text{with probability } p, \end{cases} \tag{19}$$

[11] To be precise, at least in the study of inequality phenomena, also the contractivity property could be relaxed somewhere in the "relevant region" (the square T in figure 1). The only minimum requirement is that the graphics of $f_1,\ f_2$ do not intersect inside this area and that the maps are contractions outside such area, so that the system is being attracted to the interval $[a,b]$ as time elapses.

where z_1, z_2 are any constants such that $z_1 < z_2$, then $a = z_1/(1-\alpha)$, $b = z_2/(1-\alpha)$ and (18) becomes the affine transformation

$$g_j(y) = \alpha y + (1-\alpha)\frac{z_j - z_1}{z_2 - z_1}, \qquad j = 1, 2, \tag{20}$$

which transforms IFS (19) into the similar one

$$y_t = \begin{cases} \alpha y_{t-1} & \text{with probability } 1-p \\ \alpha y_{t-1} + (1-\alpha) & \text{with probability } p \end{cases} \tag{21}$$

defined on $X = [0, 1]$. Figure 2 illustrates why interval $[0, 1]$ is the trapping region of the contractive system (21): 0 is the fixed point of the map $g_1(y) = \alpha y$ and 1 is the fixed point of the map $g_2(y) = \alpha y + (1-\alpha)$; since, at each period, the system "jumps" from one map to the other with probabilities $1-p$ and p respectively, it must eventually remain "trapped" between 0 and 1.

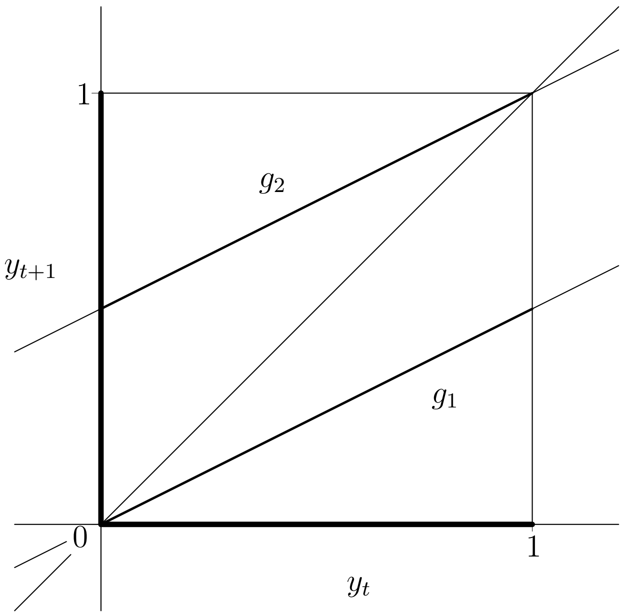


FIGURE 2: $X = [0, 1]$ is the trapping region of system (21), where $g_1(y) = \alpha y + (1-\alpha)$ and $g_2(y) = \alpha y$.

Notice that (20) provides an alternative – and more general – tool to obtain the normalized IFS (7) from the two (apparently) different systems (5) and (11) in Sections 2.1 and 2.2 respectively, where a direct change of variable has been used instead.

It is important to stress that the affine transform (20) does not affect the slope α of the maps f_j of the original linear IFS (19); in other words, the (similarity) transformation (20) *neutralizes the effect of the additive constants* z_1 and z_2. We thus have proven a general property, stated in the following lemma, which will be central in proving the main results of Section 5.

Lemma 1 *The common slope α of the maps in a* linear IFS *of the type (19) completely characterize its dynamic properties, independently of the additive constants z_1 and z_2. Accordingly, the geometric* *properties of its attractor depend uniquely on parameter α, and not on additive constants.*

Specifically, the (similar) linear IFS (19) and (21), which have the common slope α for both pairs g_j and f_j, have *similar attractors*. Thus, we are entitled to concentrate our analysis exclusively on IFS (21) – or, equivalently, on the IFS (7) – over the unit interval, $X = [0, 1]$.

To see how parameter α (and not additive constants) affects the whole geometry of IFS (21), observe that the graphs of g_1 and g_2 are two increasing parallel lines crossing the lower left and the upper right vertex of the unit square $[0, 1]^2$ respectively: the larger α (close to 1) the steeper and the closer they are, the lower α the flatter and the more apart they are. One may check (in this order) figures 3, 2 and 4(a) to grasp how these graphs change as values of α decrease.

3.4 Geometric Properties of the Attractor

It is important to emphasize some features of the attractor A of the IFS (21) – the support of its invariant distribution – which depend only on contraction factor α and are independent of probability p. This will provide a key ingredient for our definition of wealth polarization/pulverization.

A quick glance at figure 2 makes clear that the support of our IFS will be the whole interval $[0, 1]$ whenever $1/2 \le \alpha < 1$. This is because $T([0, 1]) = g_1([0, 1]) \cup g_2([0, 1]) = [0, 1]$ if the images of g_1 and g_2 overlap, that is, if $1/2 \le \alpha < 1$, as figure 3 shows. In this case we shall say that all marginal distributions ν_t, and thus also the invariant distribution ν^*, have "full support".

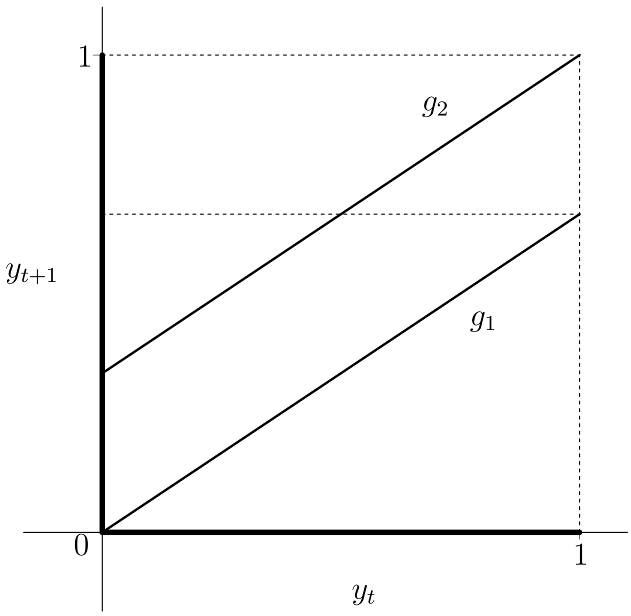


FIGURE 3: $g_1([0, 1]) \cup g_2([0, 1]) = [0, 1]$ when $1/2 \le \alpha < 1$.

More interesting is the case when images $g_1([0, 1])$ and $g_2([0, 1])$ do not overlap: this happens for $0 < \alpha < 1/2$, since $g_1([0, 1]) \cup g_2([0, 1]) = [0, \alpha] \cup [1 - \alpha, 1]$, where $[0, \alpha]$ and $[1 - \alpha, 1]$ are disjoint.

For $\alpha < 1/2$, there is a "gap" between the two image sets, with amplitude

$$h(\alpha) = 1 - 2\alpha > 0. \tag{22}$$

Note that $h(\alpha)$ is decreasing in α, and the gap "spreads" through the unit interval by successive applications of the maps g_j, reproducing itself, scaled down by a factor $1/\alpha$, in the middle of each subinterval born after each step t. Figure 4 reproduces the first three iterations of (21) starting from $[0, 1]$, generating a union of 8 ($= 2^3$) intervals of length α^3.

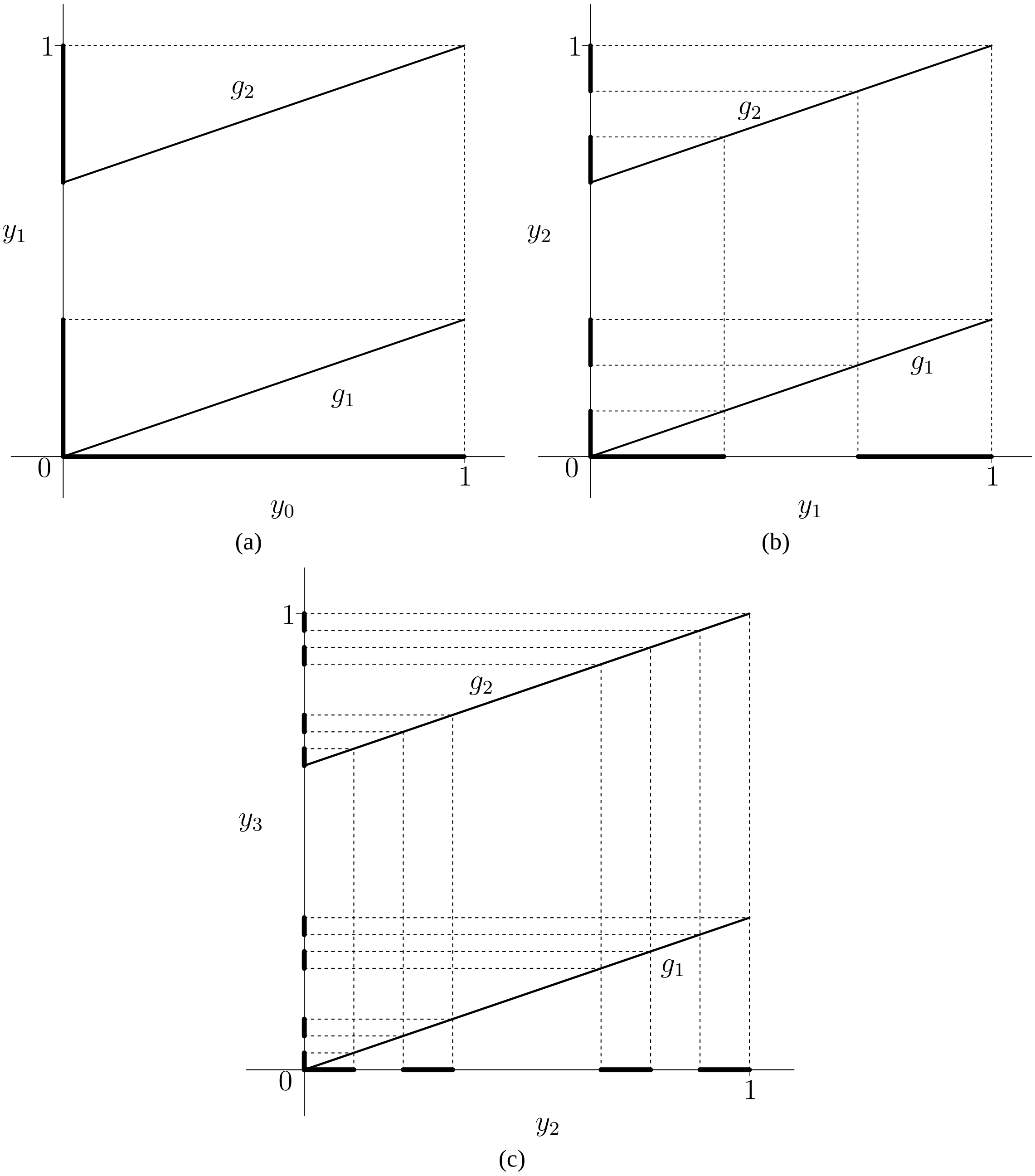


FIGURE 4: first three iterations of our IFS for $\alpha < 1/2$ starting from $[0, 1]$. The third iteration gives a union of eight intarvals of length α^3, as can be seen on the vertical axis in (c).

By pushing these iterations to the limit, we eventually find an attractor with features of the usual Cantor ternary set; in fact, for $\alpha = 1/3$, the support is precisely the Cantor ternary set. Cantor-like sets of the kind constructed by computing $\lim_{t \to \infty} T^t([0,1])$ for $0 < \alpha < 1/2$ exhibit several geometrical properties that are typical of *fractals*.

The most bewildering – and intriguing – feature of fractals is the need of a more sophisticated tool than the topological dimension – which allows only for integer values – to measure the "consistency" of their structure. Several dimensions has been constructed for this purpose, like, among others, the Hausdorff dimension, the Box-counting dimension and the Similarity dimension (for a discussion on dimensions see, for example, Falconer, 2003). All fractals have the peculiarity that their dimension is a "fraction", from which the name "fractal"; for instance, Cantor-like sets which are the attractors of (21) for $0 < \alpha < 1/2$ have Hausdorff dimension $-\ln 2/\ln \alpha$ (positive but less than 1), which, in this case, is the same as the Box-counting and the Similarity dimensions.

Whenever $\alpha < 1/2$, the attractor of (21) has dimension less than 1, which implies that it is totally disconnected; that is, between any two points there are "holes" (points laying outside the attractor). Conversely, even if dimensions less than 1 denote sets with very "disperse" points, it can be shown by means of a standard Cantor diagonal argument that Cantor-like sets contain uncountably many points, which are all pulverized across the interval itself (in the mathematical literature they are often referred as "Cantor dust"). Nonetheless, none of these points are isolated, *i.e.*, all Cantor-like sets have the paradoxical property that they are both *totally disconnected* and *perfect*. A terse and accessible discussion of the Cantor ternary set and its properties can be found in Chapter 11 in Strogatz (1994). Also Crownover (1995) is a good reference for an introductory approach.

3.5 The Invariant Distribution

Properties of the attractor A discussed before shed some light also on the limiting distribution supported on it. A subset of $\mathbb{R}$ with dimension less than 1 have Lebesgue measure zero.[12] Since A is the support of the invariant distribution ν^*, $\nu^*(A) = 1$, from which we deduce that ν^* turns out to be singular with respect to Lebesgue measure whenever $\alpha < 1/2$. However, singular invariant distributions are not confined to the case $\alpha < 1/2$, as it is widely discussed in Mitra, Montrucchio and Privileggi (2004), where singularity versus absolute continuity properties of ν^* are systematically investigated.

To have a flavor of what such an invariant distribution might look like, one may draw some iterations of Foias operator[13] M defined as in (14) starting from the uniform distribution over $[0,1]$. This, in the case $0 < \alpha < 1/2$, is equivalent to the following construction. Split a unit mass so that the right interval of $T([0,1])$ has mass p and the left interval has mass $1-p$. Then, divide the mass on each interval of $T([0,1])$ between the two subintervals of $T^2([0,1])$ in the ratio $p/(1-p)$. Continue in this way, so that the mass on each interval of $T^t([0,1])$ is divided in the ratio $p/(1-p)$ between its two subintervals in $T^{t+1}([0,1])$ (see also Example 17.1 in Falconer, 2003). Figure 5 depicts some

[12] A rigorous proof of this fact, which uses the notion of *Hausdorff measure*, can be found in Edgar (1990).

[13] The Maple code that generates plots like in figure 5 is available from the authors upon request.

iterations of M using this construction starting from the uniform distribution for $\alpha = p = 1/3$.

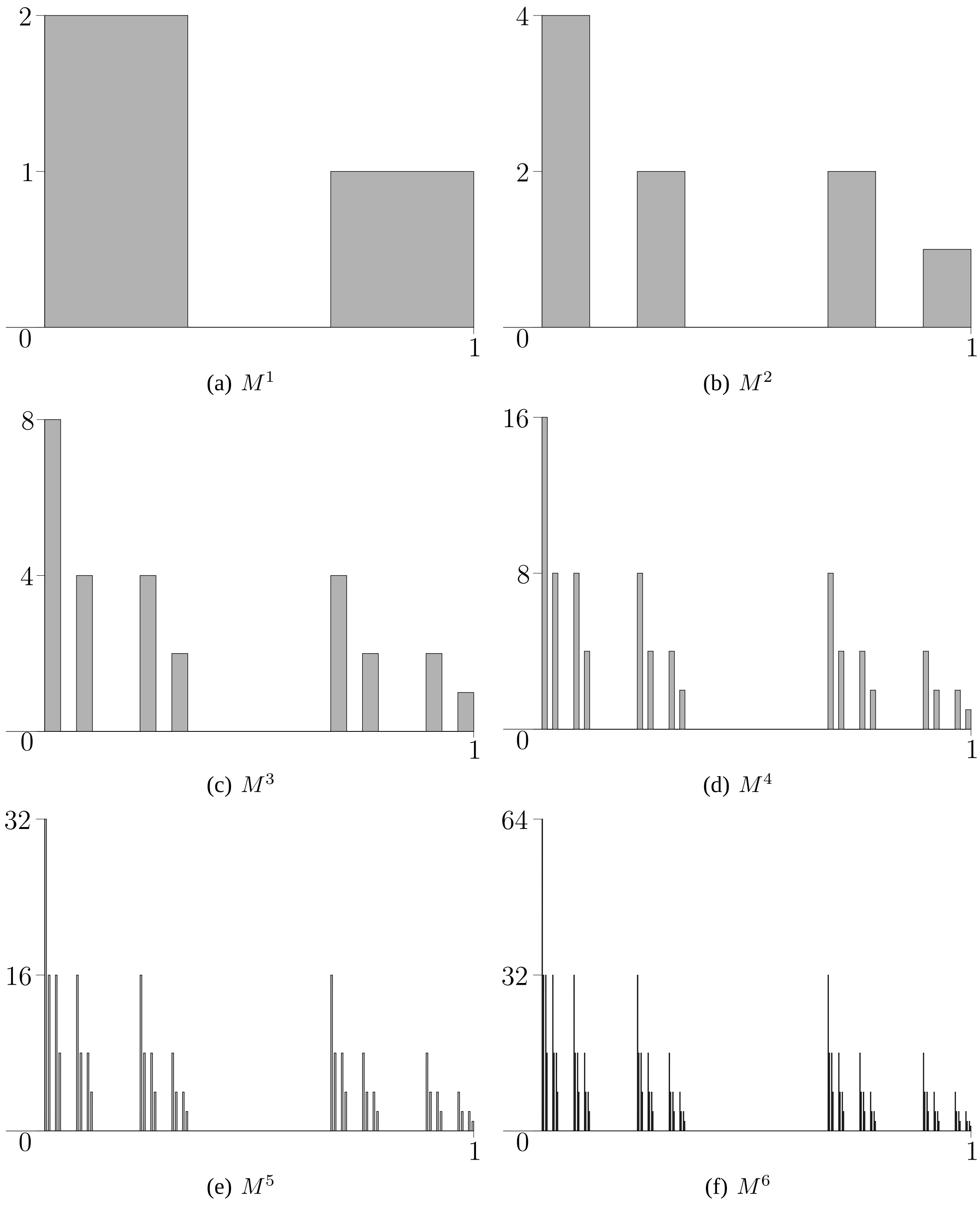
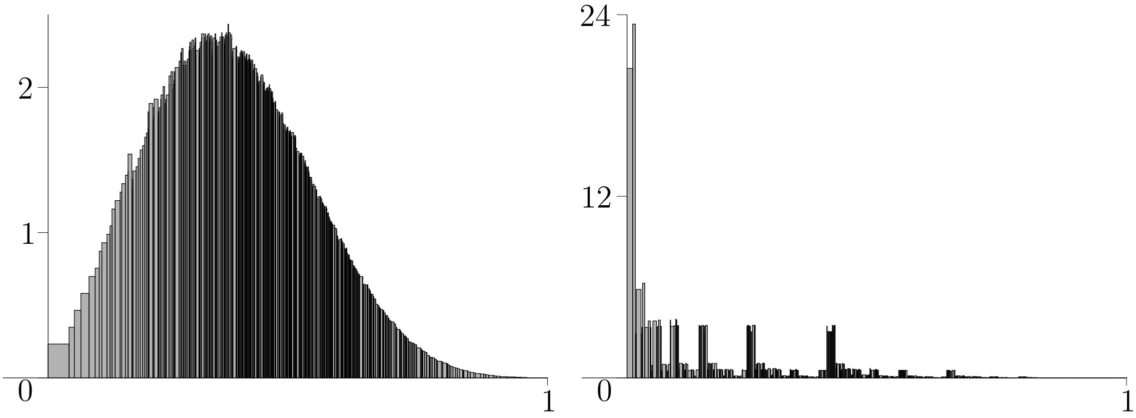


FIGURE 5: first six iterations of operator M starting from the uniform probability for $\alpha = 1/3$ and $p = 1/3$.

Figure 6 shows two examples of eight iterations of M in the overlapping case, *i.e.* for $\alpha \geq 1/2$, when the invariant distribution ν^* has full support. Note that for α close to 1 [high "degree of overlapping" of the images $g_1([0,1])$ and $g_2([0,1])$] and p sufficiently close to $1/2$, figure 6(a) suggests that ν^* will be "smooth" (absolutely continuous); while, whenever α gets closer to $1/2$ and p gets closer to the endpoints 0 or 1, as in figure 6(b), the approximation resembles the traits observed in figure 5(f), where the limiting distribution is known to be singular.



(a) $\alpha = 4/5$ and $p = 1/3$

(b) $\alpha = 3/5$ and $p = 1/8$

FIGURE 6: two examples of the first eight iterations of Foias operator M starting from the uniform probability in the overlapping case, that is, for $\alpha \geq 1/2$.

We end this section by noting that Theorem 1 applied to our IFS (21) provides also some standard information on the limiting distribution ν^*. Denote by $y^* \in [0,1]$ the random variable associated to the invariant distribution ν^*, that is, let y^* be the *random fixed point*[14] of system (21). Then, the functional equation (16) can be rewritten as

$$\nu^* \left(y^* \in B\right) = (1-p)\, \nu^* \left(\frac{y^*}{\alpha} \in B\right) + p\nu^* \left(\frac{y^*}{\alpha} - \frac{1-\alpha}{\alpha} \in B\right),$$

which allows for a direct computation of expectation and variance of y^*:

$$\mathbb{E}\left(y^*\right) = p \tag{23}$$

$$Var\left(y^*\right) = \frac{1-\alpha}{1+\alpha} p \left(1-p\right). \tag{24}$$

Note that these computations are justified thanks to weak convergence, since expectation and variance are the integrals of the identity function $f(y) = y$ and the function $f(y) = \left[y - \mathbb{E}_g(y)\right]^2$ respectively, which are both bounded and continuous on $[0,1]$.

4 Growth and Inequality

The stochastic dynamic model expressed by (7), or more generally by (21), turns out to be especially useful for a slightly different interpretation, which is the main focus of this paper. One-period probability p of individual i of successfully adopting technology A_t at the end of period t – or discovering some innovative production method in the Schumpeterian version of the model – can be seen, by the

[14]See Arnold (1998) for a detailed treatment of random dynamical systems and random fixed points.

law of large numbers, as the "average proportion of the whole population" that in the long run is able to catch the opportunity of benefitting from the (constantly evolving) new technology. In this scenario, the IFS (21) describes the evolution through time of the wealth distribution across a population of a continuum of individuals normalized to 1, which, by Theorem 1, in the long run converges to some invariant wealth distribution ν^* supported on a subset of $[0, 1]$.

From this aggregate perspective, expectation (23) can be read as the average productivity-adjusted wealth in the steady state, and variance (24) as the dispersion of individual wealths. From (23) it is immediately seen that the higher the individual probability p of exploiting technology A_t (or successfully innovating), the "richer" the economy on average; while (24) shows that low values of parameter $\alpha = \beta/\gamma$ (*i.e.*, low altruism rate β or high exogenous growth rate γ) and values of parameter p close to $1/2$, entail a dispersed invariant wealth distribution ν^*. Index (24) provides a very rough measure of wealth inequality; incidentally, note that, for any fixed value of probability of success p, the lower parameter α, the more dispersed the (steady state) wealth distribution.

In view of Section 3, we are in the position of saying much more on the steady state of such kind of economy. Specifically, we focus on the existence of a middle class, which is often considered important for growth itself, for democracy, for sociopolitical stability, and for the law and order, as quantified, among others, in the empirical analyses of Alesina and Rodrik, (1994), Perotti (1996) and Barro (1999). A strong middle class in our economy is represented by an invariant distribution ν^* that gathers a proportionally larger fraction of the population around $1/2$ than close to the endpoints 0 and 1 of the interval $[0, 1]$. Our main result, Proposition 1, provides clear-cut conditions for the converse, the *lack of a middle class*, thus characterizing economies which are polarized in terms of wealth distribution.

The self-contained description of such steady state in terms of attractor of the IFS (21) carried out in Section 3.4 makes clear the relationship between values of parameter α and the very existence of a middle class: economies featuring values $\alpha < 1/2$ for the exogenous parameter $\alpha = \beta/\gamma$ have the striking property that a middle class disappears already after one period starting from any wealth distribution ν_0 on $[0, 1]$. Such disappearance is graphically represented by the "gap" between the two disjoint image sets $g_1([0, 1])$ and $g_2([0, 1])$ in figure 4(a): already the first marginal distribution ν_1 concentrates wealth on two disjoint classes regardless of the wealth distribution ν_0 on $[0, 1]$ in $t = 0$. Furthermore, this gap is doomed to stay there forever, that is, also the limiting (steady state) wealth distribution ν^* turns out to be characterized by the same lack of a middle class. Note that, as we observed in Section 3.4, this happens independently of the probability of success p, and the size of the gap increases as parameter α decreases, which is consistent with the measure of dispersion provided by (24). Since the lack of a middle class can be seen as an extreme case of *wealth inequality*, accordingly to the literature on inequality we shall refer to it with the term *wealth polarization*.[15]

Moreover, we have seen in Section 3.4, that whenever the images sets $g_1([0, 1])$ and $g_2([0, 1])$ are

[15] We shall see in Section 6 that the term polarization becomes problematic whenever a more technical definition of polarization is needed for distributions supported on Cantor sets. Throughout most part of this paper, we shall employ the term polarization to identify whatever wealth distribution characterized by a missing middle class, as formalized in the next Definition 1.

disjoint, after the first iteration of the IFS the hole appearing in the support of the marginal wealth distribution ν_1 is being infinitely replicated on smaller scale in all supports of the successive marginal distributions ν_t, for all $t \geq 2$, leading in the limit to a support for the invariant distribution ν^* which is a Cantor-like set. This phenomenon creates a form of intra-class social disconnection that we somewhat tentatively will label *wealth pulverization*.

The discussion above leads to the following definition of wealth polarization/pulverization based on the (no) overlapping property of the image sets of the maps g_1 and g_2 of the IFS (21).

Definition 1 *Consider any economy of the type described in Section 2, where (productivity adjusted) wealth distribution through time is described by the IFS* $\{g_1, g_2, p\}$ *defined as in (21) on* $X = [0, 1]$*. We shall say that such economy is* ***polarized/pulverized*** *whenever*[16]

$$g_1(1) < g_2(0). \tag{25}$$

A direct application of Definition 1 leads to our main result.

Proposition 1 *Under Assumption 1 for the model introduced in Section 2.1, or condition (8) plus Assumption 2 for the model described in Section 2.2, if* $\gamma > 2\beta$ *the support* A *of the limit distribution* ν^* *of both economies is a Cantor-like set, and thus they are polarized/pulverized in the long run.*[17] *Moreover, the larger* γ *(and/or the smaller* β*), the larger the gap between the fractions of the population – the the "poor" and "rich" – near the endpoints of the interval* $[0, 1]$*, independently of the values of parameters* p *and* θ*.*

Proof. Since $\alpha = \beta/\gamma$, $\gamma > 2\beta \iff \alpha < 1/2$, which itself is equivalent to (25). The latter statement follows from (22), which measures the size of the gap between the "poor" and the "rich" fractions of the population as a decreasing function of $\alpha = \beta/\gamma$. ■

Proposition 1 shows that a high economic growth rate, by rewarding the successful individuals and penalizing in relative terms those who are not ready to catch the opportunities associated with the new technologies, make the middle class disappear and polarize society in two different wealth classes. Polarization becomes dramatic the larger the jump in productivity γ and the smaller the individual degree of altruism β (or, equivalently, the more selfish the individuals).

Remark 1 *It is important to highlight that a polarized wealth distribution does not mean that wealth classes are trapping the individuals: all individuals have the same opportunity to become rich or poor in this economy and it is precisely the amplitude of the* social mobility *– and not the frequency, that is the probability* p *of catching the technological opportunity, or finding some innovative production method – that generates wealth polarization.*

[16]For a discussion of the no overlap property (25) applied to stochastic optimal growth models of the Brock and Mirman (1972) type, see Mitra and Privileggi (2004, 2006 and 2009).

[17]Condition $\gamma > 2\beta$ is both necessary and sufficient for the attractor A to be a Cantor set. However it is clearly only sufficient for polarization, since, generally speaking, an invariant distribution may well have full support and at the same time exhibiting some degree of polarization, as it will be shown in Section 7.

5 Redistribution and Social Cohesion

It turns out that normalizing maps of IFS as in (19) on the interval $[0, 1]$ has important policy implications. Specifically, redistribution schemes based on lump-sum transfers from the rich to the poor aimed at doing away with social polarization/pulverization are not capable of achieving such goal, while direct wealth taxation may even make polarization worse. This happens because the "hole" that generates polarization depends only on parameters β (preferences) and γ (growth rate), as it has been widely argued in the previous sections, and cannot be affected by mere transfers of income, as the latter simply translate into different values for constants z_1, z_2 in system (19).

This result appears counter-intuitive at a first glance. We shall devote the next sections to analyze in detail whether and how alternative forms of government intervention may affect wealth polarization. First, two types of lump-sum transfers which fail to eliminate wealth polarization, one for the model described in Section 2.1 and one for the Schumpeterian version of Section 2.2, are discussed. Thereafter, such a result is being even strengthen by showing that direct wealth taxation may actually worsen polarization. In Section 5.2, however, we shall offer a fiscal solution based on random taxation of the rich that may wipe out polarization, at least in the sense of "filling the gap" in the support of a polarized invariant distribution. For simplicity, we will not assume that polarization/pulverization implies productivity losses.

5.1 Lump-Sum Transfers

In the model of Section 2.1, let us assume that the gains from success are taxed at the end of each period a proportion $0 \leq \tau < 1$ and that proceeds are redistributed lump-sum to the unluckies.[18] If all individuals exert effort e_t in order to learn technology A_t, the steady state proportion of rich families in the economy will still be p. Hence, the government in the long run will be able to collect tax revenues equals to $p\tau A_t$, which – assuming a balanced government budget every period – equals the aggregate lump sum transfer received at the end of period t by the whole poor.

Since taxation further reduces the expected benefit derived from having the opportunity of adopting technology A_t, in order to let all individuals keep putting effort e_t even under taxation and thus obtain a dynamic similar to that in (3), an upper bound on tax rate τ is needed. Let us discuss in detail how Assumption 1 needs to be modified to avoid free riding behavior due to the possibility of receiving, out of nothing, a transfer that generates a higher utility than the expected utility gain produced by putting effort e_t.

Let $0 \leq l \leq 1$ denote the fraction of the population who decides to put effort e_t in learning technology A_t. Then, at the steady state, the total amount of tax revenues is $pl\tau A_t$, and each non-successful individual i – which are both the unlucky ones who exerted effort e_t and the lazy ones who

[18]Note that, assuming lumps sum redistribution to all individuals – not only to the unluckies – would not alter the qualitative results of our analysis.

did not exert any effort, that amount to a proportion $1-pl$ of families – receives a transfer given by

$$T_t^i = \frac{pl}{1-pl}\tau A_t. \tag{26}$$

In view of (2), the individual i expected utility gain conditional to effort e_t is given by

$$\begin{aligned}\mathbb{E}\left[U\left(Y_t^i\right)|e_t\right] &= \rho\left[p\left(1-\tau\right)A_t + \left(1-p\right)T_t^i\right] - e_t \\ &= \rho\left[p\left(1-\tau\right)A_t + \left(1-p\right)\frac{pl}{1-pl}\tau A_t\right] - e_t,\end{aligned}$$

where $\rho = \left(1-\beta\right)^{1-\beta}\beta^{\beta}$, while the individual i certain utility gain obtained by exerting zero effort is given by

$$U\left(T_t^i\right) = \rho\frac{pl}{1-pl}\tau A_t.$$

In order to let all the families put the effort $e_t = \gamma^t e_0$ required to learn technology A_t, we need

$$\mathbb{E}\left[U\left(Y_t^i\right)|e_t\right] > U\left(T_t^i\right)$$

to hold for all $0 \leq l \leq 1$, which leads to

$$\left(1 - \frac{\tau}{1-pl}\right)p\rho A_0 > e_0.$$

Since the minimum of the left hand side is reached for $l = 1$, then, for each given e_0 satisfying Assumption 1, the following restriction on parameter τ guarantees that all families will always put effort e_t in learning technology A_t also under government taxation.

Assumption 3 *Assumption 1 holds and*

$$0 \leq \tau < \left(1-p\right)\left(1 - \frac{e_0}{p\rho A_0}\right), \tag{27}$$

where $\rho = \left(1-\beta\right)^{1-\beta}\beta^{\beta}$.

Hence, in view of (3), the dynamics of individual i's wealth becomes:

$$W_t^i = \begin{cases} \beta W_{t-1}^i + p\left(1-p\right)^{-1}\tau A_t & \text{with probability } 1-p \\ \beta W_{t-1}^i + \left(1-\tau\right)A_t & \text{with probability } p, \end{cases} \tag{28}$$

where $p\left(1-p\right)^{-1}\tau A_t$ represents the transfer received by a single unlucky family, *i.e.*, T_t^i in (26) with $l = 1$. By dividing both equations in (28) by A_t we get productivity-adjusted linear dynamics:

$$w_t^i = \begin{cases} \left(\beta/\gamma\right)w_{t-1}^i + p\left(1-p\right)^{-1}\tau & \text{with probability } 1-p \\ \left(\beta/\gamma\right)w_{t-1}^i + \left(1-\tau\right) & \text{with probability } p. \end{cases} \tag{29}$$

Under Assumption 3, the RHS in (27) implies $\tau < 1 - p$, which, in turn, implies $p(1-p)^{-1}\tau < (1-\tau)$. If we let $\alpha = \beta/\gamma$, $z_1 = p(1-p)^{-1}\tau$ and $z_2 = (1-\tau)$, (29) becomes as in (19), which is similar to (21), and thus a direct application of Lemma 1 immediately yields the following result.

Proposition 2 *If $\gamma > 2\beta$, polarization/pulverization never disappears for all income tax rates τ satisfying Assumption 3.*

Figure 7 shows that only the common slope of the two maps constituting the IFS affects polarization/pulverization while lump-sum transfers – which are nothing else than additive constants – have no effect in reducing inequality.

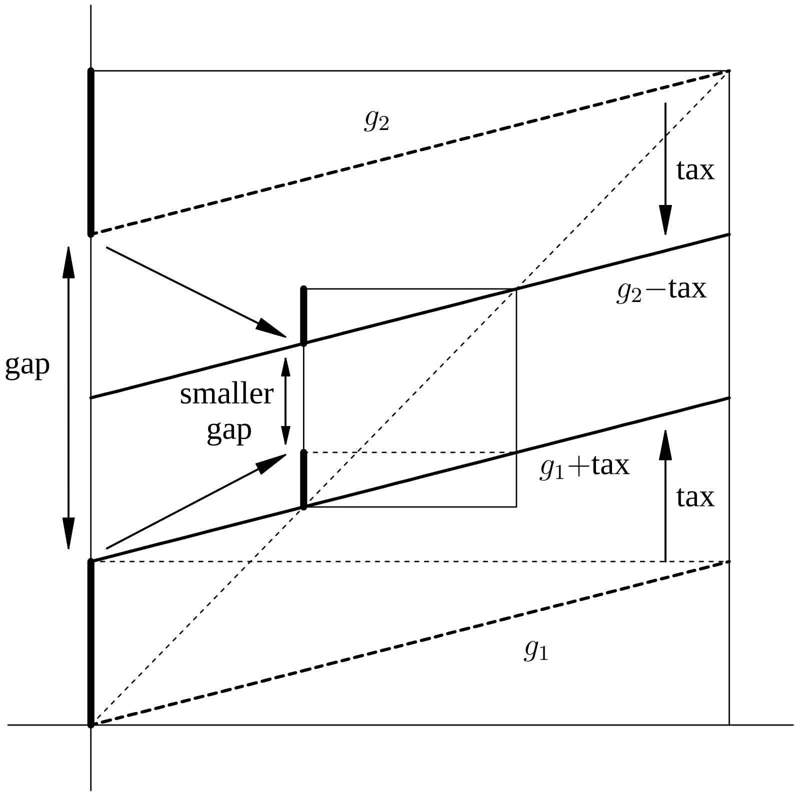


FIGURE 7: redistribution from the rich to the poor has the only effect of shrinking the size of the gap, it does not make it disappear.

There is, however, an important difference with respect to the dynamics obtained in Section 2.1. Observing the evolution through time of the supports of the marginal distributions ν_t of systems (28) or (29), it is clear that the standard of living of the poor under wealth redistribution will be bounded away from zero in the long run, that is, nobody will end up with a zero wealth in the steady state. As a matter of fact, the feasible wealths of system (28) at time t lay in some subset of the interval

$$\left[\beta b_0^i + \left(\frac{1-(\beta/\gamma)^{t+1}}{\gamma-\beta}\right)\gamma^{t+1} p(1-p)^{-1}\tau A_0, \beta b_0^i + \left(\frac{1-(\beta/\gamma)^{t+1}}{\gamma-\beta}\right)\gamma^{t+1}(1-\tau)A_0\right],$$

where the left endpoint is strictly positive and increasing over time. Therefore, although government redistribution does not affect polarization/pulverization, it still proves effective in sustaining the wealth of the poor. Clearly also the "rich side" of the population is being affected by having a reduced – by factor $(1-\tau)$ – maximum possible wealth compared to that of the original feasible region (4). Thus, the overall effect of a redistributive policy by the government is to narrow the whole absolute wealth around its mean, without changing polarization/pulverization features in relative terms.

It would be natural to endogenize fiscal policy along the lines of Alesina and Angeletos (2005a) and Alesina, Cozzi, and Mantovan (2009), with the important difference that here luck and unluck are not additive, but multiply effort. As shown by Alesina, Cozzi, and Mantovan (2009), the implied dynamics will depend on how preferences for a desired distribution are chosen: somewhat unexpectedly, the present model suggests that the introduction of the disutility of living in a polarized/pulverized society could get voters to support lower taxation.

5.2 Government Purchase of Innovations

If the effort e_t required to promote innovation is sufficiently small, in the Schumpeterian model of Section 2.2 the government could reward the innovator by purchasing the innovation and at the same time making the innovation itself immediately publicly available to everybody, as suggested by Kremer (1998).

Provided that population is normalized to 1, the society as a whole will put effort e_t in R&D for new technological projects and at the steady state there will be a fraction p of successful innovators who possess technology A_t. Suppose that the government, in order to make technology A_t publicly available in period t, buys the technological know-how from the p fraction of innovators at the lowest incentive compatible price,[19] *i.e.*, at $p^{-1}e_t$, and allows the fraction $1-p$ of unluckies to freely use it in their own firms. Assume further that the government charges all the unluckies the whole cost e_t of research through a lump-sum tax to be fully transferred to the luckies. Then, the law of motion of wealth becomes:

$$W_t^i = \begin{cases} \beta W_{t-1}^i + A_t - (1-p)^{-1} e_t & \text{with probability } 1-p \\ \beta W_{t-1}^i + A_t + p^{-1} e_t & \text{with probability } p, \end{cases} \tag{30}$$

where $(1-p)^{-1} e_t$ denotes the per capita cost of research charged to the unluckies and $p^{-1}e_t$ denotes the per capita compensation for the productivity gain loss (9). We will assume e_0 small enough to guarantee that the unluckies are better off under this forced purchase of the new technology than under *laissez faire*.

Observe that, at least for the case $p < 1/2$, which seems sufficiently realistic, system (30) can be reduced to system (7) – or (21) – through formula (20). Therefore, once again, Lemma 1 and Proposition 1 apply stating that polarization/pulverization is completely determined by condition $\gamma > 2\beta$ and a result similar to Proposition 2 holds: government financing private innovations does not

[19] Note that any price slightly higher than $p^{-1}e_t$ makes each individual strictly better off undertaking the R&D effort.

affect polarization/ pulverization.

The delicate part, as usual, is enforceability of such a policy: nobody would vote a government policy which leaves everybody worse off. The individual expected indirect utility gain is

$$\begin{aligned}\mathbb{E}\left[U\left(Y_t^i\right)\right] &= \rho\left[p\left(A_t+\frac{e_t}{p}\right)+(1-p)\left(A_t-\frac{e_t}{1-p}\right)\right]-e_t \\ &= \rho p A_t - e_t,\end{aligned}$$

and thus, the effort condition turns out to be the same as[20] in Assumption 1:

$$e_0 < \rho p A_0. \tag{31}$$

Note that, by assuming $p < 1/2$, necessarily $\rho p < 1-p$, and thus (31) implies $e_0 < (1-p)\,A_0$, which guarantees that the left endpoint of the support of the marginal probabilities of process (30), which at time t is

$$\beta b_0^i + \left(\frac{1-(\beta/\gamma)^{t+1}}{\gamma-\beta}\right)\gamma^{t+1}\left(A_0-\frac{e_0}{1-p}\right),$$

is strictly positive for all t. This means that, like in the previous section, the poorest segment of the population improves its standards of living at the same steady rate $\gamma-1$ as the richest.

5.3 Direct Wealth Taxation

Let us now consider wealth taxation (not redistributed lump-sum) for the model described in Section 2.1. If final wealth is taxed at a rate $0 < \tau_w < 1$, the dynamical system (5) becomes:

$$w_t^i = \begin{cases} (1-\tau_w)\,(\beta/\gamma)\,w_{t-1}^i & \text{with probability } 1-p \\ (1-\tau_w)\,(\beta/\gamma)\,w_{t-1}^i + (1-\tau_w) & \text{with probability } p \end{cases} \tag{32}$$

which, again in view of Lemma 1 and Proposition 1, immediately implies the following result, as can be easily established by letting $\alpha = (1-\tau_w)\,(\beta/\gamma)$, $z_1 = 0$ and $z_2 = 1-\tau_w$, so that (32) is as in (19) and thus similar to (21).

Proposition 3 *Suppose Assumption 1 holds and* $0 < \tau_w < 1-(p\rho A_0)^{-1}\,e_0$. *Then, if* $\gamma > (1-\tau_w)\,2\beta$, *polarization/pulverization emerges.*

In this case, government intervention proves effective (for the worse) in modifying polarization/pulverization as it is capable of affecting the common slope of the maps of (5) – and thus also that of the maps in (7) or (21) – besides the additive constants. Therefore, a high enough wealth tax rate can generate a polarized wealth distribution even if $\gamma < 2\beta$, that is, even if growth and altruism are such that that the private sector let alone does not generate polarization. In other words,

[20]This seems to be reasonable since utility is linear and what is taken from the unluckies goes to the luckies, leaving the expected utility gain unchanged.

somewhat paradoxically, in this model the middle class may disappear and the economy becomes polarized/pulverized as a result of an active redistributive policy. Here, to isolate the pure effect of taxation, we have not assumed any transfer from the government; recall, however, from Section 5.1, that any lump sum transfer would not have any effect on wealth polarization.

5.4 Random Taxation

We here show that a redistribution scheme based on random taxation may reduce and, in some cases, even eliminate polarization.[21] The idea is to increase the uncertainty in the model so that the two-maps IFS (28) is being replaced by a three-maps IFS in which the image set of the second map might fill the hole left by the other two images set in case of polarization.

In the framework developed in Section 2.1, let us assume that the gains from success are taxed at some rate $0 < \tau < 1$ with probability $1 - q$, with $0 < q < 1$. At each period, the successful individuals face a tax lottery such that they have to pay τA_t with probability $1 - q$ and 0 with probability q. Probability q is constant through time and is independent of the probability of success p. The government controls parameters q and τ. The total amount of proceeds are redistributed lump-sum to the unluckies.

If all individuals exert effort e_t in order to learn technology A_t, the steady state proportion of rich families in the economy will still be p. A fraction q of this proportion will be tax exempt, while the other fraction $1 - q$ will be taxed at rate τ. Hence, the government in the long run will be able to collect tax revenues equals to

$$p\,(1-q)\,\tau A_t,$$

which – assuming a balanced government budget every period – equals the aggregate lump sum transfer received at the end of period t by the whole poor.

The dynamics of individual i wealth becomes:

$$W_t^i = \begin{cases} \beta W_{t-1}^i + \dfrac{p\,(1-q)}{1-p}\tau A_t & \text{with probability } 1-p \\ \beta W_{t-1}^i + (1-\tau)\,A_t & \text{with probability } p\,(1-q) \\ \beta W_{t-1}^i + A_t & \text{with probability } pq, \end{cases}$$

where, in the first line, $p\,(1-p)^{-1}\,(1-q)\,\tau A_t$ represents the transfer received by a single unlucky family. Let $\alpha = \beta/\gamma$ and consider the productivity-adjusted dynamics:

$$w_t^i = \begin{cases} f_1\left(w_{t-1}^i\right) = \alpha w_{t-1}^i + \dfrac{p\,(1-q)}{1-p}\tau & \text{with probability } 1-p \\ f_2\left(w_{t-1}^i\right) = \alpha w_{t-1}^i + (1-\tau) & \text{with probability } p\,(1-q) \\ f_3\left(w_{t-1}^i\right) = \alpha w_{t-1}^i + 1 & \text{with probability } pq. \end{cases} \tag{33}$$

System (33) contains three (affine) contractive maps identified by parameters α, p, q and τ, where the

[21] We owe the idea of studying the effects of a random tax on polarization/pulverization to Salvador Ortigueira.

last two are decision variables for the government. We want to investigate for what values of these parameters 1) incentive compatibility holds, *i.e.*, all individuals exert effort e_t, 2) the three maps are ordered so that $f_1 < f_2 < f_3$, and 3) whether values of the parameters exist so that the image set of f_2 fills the (possible) gap left by the image sets of f_1 and f_3. The last point would mean the possibility of eliminating polarization through government redistribution under this random scheme.

With no loss of generality for the rest of this section we shall assume

$$\frac{1}{3} \leq \alpha < \frac{1}{2}.$$

The right inequality implies that the two maps in (21) exhibit polarization (their images do not overlap), while the left inequality allows for the introduction of a third affine map with the same slope α between the two given maps, so that the hole left by the two pre-existing image sets can be completely "filled". From figure 4(a), it is easily understood that maps with slope $\alpha < 1/3$ have images sets which cannot fill the whole interval $[0, 1]$. Clearly, for maps with $\alpha < 1/3$, arguments similar to the one carried out in this section can be implemented for random taxation schemes that use different tax rates. For example, if $\alpha < 1/n$, $n - 1$ tax rates, each with positive probability, are necessary.

In order to let all individuals keep putting effort e_t even under taxation, an upper bound on the tax rate τ similar to that in Assumption 3 is needed. By replacing the certain tax rate with the expected rate tax $(1 - q)\,\tau$ we are easily led to the following inequality:

$$\tau < \frac{1-p}{1-q} \left(1 - \frac{e_0}{p\rho A_0}\right), \tag{34}$$

where $\rho = (1 - \beta)^{1-\beta} \beta^{\beta}$. Moreover, in order to have $f_1 < f_2$

$$\tau < \frac{1-p}{1-pq}$$

must hold; while $f_2 < f_3$ follows from $0 < \tau < 1$. Hence, the following assumption is what we need.

Assumption 4 *Assumption 1 holds and*

$$0 < \tau < \min \left\{ \frac{1-p}{1-q} \left(1 - \frac{e_0}{p\rho A_0}\right) \,,\, \frac{1-p}{1-pq} \right\},$$

where $\rho = (1 - \beta)^{1-\beta} \beta^{\beta}$.

To analyze the possibility of eliminating polarization, let us normalize the three maps IFS (33) to the interval $[0, 1]$ along the argument discussed in Section 3.3. We shall apply formula (20) with $\alpha = \beta/\gamma$, $z_1 = p\,(1 - p)^{-1}\,(1 - q)\,\tau$ and $z_2 = 1$ to get the lower and higher maps as in (21), while the constant intercept of the map in the middle will be obtained by letting $z_j = (1 - \tau)$ in (20). Hence, we get the normalized system

$$w_t^i = \begin{cases} g_1\left(w_{t-1}^i\right) = \alpha w_{t-1}^i & \text{with probability } 1-p \\ g_2\left(w_{t-1}^i\right) = \alpha w_{t-1}^i + (1-\eta)(1-\alpha) & \text{with probability } p(1-q) \\ g_3\left(w_{t-1}^i\right) = \alpha w_{t-1}^i + (1-\alpha) & \text{with probability } pq, \end{cases} \tag{35}$$

where

$$\eta = \frac{(1-p)\,\tau}{(1-p) - p(1-q)\,\tau}.$$

Note that, under Assumption 4, $0 < \eta < 1$.

The overlapping condition for the three image sets is a straightforward computation that leads to

$$1 - 2\alpha \leq (1-\alpha)\,\eta \leq \alpha,$$

which, in terms of τ, boils down to

$$\frac{(1-2\alpha)(1-p)}{(1-p)(1-\alpha) + (1-2\alpha)\,p\,(1-q)} \leq \tau \leq \frac{\alpha\,(1-p)}{(1-p)(1-\alpha) + \alpha p\,(1-q)}. \tag{36}$$

Note that condition (36) is nonempty for $1/3 \leq \alpha < 1/2$, and coincides with a single value for τ when $\alpha = 1/3$, that is when inequalities in (36) become equalities and there is only one map g_2 in (35) whose image set can fill the whole gap left by the other two.

The left hand side of condition (36) is the most important in our analysis: it requires τ to be sufficiently large in order to eliminate polarization. However, in view of Assumption 4, we observe that τ must be not too large to let the incentive compatibility (34) be always satisfied. If this constraint is too tight, due, *e.g.*, to a high value of the ratio $e_0/(p\rho A_0)$, the left hand side in (36) might not hold, thus leaving the government with no room for applying redistributive policies against polarization.[22] Specifically, polarization is neglected if τ is chosen to be equal to the left hand side of (36) and

$$\frac{e_0}{p\rho A_0} < 1 - \frac{(1-2\alpha)(1-q)}{(1-p)(1-\alpha) + (1-2\alpha)\,p\,(1-q)}. \tag{37}$$

Note that we did not discuss any restrictions for the choice of parameter q by the government so far. Since by Assumption 1 $e_0/(p\rho A_0) < 1$, there always exist values for parameter $q < 1$, possibly close to 1, such that (37) is satisfied. In other words, there is always room for the government to eliminate polarization through a random taxation and lump-sum redistribution scheme in the sense of making the support of the steady state distribution of system (35) to be the whole interval $[0, 1]$. However, values of q close to 1 imply that almost the whole p fraction of the steady state successful

[22] Note that the other component of Assumption 4 is always satisfied since

$$\frac{(1-2\alpha)(1-p)}{(1-p)(1-\alpha) + (1-2\alpha)\,p\,(1-q)} < \frac{1-p}{1-pq}$$

is always true.

population, *i.e.*, pq out of p, is paying no taxes, while only a negligible fraction $p(1-q)$ of the successful population is paying taxes; but this amounts exactly to the "middle class" artificially created through the random taxation. Therefore, as the new middle class carries nearly no weight, polarization remains substantially unaltered in terms of "wealth distribution", even if such distribution has full support. That is, once again, a tight incentive compatibility constraint in Assumption 4 leaves little room for government intervention and substantially reduces hopes of eliminating polarization even through random taxation.

6 Inequality Versus Pulverization

So far we have used the term polarization to generically describe an extreme degree of inequality due to the disappearance of a middle class in a distribution supported on a Cantor set (as in Definition 1). A more "technical" concept of polarization assumes, besides the inequality produced by different wealth levels between groups, also a certain degree of concentration, or "clustering", of wealth within each group: if the distribution of wealth is highly gathered within groups but very diverse between groups in a population, then wealth is considered "polarized" between the groups (see, *e.g.*, Esteban and Ray, 1994, Wolfson, 1994, and, for a survey, Zhang and Kanbur, 2001). In other words, the generation of tensions possibly evolving to rebellion, revolt, or social unrest is more likely if wealth is distributed among groups which have a strong self-identity feeling.

However, we have seen in the previous sections that the striking inequality phenomenon possibly occurring after one period, the lack of a middle class, is being replicated on a smaller scale among wealth sub-clusters after each iteration of any IFS similar to (21), provided that $\gamma > 2\beta$ [see, *e.g.*, figure 5]. In Section 4 we somewhat tentatively called "pulverization" such dispersion of wealth over a Cantor set. Clearly, pulverization runs against polarization, since it may be seen as the result of a progressive erosion of the wealth concentration around the two main clusters appeared after the first period. In the limit, whenever the invariant distribution of wealth is supported over a Cantor set, all wealth groups are distinct (a Cantor set is disconnected) and each of them bear zero weight (a Cantor set has a continuum of points over which a unit mass is being spread, as we shall see in short).

All these considerations should be enough to discourage any attempt for providing meaningful polarization measures for distributions supported on Cantor sets by means of any standard index available in the literature. Nonetheless, in this section we aim at shading some light on whether pulverization may or may not affect, if not – technically speaking – polarization, at least inequality in the long run. Such goal is achieved by adapting the most popular inequality measure, the Gini coefficient, to our invariant wealth distribution when it is supported on a Cantor set.

Formally, given a finite distribution of weights $\pi_1, \ldots, \pi_n$ on wealths $W_1, \ldots, W_n$, with $\pi_i, W_i > 0$, the Gini coefficient is given by

$$G = \frac{1}{2\mu} \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i \pi_j |W_i - W_j|, \tag{38}$$

where μ denotes the mean wealth across the whole sample. Clearly (38) is meant to measure inequality by using statistical data available for societies with finite populations. The pursue of some generalization of (38) to include infinite distributions supported over fractal sets is well beyond the scope of this paper. Our goal is more modest: we just aim at checking whether pulverization affects inequality in the long run. For this purpose, the computation of the limit of G in (38) as $n \to \infty$, to see whether it remains positive or boils down to zero, should be sufficient. Such question is non trivial, as two opposite effects occur by applying formula (38) directly to our IFS in the case of real (*i.e.*, not adjusted by productivity) wealth dynamics: on one hand, the weights π_i decrease after each step, since, under our assumptions, the same unit population is being progressively spread over more and more wealth clusters, and the same does the reciprocal of the mean, $1/\mu$; on the other hand, after each period new wealth groups W_i are born and the distances between wealth clusters increase, thus raising both the number of addends in the sum and the values $|W_i - W_j|$.

Consider the dynamical system (3) discussed in Section 2.1:

$$W_t = \begin{cases} \beta W_{t-1} & \text{with probability } 1-p \\ \beta W_{t-1} + A_t & \text{with probability } p, \end{cases} \tag{39}$$

where W_t denotes some wealth amount at time t, $0 < \beta < 1$ is the degree of intergenerational altruism, $A_t = \gamma^t A_0$ is the exogenous technology with $A_0 > 0$, $\gamma > 1$, and $0 < p < 1$ represents the probability of success in the adoption of the technology. The choice of studying system (39) instead of system (7) – which is normalized on the interval $[0, 1]$ – is made to conform with the mainstream literature on inequality, where real wealth values available from statistical data are used, instead of productivity adjusted values.

Theorem 1 cannot be applied directly to the IFS (39), which has unbounded support for $t \to \infty$, however we can refer to the invariant distribution of the conjugate system (7) as the equivalent of the unique invariant distribution of (39) defined on the positive real line.[23] The system converges to this distribution starting from any initial distribution of wealth. Thus, for convenience, we may assume that the distribution at time $t = 0$ concentrates a mass $1 - p$ on some bequest $b_0 \geq 0$ inherited from the past and a mass p on $(b_0 + A_0)$; that is, $\nu_0(W) = (1-p)\,\delta_{b_0}(W) + p\delta_{b_0+A_0}(W)$, where, for any $b \in \mathbb{R}_+$, δ_b denotes the Dirac function:

$$\delta_b(W) = \begin{cases} 1 & \text{if } W = b \\ 0 & \text{otherwise.} \end{cases}$$

We may also write the initial condition for (39) as

$$W_0 = \begin{cases} b_0 & \text{with probability } 1-p \\ b_0 + A_0 & \text{with probability } p. \end{cases} \tag{40}$$

[23]Alternatively, since $0 < \beta < 1$, one may invoke Theorem 7.2 in Lasota (1995) to prove existence and uniqueness of the invariant distribution for IFS (39).

Having an initial distribution concentrating masses over a finite set of points implies that also the distribution of wealths at each date $t > 0$ concentrates masses over finite sets of points. This allows a direct application of formula (38) to the distribution of wealths at each date t. By construction, it is easily seen that, for all $t \geq 0$, the are 2^{t+1} values of wealth $W_t^1, \ldots, W_t^{2^{t+1}}$, each with weight π_t^i, $i = 1, \ldots, 2^{t+1}$. Therefore, the Gini coefficient at time t is given by

$$G_t = \frac{1}{2\mu_t} \sum_{i=1}^{2^{t+1}} \sum_{j=1}^{2^{t+1}} \pi_t^i \pi_t^j \left| W_t^i - W_t^j \right|, \tag{41}$$

where

$$\mu_t = \sum_{i=1}^{2^{t+1}} \pi_t^i W_t^i \tag{42}$$

denotes the mean of the marginal distribution ν_t for all $t \geq 0$, and, in view of (40), we may let $W_0^1 = b_0$, $W_0^2 = (b_0 + A_0)$, $\pi_0^1 = 1 - p$ and $\pi_0^2 = p$.

Since, by independence, for all $t \geq 0$, weights π_t^i have the form

$$\pi_t^i = p^{h_i} (1-p)^{t+1-h_i}, \qquad 0 \leq h_i \leq t+1, \qquad 0 < p < 1,$$

clearly $\lim_{t\to\infty} \pi_t^i \pi_t^j = 0$; in other words, masses p and $1 - p$, initially concentrated on b_0 and $(b_0 + A_0)$, are progressively spread over a set of points that eventually converge to a continuum of points and thus vanish in the limit.

Next result shows that pulverization does not annihilate inequality.

Proposition 4 *The limiting wealth distribution of the model discussed in Section 2.1 has positive Gini coefficient for all feasible values of parameters β, p, γ, b_0 and A_0 such that $\gamma > 2\beta$; specifically,*

$$\lim_{t\to\infty} G_t = \frac{(\gamma - \beta)(1-p)}{\gamma - \left[(1-p)^2 + p^2\right] \beta} > 0.$$

The proof is reported in the Appendix after some preliminary lemmas.

Proposition 4 states that, under the assumptions of Proposition 1, a unit weight progressively spread over (finite) sets of points that exponentially converge to a Cantor like set preserves inequality also in the limit whenever inequality is measured by the limit of the Gini coefficient for the finite marginal distributions. Note that such result holds for a constant (unit) population; clearly, we can conjecture that some stronger result should hold under the assumption of population growth, in which case a similar analysis might be carried out by means of some appropriate polarization index.

7 More General Processes

It is clear from section 3 that the extreme version of polarization/pulverization envisaged by Definition 1 heavily relies on the assumption of having only two states of nature; as a matter of fact, it is crucial in letting the best outcome under the low realization to be worse than the worst outcome under the high realization when the growth rate is large enough. This phenomenon quickly disappears as one allows for more realizations: the more the number of probabilistic realizations, the greater the chances that the range of the corresponding maps in the IFS will overlap. In other words, more 'degrees of success' translates into a IFS with a larger number of maps, which, in turn, would fill the holes left on the support of the marginal distributions by the iterations of only two maps, g_1 and g_2 in figure 4, thus yielding a full support, $X = [0, 1]$, for the invariant distribution. In such circumstances, neither "pulverized limits" or "disconnection" can appear, even if the overlap is only across neighboring pairs of maps (one for each realization) and not across the worst and best outcome.

Thus, all the main points of the model seem to rely on assumptions that are quite peculiar; we need to check economic relevance of our arguments in a more realistic scenario. To test robustness of our approach consider the *perturbed* system obtained by adding some 'noise' ε to the usual IFS (21):

$$y_{t+1} = g_{\varepsilon_t}(y_t) = \alpha y_t + \varepsilon_t, \tag{43}$$

where $\{\varepsilon_t\}_{t=0}^{\infty}$ is a i.i.d. stochastic process such that ε_t has a constant *density* supported on the compact interval $[0, 1-\alpha]$. The autoregressive process (43) extends our model to a completely different setting: from only two states – 'failure' or 'success' – we shift to a continuum of states governing the affine maps of the IFS, all placed between the original maps g_1, g_2, which maintain their position on the boundaries of the interval $[0, 1-\alpha]$, *i.e.*, $g_1(y) = \alpha y + \varepsilon$ when $\varepsilon = 0$ and $g_2(y) = \alpha y + \varepsilon$ for $\varepsilon = 1-\alpha$. In order to keep the basic traits of the economic models – societies that highly rewards success – discussed in the previous sections, we need to assume a *bimodal* density for the random variable ε_t; specifically, a density that concentrates most of the mass around the two endpoints $\varepsilon = 0$ and $\varepsilon = 1-\alpha$ – *i.e.*, on the two 'boundary' maps g_1 and g_2.

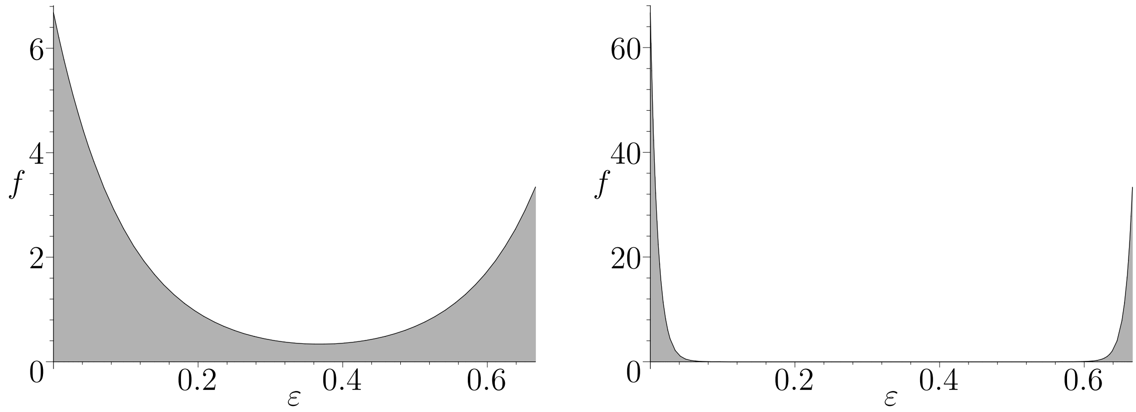
As an example, we may consider the density defined by

$$f(\varepsilon) = \frac{(1-p)\, e^{-\varepsilon/\sigma} + p e^{[\varepsilon-(1-\alpha)]/\sigma}}{\sigma\left[1 - e^{-(1-\alpha)/\sigma}\right]}, \tag{44}$$

where p and α are the same as in the previous sections and parameter σ controls its dispersion around the two boundaries $\varepsilon = 0$ and $\varepsilon = 1-\alpha$: $f(\varepsilon)$ becomes more concentrated around them for smaller values of parameter σ. Figure 8 shows that, for $\alpha = 1/3$ and $p = 1/3$, if $\sigma = 0.01$, $f(\varepsilon)$ is more concentrated on the boundaries than for $\sigma = 0.1$.

We are now in the position to provide at least some heuristic argument supporting our conjecture that – a softer than that of Definition 1, but still meaningful, notion of – polarization/pulverization is not only implied by the (extreme) assumption of having only two realizations, but rather the consequence of a strongly bimodal stylization of luck in a variety of frameworks, regardless of the process being discrete or continuous.

The Foias operator analogous to (14) when the marginal probabilities of the IFS $y_{t+1} = g_{\varepsilon}(y_t)$ are absolutely continuous and when the maps g_{ε} themselves are governed by a density $f(\varepsilon)$ can be



(a) $\alpha = 1/3, p = 1/3$ and $\sigma = 0.1$

(b) $\alpha = 1/3, p = 1/3$ and $\sigma = 0.01$

FIGURE 8: two examples of $f(\varepsilon)$ defined as in (44) for different values of σ.

written as follows (see Appendix B):

$$M_d\nu(y) = \int_0^{1-\alpha} \chi_{[0,1]}\left[g_\varepsilon^{-1}(y)\right] \nu\left[g_\varepsilon^{-1}(y)\right] \frac{\partial}{\partial y} g_\varepsilon^{-1}(y) f(\varepsilon)\, d\varepsilon, \tag{45}$$

where ν is a density on $[0,1]$ and $g_\varepsilon^{-1}(y) \in [0,1]$ denotes the preimage of $y \in [0,1]$ through g_ε for each $\varepsilon \in [0, 1-\alpha]$ and $\chi_A(\cdot)$ is the indicator function for the set A – its role in (45) is to let $\nu(\cdot) \equiv 0$ outside the interval $[0,1]$. It is easily seen that M_d maps densities on $[0,1]$ into densities on $[0,1]$; specifically, $M_d\nu(y)$ is the density associated to each point $y \in [0,1]$ after one iteration of the IFS starting from a density ν on $[0,1]$.

By invoking Theorem 1.1 in Diaconis and Freedman (1999) (see also Section 6.1, p. 64, in the same paper), it can be shown that the sequence of marginal densities $\nu_t = M_d\left(M_d^{t-1}\nu_{t-1}\right) = M_d^t\nu_0$ converges weakly to a unique invariant density ν^* – such that $\nu^* = M_d\nu^*$ – starting from any density ν_0 on $[0,1]$, provided that all g_ε are Lipschitz with Lipschitz constants K_ε satisfying the following "average contraction" condition:

$$\int_0^{1-\alpha} \ln K_\varepsilon f(\varepsilon)\, d\varepsilon < 0. \tag{46}$$

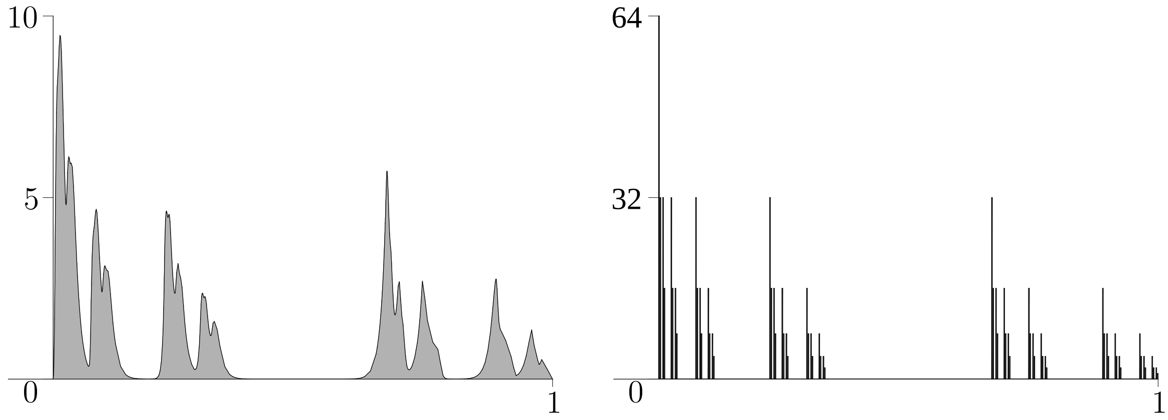
In other words, Theorem 1.1 in Diaconis and Freedman (1999) generalizes Theorem 1 reported in Section 3.1 to IFS constituted by infinitely many maps (see also the references reported there).

Since $K_\varepsilon \equiv \alpha < 1$ for all $\varepsilon \in [0, 1-\alpha]$, property (46) certainly holds for the IFS $y_{t+1} = \alpha y_t + \varepsilon$ defined in (43), which thus has a unique invariant density ν^*. By using the change of variable formula (see Appendix B), (45) becomes

$$M_d\nu(y) = \int_{\max\{1-(1-y)/\alpha,0\}}^{\min\{y/\alpha,1\}} \nu(x) f(y - \alpha x)\, dx, \tag{47}$$

which can be approximated by numerical methods.

Figure 9(a) approximates the first six iterations of Foias operator M_d as defined in (47) for $\alpha = 1/3$



(a) M_d^6 as in (47) for $\alpha = 1/3, p = 1/3, \sigma = 0.01$ (b) M^6 as in (14) for $\alpha = 1/3, p = 1/3$

FIGURE 9: first six iterations of Foias operator starting from the uniform density; a) IFS with a continuum of maps with density $f(\varepsilon)$ defined in (44), b) IFS with only two maps g_1, g_2 of the type (21).

and $f(\varepsilon)$ as in (44) with $p = 1/3$ and $\sigma = 0.01$ starting from the uniform density. This is achieved by numerical integration[24] over a partition of 500 subintervals of $[0, 1]$. Recall that Foias operator converges at a geometric rate, therefore figure 9(a) provides a reliable picture of what the invariant density ν^* might look like. Even if it is a density, it clearly exhibits a pattern very similar to the distribution in figure 9(b), which is the same as figure 5(f), where the first six iteration of the Foias operator in the case of the IFS with only two maps – g_1 with probability $1 - p$ and g_2 with probability p – is plotted. Not only a lack of the middle class, but also the replication of the same phenomenon at smaller scale in each cluster of wealth after each iteration appear. Clearly, in figure 9(a) peaks are shorter (below 10) than those in figure 9(b) (up to 64); also, self similarity on smaller scale tends to blur in figure 9(a), due to the smoothing of the density f around the two 'boundary' maps corresponding to the former g_1 and g_2 after each iteration. At any rate, however, the distributions portrayed in figures 9(a) and 9(b) respectively exhibit very close qualitative traits, at least in terms of – a broader meaning of – polarization/pulverization.

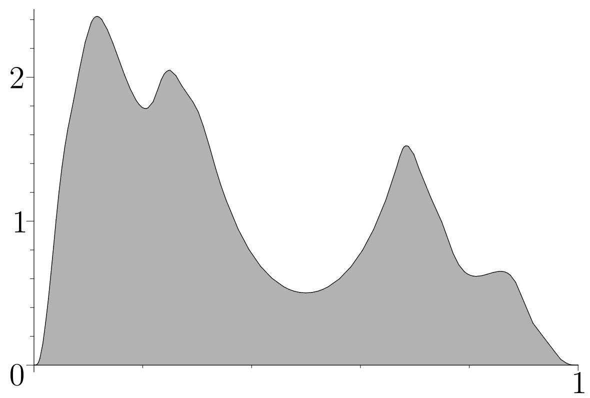


FIGURE 10: first six iterations of M_d as in (47) starting from the uniform density in case of a IFS with a continuum of maps with density $f(\varepsilon)$ as in (44) for $\alpha = 1/3$, $p = 1/3$, $\sigma = 0.1$.

[24]The Maple code that generates plots like in figure 9(a) and figure 10 is available from the authors upon request.

Figure 10 shows the same first six iterations of M_d starting from the uniform density as in figure 9(a) but with a density $f(\varepsilon)$ more dispersed around the boundaries $\varepsilon = 0$ and $\varepsilon = 1-\alpha$, characterized by $\sigma = 0.1$ [see figure 8(a)]. It is remarkable that also when 'success' is more evenly distributed, less weight on intermediate degrees of success still translates into some degree of wealth polarization due to a smaller middle class – corresponding to the large hollow in the middle of the graph – compared to the poor and the rich. Our conjecture is that a more general notion of wealth inequality, determined by a smaller size for the middle class with respect to the poor and the rich, is a direct consequence of assuming a bimodal distribution of success. This will be the topic of future research.

8 Concluding Remarks

In this paper we have pointed out how wealth polarization/pulverization is not to be contrasted with equal opportunities characterizing economies with a high degree of social mobility, but instead it can be exactly the effect of a large amplitude of mobility itself. What really matters for polarization/pulverization is the reward from being successful, which is increasing in the size of the technological jump. Private investment in the human capital necessary to adopt an exogenous innovation stream can be one cause; private investment in research aimed at improving everybody's productivity can be another cause. Despite the differences between these two engines of growth, both induce the disappearance of the middle class due to the fractal properties of the support of the invariant wealth distribution, provided that the growth rate of the economy is higher than a common threshold.

We have shown that in this framework polarization and pulverization cannot be eliminated by fiscal measures such as wealth redistribution through taxation of the successful people with tax revenues lump-sum redistributed to the unsuccessful ones, while wealth taxation can even create polarization. Some more sophisticated device is required. A random taxation scheme may be able to reintroduce an artificial middle class, but unlikely gives it enough strength, especially if the incentive compatibility constraint is tight.

Hence, there seems to be a general lesson one can learn from the direct relationship between high growth rates and inequality emerged by applying the IFS approach to wealth dynamics in a society characterized by equal opportunities and fast social mobility: the goal of containing inequality may be better achieved through policies aimed at tackling the growth rate itself – *e.g.*, by means of monetary policies devised to "cool down" the economy – rather than resorting on redistributive devices. Our proof of this new effect of growth on wealth distribution suggests future works in which fiscal policy is endogenized and polarization/pulverization is incorporated in people's preferences, linking social mobility to the demand for redistribution.[25]

In view of recent works on optimal growth theory (see, *e.g.*, Mitra and Privileggi, 2004, 2006, 2009), further investigation on wealth inequality may be pursued by means of models characterized by an infinitely lived representative agent, as well as models whose wealth dynamics can be described by non-linear IFS – note that the second part of our Definition 1 is readily applicable to such cases.

[25] See, *e.g.*, Alesina and Glaeser (2005), Alesina and Giuliano (2009) and Alesina, Cozzi, and Mantovan (2009).

Also IFS with state-dependent probabilities might be worth considering, as they can introduce a "damping effect" on social mobility – for example through a higher probability for both the poor and the rich to remain in the same wealth cluster and a lower probability to switch from one class to the other – which may seem closer to reality. For example, the poor might find educational costs unbearable or access to credit market precluded, thus indirectly reducing their probability of success, while for the rich an easier access to education and credit markets improves their probability of being rich also in the future. These observations suggest that models on wealth inequality from the traditional stream of research, like the ones in Galor and Zeira (1993) or in Aghion and Bolton (1997) (see also the whole literature cited in the introduction), which assume imperfect capital markets, may easily fit our framework with the necessary modifications.

In so far as people of similar wealth levels tend to live together, gaps in wealth levels imply gaps in location, and therefore geographic segregation. The residential segregation associated with wealth polarization implies that the "city" partitions itself into a (polarized) fractal as a result of fast growth. Adding state-dependence would be natural, as residential segregation entails educational segregation: our results may then be extended and would contribute to the literature on segregation, such as Benabou (1993, 1996a), and Sethi and Somanathan (2004).

Appendix

A Gini Coefficient and Cantor-like sets

This appendix is devoted to the proof of Proposition 4 in Section 6. Since both wealths W_t^i and weights π_t^i have a recursive formulation generated by dynamic (39), it is convenient to write formula (41) in a form more suitable for direct handling.

Lemma 2 *For each $t \geq 0$, label the set of wealths so that they are ordered: $W_t^1 < W_t^2 < \cdots < W_t^{2^{t+1}}$. Then formula (41) can be rewritten as follows:*

$$G_t = \frac{1}{\mu_t} \sum_{j=1}^{2^{t+1}-1} \sum_{i=1+j}^{2^{t+1}} \pi_t^i \pi_t^j \left(W_t^i - W_t^j \right), \tag{48}$$

where μ_t is given by (42).

Proof. If the initial condition for system (39) is given by (40), in the sum (41) there are $\left(2^{t+1} - 1\right) 2^t$ non-zero addends of the form $\left| W_t^i - W_t^j \right|$, with $i \neq j$, and each of them is counted twice. By summing up all ordered non-zero differences $W_t^i - W_t^j$, with $W_t^i > W_t^j$, we get

$$G_t = \frac{1}{2\mu_t} \left[2 \sum_{j=1}^{2^{t+1}-1} \sum_{i=1+j}^{2^{t+1}} \pi_t^i \pi_t^j \left(W_t^i - W_t^j \right) \right],$$

which is (48). ■

It is convenient to label the sum on the RHS of (48) by D_t, so we can use the shorthand

$$G_t = \frac{D_t}{\mu_t}.$$

The next three lemmas provide a recursive formulation for both the mean μ_t and the sum D_t which allow to compute G_t directly in terms of parameters and initial conditions.

Lemma 3 *The mean μ_t has the following recursive formulation:*

$$\mu_{t+1} = \beta\mu_t + pA_{t+1}, \tag{49}$$

thus,

$$\mu_t = \beta^t b_0 + \frac{1-(\beta/\gamma)^{t+1}}{\gamma-\beta}\gamma^{t+1}pA_0. \tag{50}$$

Proof. The construction of the t^{th} marginal distribution ν_t through system (39) implies that each point W_t^i with associated weight π_t^i at time t is being split into two wealth values $W_{t+1}^{i_L} = \beta W_t^i$ and $W_{t+1}^{i_U} = \beta W_t^i + A_{t+1}$ with weights $\pi_{t+1}^{i_L} = (1-p)\,\pi_t^i$ and $\pi_{t+1}^{i_U} = p\pi_t^i$ respectively at time $t+1$, for $i = 1, \ldots, 2^{t+1}$. Therefore, all 2^{t+2} terms in the sum defining μ_{t+1} as in (42) can be grouped into 2^{t+1} pairs, each of them generated by a single term in the sum defining μ_t; thus all such pairs can be written as functions of W_t^i and π_t^i as follows:

$$\begin{aligned}
\mu_{t+1} &= \sum_{i=1}^{2^{t+2}} \pi_{t+1}^i W_{t+1}^i = \sum_{i_L=1}^{2^{t+1}}\sum_{i_U=1}^{2^{t+1}} \left(\pi_{t+1}^{i_L} W_{t+1}^{i_L} + \pi_{t+1}^{i_U} W_{t+1}^{i_U}\right) = \sum_{i_L=1}^{2^{t+1}} \pi_{t+1}^{i_L} W_{t+1}^{i_L} + \sum_{i_U=1}^{2^{t+1}} \pi_{t+1}^{i_U} W_{t+1}^{i_U} \\
&= \sum_{i=1}^{2^{t+1}} (1-p)\,\pi_t^i \beta W_t^i + \sum_{i=1}^{2^{t+1}} p\pi_t^i \left(\beta W_t^i + A_{t+1}\right) = \sum_{i=1}^{2^{t+1}} \left[(1-p)\,\pi_t^i \beta W_t^i + p\pi_t^i \left(\beta W_t^i + A_{t+1}\right)\right] \\
&= \sum_{i=1}^{2^{t+1}} \pi_t^i \beta W_t^i + pA_{t+1} \sum_{i=1}^{2^{t+1}} \pi_t^i = \beta\mu_t + pA_{t+1},
\end{aligned}$$

where in the second and third equalities we have indexed by i_L terms of the type $\pi_{t+1}^{i_L} W_{t+1}^{i_L} = (1-p)\,\pi_t^i \beta W_t^i$ (corresponding to the lower branch of a term $\pi_t^i W_t^i$ in t) and by i_U terms of the type $\pi_{t+1}^{i_U} W_{t+1}^{i_U} = p\pi_t^i\left(\beta W_t^i + A_{t+1}\right)$ (corresponding to the upper branch of a term $\pi_t^i W_t^i$ in t), while in the last equality $\sum_{i=1}^{2^{t+1}} \pi_t^i = 1$ holds, as population is normalized to 1. Hence, (49) is established, and, as $\mu_0 = b_0 + pA_0$, (50) follows accordingly. ■

Before giving a recursive formula for D_t, we need the following lemma which states that, under the assumption that a middle class disappears after one iteration of (39) as prescribed by Proposition 1, the poorest individual at time t which is successful at time $t+1$ becomes richer than the richest individual at time t which is not successful at time $t+1$. Recall that, under the (ordered) labeling as in Lemma 2, W_t^1 and $W_t^{2^{t+1}}$ denote the smallest and the largest wealth at time t respectively.

Lemma 4 *Let $W_{t+1}^{1_U} = \beta W_t^1 + A_{t+1}$ denote the wealth of the individual which is the poorest at time t but becomes successful at time $t+1$ and $W_{t+1}^{2_L^{t+1}} = \beta W_t^{2^{t+1}}$ denote the wealth of the individual which is the richest at time t but becomes unsuccessful at time $t+1$. Then, if $\gamma > 2\beta$, $W_{t+1}^{1_U} > W_{t+1}^{2_L^{t+2}}$ for all $t \geq 0$.*

Proof. It is easily seen that $W_t^1 = \beta^t b_0$ and $W_t^{2^{t+1}} = \beta^t b_0 + (\gamma - \beta)^{-1} \left[1 - (\beta/\gamma)^{t+1}\right] \gamma^{t+1} A_0$. Hence,

$$\begin{aligned} W_{t+1}^{1_U} &= \beta^{t+1} b_0 + \gamma^{t+1} A_0 \\ &> \beta^{t+1} b_0 + \beta \frac{1 - (\beta/\gamma)^{t+1}}{\gamma - \beta} \gamma^{t+1} A_0 = \beta \left[\beta^t b_0 + \frac{1 - (\beta/\gamma)^{t+1}}{\gamma - \beta} \gamma^{t+1} A_0\right] = W_{t+1}^{2_L^{t+2}}, \end{aligned}$$

where the inequality follows from $\gamma > 2\beta$. ■

An immediate consequence of Lemma 4 is the following Corollary.

Corollary 1 *Under the assumption $\gamma > 2\beta$, if $W_t^i > W_t^j$, then $W_{t+1}^{j_U} = \beta W_t^j + A_{t+1}$ and $W_{t+1}^{i_L} = \beta W_t^i$ are such that $W_{t+1}^{j_U} > W_{t+1}^{i_L}$.*

Lemma 5 *Under the assumption $\gamma > 2\beta$ the sum*

$$D_t = \sum_{j=1}^{2^{t+1}-1} \sum_{i=1+j}^{2^{t+1}} \pi_t^i \pi_t^j \left(W_t^i - W_t^j\right) \tag{51}$$

in (48) has the following recursive formulation:

$$D_{t+1} = \left[(1-p)^2 + p^2\right] \beta D_t + p(1-p) A_{t+1}, \tag{52}$$

thus,

$$D_t = \frac{1 - (s/\gamma)^{t+1}}{\gamma - s} \gamma^{t+1} p(1-p) A_0, \tag{53}$$

where $s = \left[(1-p)^2 + p^2\right] \beta$.

Proof. We follow an argument parallel to that in the proof of Lemma 3. For $j = 1, \ldots, 2^{t+1} - 1$ and $i = 1 + j, \ldots, 2^{t+1}$, each addend $\pi_t^i \pi_t^j \left(W_t^i - W_t^j\right)$ in (51) at time t contains two wealth values, W_t^i and W_t^j, such that $W_t^i > W_t^j$, with associated weights π_t^i and π_t^j respectively. The construction of the t^{th} marginal distribution ν_t through system (39) implies that both such terms are being split into two wealth values at time $t+1$, for a total of four terms, that we can label as follows:

$$\begin{array}{ll} W_{t+1}^{j_L} = \beta W_t^j & \text{with weight } \pi_{t+1}^{j_L} = (1-p)\,\pi_t^j, \\ W_{t+1}^{j_U} = \beta W_t^j + A_{t+1} & \text{with weight } \pi_{t+1}^{j_U} = p\pi_t^j, \\ W_{t+1}^{i_L} = \beta W_t^i & \text{with weight } \pi_{t+1}^{i_L} = (1-p)\,\pi_t^i, \\ W_{t+1}^{i_U} = \beta W_t^i + A_{t+1} & \text{with weight } \pi_{t+1}^{i_U} = p\pi_t^i. \end{array}$$

Hence, each addend $\pi_t^i \pi_t^j \left(W_t^i - W_t^j\right)$ in D_t at time t corresponds to the following $(2^2 - 1)\, 2 = 6$ positive addends in D_{t+1} at time $t+1$:

$$
\begin{aligned}
&\pi_{t+1}^{i_L} \pi_{t+1}^{j_L} \left(W_{t+1}^{i_L} - W_{t+1}^{j_L}\right) = (1-p)^2 \pi_t^i \pi_t^j \beta \left(W_t^i - W_t^j\right) \\
&\pi_{t+1}^{j_U} \pi_{t+1}^{j_L} \left(W_{t+1}^{j_U} - W_{t+1}^{j_L}\right) = (1-p)\, p \left(\pi_t^j\right)^2 A_{t+1} \\
&\pi_{t+1}^{i_U} \pi_{t+1}^{j_L} \left(W_{t+1}^{i_U} - W_{t+1}^{j_L}\right) = (1-p)\, p \pi_t^i \pi_t^j \left[\beta \left(W_t^i - W_t^j\right) + A_{t+1}\right] \\
&\pi_{t+1}^{j_U} \pi_{t+1}^{i_L} \left(W_{t+1}^{j_U} - W_{t+1}^{i_L}\right) = (1-p)\, p \pi_t^i \pi_t^j \left[\beta \left(W_t^j - W_t^i\right) + A_{t+1}\right] \\
&\pi_{t+1}^{i_U} \pi_{t+1}^{i_L} \left(W_{t+1}^{i_U} - W_{t+1}^{i_L}\right) = (1-p)\, p \left(\pi_t^i\right)^2 A_{t+1} \\
&\pi_{t+1}^{i_U} \pi_{t+1}^{j_U} \left(W_{t+1}^{i_U} - W_{t+1}^{j_U}\right) = p^2 \pi_t^i \pi_t^j \beta \left(W_t^i - W_t^j\right),
\end{aligned}
\tag{54}
$$

each of them defined as functions of W_t^i, W_t^j, π_t^i and π_t^j. Note that all such terms are positive provided that $\gamma > 2\beta$, which, by Corollary 1, guarantees that also the fourth term (on the LHS of the equation) is positive.

Therefore, all $(2^{t+2} - 1)\, 2^{t+1}$ terms in the sum defining D_{t+1} as in (51) can be gathered into $(2^{t+1} - 1)\, 2^t$ groups of six addends, with each group generated by a single term in the sum defining D_t, as follows:

$$
\begin{aligned}
D_{t+1} &= \sum_{j=1}^{2^{t+2}-1} \sum_{i=1+j}^{2^{t+2}} \pi_{t+1}^i \pi_{t+1}^j \left(W_{t+1}^i - W_{t+1}^j\right) \\
&= \sum_{j_L=1}^{2^{t+1}-1} \sum_{i_L=1+j_L}^{2^{t+1}} \pi_{t+1}^{i_L} \pi_{t+1}^{j_L} \left(W_{t+1}^{i_L} - W_{t+1}^{j_L}\right) + \sum_{j_L=1}^{2^{t+1}-1} \sum_{j_U=1+j_L}^{2^{t+1}} \pi_{t+1}^{j_U} \pi_{t+1}^{j_L} \left(W_{t+1}^{j_U} - W_{t+1}^{j_L}\right) \\
&\quad + \sum_{j_L=1}^{2^{t+1}-1} \sum_{i_U=1+j_L}^{2^{t+1}} \pi_{t+1}^{i_U} \pi_{t+1}^{j_L} \left(W_{t+1}^{i_U} - W_{t+1}^{j_L}\right) + \sum_{i_L=1}^{2^{t+1}-1} \sum_{j_U=1+i_L}^{2^{t+1}} \pi_{t+1}^{j_U} \pi_{t+1}^{i_L} \left(W_{t+1}^{j_U} - W_{t+1}^{i_L}\right) \\
&\quad + \sum_{i_L=1}^{2^{t+1}-1} \sum_{i_U=1+i_L}^{2^{t+1}} \pi_{t+1}^{i_U} \pi_{t+1}^{i_L} \left(W_{t+1}^{i_U} - W_{t+1}^{i_L}\right) + \sum_{j_U=1}^{2^{t+1}-1} \sum_{i_U=1+j_U}^{2^{t+1}} \pi_{t+1}^{i_U} \pi_{t+1}^{j_U} \left(W_{t+1}^{i_U} - W_{t+1}^{j_U}\right) \\
&= \left[(1-p)^2 + p^2\right] \beta \sum_{j=1}^{2^{t+1}-1} \sum_{i=1+j}^{2^{t+1}} \pi_t^i \pi_t^j \left(W_t^i - W_t^j\right) \\
&\quad + p\,(1-p)\, A_{t+1} \left[\sum_{k=1}^{2^{t+1}} \left(\pi_t^k\right)^2 + 2 \sum_{j=1}^{2^{t+1}-1} \sum_{i=1+j}^{2^{t+1}} \pi_t^i \pi_t^j\right] \\
&= \left[(1-p)^2 + p^2\right] \beta D_t + p\,(1-p)\, A_{t+1},
\end{aligned}
$$

where in the second to the sixth lines we have substituted terms as in (54) and simplified terms, while the last line holds since π_t^ks add up to 1 and

$$
\sum_{k=1}^{2^{t+1}} \left(\pi_t^k\right)^2 + 2 \sum_{j=1}^{2^{t+1}-1} \sum_{i=1+j}^{2^{t+1}} \pi_t^i \pi_t^j = \left(\sum_{k=1}^{2^{t+1}} \pi_t^k\right)^2 = 1.
$$

Hence, (52) is established, and, since $G_0 = p\,(1-p)\, A_0$, (53) follows accordingly. ■

Proof of Proposition 4. By Lemmas 2 – 5,

$$\lim_{t\to\infty} G_t = \lim_{t\to\infty} \frac{D_t}{\mu_t} = \lim_{t\to\infty} \frac{\dfrac{1-(s/\gamma)^{t+1}}{\gamma - s}\gamma^{t+1}p\,(1-p)\,A_0}{\beta^t b_0 + \dfrac{1-(\beta/\gamma)^{t+1}}{\gamma-\beta}\gamma^{t+1}pA_0} = \frac{(\gamma-\beta)\,(1-p)}{\gamma - \left[(1-p)^2+p^2\right]\beta},$$

where $s = \left[(1-p)^2 + p^2\right]\beta$, and the proof is complete. ■

B The Foias Operator for Densities

We first construct formula (45) for the Foias operator when the IFS is of the kind (43), $g_\varepsilon(y) = \alpha y + \varepsilon$, that is, it has a continuum of maps each chosen by means of a density $f(\varepsilon)$ on the interval interval $[0, 1-\alpha]$.

If X and Y denote two random variables with densities ν and $M_d\nu$ on $[0,1]$ respectively, then:

$$\Pr(Y \in B) = \int_0^{1-\alpha} \Pr\left[X \in g_\varepsilon^{-1}(B)\right] f(\varepsilon)\, d\varepsilon.$$

For $B = [0, y]$ this is equivalent to

$$\int_0^y M_d\nu(u)\, du = \int_0^{1-\alpha} \chi_{[0,1]}\left[g_\varepsilon^{-1}(y)\right] \int_{g_\varepsilon^{-1}(0)}^{g_\varepsilon^{-1}(y)} \nu(u)\, f(\varepsilon)\, du d\varepsilon,$$

which, since $M_d\nu(y) = (\partial/\partial y) \int_0^y M_d\nu(u)\, du$, leads to

$$\begin{aligned} M_d\nu(y) &= \frac{\partial}{\partial y} \int_0^{1-\alpha} \chi_{[0,1]}\left[g_\varepsilon^{-1}(y)\right] \int_{g_\varepsilon^{-1}(0)}^{g_\varepsilon^{-1}(y)} \nu(u)\, f(\varepsilon)\, du d\varepsilon \\ &= \int_0^{1-\alpha} \chi_{[0,1]}\left[g_\varepsilon^{-1}(y)\right] \left[\frac{\partial}{\partial y} \int_{g_\varepsilon^{-1}(0)}^{g_\varepsilon^{-1}(y)} \nu(u)\, du\right] f(\varepsilon)\, d\varepsilon, \end{aligned}$$

which is (45).

Noting that $(\partial/\partial y)\, g_\varepsilon^{-1}(y) \equiv 1/\alpha$ for all $\varepsilon \in [0, 1-\alpha]$ and by using the change of variable $x = g_\varepsilon^{-1}(y) = (y-\varepsilon)/\alpha$, which is a strictly decreasing transformation of variable ε, (45) can easily be transformed into (47):

$$\begin{aligned}
M_d\nu\left(y\right) &= \frac{1}{\alpha}\int_0^{1-\alpha}\chi_{[0,1]}\left[g_\varepsilon^{-1}\left(y\right)\right]\nu\left[g_\varepsilon^{-1}\left(y\right)\right]f\left(\varepsilon\right)d\varepsilon \\
&= \frac{1}{\alpha}\int_{1-(1-y)/\alpha}^{y/\alpha}\chi_{[0,1]}\left(x\right)\nu\left(x\right)f\left(y-\alpha x\right)\alpha dx \\
&= \int_{1-(1-y)/\alpha}^{y/\alpha}\chi_{[0,1]}\left(x\right)\nu\left(x\right)f\left(y-\alpha x\right)dx \\
&= \int_{\max\{1-(1-y)/\alpha,0\}}^{\min\{y/\alpha,1\}}\nu\left(x\right)f\left(y-\alpha x\right)dx,
\end{aligned}$$

where in the last equality we translated the bounds given by the indicator $\chi_{[0,1]}\left(\cdot\right)$ into the limits of integration.

References

[1] Acemoglu, D., 1999. Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *Quarterly Journal of Economics* **113**: 1055-1090.

[2] Acemoglu, D., 2002. Technical Change, Inequality and the Labor Market. *Journal of Economic Literature* **40**: 7-72.

[3] Acemoglu, D., 2007. Equilibrium Bias of Technology. *Econometrica*, **75**: 1371-1410.

[4] Acemoglu, D., 2008. *Introduction to Modern Economic Growth*. Princeton University Press, Princeton, NJ.

[5] Aghion, P., Bolton, P., 1997. A Theory of Trickle-Down Growth and Development. *Review of Economics Studies* **64**: 151-172.

[6] Aghion, P., Howitt, P., 1998. *Endogenous Growth Theory*. The MIT Press, Cambridge, MA.

[7] Aghion, P., Caroli, E., Garcia-Penalosa, C., 1999. Inequality and Economic Growth: The Perspective of the New Growth Theories. *Journal of Economic Literature* **37**: 1615-1660.

[8] Alesina, A., Angeletos, G. M., 2005a. Fairness and Redistribution: US vs. Europe. *American Economic Review* **95**: 960-978.

[9] Alesina, A., Angeletos, G. M., 2005b. Redistribution, Corruption and Fairness. *Journal of Monetary Economics* **52**: 1227-1244.

[10] Alesina, A., Cozzi, G., Mantovan, N., 2009. *The Evolution of Ideology, Fairness, and Redistribution*. Unpublished Manuscript.

[11] Alesina, A., Giuliano, P., 2009. Preferences for Redistribution. *NBER Working Paper Series* **14825**.

[12] Alesina, A., Glaeser, E., 2004. *Fighting Poverty in the U.S. and Europe: A World of Differences*. Oxford University Press, New York.

[13] Alesina, A., La Ferrara, E., 2005. Preferences for Redistribution in the Land of Opportunities. *Journal of Public Economics* **89**: 897-931.

[14] Alesina, A., Rodrik, D., 1994. Distributive Politics and Economic Growth. *Quarterly Journal of Economics* **109**: 465-490.

[15] Andreoni, J., 1989. Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy* **6**: 1447-1458.

[16] Arnold L., 1998. *Random Dynamical Systems*. Springer-Verlag, Berlin.

[17] Autor, D., Katz, F., Kearney, M., 2006. The Polarization of the US Labour Market. *AEA Papers and Proceedings* **96**: 189-194.

[18] Banerjee A., Newmann, A., 1993. Occupational Choice and the Process of Development. *Journal of Political Economy* **101**: 274-298.

[19] Barnsley, M. F., Demko, S., 1985. Iterated Function Systems and the Global Construction of Fractals. *Proceeding Royal Society London Series* **A 399**: 243-275.

[20] Barro, R., 1999. Inequality, Growth, and Investment. *NBER Working Paper* **7038**.

[21] Barro, R., 2000. Inequality and Growth in a Panel of Countries. *Journal of Economic Growth* **5**: 5-30.

[22] Benabou, R., 1993. Workings of a City. *Quarterly Journal of Economics* **108**: 619-52.

[23] Benabou, R., 1996a. Equity and Efficiency in Human Capital Investment: The Local Connection. *Review of Economic Studies* **63**: 237-64.

[24] Benabou, R., 1996b. Inequality and Growth. *NBER Macroeconomics Annual*. The MIT Press, Cambridge, MA.

[25] Benabou, R., 2000. Unequal Societies: Income Distribution and the Social Contract. *American Economic Review* **90**: 96-129.

[26] Benhabib, J., Rustichini, A., 1996. Social Conflict and Growth. *Journal of Economic Growth* **1**: 125-142.

[27] Benhabib, J., Spiegel, M., 1997. Cross-Country Growth Regressions. *C.V. Starr Center Research Report* **9720**.

[28] Bhattacharya, R., Majumdar, M., 1999a. Convergence to Equilibrium of Random Dynamical Systems Generated by IID Monotone Maps, With Applications to Economics, in: Ghosh, S. (Ed.), *Asymptotics, Nonparametrics, and Time Series*. Marcel Dekker Inc., New York, pp. 713-741.

[29] Bhattacharya, R., Majumdar, M., 1999b. On a Theorem of Dubins and Freedman. *Journal of Theoretical Probability* **12**: 1067-1087.

[30] Bhattacharya, R., Majumdar, M., 2001. On a Class of Stable Random Dynamical Systems: Theory and Applications. *Journal of Economic Theory* **96**: 208-229.

[31] Bhattacharya, R., Majumdar, M., 2007. *Random Dynamical Systems: Theory and Applications*. Cambridge University Press, New York.

[32] Brock, W. A., J. Mirman, L., 1972. Optimal Economic Growth and Uncertainty: the Discounted Case, *Journal of Economic Theory* **4**: 479-513.

[33] Bucci, A., Kunze, H. A., La Torre, D., 2008. Parameter Identification, Population and Economic Growth in an Extended Lucas and Uzawa-type Two Sector Model. *Università degli Studi di Milano, Dipartimento di Scienze Economiche Aziendali e Statistiche Working Paper Series* **2008-34** (http://www.economia.unimi.it/typo3conf/ext/wp_collection/res/download.php?file=Deas2008_34wp.pdf).

[34] Card, D., Di Nardo, J., 2002. Skill-Based Technological Change And Rising Wage Inequality: Some Problems and Puzzles. *Journal of Labor Economics*, **20**: 733-783.

[35] Cozzi, G., Privileggi, F., 2002. Wealth Polarization and Pulverization in Fractal Societies. *ICER Working Paper* **39** (http://www.icer.it/docs/wp2002/privileggi39-02.pdf).

[36] Crownover, R. M., 1995. *Introduction to Fractals and Chaos*. Jones & Bartlett Publishers, Boston.

[37] Cunha, F., Heckman, J., 2006. Investing in our Young People. Manuscript, Department of Economics, University of Chicago (http://www-news.uchicago.edu/releases/06/061115.education.pdf).

[38] Cunha, F., Heckman, J., 2007. The Technology of Skill Formation. *American Economic Review* **97**: 31-47.

[39] Cunha, F., Heckman, J., 2009. The Economics and Psychology of Inequality and Human Development. *NBER Working Paper* **14695**.

[40] D'Ambrosio, C., Wolff, E., 2008. The Distribution of Wealth in the United States from 1983 to 2004: Inequality and Polarization. Paper presented at the Workshop *Income Polarization: Measurement, Determinants and Implications*, May 26-28 2008, Israel (http://www.biu.ac.il/soc/ec/conference/papers/D%27Ambrosio%20Wolff.pdf).

[41] Drew-Becker, I., Gordon, R. J., 2007. Unresolved Issues in the Rise of American Inequality. *Brookings Papers on Economic Activity* **38**.

[42] Diaconis, P., Freedman, D., 1999. Iterated Random Functions. *SIAM Review* **41**: 45-76.

[43] Edgar, G. A., 1990. *Measure, Topology and Fractal Geometry*. Springer-Verlag, New York.

[44] Esteban, J. M., D. Ray, 1994. On the Measurement of Polarization. *Econometrica* **62**: 819-851.

[45] Falconer, K. J., 1997. *Techniques in Fractal Geometry*. John Wiley & Sons, New York.

[46] Falconer, K. J., 2003. *Fractal Geometry*. John Wiley & Sons, Chichester.

[47] Forbes K. J., 2000. A Reassessment of the Relationship Between Inequality and Growth. *American Economic Review* **90**: 869-887.

[48] Galor, O., Zeira, J., 1993. Income Distribution and Macroeconomics. *Review of Economic Studies* **60**: 35-52.

[49] Heckman, J., 2008. Schools, Skills, and Synapses. *Economic Inquiry* **46**, 289-324.

[50] Helpman, E. (Ed.), 1998. *General Purpose Technologies and Economic Growth*. The MIT-Press, Cambridge, MA.

[51] Howitt, P., 1999. Steady Endogenous Growth with Population and R&D Inputs Growing. *Journal of Political Economy* **107**: 715-730.

[52] Hutchinson, J., 1981. Fractals and Self-similarity. *Indiana University Mathematics Journal* **30**: 713-747.

[53] Jones, C., 1995. R&D-Based Models of Economic Growth. *Journal of Political Economy* **103**: 759-784.

[54] Kremer, M., 1998. Patent Buy-outs: a Mechanism for Encouraging Innovation. *Quarterly Journal of Economics* **113**: 1137-1167.

[55] Kunze, H. A., La Torre, D., Vrscay, E. R., 2007a. Contractive Multifunctions, Fixed Point Inclusions and Iterated Multifunction Systems. *Journal of Mathematical Analysis and Applications* **330**: 159-173.

[56] Kunze, H. A., La Torre, D., Vrscay, E. R., 2007b. Random Fixed Point Equations and Inverse Problems Using “Collage Method” for Contraction Mappings. *Journal of Mathematical Analysis and Applications* **334**: 1116-1129.

[57] Kunze, H. A., La Torre, D., Vrscay, E. R., 2009. Inverse Problems for Random Differential Equations Using the Collage Method for Random Contraction Mappings. *Journal of Computational and Applied Mathematics* **223**: 853-861.

[58] La Torre, D., Mendivil, F., 2008. Iterated Function Systems on Multifunctions and Inverse Problems. *Journal of Mathematical Analysis and Applications* **340**: 1469-1479.

[59] Lasota, A., Mackey, M. C., 1994. *Chaos, Fractals, and Noise, Stochastic Aspects of Dynamics*. Springer-Verlag, New York.

[60] Lasota, A., 1995. From Fractals to Stochastic Differential Equations, in: Garbaczewski P., Wolf, M., Weron, A. (Eds.), *Chaos - The Interplay Between Stochastic and Deterministic Behaviour, Proceedings of the XXXIst Winter School of Theoretical Physics*, Springer-Verlag, Berlin Heidelberg, pp. 235-255.

[61] Loury, G., 1981. Intergenerational Transfers and the Distribution of Earnings. *Econometrica* **49**: 843-867.

[62] Lundberg, M., Squire, L., 2003. The simultaneous evolution of growth and inequality. *Economic Journal* **113**, 326-344.

[63] Mitra, T., Montrucchio, L., Privileggi, F., 2004. The Nature of the Steady State in Models of Optimal Growth Under Uncertainty. *Economic Theory* **23**: 39-71.

[64] Mitra, T., Privileggi, F., 2004. Cantor Type Invariant Distributions in the Theory of Optimal Growth Under Uncertainty. *Journal of Difference Equations and Applications* **10**: 489-500.

[65] Mitra, T., Privileggi, F., 2006. Cantor Type Attractors in Stochastic Growth Models. *Chaos, Solitons and Fractals* **29**: 626-637.

[66] Mitra, T., Privileggi, F., 2009. On Lipschitz Continuity of the Iterated Function System in a Stochastic Optimal Growth Model. *Journal of Mathematical Economics* **45**: 185-198.

[67] Montrucchio, L., Privileggi, F., 1999. Fractal Steady States in Stochastic Optimal Control Models. *Annals of Operations Research* **88**: 183-197.

[68] Mookherjee, D., Ray, D., 2002. Is Equality Stable?. *American Economic Review* **92**, 253-259.

[69] Mookherjee, D., Ray, D., 2003. Persistent Inequality. *Review of Economic Studies* **70**, 369-393.

[70] Mookherjee, D., Ray, D., 2005. Occupational Diversity and Endogenous Inequality. *Boston University - Department of Economics - Working Papers Series* **WP2005-022**.

[71] Perloff, J., Wu, X., 2005. China's Income Distribution, 1985-2001. *Institute for Research on Labor and Employment Working Paper Series iirwps* **117-05**, University of California, Berkeley (http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1115&context=iir).

[72] Perotti, R., 1996. Growth, Income Distribution, and Democracy: What the Data Say. *Journal of Economic Growth* **1**: 149-187.

[73] Persson, T., Tabellini, G., 1994. Is Inequality Harmful for Growth? Theory and Evidence. *American Economic Review* **84**: 600-621.

[74] Piketty, T., 1997. The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing. *Review of Economic Studies* **64**: 173-189.

[75] Roine, J., Vlachos, J., Waldenström D., 2007. What Determines Top Income Shares? Evidence from the Twentieth Century. *Research Institute of Industrial Economics (IFN) Working Paper* **721**.

[76] Schumpeter, J. A., 1934. *The Theory of Economic Development*. Harvard University Press, Cambridge, MA.

[77] Schumpeter, J. A., 1939. *Business Cycles. A Theoretical, Historical and Statistical Analysis of the Capitalist Process*. Mc Graw Hill, New York.

[78] Sethi, R., Somanathan, R., 2004. Inequality and Segregation. *Journal of Political Economy* **112**: 1296-1321.

[79] Stark, J., Bressloff, P., 1993. Iterated Function Systems and their Applications, in: Farge, M., Hunt, J. C. R., Vassilicos, J. C. (Eds.), *Wavelets, Fractals and Fourier Transforms*. Clarendon Press, Oxford, pp. 65-90.

[80] Stokey, N. L., Lucas, R. E., 1989. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, MA.

[81] Strogatz, S. H., 1994. *Nonlinear Dynamics and Chaos*. Addison-Wesley, New York.

[82] Vrscay, E. R., 1991. Iterated Function Systems: Theory, Applications and the Inverse Problem, in: Bélair, J., Dubuc, S. (Eds.), *Fractal Geometry and Analysis*. Kluwer Academic Publishers, Dordrecht, pp. 405-468.

[83] Wolff, E., 2007. Recent Trends in Household Wealth in the United States: Rising Debt and the Middle-Class Squeeze. *Levy Economics Institute Working Paper* **502**.

[84] Wolfson, M., 1994. When Inequalities Diverge?. *American Economic Review* **84**: 353-358.

[85] Zhang, X., Kanbur, R., 2001. What Difference do Polarization Measures Make? An Application to China. *Journal of Development Studies* **37**: 85-98.