.

Note: this document contains two (related) papers:

Social Mobility and the Demand for Redistribution: the POUM Hypothesis

and

Mobility as Progressivity: Ranking Income Processes According to Equality of Opportunity

by

Roland Bénabou
Princeton University

and

Efe A. Ok
New York University

Social Mobility and the Demand for Redistribution: the POUM Hypothesis[1]

Roland Bénabou[2] and Efe A. Ok[3]

This version: January 2000

[1]We thank Abhijit Banerjee, Jess Benhabib, Ed Glaeser, Levent Koçkesen, Ignacio Ortuño–Ortin, three anonymous referees, and seminar participants at NYU, Maryland, MIT, Michigan, Pennsylvania, Pompeu Fabra, Princeton and Toulouse for their helpful comments.

[2]Princeton University, NBER, CEPR and IRP. Financial support from the National Science Foundation, the MacArthur Foundation and the C.V. Starr Center is gratefully acknowledged.

[3]New York University. Financial support from the C.V. Starr Center is gratefully acknowledged.

Abstract

Even relatively poor people oppose high rates of redistribution because of the anticipation that they, or their children, may make it up the income ladder. This "prospect of upward mobility" (POUM) hypothesis is often advanced as one of the factors limiting the extent of redistribution in democracies. But is it compatible with all voters holding rational expectations? This paper establishes the formal basis for the POUM mechanism. There is a range of incomes below the mean where agents oppose lasting redistributions, if (and, in a sense, only if) tomorrow's expected income is increasing and concave in today's income. The coalition against more redistributive policies is larger, the more concave the transition function, and the longer the policy horizon. We illustrate the general analysis with an example where, in every period, 3/4 of families are poorer than average, yet a 2/3 majority has expected future incomes above the mean, and therefore desires low tax rates for all future generations. Using mobility matrices from the PSID, we also make a first pass at an empirical assessment of the POUM mechanism. We find that this effect is indeed present in the data, but probably dominated by the demand for insurance.

Keywords: Social Mobility, Income Distribution, Political Economy, Inequality, Taxation.
JEL classification: D31, D72, P16, H20.

"In the future, everyone will be world-famous for fifteen minutes."
Andy Warhol (1968)

Introduction

The following argument is among those commonly advanced to explain why democracies, where a relatively poor majority holds the political power, do not engage in large-scale expropriation and redistribution. Even people with income below average, it is said, will not support high tax rates because of the prospect of upward mobility: they take into account the fact that they, or their children, may move up in the income distribution and therefore be hurt by such policies.[1] For instance, Okun (1975) relates that:

> *"In 1972 a storm of protest from blue–collar workers greeted Senator McGovern's proposal for confiscatory estate taxes. They apparently wanted some big prizes maintained in the game. The silent majority did not want the yacht clubs closed forever to their children and grandchildren while those who had already become members kept sailing along."*

The question we ask in this paper is simple: does this story make sense with economic agents who hold rational expectations over their income dynamics, or does it require that the poor systematically overestimate their chances of upward mobility –a form of what Marxist writers refer to as "false consciousness"?

The "prospect of upward mobility" (POUM) hypothesis has, to the best of our knowledge, never been formalized, which is rather surprising for such a recurrent theme in the political economy of redistribution. There are three implicit premises behind this story. The first is that policies chosen today will, to some extent, persist into future periods. Some degree of inertia or commitment power in the setting of fiscal policy seems quite reasonable. The second assumption is that agents are not too risk–averse, for otherwise they must realize that redistribution provides valuable insurance against the fact that their income may go down as well as up. The third and key premise is that individuals or families who are currently *poorer than average* –for instance, the median voter– expect to become *richer than average.* This "optimistic" view clearly cannot be true for everyone below the mean, barring the implausible case of negative serial correlation. Moreover, a standard mean–reverting income process would seem to imply that tomorrow's expected income lies somewhere between today's income and the mean. This would leave the poor of today still poor in relative terms tomorrow, and therefore demanders of redistribution. And even if a positive fraction of agents below the mean today can somehow expect to be above it tomorrow, the expected incomes

[1] See for example Roemer (1998) or Putterman (1996). The prospect of upward mobility hypothesis is also related to Hirschman's (1973) famous "tunnel effect," although his argument is more about how people make inferences about their mobility prospects from observing the experience of others. There are of course several other explanations for the broader question of why the poor do not expropriate the rich. These include the deadweight loss from taxation (e.g., Meltzer and Richard (1981)), and the idea that the political system is biased against the poor (Peltzman (1980), Bénabou (2000). Putterman, Roemer, and Sylvestre (1999) provide a review.

of those who are currently richer than them must be even higher. Does this not then require that the number of people above the mean be forever rising over time, which cannot happen in steady–state? It thus appears –and economists have often concluded– that the intuition behind the POUM hypothesis is flawed, or at least incompatible with everyone holding realistic views of their income prospects.[2]

The contribution of this paper is to formally examine the "prospect of upward mobility" hypothesis, asking whether and when it can be valid. The answer turns out to be surprisingly simple, yet a bit subtle. We show that there exists a range of incomes below the mean where agents oppose lasting redistributions if (and, in a sense, only if) tomorrow's expected income is an increasing and *concave* function of today's income. The more concave the transition function, and the longer the length of time for which taxes are pre-set, the lower the demand for redistribution. Even the median voter –in fact, even an *arbitrarily poor* voter– may oppose redistribution if either of these factors is large enough. We also explain how the *concavity* of the expected transition function and the *skewness* of idiosyncratic income shocks interact to shape the long–run distribution of income. We construct, for instance, a simple Markov process whose steady–state distribution has three quarters of the population below mean income, so that they would support purely contemporary redistributions. Yet when voters look ahead to the next period, two thirds of them have expected incomes above the mean, and this super–majority will therefore oppose (perhaps through constitutional design) any redistributive policy that bears primarily on future incomes.

Concavity of the expected transition function is a rather natural property, being simply a form of *decreasing returns:* as current income rises, the odds for future income improve, but at a decreasing rate. While this requirement is stronger than simple mean reversion or convergence of individual incomes, concave transition functions are ubiquitous in economic models and econometric specifications. They arise for instance when current resources affect investment due to credit constraints and the accumulation technology has decreasing returns; or when some income–generating individual characteristic, such as ability, is passed on to children according to a similar "technology". In particular, the specification of income dynamics most widely used in theoretical and empirical work, namely the loglinear ar(1) process, has this property.

Let us now explain the key role played by concavity in the POUM mechanism. For maximum simplicity (but minimum realism), let agents decide today between "laissez-faire" and complete sharing with respect to next period's income, which is a deterministic function of current income: $y' = f(y)$, for all y in some interval $[0, \bar{y}]$. Without loss of generality, normalize f so that someone with income equal to the average, μ, maintains that same level tomorrow ($f(\mu) = \mu$). As shown on Figure 1, everyone who is initially poorer will then see their income rise, and conversely all those who are initially richer will experience a decline. The concavity of f –more specifically, Jensen's inequality– means that the losses of the rich sum to more than the gains of the poor; therefore tomorrow's per capita income μ' is below μ. An agent with mean initial income, or even somewhat

[2] For instance, Putterman, Roemer, and Sylvestre (1999) state that *"voting against wealth taxation to preserve the good fortune of one's family in the future cannot be part of a rational expectations equilibrium, unless the deadweight loss from taxation is expected to be large or voters are risk loving over some range"*.

poorer, can thus *rationally* expect to be richer than average in the next period, and will therefore oppose future redistributions.

To provide an alternative interpretation, let us now normalize the transition function so that tomorrow's and today's mean incomes coincide, $\mu' = \mu$. The concavity of f can then be interpreted as saying that y' is obtained from y through a progressive, balanced budget, redistributive scheme, which *shifts the Lorenz curve* upwards and *reduces the skewness* of the income distribution. As is well known, such progressivity leaves the individual with average endowment better off than under "laissez-faire", because income is taken disproportionately from the rich. This means that the expected income y' of a person with initial income μ is strictly greater than μ, hence greater than the average of y' across agents. This person, and those with initial incomes not too far below, will therefore be hurt if future incomes are redistributed.[3]

Extending the model to a more realistic stochastic setting brings to light another important element of the story, namely the *skewness* of idiosyncratic income shocks. The notion that life resembles a lottery where a lucky few will "make it big" is somewhat implicit in casual descriptions of the POUM hypothesis –such as Okun's. But, in contrast to concavity, skewness in itself does nothing to reduce the demand for redistribution; in particular, it clearly does not affect the distribution of expected incomes. The real role played by such idiosyncratic shocks, as we show, is to offset the skewness–reducing effect of concave expected transitions functions, so as to maintain a positively skewed distribution of income *realizations* (especially in steady–state). The balance between the two forces of concavity and skewness is what allows us to rationalize the apparent risk-loving behavior, or over–optimism, of poor voters who consistently vote for low tax rates due to the slim prospects of upward mobility.

With the important exceptions of Hirschman (1973) and Piketty (1995a, 1995b), the economic literature on the implications of social mobility for political equilibrium and redistributive policies is very sparse. For instance, mobility concerns are completely absent from the many papers recently devoted to the links between income inequality, redistributive politics, and growth (e.g., Alesina and Rodrik (1994), Persson and Tabellini (1994)). A key mechanism in this class of models is that of a poor median voter who chooses high tax rates or other forms of expropriation, which in turn discourage accumulation and growth. We show that when agents vote not just on the current fiscal policy but on one that will remain in effect for some time, even a poor median voter may choose a low tax rate –independently of any deadweight loss considerations.

While sharing the same general motivation as Piketty (1995a, 1995b), our approach is quite different. Piketty's main concern is to explain persistent differences in attitudes towards redistribution. He therefore studies the inference problem of agents who care about a common social

[3]The concavity of f implies that $f(x)/x$ is decreasing, which corresponds to "tax" progressivity and Lorenz equalization, on any interval $[\underline{y}, \bar{y}]$ such that $\underline{y} f'(\underline{y}) \leq f(\underline{y})$. This clearly applies in the present case, where $\underline{y} = 0$ and $f(0) \geq 0$ since income is non–negative. Where the boundary condition does not hold, concavity is consistent with (local or global) regressivity. At the most general level, a concave scheme is thus one that *redistributes from the extremes* towards the mean. This is the economic meaning of Jensen's inequality, given the normalization $\mu' = \mu$. In practice, however, most empirical mobility processes are clearly progressive (in expectation). The progressive case discussed above and illustrated on Figure 1 is thus really the relevant one for conveying the key intuition.

welfare function, but learn about the determinants of economic success only through personal or dynastic experimentation. Because this involves costly effort, they may end up with different long–run beliefs over the incentive costs of taxation. We focus instead on agents who know the true (stochastic) mobility process and whose main concern is to maximize the present value of their aftertax incomes, or that of their progeny. The key determinant of their vote is therefore how they assess their prospects for upward and downward mobility, relative to the rest of the population.

The paper will formalize the intuitions presented above linking these relative income prospects to the concavity in the mobility process, then examine their robustness to aggregate uncertainty, longer horizons, discounting, risk–aversion, and nonlinear taxation. It will also present an analytical example which demonstrates how a large majority of the population can be simultaneously below average in terms of current income and above average in terms of expected future income, *even though the income distribution remains invariant.*[4] Interestingly, a simulated version of this simple model fits some of the main features of the US income distribution and intergenerational persistence rather well. It also suggests, on the other hand, that the POUM effect can have a significant impact on the political equilibrium only if agents have relatively low degrees of risk–aversion.

Finally, the paper also makes a first pass at the empirical assessment of the POUM hypothesis. Using interdecile mobility matrices from the PSID, we compute over different horizons the proportion of agents who have expected future incomes above the mean. Consistent with the theory, we find that this "laissez–faire" coalition grows with the length of the forecast period, to reach a majority for a horizon of about twenty years. We also find, however, that these expected income gains of the middle class are likely to be dominated, under standard values of risk–aversion, by the desire for social insurance against the risks of downward mobility or stagnation.

1 Preliminaries

We consider an endowment economy populated by a continuum of individuals indexed by $i \in [0, 1]$, whose initial levels of income lie in some interval $X \equiv [0, \bar{y}]$, $0 < \bar{y} \leq \infty$.[5] An **income distribution** is defined as a continuous function $F : X \to [0, 1]$ such that $F(0) = 0$, $F(\bar{y}) = 1$ and $\mu_F \equiv \int_X y\, dF < \infty$. We shall denoted by $\mathcal{F}$ the class of all such distributions, and by $\mathcal{F}_+$ the subset of those whose median, $m_F \equiv F^{-1}(1/2)$, is below their mean. We shall refer to such distributions as positively skewed, and more generally we shall measure "skewness" in a random variable as the proportion of realizations below the mean (minus a half), rather than by the usual normalized third moment.

A **redistribution scheme** is defined as a function $r : X \times \mathcal{F} \to \mathbb{R}_+$ which associates to each pretax income and initial distribution a level a disposable income $r(y; F)$, while preserving total income: $\int_X r(y; F)\, dF(y) = \mu_F$. We thus abstract from any deadweight losses which such a scheme might realistically entail, in order to better highlight the different mechanism which is our focus. Both represent complementary forces reducing the demand for redistribution, and could potentially

[4] Another analytical example is the loglinear, lognormal ar(1) process commonly used in econometric studies. Complete closed–form solutions to the model under this specification are provided in Bénabou and Ok (1998).

[5] All the results extend to income supports of the form $[\underline{y}, \bar{y}]$, with $\underline{y} \geq 0$. We choose $\underline{y} = 0$ for notational simplicity.

be combined into a common framework.

The class of redistributive schemes used in a vast majority of political economy models is that of *proportional* schemes, where all incomes are taxed at the rate τ and the collected revenue is redistributed in a lump-sum manner.[6] We denote this class as $\mathcal{P} \equiv \{r_\tau \,|\, 0 \leq \tau \leq 1\}$, where $r_\tau(y; F) \equiv (1-\tau)y + \tau\mu_F$ for all $y \in X$ and $F \in \mathcal{F}$. We shall mostly work with just the two extreme members of $\mathcal{P}$, namely, r_0 and r_1. Clearly, r_0 corresponds to the "*laissez-faire*" policy, whereas r_1 corresponds to "*complete equalization*".

Our focus on these two polar cases is not nearly as restrictive as might initially appear. First, the analysis immediately extends to the comparison between any two proportional redistribution schemes, say r_τ and $r_{\tau'}$, with $0 \leq \tau < \tau' \leq 1$. Second, r_0 and r_1 are in a certain sense "focal" members of $\mathcal{P}$ since, in the simplest framework where one abstracts from taxes' distortionary effects as well as their insurance value, these are the only candidates in this class that can be Condorcet winners. In particular, for any distribution with median income below the mean, r_1 beats every other linear scheme under majority voting if agents care only about their *current* disposable income. We shall see that this conclusion may be dramatically altered when individuals' voting behavior also incorporates concerns about their *future* incomes. Finally, in Section 4 we shall extend the analysis to nonlinear (progressive or regressive) schemes, and show that our main results remain valid.

As pointed out earlier, mobility considerations can enter into voter preferences only if current policy has lasting effects. Such persistence is quite plausible given the many sources of inertia and status quo bias that characterize the policy–making process, especially in an uncertain environment. These include constitutional limits on the frequency of tax changes, the costs of forming new coalitions and passing new legislation, the potential for prolonged gridlock, and the advantage of incumbent candidates and parties in electoral competitions. We shall therefore take such persistence as given, and formalize it by assuming that tax policy must be set one period in advance, or more generally preset for T periods. We will then study how the length of this commitment period affects the demand for redistribution.[7]

The third and key feature of the economy is the **mobility process.** We shall study economies where individual incomes or endowments y_t^i evolve according to a law of motion of the form

$$y_{t+1}^i = f(y_t^i, \theta_{t+1}^i), \quad t = 0, ..., T-1, \tag{1}$$

[6] See, for instance, Meltzer and Richard (1981), Persson and Tabellini (1991), or Alesina and Rodrik (1994). Proportional schemes reduce the voting problem to a single-dimensional one, thereby allowing the use of the median voter theorem. By contrast, when unrestricted nonlinear redistributive schemes are allowed there is generally no voting equilibrium (in pure strategies): the core of the voting game is empty.

[7] Another possible channel through which current tax decisions might incorporate concerns about future redistributions is if voters try to influence future political outcomes by affecting the evolution of the income distribution, through the current tax rate. This strategic voting idea has little to do with the POUM hypothesis as discussed in the literature (see references in footnote 1, as well as Okun's citation). Moreover, these dynamic voting games are notoriously intractable, so the nearly universal practice in political economy models is to assume "myopic" voters. In our model the issue does not even arise since we focus on endowment economies, where income dynamics are exogenous.

where f is a **stochastic transition function** and θ^i_{t+1} is the realization of a random shock Θ^i_{t+1}.[8] We require this stochastic process to have the following properties:

(i) The random variables Θ^i_t, $(i,t) \in [0,1] \times \{1, \ldots, T\}$, have a common probability distribution function P, with support   .

(ii) The function $f : X \times \;\; \rightarrow X$ is continuous, with a well–defined expectation $\mathsf{E}_\Theta[f(\;;\Theta)]$ on X.

(iii) Future income increases with current income, in the sense of *first–order stochastic dominance*: for any $(y, y') \in X^2$, the conditional distribution $M(y'|\, y) \equiv \mathsf{Prob}(\{\theta \in \;\; | \, f(y;\theta) \leq y'\})$ is decreasing in y, with strict monotonicity on some non–empty interval in X.

The first condition means that everyone faces the same uncertain environment, which is stationary across periods. Put differently, current income is the only individual-level state variable which helps predict future income. While this focus on unidimensional processes follows a long tradition in the study of socioeconomic mobility (e.g., Atkinson (1983), Shorrocks (1978), Conlisk (1990), Dardanoni (1993)), one should be aware that it is fairly restrictive, especially in an intragenerational context. It means for instance that one abstracts from life–cycle earnings profiles and other sources of lasting heterogeneity such as gender, race, or occupation, which would introduce additional state variables into the income dynamics. This becomes less of a concern when dealing with intergenerational mobility, where one can essentially think of the two–period case, $T = 1$, as representing overlapping generations. Note, finally, that condition (i) puts no restriction on the correlation of shocks across individuals: it allows for purely aggregate shocks ($\Theta^i_t = \Theta^j_t$ for all i, j in $[0, 1]$), purely idiosyncratic shocks (the Θ^i_t's are independent across agents and sum to zero), and all cases in between.[9]

The second condition is a minor technical requirement. The third condition implies that expected income $\mathsf{E}_\Theta[f(\cdot\,;\Theta)]$ rises with current income, which is what we shall actually use in the results. We impose the stronger distributional monotonicity for realism, as all empirical studies of mobility (intra or intergenerational) find income to be positively serially correlated and transition matrices to be monotone. Thus, given the admittedly restrictive assumption (i), (ii) is a natural requirement to impose.

We shall initially focus the analysis on *deterministic* income dynamics (where θ^i_t is just a constant, and therefore dropped from the notation), then incorporate random shocks. While the *stochastic* case is obviously of primary interest, the deterministic one makes the key intuitions more transparent, and provides useful intermediate results. This two–step approach will also help highlight the fundamental dichotomy between the roles of concavity in expectations and skewness in realizations.

[8] We thus consider only endowment economies, but the POUM mechanism remains operative when agents make effort and investment decisions, and the transition function varies endogenously with the chosen redistributive policies. Bénabou (1999, 2000) develops such a model, using specific functional and distributional assumptions.

[9] Throughout the paper we shall follow the common practice of ignoring the subtle mathematical problems involved with continua of independent random variables, and thus treat each Θ^i_t as jointly measurable in i, for any t. Consequently, the law of large numbers and Fubini's theorem are applied as usual.

2 Income Dynamics and Voting under Certainty

It is thus assumed for now that individual pre-tax incomes or endowments evolve according to a deterministic transition function f, which is continuous and strictly increasing. The income stream of an individual with initial endowment $y \in X$ is then y, $f(y)$, $f^2(y), \dots$, $f^t(y), \dots$, and for any initial $F \in \mathcal{F}$ the cross–sectional distribution of incomes in period t is $F_t \equiv F \circ f^{-t}$. A particularly interesting class of transition functions for the purposes of this paper is the set of all *concave* (but not affine) transition functions; we denote this set by $\mathcal{T}$.

2.1 Two-Period Analysis

To distill our main argument to its most elementary form, we focus first on a two-period (or overlapping generations) scenario, where individuals vote "today" (date 0) over alternative redistribution schemes which will be enacted only "tomorrow" (date 1). For instance, the predominant motive behind agents' voting behavior could be the well-being of their offspring, who will be subject to the tax policy designed by the current generation. Accordingly, agent $y \in X$ votes for r_1 over r_0 if she expects her period one earnings to be below the per capita average:

$$f(y) < \int_X f \, dF_0 = \mu_{F_1} \, . \tag{2}$$

Suppose now that $f \in \mathcal{T}$, that is, it is concave but not affine. Then, by Jensen's inequality,

$$f(\mu_{F_0}) = f\left(\int_X y \, dF_0\right) > \int_X f \, dF_0 = \mu_{F_1}, \tag{3}$$

so the agent with *average income* at date zero will *oppose* date one redistributions. On the other hand, it is clear that $f(0) < \mu_{F_1}$, so there must exist a unique y^*_f in $(0, \mu_{F_0})$ such that $f(y^*_f) \equiv \mu_{F_1}$. Of course $y^*_f = f^{-1}(\mu_{F_0 \circ f^{-1}})$ also depends on F_0 but, for brevity, we do not make this dependence explicit in the notation. Since f is strictly increasing, it is clear that y^*_f acts as a *tipping point* in agents' attitudes towards redistributions bearing on future income. Moreover, since Jensen's inequality –with respect to *all* distributions F_0– characterizes concavity, the latter is both necessary and sufficient for the "prospect of upward mobility" hypothesis to be valid, under any linear redistribution scheme.[10]

Proposition 1 *The following two properties of a transition function f are equivalent:*

(a) f is concave (but not affine), i.e. $f \in \mathcal{T}$.

(b) For any income distribution $F_0 \in \mathcal{F}$ there exists a unique $y^*_f < \mu_{F_0}$ such that all agents in $[0, y^*_f)$ vote for r_1 over r_0, while all those in $(y^*_f, \bar{y}]$ vote for r_0 over r_1.

[10] For any r_τ and $r_{\tau'}$ in $\mathcal{P}$ such that $0 \leq \tau < \tau' \leq 1$, agent $y \in X$ votes for $r_{\tau'}$ over r_τ iff $(1-\tau)f(y) + \tau\mu_{F_1} < (1-\tau')f(y) + \tau'\mu_{F_1}$, which in turn holds iff (2) holds. Thus, as noted earlier, nothing is lost by focusing only on the two extreme schemes in $\mathcal{P}$, namely r_0 and r_1.

Yet another way of stating the result is that $f \in T$ if it is *skewness–reducing:* for any initial F_0, next period's distribution $F_1 = F_0 \circ f^{-1}$ is such that $F_1(\mu_{F1}) < F_0(\mu_{F_0})$. Compared to the standard case where individuals base their votes solely on how taxation affects their current disposable income, popular support for redistribution thus *falls by a measure* $F_0(\mu_{F_0}) - F_1(\mu_{F1}) = F_0(\mu_{F_0}) - F_0(y_f^*) > 0$. The underlying intuition also suggests that the more concave is the transition function, the fewer people should vote for redistribution. This simple result will turn out to be very useful in establishing some of our main propositions on the outcome of majority voting and on the effect of longer political horizons.

We shall say that $f \in \mathcal{T}$ is **more concave than** $g \in \mathcal{T}$, and write $f \succ g$, if and only if f is obtained from g through an increasing and concave (not affine) transformation, that is, if there exists an $h \in \mathcal{T}$ such that $f = h \circ g$. Put differently, $f \succ g$ if and only if $f \circ g^{-1} \in \mathcal{T}$.

Proposition 2 *Let $F_0 \in F$ and $f, g \in T$. Then $f \succ g$ implies that $y_f^* < y_g^*$.*

The underlying intuition is, again, straightforward: the demand for future fiscal redistribution is lower under the transition process which reduces skewness by more. Can prospects of upward mobility be favorable enough for r_0 to beat r_1 under *majority voting?* Clearly, the outcome of the election depends on the particular characteristics of f and F_0. One can show, however, that for any given pre-tax income distribution F_0 there exists a transition function f which is "concave enough" that a majority of voters choose "laissez-faire" over redistribution.[11] When combined with Proposition 2 it allows us to show the following, more general result.

Theorem 1 *For any $F_0 \in \mathcal{F}_+$, there exists an $f \in \mathcal{T}$ such that r_0 beats r_1 under pairwise majority voting for all transition functions that are more concave than f, and r_1 beats r_0 for all transition functions that are less concave than f.*

This result is subject to an obvious caveat, however: for a majority of individuals to vote for "laissez-faire" at date zero, the transition function must be sufficiently concave to make the date one income distribution F_1 negatively skewed. Indeed, if $y_f^* = f^{-1}(\mu_{F_1}) < m_{F_0}$, then $\mu_{F_1} < f(m_{F_0}) = m_{F_1}$. There are two reasons why this is far less problematic than might initially appear. First and foremost, it simply reflects the fact that we are momentarily abstracting from idiosyncratic shocks, which typically contribute to reestablishing positive skewness. Section 3 will thus present a stochastic version of Theorem 1 where F_1 can remain as skewed as one desires. Second, it may in fact not be necessary that the cutoff y_f^* fall all the way below the median for redistribution to be defeated. Even in the most developed democracies it is empirically well documented that poor individuals have lower propensities to vote, contribute to political campaigns, and otherwise participate in the political process, than rich ones. The general message of our results can then

[11] In this case, r_0 is the unique Condorcet winner in $\mathcal{P}$. Note also that Theorem 1 –like every other result in the paper concerning median income m_{F_0}– holds in fact for any *arbitrary income cutoff* below μ_{F_0} (see the proof in the appendix).

be stated as follows: the more concave the transition function, the smaller the departure from the "one person, one vote" ideal needs to be for redistributive policies, or parties advocating them, to be defeated.

2.2 Multi-Period Redistributions

In this section we examine how the length of the horizon over which taxes are set and mobility prospects evaluated affects the political support for redistribution. We thus make the more realistic assumption that the tax scheme chosen at date zero will remain in effect during periods $t = 0, \dots, T$, and that agents care about the *present value* of their disposable income stream over this entire horizon. Given a transition function f and a discount factor $\delta \in (0, 1]$, agent $y \in X$ votes for "laissez–faire" over "complete equalization" if

$$\sum_{t=0}^{T} \delta^t f^t(y) > \sum_{t=0}^{T} \delta^t \mu_{F_t}, \tag{4}$$

where we recall that f^t denotes the t-th iterate of f and $F_t \equiv F_0 \circ f^{-t}$ is the period t income distribution, with mean μ_{F_t}.

We shall see that there again exists a unique tipping point $y_f^*(T)$ such that all agents with initial income less than $y_f^*(T)$ vote for r_1, while all those richer than $y_f^*(T)$ vote for r_0. When the policy has no lasting effects, this point coincides with the mean: $y_f^*(0) = \mu_{F_0}$. When future incomes are factored in, the coalition in favor of "laissez-faire" expands: $y_f^*(T) < \mu_{F_0}$ for $T \geq 1$. In fact, *the more farsighted* voters are, or *the longer the duration* of the proposed tax scheme, *the less support* for redistribution there will be: $y_f^*(T)$ is strictly decreasing in T. If agents care enough about future incomes, the increase in the vote for r_0 can be enough to ensure its victory over r_1.

Theorem 2 *Let $F \in \mathcal{F}_+$ and $\delta \in (0, 1]$.*

(a) *For all $f \in \mathcal{T}$, the longer is the horizon T, the larger is the share of the votes that go to r_0.*

(b) *For all δ and T large enough, there exists an $f \in \mathcal{T}$ such that r_0 ties with r_1 under pairwise majority voting. Moreover, r_0 beats r_1 if the duration of the redistribution scheme is extended beyond T, and is beaten by r_1 if this duration is reduced below T.*

Simply put, longer horizons magnify the strength of the "prospect of upward mobility" effect, whereas discounting works in the opposite direction. The intuition is very simple, and related to Proposition 2: when forecasting incomes further into the future the one–step transition f gets compounded into $f^2, \dots, f^T$, etc., and each of these functions is more concave than its predecessor.[12]

[12] The reason why δ and T must be large enough in part (b) of Theorem 2 is that redistribution is now assumed to be implemented right away, starting in period 0. If it takes effect only in period 1, as in the previous section, the results apply for all δ and $T \geq 1$.

3 Income Dynamics and Voting under Uncertainty

The assumption that individuals know their future incomes with certainty is obviously unrealistic. Moreover, in the absence of idiosyncratic shocks the cross-sectional distribution becomes more equal over time, and eventually converges to a single mass-point. In this section we therefore extend the analysis to the stochastic case, while maintaining risk–neutrality. The role of insurance will be considered later on.

Income dynamics are now governed by a stochastic process $y^i_{t+1} = f(y^i_t; \theta^i_{t+1})$ satisfying the basic requirements (i)–(iii) discussed in Section 1, namely stationarity, continuity and monotonicity. In the deterministic case the validity of the POUM conjecture was seen to hinge upon the concavity of the transition function. The most strict extension of this property to the stochastic case is that it should hold with probability one. Let therefore $\mathcal{T}_P$ be the set of transition functions such that $\mathsf{Prob}[\{\theta \mid f(\cdot\,;\theta) \in \mathcal{T}\ \}] = 1$. It is clear that, for any f in $\mathcal{T}_P$:

(iv) The expectation $\mathsf{E}_\Theta[f(\cdot\,;\Theta)]$ is concave (but not affine) on X.

For some of our purposes the requirement that $f \in \mathcal{T}_P$ will be too strong, so we shall develop our analysis for the larger set of mobility processes which simply satisfy *concavity in expectation.* We shall denote as $\mathcal{T}^*_P$ the set of transition functions which satisfy conditions (i) to (iv).

3.1 Two-Period Analysis

We first return to the basic case where risk–neutral agents vote in period 0 over distributing period 1 incomes. Agent $y^i \in X$ then prefers r_0 to r_1 if and only if

$$\mathsf{E}_{\Theta^i}\left[f(y^i;\Theta^i)\right] > \mathsf{E}\left[\mu_{F_1}\right], \tag{5}$$

where the subscript Θ^i on the left–hand side indicates that the expectation is taken only with respect to Θ^i, for given y^i. When shocks are purely idiosyncratic, the future mean μ_{F_1} is deterministic due to the law of large numbers; with aggregate shocks it remains random. In any case, the *expected mean* income at date one is the *mean expected* income across individuals:

$$\mathsf{E}\left[\mu_{F_1}\right] = \mathsf{E}\left[\int_0^1 f(y^j\,;\Theta^j)\,dj\right] = \int_0^1 \mathsf{E}[f(y^j\,;\Theta^j)]\,dj = \int_X \mathsf{E}_{\Theta^i}\left[f(y\,;\Theta^i)\right] dF_0(y),$$

by Fubini's theorem. This is less than the expected income of an agent whose initial endowment is equal to the mean level μ_{F_0}, whenever $f(y;\theta)$ –or, more generally, $\mathsf{E}_{\Theta^i}[f(y;\Theta^i)]$– is concave in y :

$$\int_X \mathsf{E}_{\Theta^i}\left[f(y\,;\Theta^i)\right] dF_0(y) < \mathsf{E}_{\Theta^i}\left[f(\mu_{F_0};\Theta^i)\right]. \tag{6}$$

Consequently, there must again exist a nonempty interval $[y^*_f, \mu_{F_0}]$ of incomes in which agents will oppose redistribution, with the cutoff y^*_f defined by

$$\mathsf{E}_{\Theta^i}\left[f(y_f^*;\Theta^i)\right] = \mathsf{E}\left[\mu_{F_1}\right].$$

The basic POUM result thus holds for risk–neutral agents whose incomes evolve stochastically. To examine whether an appropriate form of concavity still affects the cutoff monotonically, and whether enough of it can still cause r_0 to beat r_1 under majority voting, observe that the inequality in (6) involves only the *expected transition function* $\mathsf{E}_{\Theta^i}[f(y;\Theta^i)]$, rather than f itself. This leads us to replace the "more concave than" relation with a "more concave *in expectation* than" relation. Given any probability distribution P, we define this ordering on the class $\mathcal{T}_P^*$ as

$$f \succ_P g \quad \text{if and only if} \quad \mathsf{E}_\Theta[f(\cdot\,;\Theta)] \succ \mathsf{E}_\Theta[g(\cdot\,;\Theta)],$$

where Θ is any random variable with distribution P.[13] It is easily shown that $f \succ_P g$ implies $y_f^* < y_g^*$. In fact, making f concave enough in expectation will, as before, drive the cutoff y_f^* below the median m_{F_0}, or even below any chosen income level. Most importantly, since this condition bears only on the *mean* of the random function $f(\cdot\,;\Theta)$, it puts essentially *no restriction on the skewness* of the period 1 income distribution F_1, in sharp contrast to what occurred in the deterministic case. In particular, a sufficiently skewed distribution of shocks will ensure that $F_1 \in \mathcal{F}_+$ without affecting the cutoff y_f^*. This *dichotomy between expectations and realizations* is the second key component of the POUM mechanism, and allows us to establish a stochastic generalization of Theorem 1.[14]

Theorem 3 *For any $F_0 \in \mathcal{F}_+$ and any $\sigma \in (0,1)$, there exists a mobility process (f,P) with $f \in \mathcal{T}_P^*$ such that $F_1(\mu_{F_1}) \geq \sigma$ and, under pairwise majority voting, r_0 beats r_1 for all transition functions in $\mathcal{T}_P^*$ that are more concave than f in expectation, while r_1 beats r_0 for all those that are less concave than f in expectation.*

Thus, once random shocks are incorporated we reach essentially the same conclusions as in Section 2, but with much greater realism. Concavity of $\mathsf{E}_\Theta[f(\cdot\,;\Theta)]$ is necessary and sufficient for the political support behind the "laissez-faire" policy to increase when individuals' voting behavior takes into account their future income prospects. If f is concave enough in expectation, then r_0 can even be the preferred policy of a majority of voters.

[13] Interestingly, $\succ$ and $\succ_P$ are logically independent orderings. Even if there exists some $h \in \mathcal{T}$ such that $f(\cdot,\theta) = h(g(\cdot,\theta))$ for all θ, it need not be that $f \succ_P g$.

[14] The simplest case where the distribution of expectations and the distribution of realizations differ is that of a lottery. The first distribution reduces to a single mass-point (everyone has the same expected payoff), whereas the second is extremely unequal (there is only one winner). Note, however, that this income process does not have the POUM property (instead, everyone's expected income coincides with the mean), precisely because it is not concave.

3.2 Steady–State Distributions

The presence of idiosyncratic uncertainty is not only realistic, but is also required to ensure a non–degenerate long–run income distribution. This, in turn, is essential to show that our previous findings describe not just transitory, short-run effects, but stable, permanent ones as well.

Let P be a probability distribution of idiosyncratic shocks and f a transition function in $\mathcal{T}_P^*$. An invariant or steady–state distribution of this stochastic process is an $F \in \mathcal{F}$ (not necessarily positively skewed) such that

$$F(y) = \int \int_X \mathbf{1}_{\{f(x,\theta) \leq y\}} \, dF(x) \, dP(\theta) \text{ for all } y \in X,$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. Since the basic result that the coalition opposed to lasting redistributions includes agents poorer than the mean holds for *all* distributions in $\mathcal{F}$, it applies to invariant ones in particular: thus $y_f^* < \mu_F$.[15]

This brings us back to the puzzle mentioned in the introduction. How can there be a stationary distribution F where a positive fraction of agents below the mean μ_F have expected incomes greater than μ_F, as do all those who start above this mean, given that the number of people on either side of μ_F must remain invariant over time?

The answer is that even though everyone makes *unbiased forecasts,* the number of agents with *expected* income above the mean, $1 - F(y_{f,F}^*)$, strictly exceeds the number who actually end up with *realized* incomes above the mean, $1 - F(\mu_F)$, whenever f is concave in expectation. This result is apparent on Figure 2, which provides additional intuition by plotting each agent's expected income path over future dates, $\mathsf{E}\left[y_t^i \,|\, y_0^i\right]$. In the long run everyone's expected income converges to the population mean μ_F, but this *convergence is non–monotonic* for all initial endowments in some interval $(\underline{y}_F, \bar{y}_F)$ around μ_F. In particular, for $y_0^i \in (\underline{y}_F, \mu_F)$ expected income first crosses the mean from below, then converges back to it from above. While such non-monotonicity may seem surprising at first, it follows from our results that *all* concave (expected) transition functions must have this feature.

This still leaves us with one of the most interesting questions: can one find income processes whose stationary distribution is *positively skewed*, but where a *strict majority* of the population nonetheless *opposes redistribution?* The answer is affirmative, as we shall demonstrate through a simple Markovian example. Let income take one of three values: $X = \{a_1, a_2, a_3\}$, with $a_1 < a_2 < a_3$. The transition probabilities between these states are independent across agents, and given by

[15] If the inequality $y_f^* < \mu_F$ is required to hold *only* for the steady–state distribution(s) F induced by f and P, rather than for *all* initial distributions, concavity of $\mathsf{E}_\Theta[f(\cdot;\Theta)]$ is still a sufficient condition, but no longer a necessary one. Nonetheless, some form of concavity "on average" is still required, so to speak: if $\mathsf{E}_\Theta[f(\cdot;\Theta)]$ were linear or convex, we would have $y_f^* \geq \mu_F$ for all distributions, including stationary ones.

the Markov matrix:

$$M = \begin{bmatrix} 1-r & r & 0 \\ ps & 1-s & (1-p)s \\ 0 & q & 1-q \end{bmatrix}, \tag{7}$$

where $(p,q,r,s) \in (0,1)^4$. The invariant distribution induced by M over $\{a_1, a_2, a_3\}$ is found by solving $\pi M = \pi$. It will be denoted by $\pi = (\pi_1, \pi_2, \pi_3)$, with mean $\mu \equiv \pi_1 a_1 + \pi_2 a_2 + (1 - \pi_1 - \pi_2) a_3$. We require that the mobility process and associated steady–state satisfy the following conditions:

(a) Next period's income $y^i_{t+1}$is stochastically increasing in current income y^i_t;[16]

(b) The median income level is $a_2 : \pi_1 < 1/2 < \pi_1 + \pi_2$;

(c) The median agent is poorer than the mean: $a_2 < \mu$;

(d) The median agent has expected income above the mean: $\mathsf{E}\left[y^i_{t+1} \mid y^i_t = a_2\right] > \mu$.

Conditions (b) and (c) together ensure that a strict majority of the population would *vote for current redistribution,* while (b) and (d) together imply that a strict majority will *vote against future redistribution.* In Bénabou and Ok (1998) we provide sufficient conditions on $(p, q, r, s; a_1, a_2, a_3)$ for (a)–(d) to be satisfied, and show them to hold for a wide set of parameters. In the steady–state of such an economy the distribution of expected incomes is negatively skewed, even though the distribution of actual incomes remains *positively* skewed and every one has rational expectations.

Granted that such income processes exist, one might still ask: are they at all empirically plausible? We shall present two specifications which match the broad facts of the US income distribution and intergenerational persistence reasonably well. First, let $p = .55$, $q = .6$, $r = .5$, and $s = .7$, leading to the transition matrix

$$M = \begin{bmatrix} .5 & .5 & 0 \\ .385 & .3 & .315 \\ 0 & .6 & .4 \end{bmatrix}$$

and the stationary distribution $(\pi_1, \pi_2, \pi_3) = (.33, .44, .23)$. Thus, 77% of the population is always poorer than average, yet 67% always have expected income above average. In each period, however, only 23% actually end up with realized incomes above the mean, thus replicating the invariant distribution. Choosing $(a_1, a_2, a_3) = (16000, 36000, 91000)$, we obtain a rather good fit with the data, especially in light of the model's extreme simplicity; see **Table 1,** columns 1 and 2. This income process also has more persistence for the lower and upper income groups than for the middle class, which is consistent with the findings of Cooper, Durlauf and Johnson (1994). But most striking is its main political implication: a *two–thirds majority* of voters will support a policy or constitution designed to implement a zero tax rate for all future generations, even though:

[16] Put differently, we posit that $M = [m_{kl}]_{3\times3}$ is a *monotone* transition matrix, requiring row $k+1$ to stochastically dominate row k : $m_{11} \geq m_{21} \geq m_{31}$ and $m_{11} + m_{12} \geq m_{21} + m_{22} \geq m_{31} + m_{32}$. Monotone Markov chains were introduced by Keilson and Ketser (1977), and applied to the analysis of income mobility by Conlisk (1990) and Dardanoni (1993).

– *no deadweight loss* concern enters into voters' calculations;

– *three quarters* of the population is always *poorer than average;*

– the pivotal middle class, which accounts for most of the "laissez–faire" coalition, knows that its children have *less than a one in three* chance of "making it" into the upper class. .

The last column of Table 1 presents the results for a slightly different specification, which also does a good job of matching the key features of the data, and which we shall use later on when studying the effects of risk–aversion. With $(p, q, r, s) = (.45, .6, .3, .7)$, and $(a_1, a_2, a_3) = (20000, 35000, 90000)$, the transition matrix is now

$$M = \begin{bmatrix} .7 & .3 & 0 \\ .315 & .3 & .385 \\ 0 & .6 & .4 \end{bmatrix},$$

and the invariant distribution is $(\pi_1,\ \pi_2, \pi_3) = (.39, .37, .24)$. The majority opposing future redistributions is a bit lower, but still 61%. Middle class children now have about a 40% chance of upward mobility, and this will make a difference when we introduce risk–aversion later on. Note that the source of these greater expected income gains is increased concavity in the transition process, relative to the first specification.

3.3 Multi–Period Redistributions

We now extend the analysis of the general model to multi–period redistributions under uncertainty, maintaining the assumption of risk–neutrality (or complete markets). Agents thus care about the *expected present value* of their net income over the $T + 1$ periods during which the chosen tax scheme is to remain in place. For any individual i, we denote by $\underline{\Theta}_t^i \equiv (\Theta_1^i, ..., \Theta_t^i)$ the random sequence of shocks which she receives up to date t, and by $\underline{\theta}_t^i \equiv (\theta_1^i, ..., \theta_t^i)$ a sample realization. Given a one–step transition function $f \in \mathcal{T}_P$, her income in period t is:

$$y_t^i = f(..., f(f(y_0^i; \theta_1^i); \theta_2^i); ...; \theta_t^i) \equiv f^t(y_0^i; \underline{\theta}_t^i), \quad t = 1, ..., T, \tag{8}$$

where $f^t(y_0^i; \underline{\theta}_t^i)$ now denotes the $t-$step transition function. Under "laissez-faire," the expected present value of this income stream over the political horizon is

$$V^T(y_0^i) \equiv \mathsf{E}_{\Theta_1} \cdots \mathsf{E}_{\Theta_T} \left[\sum_{t=0}^{T} \delta^t y_t^i \mid y_0^i \right] = \sum_{t=0}^{T} \delta^t \mathsf{E}_{\underline{\Theta}_t^i} f^t(y; \underline{\Theta}_t^i) = \sum_{t=0}^{T} \delta^t \mathsf{E}_{\underline{\Theta}_t^i} f^t(y; \underline{\Theta}_t^i),$$

where we suppressed the index i on the random variables $\underline{\Theta}_t^i$ since they all have the same probability distribution $P^t(\underline{\theta}_t) \equiv \prod_{k=1}^{t} P(\theta_k)$ on ^{t}. Under the policy r_1, on the other hand, agent i's expected income at each t is the expected mean $\mathsf{E}_{\underline{\Theta}_t} \left[\mu_{F_t} \right]$, which by Fubini's theorem is also the mean expected income. The resulting payoff is

$$\sum_{t=0}^{T} \delta^t \mathsf{E}[\mu_{F_t}] = \sum_{t=0}^{T} \delta^t \left(\int_0^1 \mathsf{E}_{\underline{\Theta}_t^j} \left[f^t(y_0^j; \underline{\Theta}_t^j) \right] dj \right) = \sum_{t=0}^{T} \delta^t \int_X \mathsf{E}_{\underline{\Theta}_t} \left[f^t(y; \underline{\Theta}_t) \right] dF_0(y) \,,$$

so that agent i votes for r_1 over r_0 if and only if

$$V^T(y_0^i) > \int_X V^T(y) \, dF_0(y). \tag{9}$$

It is easily verified that for transition functions which are concave (but not affine) in y with probability one, that is, for $f \in \mathcal{T}_P$, every function $f^t(y; \underline{\theta}_t^i)$, $t \geq 1$, inherits this property. Naturally, so do the weighted average $\sum_{t=0}^{T} \delta^t f^t(y; \underline{\theta}_t^i)$ and its expectation $V^T(y)$, for $T \geq 1$. Hence, in this quite general setup, the now familiar result:

Proposition 3 *Let $F_0 \in \mathcal{F}$, $\delta \in (0,1]$, $T \geq 1$. For any mobility process (f, P) with $f \in \mathcal{T}_P$, there exists a unique $y_f^*(T) < \mu_{F_0}$ such that all agents in $[0, y_f^*(T))$ vote for r_1 over r_0, while all those in $(y_f^*(T), 1]$ vote for r_0 over r_1.*

Note that Proposition 3 does not cover the larger class of transition processes $\mathcal{T}_P^*$ defined earlier, since $V^T(y)$ need not be concave if f is only concave in expectation. Can one obtain a stronger result, similar to that of the deterministic case, namely that the tipping point decreases as the horizon lengthens? While this seems quite intuitive, and will indeed occur in the data analyzed in Section 5, it may in fact not hold without relatively strong additional assumptions. The reason is that the expectation operator does not, in general, preserve the "more concave than" relation. A sufficient condition which insures this result is that the $t+1$–step transition function be more concave in expectation than the t–step transition function.

Proposition 4 *Let $F_0 \in \mathcal{F}$, $\delta \in (0,1]$, $T \geq 1$, and let (f, P) be a mobility process with $f \in \mathcal{T}_P^*$. If, for all t, $f^{t+1}(\cdot\,; \underline{\Theta}_{t+1}) \succ_{P^{t+1}} f^t(\cdot\,; \underline{\Theta}_t)$, that is,*

$$\mathsf{E}_{\Theta_1} \cdots \mathsf{E}_{\Theta_{t+1}} \left[f^{t+1}(\cdot\,; \Theta_1, ..., \Theta_{t+1}) \right] \succ \mathsf{E}_{\Theta_1} \cdots \mathsf{E}_{\Theta_t} \left[f^t(\cdot\,; \Theta_1, ..., \Theta_t) \right],$$

then the larger the political horizon T, the larger the share of the votes that go to r_0.

The interpretation is the same as that of Theorem 2(a): the more forward–looking the voters, or the more long–lived the tax scheme, the lower is the political support for redistribution. An immediate corollary is that this monotonicity holds when the transition function is of the form $f(y, \theta) = y^\alpha \phi(\theta)$, where $\alpha \in (0,1)$ and ϕ can be an arbitrary function. This is the familiar loglinear model of income mobility, $\ln y_{t+1}^i = \alpha \ln y_t^i + \ln \varepsilon_{t+1}^i$, which is widely used in the empirical literature.

4 Extending the Basic Framework

4.1 Risk–Aversion

When agents are risk-averse, the fact that redistributive policies provide insurance against idiosyncratic shocks increases their attractiveness, hence the breadth of their political support. Consequently, the cutoff separating those who vote for r_0 from those who prefer r_1 may be above or below the mean, depending on the relative strength of the "prospect of mobility" and the risk–aversion effects. While the tension between these two forces is very intuitive, no general characterization of the cutoff in terms of the relative concavity of the transition and utility functions can be provided. To understand why, consider again the simplest setup where agents vote at date 0 over the tax scheme for date 1. Denoting by U their utility function, the cutoff falls below the mean if

$$\mathsf{E}_{\Theta}U\left(f\left(\mathsf{E}_{F_0}[y];\Theta\right)\right) > U\left(\mathsf{E}_{\Theta}\mathsf{E}_{F_0}[f(y;\theta)]\right),$$

where $\mathsf{E}_{F_0}[y] = \mu_{F_0}$ denotes the expectation with respect to the initial distribution F_0. Observe that $f(\cdot,\theta) \in \mathcal{T}$ if and only if the left hand side is greater (for all U and P) than $\mathsf{E}_{\Theta}\left[U\left(\mathsf{E}_{F_0}[f(y;\Theta)]\right)\right]$. But the concavity of U, namely risk-aversion, is equivalent to the fact that this latter expression is also smaller than the right hand side of the above inequality. The curvatures of the transition and utility functions clearly work in opposite directions, but the cutoff is not determined by any simple composite of the two.[17]

One can, on the other hand, assess the outcome of this *"battle of the curvatures"* quantitatively. To that effect, let us return to the simulations of the Markovian model reported in Table 1, and ask the following question: *how risk averse* can the agents with median income a_2 be, and still vote against redistribution of future incomes based on the prospects of upward mobility? Assuming CRRA preferences and comparing expected utilities under r_0 and r_1, we find that the maximum degree of risk–aversion is only 0.35 under the specification of column 2, but rises to 1 under that of column 3. The second number is well within the range of plausible estimates, albeit still somewhat on the low side.[18] While the results from such a simple model need to be interpreted with caution, these numbers do suggest that, with empirically plausible income processes, the POUM mechanism will sustain only moderate degrees of risk aversion. The underlying intuition is simple: to offset risk–aversion, the expected income gain from the POUM mechanism has to be larger, which means that the expected transition function must be more concave. In order to maintain a realistic invariant distribution, the skewed idiosyncratic shocks must then be more important, which is of course disliked by risk–averse voters.

[17] One case where complete closed–form results can be obtained is for the loglinear specification with lognormal shocks; see Bénabou and Ok (1998).

[18] Conventional macroeconomic estimates and values used in calibrated models range from .5 to 4, but most cluster between .5 and 2. In a recent detailed study of the income and consumption profiles of different education and occupation groups, Gourinchas and Parker (1999) estimate risk aversion to lie between 0.5 and 1.0.

4.2 Non-Linear Taxation

Our analysis so far has mostly focussed on an all-or-nothing policy decision, but we explained earlier that it immediately extends to the comparison of any two linear redistribution schemes. We also provided results on voting equilibrium within this class of linear policies. In practice, however, fiscal policy often involves progressivity in taxes and benefits, and thus departs from linearity. In this subsection we shall therefore extend the model to the comparison of arbitrary progressive and regressive schemes, and demonstrate that our main conclusions remain essentially unchanged.

When departing from linear taxation (and before mobility considerations are even introduced), one is inevitably confronted with the nonexistence of a voting equilibrium in a multidimensional policy space. Even the simplest cases are subject to this well known problem. For instance, if the policy space is that of piecewise linear, balanced schemes with just two tax brackets (which has dimension three), there is never a Condorcet winner. Even if one restricts the dimensionality further by imposing a zero tax rate for the first bracket, which then corresponds to an exemption, the problem remains: there exist many (positively skewed) distributions for which there is no equilibrium.

One can still, however, restrict voters to a binary policy choice (e.g., a new policy proposal versus the status quo), and derive results on *pairwise contests* between alternative nonlinear schemes. This is consistent with our earlier focus on pairwise contests between linear schemes, and will most clearly demonstrate the main new insights. In particular, we will show how Marhuenda and Ortuño-Ortin's (1995) result for the static case may be reversed once mobility prospects are taken into account.

Recall that a general redistribution scheme was defined as a mapping $r : X \times \mathcal{F} \to \mathbb{R}_+$ which preserves total income. In what follows we shall take the economy's initial income distribution as given, and denote disposable income $r(y; F)$ as simply $r(y)$. A redistribution scheme r can then equivalently be specified by means of a **tax function** $T : X \to \mathbb{R}$, with collected revenue $\int_X T(y)\, dF$ rebated lump-sum to all agent and the normalization $T(0) \equiv 0$ imposed without loss of generality. Thus:

$$r(y) = y - T(y) + \int_X T(y) dF, \quad \text{for all } y \in X. \tag{10}$$

We shall confine our attention to redistribution schemes with the following standard properties: T and r are increasing and continuous on X, with $\int_X T\, dF \geq 0$ and $r \geq 0$. We shall say that such a redistribution scheme is **progressive** (resp. **regressive**) if its associated tax function is convex (resp. concave).[19]

To analyze how non-linear taxation interacts with the POUM mechanism, we maintain the simple two-period setup of Sections 2.1 and 3.1, and consider voters faced with the choice between a progressive redistribution scheme and a regressive one (which could be laissez–faire). The key question is then *whether political support for the progressive scheme is lower* when the proposed policies are to be enacted next period, rather than in the current one. We shall show that the

[19] Recall that with $T(0) = 0$, if T is convex then $T(x)/x$ is decreasing, while if T is concave then $T(x)/x$ is increasing.

answer is positive, provided the transition function f *is concave enough* relative to the *curvature* of the proposed (net) redistributive policy.

To be more precise, let r_{prog} and r_{reg} be any progressive and regressive redistribution schemes, with associated tax functions T_{reg} and T_{prog}. Let F_0 denote the initial income distribution, and y_{F_0} the income level where agents are indifferent between implementing r_{reg} or r_{prog} today. Given a mobility process $f \in \mathcal{T}_P$ let y_f^* denote, as before, the income level where agents are indifferent between implementing r_{reg} or r_{prog} next period. The basic POUM hypothesis is that $y_f^* < y_{F_0}$, so that expectations of mobility reduce the political support for the more redistributive policy by $F_0(y_{F_0}) - F_0(y_f^*)$. We shall establish two propositions which provide sufficient conditions for an even stronger result, namely $y_{F_0} \geq \mu_{F_0} > y_f^*$.[20] In other words, whereas the coalition *favoring progressivity in current fiscal policy* extends to agents even *richer* than the mean (Marhuenda and Ortuño–Ortin (1995)), the coalition *opposing progressivity in future fiscal policy* extends to agents even *poorer* than the mean.

The first proposition places no restrictions on F_0 but focuses on the case where the progressive scheme T_{prog} involves (weakly) higher tax rates than T_{reg} at every income level. In particular, it raises more total revenue.

Proposition 5 *Let $F_0 \in \mathcal{F}$, let (f, P) be a mobility process with $f \in \mathcal{T}_P^*$, and let T_{prog}, T_{reg} be progressive and regressive tax schemes with $T'_{prog}(0) \geq T'_{reg}(0)$. If*

$$\mathsf{E}_\Theta \left[(T_{prog} - T_{reg})(f(y, \Theta)) \right] \text{ is concave (not affine),}$$

the political cutoffs for current and future redistributions are such that $y_{F_0} \geq \mu_{F_0} > y_f^*$.

Note that $T_{\text{prog}} - T_{\text{reg}}$ is a convex function, so the key requirement is that f be concave enough, on average, to dominate this curvature. The economic interpretation is straightforward: the differential in an agent's (or her child's) future *expected tax bill* must be concave in her current income.

Given the other assumptions, the boundary condition $T'_{\text{prog}}(0) \geq T'_{\text{reg}}(0)$ implies that the tax differential $T_{\text{prog}} - T_{\text{reg}}$ is always positive and increases with pretax income. This requirement is always satisfied when $T_{\text{reg}} \equiv 0$, which corresponds to the laissez-faire policy. On the other hand, it may be too strong if the regressive policy taxes the poor more heavily. It is dropped in the next proposition, which focuses on economies in steady–state and tax schemes that raise equal revenues.

Proposition 6 *Let (f, P) be a mobility process with $f \in \mathcal{T}_P^*$, and denote by F the resulting steady–state income distribution. Let T_{prog} and T_{reg} be progressive and regressive tax schemes raising the same amount of steady–state revenue: $\int_X T_{prog}\, dF = \int_X T_{reg}\, dF$. If*

$$\mathsf{E}_\Theta \left[(T_{prog} - T_{reg})(f(y, \Theta)) \right] \text{ is concave (not affine)}$$

[20] The first inequality is strict unless $T_{\text{reg}} - T_{\text{prog}}$ happens to be linear. The above discussion implicitly assumes that voter preferences satisfy a single–crossing condition, so that y_{F_0} and y_f^* exist and are indeed tipping points. Such will be the case in our formal analysis.

with $\mathsf{E}_\Theta\left[(T_{prog} - T_{reg})(f(\bar{y}, \Theta))\right] \geq 0$, *then the political cutoffs for current and future redistributions are such that* $y_F \geq \mu_F > y_f^*$.

The concavity requirement and its economic interpretation are the same as above. The boundary condition now simply states that agents close to the maximum income $\bar{y}$ face a higher expected future tax bill under T_{prog} than under T_{reg}. This requirement is both quite weak and very intuitive.

Propositions 5 and 6 demonstrate that the essence of our previous analysis remains intact when we allow for plausible non-linear taxation policies. At most, progressivity may increase the extent of concavity in f required for the POUM mechanism to be effective.[21] Note also that all previous results comparing complete redistribution and laissez–faire ($T_{\text{prog}}(y) \equiv y, T_{\text{reg}}(y) \equiv 0$), or more generally two linear schemes with different marginal rates, immediately obtain as special cases.

5 Measuring POUM in the Data

The main objective of this paper was to determine whether the POUM hypothesis is theoretically sound, in spite of its apparently paradoxical nature. As we have seen, the answer is affirmative. The next question which naturally arises is whether the POUM effect is present in the *actual data,* and if so, whether it is large enough to matter for redistributive politics.

Our purpose here is not to carry out a large–scale empirical study or detailed calibration, but to show how the POUM effect can be measured quite simply from income mobility and inequality data. We shall again start with the benchmark case of risk–neutral agents, then introduce risk–aversion. The first question that we ask is thus the following: at any given horizon, *what is the proportion of agents who have expected future incomes strictly above the mean?* In particular, *does it increase with the length of the forecast horizon, and does it eventually rise above 50%?*[22]

Rather than impose a specific functional form on the income process, we shall use here the more flexible description provided by empirical mobility matrices. These are often estimated for transitions between income quintiles, which is too coarse a grid for our purposes, especially around the median. We shall therefore use the more disaggregated data compiled by Hungerford (1993) from the Panel Study on Income Dynamics (PSID), namely:

(a) *interdecile mobility matrices* for the periods 1969–1976 and 1979–1986, denoted M_{69}^{76} and M_{79}^{86} respectively. Each of those is in fact computed in two different ways: using the straight data on annual family incomes, and using five–year averages centered on the first and last years of the transition period, so as to reduce measurement error.

(b) *mean income for each decile*, in 1969 and 1979. We shall treat each decile as homogenous, and denote the vectors of relative incomes as a_{69} and a_{79}. A "prime" will denote transposition.

[21] Recall that Propositions 5 and 6 provide *sufficient* conditions for $y_f^* < \mu_{F_0} \leq y_{F_0}$. They are thus probably stronger than actually necessary, especially for the basic POUM result, which is only that $y_f^* < y_{F_0}$.

[22] Recall that there is *no reason a priori* (i.e., absent some concavity in the transition function) why either effect should be observed in the data, since these are not general features of stationary processes. Rather than just "mean–reverting", the expected income dynamics must be "mean–crossing from below" over some range, as in Figure 2.

Let us start by examining these two income distributions:

$$
\begin{aligned}
a'_{69} &= (.211\ \ .410\ \ .566\ \ .696\ \ .822\ \ .947\ \ 1.104\ \ 1.302\ \ 1.549\ \ 2.393)\,, \\
a'_{79} &= (.179\ \ .358\ \ .523\ \ .669\ \ .801\ \ .933\ \ 1.084\ \ 1.289\ \ 1.588\ \ 2.576)\,.
\end{aligned}
$$

In both years the median group earned approximately 80% of mean income, while those with the average level of resources were located somewhere between the 60th and 70th percentiles. More precisely, by linear interpolation we can estimate the size of the redistributive coalition to be 63.4% in 1969 and 64.4% in 1979.

Next, we apply the appropriate empirical transition matrix to compute the vector of conditionally expected relative incomes $t \times 7$ years ahead, namely $\left(M_{69}^{76}\right)^t \cdot a_{69}$ or $\left(M_{79}^{86}\right)^t \cdot a_{79}$, for $t = 1, ...3$. The estimated rank of the *cutoff* y^*_{ft} where expected future income equals the population mean is then obtained by linear interpolation of these decile values. By iterating a 7–year transition matrix to compute mobility over 14 and 21 years we are, once again, treating the transition process as stationary. Similarly, by using the initial income distribution vectors we are abstracting from changes in the deciles' relative incomes during the transition period. These are obviously simplifying approximations, imposed by the limitations of the data.[23]

The results, presented in **Table 2a,** are consistent across all specifications: the POUM effect appears to be *a real feature of the process of socioeconomic mobility* in the United States –even at relatively short horizons, but especially over longer ones.[24] It affects approximately 3.5% of the population over 7 years, and 10% over 14 years. This is far from negligible, especially since the differentials rates of political participation according to socioeconomic class observed in the U.S. imply that the pivotal agent is almost surely located above the 55th percentile, and probably above the 60th.[25] Over a horizon of approximately 20 years mobility prospects offset the entire 13-15% point interval between mean and median incomes, so that if agents are risk–neutral a strict majority will oppose such long–run redistributions. Thus, in both 1969 and 1979, 64% of the population was poorer than average in terms of current income, and yet 51% could rationally see themselves as richer than average in terms of expected income two decades down the road.

Upwards mobility prospects for the poor, however, represent only one of the forces which deter-

[23] As a basic robustness check, we used the composite matrix $M_{69}^{76} \cdot M_{79}^{86}$ to recompute the 14 year transitions. We also apply the transition matrices M_{69}^{76}, $M_{69}^{76} \cdot M_{79}^{86}$ and relevant iterates to the income distribution a_{79} instead of a_{69}. The results (not reported here) did not change.

[24] Tracing the effect back to its source, one can also examine to what extent expected future income is concave in current income. The expected transition is in fact concave over most, but not all, of its domain: of the nine slopes defined by the ten decile values, only three are larger than their predecessor when we use $(M_{69}^{76}; a_{69})$, and only two when we use $(M_{79}^{86}; a_{79})$. Recall that while concavity at every point is always a sufficient condition for the POUM effect, it is a necessary one only if one requires $y^*_f < \mu_{F_0}$ to hold for *any* initial F_0. For a given initial distribution, such as the one observed in the data, there must be simply "enough" concavity on average, so that Jensen's inequality is satisfied. This is clearly the situation encountered here.

[25] Using data on how the main forms of political participation (voting, trying to influence others, contributing money, participating in meetings and campaigns, etc.) vary with income and education, Bénabou (2000) computes the resulting bias with respect to the median. It is found to vary between 6% for voting propensities and 24% for propensities to contribute money, with most values being above 10%.

mine the equilibrium rate of redistribution –alongside with deadweight losses, the political system, and especially risk–aversion, which was seen to work in the opposite direction. The second question we consider is therefore how the POUM effect compares in magnitude to the demand for *social insurance.* To that end, let us now assume that agents have constant–relative risk–aversion $\beta > 0$. Using the same procedure and horizons of $t \times 7$ years as before, we now compare each decile's expected utility under laissez–faire to the utility of receiving the mean income for sure. The threshold where they coincide is again computed by linear interpolation. The results, presented in **Tables 2b to 2d,** indicate that even small amounts of *risk–aversion will dominate upwards mobility prospects* in the expected utility calculations of the middle class. Beyond values of about $\beta = 0.25$ the threshold ceases to decrease with T, becoming instead either U–shaped or even monotonically increasing.

Our findings can thus be summarized as follows:

(a) a sizeable fraction of the middle class can rationally look forward to expected incomes that rise above average over a horizon of 10 to 20 years;

(b) on the other hand, these expected income gains are small enough, compared to the risks of downward mobility or stagnation, that they are not likely to have a significant impact on the political outcome, unless voters have very low risk aversion and discount rates.

Being drawn from such a limited empirical exercise, these conclusions should of course be taken with caution. A first concern might be measurement error in the PSID data underlying Hungerford's (1979) mobility tables. The use of family income rather than individual earnings, and the fact that replacing yearly incomes with five–year averages does not affect our results, suggest that this is probably not a major problem. A second caveat is that the data obviously pertain to a particular country, namely the United states, and a particular period. One may note, on the other hand, that the two sample periods 1969–1976 and 1979–1986 witnessed very different evolutions in the distribution of incomes (relatively stable inequality in the first, exploding inequality in the second), yet lead to very similar results. Still, things might well be different elsewhere, such as in a developing economy on its transition path. Finally, there is our compounding the 7–year mobility matrices to obtain forecasts at longer horizons. This was dictated by the lack of sufficiently detailed data on longer transitions, but may well presume too much stationarity in families' income trajectories.

The above exercise should thus be taken as representing *only a first pass* at testing for the POUM effect in actual mobility data. More empirical work will hopefully follow, using different and/or better data to further investigate the issue of how people's subjective and, especially, objective income prospects relate to their attitudes vis–a–vis redistribution. Ravallion and Lokshin (1999) and Graham and Pettinato (1999) are recent examples based on survey data from post-transition Russia and post–reform Latin American countries, respectively. Both studies find a significant correlation between (self–assessed) mobility prospects and attitudes towards laissez–faire versus redistribution.

6 Conclusion

In spite of its apparently overoptimistic flavor, the "prospect of upward mobility" hypothesis often encountered in discussions of the political economy is perfectly compatible with rational expectations, and fundamentally linked to an intuitive feature of the income mobility process, namely concavity. Voters poorer than average may nonetheless opt for a low tax rate if the policy choice bears sufficiently on future income, and if the latter's expectation is a concave function of current income. The political coalition in favor of redistribution is smaller, the more concave the expected transition function, the longer the duration of the proposed tax scheme, and the more farsighted the voters. The POUM mechanism is, however, subject to several important limitations. First, there must be sufficient inertia or commitment power in the choice of fiscal policy, governing parties or institutions. Second, the other potential sources of curvature in voters' problem, namely risk–aversion and non–linearities in the tax system, must not be too large compared to the concavity of the transition function.

With theoretical puzzle resolved, the issue now becomes a purely empirical one, namely whether the POUM effect is large enough to significantly affect the political equilibrium. We made a first pass at this question and found that this effect is indeed present in the US mobility data, but likely to be dominated by the value of redistribution as social insurance, unless voters have very low degrees of risk aversion. Due to the limitations inherent in this simple exercise, however, only more detailed empirical work on mobility prospects (in the US and other countries) will provide a definite answer.

Appendix

Proof of Proposition 2

By using Jensen's inequality, we observe that $f \succ g$ implies

$$f(y_f^*) = \int_X f\, dF_0 = \int_X h(g)\, dF_0 < h\left(\int_X g\, dF_0\right) = h(g(y_g^*)) = f(y_g)$$

for some $h \in \mathcal{T}$. The proposition follows from the fact that f is strictly increasing. $\|$

Proof of Theorem 1

Let $F_0 \in \mathcal{F}_+$, so that the median m_{F_0} is below the mean μ_{F_0}. More generally, we shall be interested in any income cutoff $\eta < \mu_{F_0}$. Therefore, define for any $\alpha \in [0,1]$ and any $\eta \in (0, \mu_{F_0})$, the function

$$g_{\eta,\alpha}(y) \equiv \min\{y, \eta + \alpha(y - \eta)\}, \tag{A.1}$$

which clearly is an element of $\mathcal{T}$. It is clear that

$$\int_X g_{\eta,0}\, dF_0 < \eta < \mu_{F_0} = \int_X g_{\eta,1}\, dF_0,$$

so by continuity there exists a unique $\alpha(\eta) \in (0,1)$ which solves:

$$\int_X g_{\eta,\alpha(\eta)}\, dF_0 = \eta \tag{A.2}$$

Finally, let $f = g_{\eta,\alpha(\eta)}$, so that $\mu_{F_1} = \eta$ by (A.2). Adding the constant $\mu_{F_0} - \eta$ to the function f so as to normalize $\mu_{F_1} = \mu_{F_0}$ would of course not alter any of what follows. It is clear from (A.1) that everyone with $y < \eta$ prefers r_0 to r_1, while the reverse is true for everyone with $y > \eta$, so $y_f^* = \eta$. By Proposition 2, therefore, the fraction of agents who support redistribution is greater (respectively, smaller) than $F(\eta)$ for all $f \in \mathcal{T}$ which are more (respectively, less) concave than f. In particular, choosing $\eta = m_{F_0} < \mu_{F_0}$ yields the claimed results for majority voting. $\|$

Proof of Theorem 2

For each $t = 1, ..., T$, $f^t \in \mathcal{T}$, so by Proposition 1 there is a unique $y_{f^t}^* \in (0, \mu_F)$ such that $f^t(y_{f^t}^*) = \mu_{F_t}$. Moreover, since $f^T \succ f^{T-1} \succ \cdots \succ f$, Proposition 2 implies that $y_{f^T}^* < y_{f^{T-1}}^* < \cdots < y_f^* < y_{f^0}^* \equiv \mu_{F_0}$. The concavity of f also implies $f(\mu_{F_t}) > \mu_{F_{t+1}}$ for all t, from which it follows by a simple induction that

$$f^t(\mu_{F_0}) \geq \mu_{F_t} = f^t(y_{f^t}^*) \tag{A.3}$$

with strict inequality for $t > 1$ and $t < T$ respectively. Let us now define the operators $V^T : \mathcal{T} \to \mathcal{T}$ and $W^T : \mathcal{T} \to \mathbb{R}$ as follows:

$$V^T(f) \equiv \sum_{t=0}^{T} \delta^t f^t \quad \text{and} \quad W^T(f) \equiv \int_X V^T(f) dF_0 = \sum_{t=0}^{T} \delta^t \mu_{F_t}. \tag{A.4}$$

Agent y achieves utility $V^T(f)(y)$ under "laissez-faire", and utility $W^T(f)$ under the redistributive policy. Moreover, (A.3) implies that

$$V^T(f)(\mu_{F_0}) = \sum_{t=0}^{T} \delta^t f^t(\mu_{F_0}) > \sum_{t=0}^{T} \delta^t \mu_{F_t} > \sum_{t=0}^{T} \delta^t f^t(y^*_{f^t}) = V^T(f)(y^*_{f^T})$$

for any $T \geq 1$. Since $V^T(f)(\cdot)$ is clearly continuous and increasing, there must therefore exist a unique $y^*_f(T) \in (y_{f^T}, \mu_{F_0})$ such that

$$V^T(f)(y^*_f(T)) = \sum_{t=0}^{T} \delta^t \mu_{F_t} = W^T(f). \tag{A.5}$$

But since $y^*_{f^{T+1}} < y^*_{f^T}$, we have $y^*_{f^{T+1}} < y^*_f(T)$. This implies that $\mu_{F_{T+1}} = f^{T+1}(y^*_{f^{T+1}}) < f^{T+1}(y^*_f(T))$, and hence

$$V^{T+1}(f)(y^*_f(T)) = \sum_{t=0}^{T+1} \delta^t f^t(y^*_f(T)) > \sum_{t=0}^{T+1} \delta^t \mu_{F_t} = V^{T+1}(f)(y^*_f(T+1)).$$

Therefore, $y^*_f(T+1) < y^*_f(T)$ must hold. By induction, we conclude that $y^*_f(T') < y^*_f(T)$ whenever $T' > T$; part (a) of the theorem is proved.

To prove part (b), we shall use again the family of piecewise linear functions $g_{\eta,\alpha}$ defined in (A.1). Let us first observe that the t–th iterate of $g_{\eta,\alpha}$ is simply:

$$(g_{\eta,\alpha})^t(y) \equiv \min\{y, \eta + \alpha^t(y - \eta)\} = g_{\eta,\alpha^t}(y). \tag{A.6}$$

In particular, both $g_{\eta,1} : y \mapsto y$ and $g_{\eta,0} : y \mapsto \min\{y, \eta\}$ are idempotent. Therefore:

$$V^T(g_{\eta,1})(\eta) = \sum_{t=0}^{T} \delta^t \eta < \sum_{t=0}^{T} \delta^t \mu_{F_0} = W^T(g_{\eta,1}).$$

On the other hand, when the transition function is $g_{\eta,0}$, the voter with initial income η prefers r_0 (under which she receives η in each period) to r_1, if and only if

$$\eta + \sum_{t=1}^{T} \delta^t \eta = V^T(g_{\eta,0})(\eta) > W^T(g_{\eta,0}) = \mu_{F_0} + \sum_{t=1}^{T} \delta^t \int_X \min\{y, \eta\} dF_0,$$

or equivalently

$$\frac{\mu_{F_0} - \eta}{\eta - \int_X \min\{y, \eta\} dF_0} < \sum_{t=1}^{T} \delta^t = \frac{\delta(1 - \delta^T)}{1 - \delta}. \tag{A.7}$$

This last inequality is clearly satisfied for $(\delta, 1/T)$ close enough to $(1, 0)$. In that case, we have $W^T(g_{\eta,1}) > \sum_{t=0}^{T} \delta^t \eta > W^T(g_{\eta,0})$. Next, it is clear from (A.4) and (A.1) that $W^T(g_{\eta,\alpha})$ is continuous and strictly increasing in α. Therefore, there exists a unique $\alpha(\eta) \in (0, 1)$ such that $W^T(g_{\eta,\alpha(\eta)}) =$

$\sum_{t=0}^{T}\delta^t\eta$. This means that, under the transition function $f \equiv g_{\eta,\alpha(\eta)}$, we have

$$W^T(f) = \sum_{t=0}^{T}\delta^t f^t(\eta) = V^T(f)(\eta),$$

so that an agent with initial income η is just indifferent between receiving her "laissez–faire" income stream, equal to η in every period, and the stream of mean incomes μ_{F_t}. Moreover, under "laissez–faire" an agent with initial $y < \eta$ receives y in every period, while an agent with $y > \eta$ receives $\eta + \alpha^t(y-\eta) > \eta$. Therefore, η is the cutoff $y_f^*(T)$ separating those who support r_0 from those who support r_1, given $f \equiv g_{\eta,\alpha(\eta)}$. This proves the first statement in part (b) of the theorem.

Finally, by part (a) of the theorem, increasing (decreasing) the horizon T will reduce (raise) the cutoff $y_f^*(T)$ below (above) η. Applying these results to the particular choice of a cutoff equal to median income, $\eta \equiv m_{F_0}$, completes the proof.$\|$

Proof of Theorem 3

As in the proof Theorem 1, let $F_0 \in \mathcal{F}_+$ and consider any income cutoff $\eta < \mu_{F_0}$. Recall the function $g_{\eta,\alpha(\eta)}(y)$ which was defined by (A.1) and (A.2) so as to ensure that $\mu_{F_1} = \eta$. (Once again, adding any positive constant to f would not change anything). For brevity, we shall now denote $\alpha(\eta)$ and $g_{\eta,\alpha(\eta)}$ as just α and g. Let us now construct a stochastic transition function whose expectation is g and which, together with F_0, results in a positively skewed F_1. Let $p \in (0,1)$ and let Θ be a random variable taking values 0 and 1 with probabilities p and $1-p$. For any $\varepsilon \in (0,\eta)$, we define $f : X \times \quad \to \mathbb{R}_+$ as follows:

- if $0 \leq y \leq \eta - \varepsilon$, $\quad f(y;\theta) \equiv y \qquad$ for all θ
- if $\eta - \varepsilon < y \leq \eta$, $\quad f(y;\theta) \equiv \begin{cases} \eta - \varepsilon & \text{if } \theta = 0 \quad (\text{ probability } p) \\ \dfrac{y - p(\eta-\varepsilon)}{1-p} & \text{if } \theta = 1 \quad (\text{ probability } 1-p) \end{cases}$
- if $\eta \leq y \leq \bar{y}$, $\quad f(y;\theta) \equiv \begin{cases} \eta - \varepsilon & \text{if } \theta = 0 \quad (\text{ probability } p) \\ \eta + \dfrac{\alpha(y-\eta)+p\varepsilon}{1-p} & \text{if } \theta = 1 \quad (\text{ probability } 1-p) \end{cases}$

By construction, $\mathsf{E}_\Theta\left[f(y;\Theta)\right] = g(y)$ for all $y \in X$, therefore, $\mathsf{E}_\Theta\left[f(\cdot\,;\Theta)\right] = g \in \mathcal{T}$. It remains to be checked that $f(y;\Theta)$ is strictly stochastically increasing in y. In other words, for any $x \in X$, the conditional distribution $M(x|y) \equiv P(\{\theta \in \quad | f(y;\theta) \leq x\})$ must be decreasing in y on X, and strictly increasing on a nonempty subinterval of X. But this is equivalent to saying that $\int_X h(x)$ $dM(x|y)$ must be (strictly) increasing in y, for any (strictly) increasing function $h : X \to \mathbb{R}$; this latter form of the property is easily verified from the above definition of $f(y;\theta)$.

Because $\mathsf{E}_\Theta\left[f(\cdot\,;\Theta)\right] = g$, so that $\mu_{F_1} = \eta$ by (A.2), the cutoff between the agents who prefer r_0 and those who prefer r_1 is $y_f^* = \eta$. This tipping point can be set to any value below μ_{F_0} (i.e., $1 - F_0(\eta)$ can be made arbitrarily small), while simultaneously ensuring that $F_1(\mu_{F_1}) > \sigma$. Indeed,

$$F_1(\mu_{F_1}) = F_1(\eta) = p\int_X \mathbf{1}_{\{x: f(x,0)\leq\eta\}}dF_0(x) + (1-p)\int_X \mathbf{1}_{\{x: f(x,1)\leq\eta\}}dF_0(x) = p + (1-p)F_0(\eta - p\varepsilon),$$

so by choosing p close to 1 this expression can be made arbitrarily close to 1, for any given η.

To conclude the proof it only remains to observe that a transition function $f_* \in \mathcal{T}$ is more concave than f in expectation if and only if $\mathsf{E}_\Theta[f_*(\cdot\,;\Theta)] \succ \mathsf{E}_\Theta[f(\cdot\,;\Theta)] = g$. Proposition 2 then implies that the fraction of agents who support redistribution under f_* is greater than $F_0(\eta)$. The reverse inequalities hold whenever $f \succ_P f_*$. As before, choosing the particular cutoff $\eta = m_F < \mu_F$ yields the claimed results pertaining to majority voting, for any distribution $F_0 \in \mathcal{F}_+$. ‖

Proof of Proposition 4

Define $h_t \equiv \mathsf{E}_{\Theta_1} \cdots \mathsf{E}_{\Theta_t} f^t(\cdot\,;\Theta_1,...,\Theta_t)$, $t = 1,...,$ and observe that $h_t \in \mathcal{T}$ and $h_{t+1} \succ h_t$ for all t under the hypotheses of the proposition. The proof is thus identical to part (a) of Theorem 2, with h_t playing the role of f^t. ‖

Proof of Proposition 5

1) Properties of y_{F_0}. By the mean value theorem, there exists a point $y_{F_0} \in X$ such that

$$(T_{\text{prog}} - T_{\text{reg}})(y_{F_0}) = \int_X (T_{\text{prog}} - T_{\text{reg}})\, dF_0, \tag{A.8}$$

which by (10) means that $r_{\text{prog}}(y_{F_0}) = r_{\text{reg}}(y_{F_0})$. Now, since $T_{\text{prog}} - T_{\text{reg}}$ is convex, Jensen's inequality allows us to write:

$$(T_{\text{prog}} - T_{\text{reg}})(\mu_{F_0}) \leq \int_X (T_{\text{prog}} - T_{\text{reg}})\, dF_0 = (T_{\text{prog}} - T_{\text{reg}})(y_{F_0}). \tag{A.9}$$

With $(T_{\text{prog}} - T_{\text{reg}})'(0) \geq 0$, the convexity of $T_{\text{prog}} - T_{\text{reg}}$ implies that it is a non–decreasing function on $[0, \bar{y}]$. In the trivial case where it is identically zero, indifference obtains at every point, so choosing $y_F = \mu_{F_0}$ immediately yields the results. Otherwise, two cases are possible. Either $T_{\text{prog}} - T_{\text{reg}}$ is a strictly increasing linear function; or else it is convex (not affine) on some subinterval, implying that the inequality in (A.9) is strict. Under either scenario (A.8) implies that y_{F_0} is the unique cutoff point between the supporters of the two (static) policies, and (A.9) requires that $y_{F_0} \geq \mu_{F_0}$.

2) Properties of y_f^*. By the mean value theorem, there exists a point $y_f^* \in X$ such that

$$\mathsf{E}_\Theta\left[(T_{\text{prog}} - T_{\text{reg}})(f(y_f^*, \Theta))\right] = \int_X \mathsf{E}_\Theta\left[(T_{\text{prog}} - T_{\text{reg}})(f(y, \Theta))\right] dF_0(y), \tag{A.10}$$

or equivalently $\mathsf{E}_\Theta\left[r_{\text{prog}}(f(y_f^*, \Theta))\right] = \mathsf{E}_\Theta\left[r_{\text{reg}}(f(y_f^*, \Theta))\right]$. Moreover:

$$\int_X \mathsf{E}_\Theta\left[(T_{\text{prog}} - T_{\text{reg}})(f(y, \Theta))\right] dF_0(y) < \mathsf{E}_\Theta\left[(T_{\text{prog}} - T_{\text{reg}})(f(\mu_{F_0}, \Theta))\right], \tag{A.11}$$

by Jensen's inequality for the concave (not affine) function $\mathsf{E}_\Theta[(T_{\text{prog}} - T_{\text{reg}}) \circ f(\cdot, \Theta)]$. This function is also non–decreasing, since $T_{\text{prog}} - T_{\text{reg}}$ has this property and $f \in \mathcal{T}^P$ is increasing in y with probability one. Therefore y_f^* is indeed the cutoff point such that $\mathsf{E}_\Theta\left[r_{\text{prog}}(f(y, \Theta))\right] \gtrless \mathsf{E}_\Theta\left[r_{\text{reg}}(f(y, \Theta)\right]$ as $y \lessgtr y_f^*$, and moreover (A.11) requires that $\mu_{F_0} > y_f^*$.‖

Proof of Proposition 6

1) Properties of y_F. The proofs of (A.8) and (A.9) proceed as above, except that the right–hand side of (A.8) is now equal to zero due to the equal–revenue constraint. Since $T_{\text{prog}} - T_{\text{reg}}$ is convex, equal to zero at $y = 0$, and has a zero integral, is must be (barring again the trivial case where it is zero everywhere) that it is first negative on $(0, y_F)$, then positive on $(y_F, \bar{y})$. Therefore y_F is indeed the threshold such that $r_{\text{prog}}(y) \gtreqless r_{\text{reg}}(y)$ as $y \lesseqgtr y_F$, and moreover by (A.8) it must be that $\mu_F \leq y_F$.

2) Properties of y_f^*. Since the distribution F is, by definition, invariant under the mobility process (f, P), so is the revenue raised by either tax scheme. This implies that the function $\mathsf{E}_\Theta \left[(T_{\text{prog}} - T_{\text{reg}})(f(y_f^*, \Theta))\right]$, like $T_{\text{prog}} - T_{\text{reg}}$, sums to zero over $[0, \bar{y}]$:

$$\begin{aligned} \int_X \mathsf{E}_\Theta \left[(T_{\text{prog}} - T_{\text{reg}})(f(y, \Theta))\right] dF(y) &= \int_X \int (T_{\text{prog}} - T_{\text{reg}})(f(y, \theta))\, dF(y) dP(\theta) \\ &= \int_X (T_{\text{prog}} - T_{\text{reg}})(x)\, dF(x) = 0. \qquad \text{(A.12)} \end{aligned}$$

Therefore, by the mean value theorem there is at least one point y_f^* where

$$\mathsf{E}_\Theta \left[(T_{\text{prog}} - T_{\text{reg}})(f(y_f^*, \Theta))\right] = 0.$$

Furthermore, since this same function is concave (not affine), Jensen's inequality implies:

$$\mathsf{E}_\Theta \left[(T_{\text{prog}} - T_{\text{reg}})(f(\mu_F, \Theta))\right] > \int_X \mathsf{E}_\Theta \left[(T_{\text{prog}} - T_{\text{reg}})(f(y, \Theta))\, dF(y)\right] = 0. \qquad \text{(A.13)}$$

Finally, observe that since the function $\mathsf{E}_\Theta \left[(r_{\text{prog}} - r_{\text{reg}})(f(\cdot, \Theta))\right]$ is convex (not affine) on $[0, \bar{y}]$, equal to zero at y_f^* and non–negative at the right boundary $\bar{y}$, it must be first negative on $(0, y_f^*)$, then positive on $(y_f^*, \bar{y})$. Therefore y_f^* is indeed the unique tipping point, meaning that $\mathsf{E}_\Theta \left[r_{\text{prog}}(f(y, \Theta))\right] \gtreqless \mathsf{E}_\Theta \left[r_{\text{reg}}(f(y, \Theta))\right]$ as $y \lesseqgtr y_f^*$; moreover, by (A.13) we must have $\mu_F > y_f^*$.$\|$

References

[1] Alesina, A. and Rodrik, D. (1994), "Distributive Politics and Economic Growth," *Quarterly Journal of Economics,* 109, 465-490.

[2] Atkinson, A. B. (1983), "The Measurement of Economic Mobility," in *Social Justice and Public Policy.* Cambridge, MIT Press..

[3] Bénabou, R. (1999), "Tax and Education Policy in a Heterogeneous Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?", NBER Working Paper 7132 May 1999.

[4] Bénabou, R. (2000), "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review,* forthcoming

[5] Bénabou, R. and Ok, E. A. (1998), "Social Mobility and the Demand for Redistribution: the POUM Hypothesis" NBER Working Paper 6795, November.

[6] Conlisk, J. (1990), "Monotone Mobility Matrices." *Journal of Mathematical Sociology,* 15, 173-191.

[7] Cooper, S., Durlauf, S. and Johnson, P. (1994), "On the Evolution of Economic Status Across Generations. " *American Statistical Association, (Business and Economics Section), Papers and Proceedings,* 50-58.

[8] Dardanoni, V. (1993), "Measuring Social Mobility." *Journal of Economic Theory*, 61, 372-394.

[9] Gourinchas, P.O. and Parker, J. (1999), "Consumption Over the Life Cycle," NBER Working Paper No. 7271.

[10] Graham C. and Pettinato S. (1999), "Assessing Hardship and Happiness: Trends in Mobility and Expectations in the New Market Economies," mimeo, Brookings Institution.

[11] Hirschman, A. O. (1973), "The Changing Tolerance for Income Inequality in the Course of Economic Development (with a Mathematical Appendix by Michael Rothschild)" *Quarterly Journal of Economics,* 87, 544–566.

[12] Hungerford, T.L. (1993), "U.S. Income Mobility in the Seventies and Eighties." *Review of Income and Wealth*, 31, 403–417.

[13] Keilson, J. and Ketser, A. (1977), "Monotone Matrices and Monotone Markov Chains." *Stochastic Processes and their Applications,* 5, 231-241.

[14] Marhuenda, F. and Ortuño-Ortin I. (1995), "Popular Support for Progressive Taxation." *Economics Letters*, 48, 319-324.

[15] Meltzer, A.H. and Richard, S.F. (1981), "A Rational Theory of the Size of Government." *Journal of Political Economy*, 89, 914-927.

[16] Mulligan, C. (1995), "Some Evidence on the Role of Imperfect Markets for the Transmission of Inequality." mimeo, University of Chicago.

[17] Okun, A. (1975), *"Equality and Efficiency: The Big Tradeoff."* Washington, D.C: The Brookings Institution.

[18] Peltzman, S. (1980), "The Growth of Government," *Journal of Law and Economics,* 23, 209–287.

[19] Persson, T. and Tabellini, G. (1994), "Is Inequality Harmful for Growth? Theory and Evidence." *American Economic Review*, 48, 600-621.

[20] Piketty, T. (1995a), "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics, 110, 551–583.*

[21] Piketty, T. (1995b), "Redistributive Responses to Distributive Trends," mimeo, MIT.

[22] Putterman, L. (1996), "Why Have the Rabble Not Redistributed the Wealth? On the Stability of Democracy and Unequal Property," in *Property Relations, Incentives and Welfare,* J. E. Roemer ed. London: MacMillan.

[23] Putterman, L., Roemer, J. and Sylvestre, J. (1996), "Does Egalitarianism have a Future?", *Journal of Economic Literature*, *36*, 861–902.

[24] Ravallion, M. and Lokshin, M. (1999), "Who Wants to Redistribute? Russia's Tunnel Effect in the 1990's," World Bank Policy Research Paper 2150.

[25] Roemer, J. (1998), "Why the Poor Do Not Expropriate the Rich in Democracies: An Old Argument in New Garb," *Journal of Public Economics, 70*, 399–426.

[26] Shorrocks, A. (1978), "The Measurement of Mobility," *Econometrica* 45, 1013-1024.

[27] Solon, Gary (1992), "Intergenerational Income Mobility in the United States." *American Economic Review,* 82, 393-408.

[28] United States Bureau of the Census, *"Current Population Reports"* and data available at http://www.census.gov/ftp/pub/hhes/www/income.html. U.S. Department of Commerce, Income Statistics Branch, HHES Division, Washington, D.C. 20233-8500.

[29] Zimmerman, D. (1992), "Regression Toward Mediocrity in Economic Stature," *American Economic Review,* 82, 409-429.

	Data (1990)	Model 1	Model 2
Median family income ($)	35,353	$36,000$	$35,000$
Mean family income ($)	42,652	$41,872$	$42,260$
Standard deviation of family incomes ($)	29,203	$28,138$	$27,499$
Share of bottom $100 * \pi_1 = 33\%$ of population (%)	11.62	12.82	–
Share of middle $100 * \pi_2 = 44\%$ of population (%)	39.36	37.46	–
Share of top $100 * \pi_3 = 24\%$ of population (%)	49.02	49.72	–
Share of bottom $100 * \pi_1\% = 39\%$ of population (%)	14.86	–	18.5
Share of middle $100 * \pi_2\% = 37\%$ of population (%)	36.12	–	30.8
Share of top $100 * \pi_3\% = 24\%$ of population (%)	49.02	–	50.8
Intergenerational correlation of log–incomes	0.35 to 0.55	. 45.	0.51

Table I Distribution and Persistence of Income in the United States

Sources: median and mean income are from the 1990 US Census (Table F-5). The shares presented here are obtained by linear interpolation from the shares of the five quintiles (respectively 4.6, 10.8, 16.6, 23.8, and 44.3 percent) given for 1990 by the US Census Bureau (Income Inequality Table 1). The variance is computed from the average income levels of each quintile in 1990 (Table F–3). Estimates of the intergenerational correlation from PSID or NLS data are provided by Solon (1992), Zimmerman, (1992) and Mulligan (1995), among others.

Forecast horizon (years)	0	7	14	21
1. Mobility matrix: M_{69}^{76}				
Annual incomes	63.4	61.8	54.2	48.8
Five–year averages	63.4	60.8	56.4	52.9
2. Mobility matrix: M_{79}^{86}				
Annual incomes	64.4	60.9	51.3	48.1
Five–year averages	64.4	58.8	54.3	51.4

Table II a

Income Percentile of the Political Cutoff: Risk–Neutral Agents

Horizon (years)	0	7	14	21
1. Mobility Matrix: M_{69}^{76}				
Annual incomes	63.4	62.5	60.8	60.2
Five–year averages	63.4	61.4	58.5	56.5
2. Mobility Matrix: M_{79}^{86}				
Annual incomes	64.4	61.7	56.6	53.0
Five–year averages	64.4	59.4	55.	53.6

Table II b

Income Percentile of the Political Cutoff: Risk–Aversion = 0.10

Horizon (years)	0	7	14	21
1. Mobility Matrix: M_{69}^{76}				
Annual incomes	63.4	63.4	63.2	65.0
Five–year averages	63.4	62.4	61.3	61.5
2. Mobility Matrix: M_{79}^{86}				
Annual incomes	64.4	63.0	62.0	64.4
Five–year averages	64.4	60.3	57.6	56.9

Table II c

Income Percentile of the Political Cutoff: Risk–Aversion = 0.25

Horizon (years)	0	7	14	21
1. Mobility Matrix: M_{69}^{76}	63.4	65.1	67.3	81.3
Annual incomes	63.4	64.3	65.63	69.1
Five–year averages				
2. Mobility Matrix: M_{79}^{86}				
Annual incomes	64.4	65.5	67.9	73.3
Five–year averages	64.4	62.0	61.3	64.3

Table II d

Income Percentile of the Political Cutoff: Risk–Aversion = 0.50

.

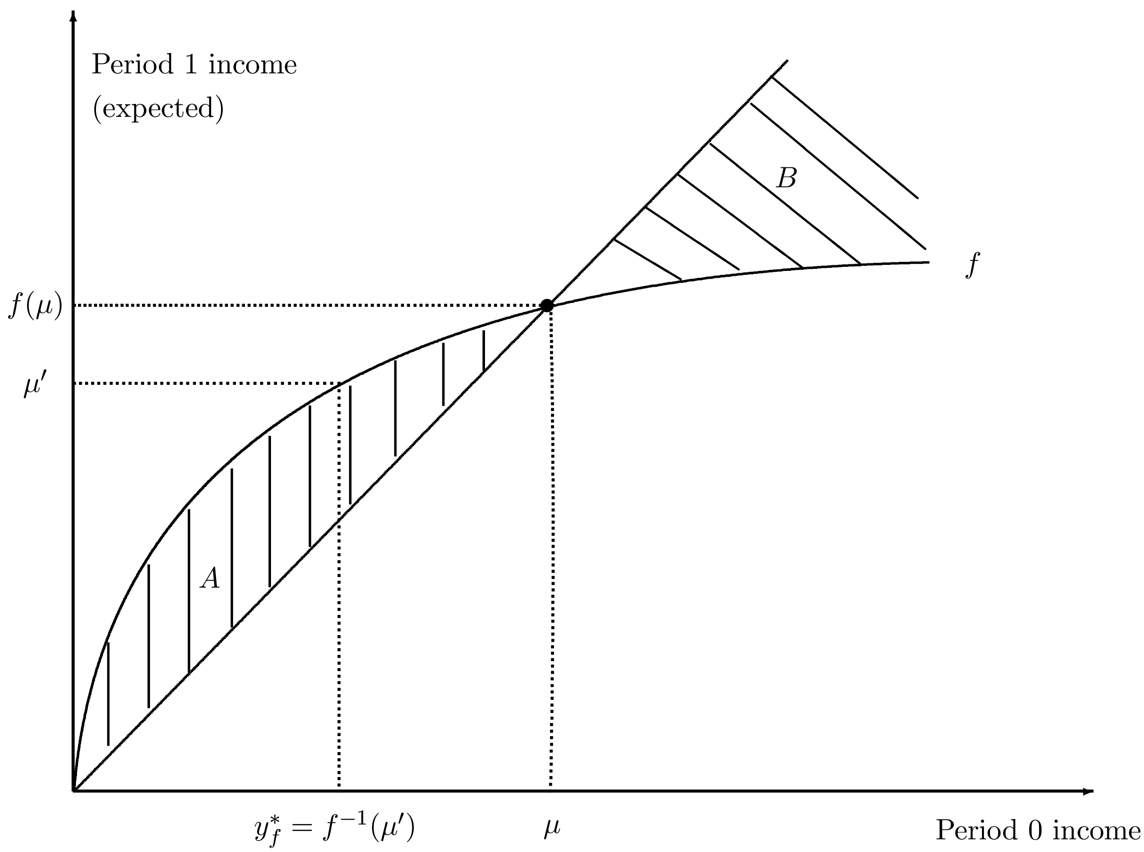

Figure I: Concave transition function
Note that $A < B,$ therefore $\mu' < f(\mu)$.
The figure also applies to the stochastic case, with f replaced by Ef everywhere.

.

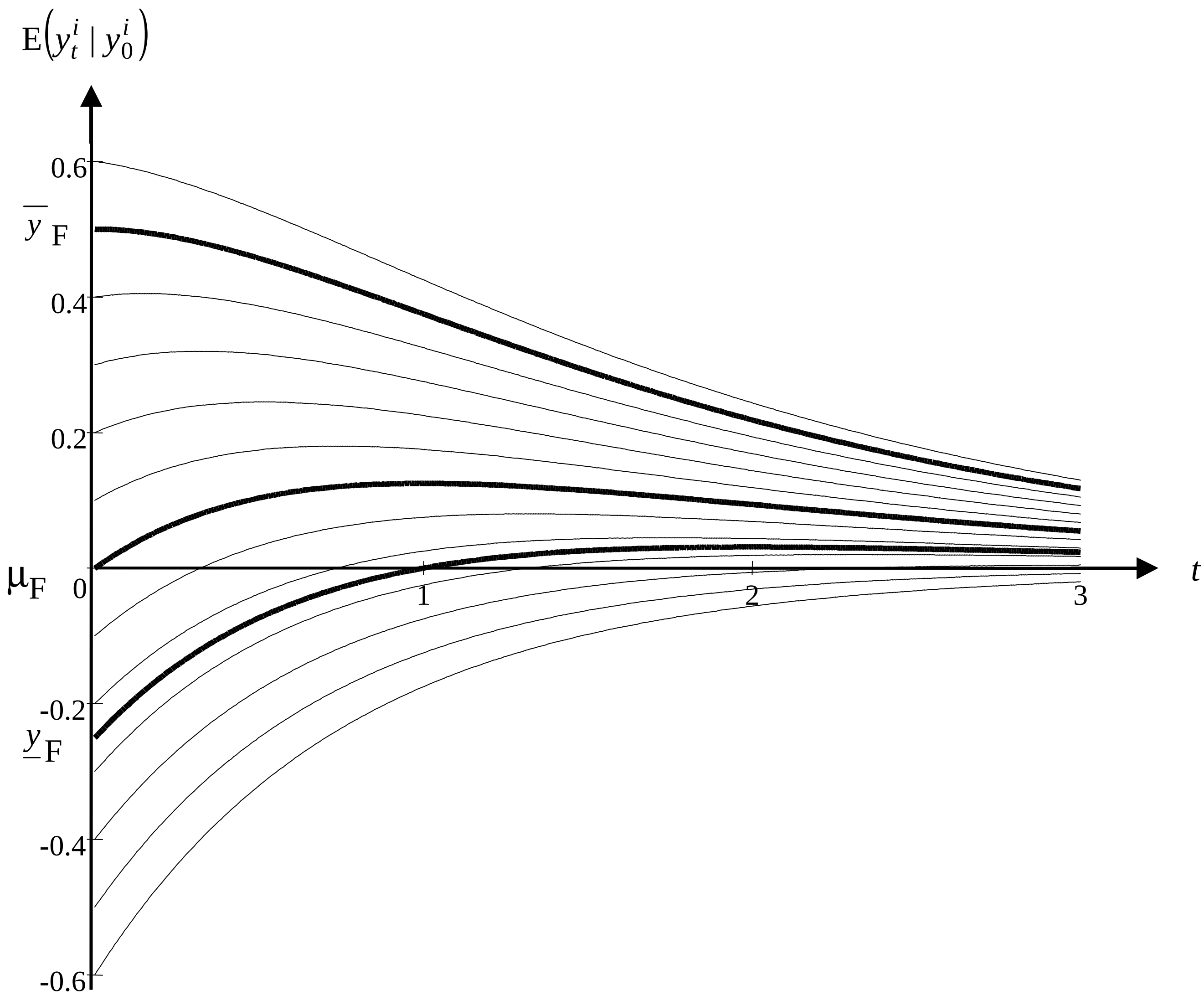

Figure II

Expected future income under a concave transition function (semi-logarithmic scale)

Mobility as Progressivity: Ranking Income Processes According to Equality of Opportunity [1]

Roland Bénabou[⋆2] and Efe A. Ok[♯3]

First draft: August 1998
This version: August 2000

[1]We wish to thank Andrea Ichino for providing us with the data on Italy and the United States.

[2⋆] Princeton University. Financial support from the National Science Foundation and the MacArthur Foundation is gratefully acknowledged.

[3♯]New York University. Financial support from the C.V. Starr Center is gratefully acknowledged.

Abstract

Interest in economic mobility stems largely from its perceived role as an equalizer of opportunities, though not necessarily of outcomes. In this paper we show that this view leads very naturally to a methodology for the measurement of social mobility which has strong parallels with the theory of progressive taxation. We characterize opportunity–equalizing mobility processes, and provide simple criteria to determine when one process is more equalizing than another. We then explain how this mobility ordering relates to social welfare analysis, and how it differs from existing ones. We also extend standard indices of tax progressivity to mobility processes, and illustrate our general methodology on intra- and intergenerational mobility data from the United States and Italy.

Keywords: Social Mobility, Income Distribution, Inequality, Equality of Opportunity, Progressive Taxation.
JEL classification: D31, D63, H20, J62.

Introduction

Since the work of Kolm (1960), Atkinson (1970), Dasgupta, Sen and Starett (1973) and others, the measurement of income inequality is largely a settled issue. By contrast, the measurement of mobility remains in a state of flux, with the literature showing a somewhat bewildering array of approaches and concepts (see Fields and Ok (1999a) for a recent survey).[1] All emphasize different yet overlapping facets aspects of distributional dynamics, and without a prior view of *why* mobility is important, none stands out as uniquely compelling.

The point of this paper is that *equality of opportunity* provides a very natural approach to the evaluation of mobility processes, so natural that there is in fact no need for special concepts or indices to measure it. One cares about mobility not because income movements are intrinsically valuable, but primarily because of the view –or the hope– that it helps attenuate the effects of disparities in initial endowments, or social origins, on future income prospects (e.g., Stokey (1998)). From this view of mobility as an equalizer of opportunities (but not necessarily of outcomes), it follows quite naturally that one should measure it precisely by the extent to which it achieves such levelling. This, in turn, corresponds to a notion of progressivity quite similar to that used to assess tax functions in public finance. In other words, (desirable) mobility *is* progressivity, in the mapping between initial incomes and future opportunities. Measures of pure persistence and other existing mobility indices are of course not unrelated to this notion of opportunity–equalization, but none directly corresponds to it. In particular, movements in relative incomes may in general be equalizing or disequalizing, and yet many mobility criteria proposed in the literature fail to distinguish between the two.

The main insight underlying this paper is thus to view income mobility as just another form of *redistribution* –albeit a stochastic one.[2] Just like a tax scheme maps pre-tax incomes into post-tax incomes, a mobility process maps initial incomes into expected future incomes, or more generally into expected levels of intertemporal welfare.[3] The extent to which the terminal distribution is equalized compared to the initial one is then precisely measured by the degree to which the mapping is progressive, in the sense of having decreasing average "tax" rates. This simple idea allows us to build on the theory of progressivity measurement and Lorenz dominance (Jakobsson (1976), Fellman (1976)) and easily obtain a number of useful results for the study of income mobility. We characterize progressive or opportunity–equalizing mobility processes and their main properties, and provide simple criteria to determine when one process is more progressive than another. We also explain how this ordering relates to social welfare analysis, and discuss how it differs from

[1] The available panoply ranges from purely statistical measures of persistence based on the eigenvalues of transition matrices (see Conlisk (1989)), average numbers of changes in ranks (Bartholomew (1982)), or variations in relative incomes (Fields and Ok (1999b)), to ratios between single and multi–period measures of inequality (Shorrocks (19878a)), and finally to a number of more explicitly welfare–based criteria (Kanbur and Stiglitz (1986), Dardanoni (1993), (1995)).

[2] This view was first put forward in a political economy context in Bénabou and Ok (2000), which studies how mobility expectations affect the political support for redistributive policies.

[3] The relevant definition of "opportunity" naturally depends on agents' horizons and attitudes towards risk; our approach will be quite flexible in that respect.

some of those found in the literature. Since it is of course only a partial ordering, we also show how to extend standard indices of tax progressivity to dynamic processes, so as to obtain mobility indices consistent with the opportunity–equalization ordering. Finally, we illustrate the proposed methodology with intra– and inter–generational mobility from the United States and Italy.

As mentioned above, we are interested in mobility as an equalizer of ex–ante opportunities (or welfare), not of ex–post outcomes. Thus, unlike Kanbur and Stromberg (1988) and Dardanoni (1995), our concern is not whether future *realized* income distributions will be more or less equal than the current one. They could be much more unequal, but if this primarily reflects shocks which were unpredictable on the basis of initial conditions there is little disparity of opportunity. Moreover, in steady-state the income distributions in different periods must coincide in any case, making outcome–based comparisons inapplicable.

Our approach of assessing mobility from an "ex–ante" perspective is more closely related to the papers which take a welfare–based view, such as Kanbur and Stiglitz (1986) and, especially, Dardanoni (1993). Like them, we judge a mobility process to be more equal than another one if it results in a higher Lorenz curve for the distribution of individual levels of intertemporal welfare. Their approach, however, is restricted to mobility processes described by transition matrices between discrete income states. Most importantly, it only allows two processes to be compared if they have the same steady–state, and start in that steady–state. This may seem less of a problem for mobility analyses based on fractile matrices (which specify only rank transitions), since all bistochastic matrices have the same steady–state, namely the uniform distribution. As we explain later on, however, these matrices provide a very incomplete picture of income mobility because, in practice, interfractile income differences vary considerably from one fractile to the next, as well as across countries which have different income distributions. Furthermore, there is really no reason to limit the analysis to steady–states in the first place.[4]

The approach developed in this paper, by contrast, is applicable to continuous as well as discrete processes (or mixtures of the two), and to economies both in and out of steady–state. Moreover, the view of mobility as just another form of redistribution provides a natural "metric" to assess income processes, namely that of tax policy: one can compare the implied residual elasticities, progressivity indices, and marginal tax rates to those which are familiar from fiscal redistributions. Our approach also relates very naturally to the macroeconomic literature on "convergence," since progressivity in a mobility process correspond (in the simplest case) to the property that an agent's expected rate of income growth declines with her initial level of income.

The paper is organized as follows. Section 1 introduces basic concepts and notations. Section 2 presents our mobility ordering and a characterization result for general (Markovian) mobility processes. Section 3 focuses on the case where mobility is represented by transition matrices between discrete income states, offering in particular a very simple test of matrix dominance. Section 4 shows how summary indices of equalizing (or progressive) mobility follow naturally from

[4]This last point is also raised by Formby, Smith and Zheng (1995), who relax the assumption of a common steady–state vector. Their mobility ranking remains conditioned on a particular initial distribution, however, and still applies only within the context of transition matrices.

the general ordering. Section 5 applies the theory to mobility data from the United States and Italy. All proofs are gathered in the appendix.

1 Preliminaries

An *income distribution* is identified by a cumulative distribution function (cdf) $F : \mathbb{R}_+ \to [0,1]$, with finite mean μ_F.[5] Since one often needs to restrict the set of feasible incomes in an economy (the available data may also impose such a restriction), we shall denote as $\mathcal{F}(X)$ the class of all income distributions whose support is contained in a given subset $X \subseteq \mathbb{R}_+$.

The generalized inverse of a distribution $F \in \mathcal{F}(X)$ is defined as

$$F^{-1}(p) \equiv \inf\{y \in X : F(y) \geq p\}, \quad 0 \leq p \leq 1,$$

which corresponds to the income of the person whose rank in the distribution is p. The *Lorenz curve* associated with F can then be defined as the graph of the function:

$$L_F(p) \equiv \frac{1}{\mu_F} \int_0^p F^{-1}(q)\, dq, \quad 0 \leq p \leq 1.$$

Thus $L_F(p)$ is the proportion of the total resources owned by the poorest $100\,p$ percent of individuals. An income distribution F Lorenz–dominates another distribution G when

$$L_F(p) \geq L_G(p) \quad \text{for all } p \in [0,1],$$

which we denote as $F \succsim_L G$. The dominance is strict, and denoted $F \succ_L G$, when $F \succsim_L G$ and $L_F(p) > L_G(p)$ for some $p \in [0,1]$. It is a well-established tradition in welfare economics (especially since Atkinson, (1970)) to regard one income distribution as unambiguously more equal than another, whenever the former strictly Lorenz–dominates the latter.[6]

Let $X \subseteq \mathbb{R}_+$ stand for the set of all feasible income levels. A *mobility process* on X is a function $M : \mathbb{R}_+ \times X \to [0,1]$ such that $M(\cdot \mid y) \in \mathcal{F}(X)$ for all $y \in X$. Thus $M(x \mid y)$ is the probability that an individual with income y today will earn at most x tomorrow. Empirical plausibility requires that future income prospects increase smoothly with the current level, in the sense of *first order stochastic dominance*. We thus restrict attention to processes which are continuous and strictly *monotone:* for any $y_1, y_2 \in X$ with $y_2 > y_1$,

$$M(x \mid y_1) \geq M(x \mid y_2) \quad \text{for all } x \in \mathbb{R}_+,$$

with strict inequality for some x. The set of all such mobility processes on X is denoted $\mathcal{M}(X)$.

An *economy* will be defined as a triplet (X, F, M) consisting of set of feasible income levels $X \subseteq \mathbb{R}_+$, an initial income distribution $F \in \mathcal{F}(X)$, and a mobility process $M \in \mathcal{M}(X)$.

[5] By definition, F is increasing and right-continuous, with $F(0) = 0$, $F(\infty) = 1$ and $\mu_F \equiv \int_0^\infty y\, dF > 0$. Its support will be denoted $supp(F) \equiv \{x \geq 0 : F(x+\varepsilon) - F(x-\varepsilon) > 0 \text{ for all } \varepsilon > 0\}$.

[6] For exhaustive reviews of the literature on the Lorenz ordering and the theory of inequality measurement at large, we refer the reader to Foster (1985) and Sen (1997).

2 Opportunity–Equalizing Processes

2.1 A Mobility Ordering According to Equality of Opportunity

The following two questions summarize the main inquiry of this paper:

[Q1] When would we say that mobility over time reduces inequality of opportunity, relative to initial endowments or social origins, in an economy (X, F, M)?

[Q2] When would we say that mobility is a more powerful equalizer of opportunities in an economy (X, F, M) than in another economy (X, G, N)?

The simplest framework where opportunities can be distinguished from initial conditions and ex–post outcomes is a two–period stochastic framework. For an agent with income $y \in X$, future opportunities are then fully described by the conditional distribution $M(\cdot|y)$. To simplify further, we shall initially summarize these uncertain prospects by their conditional expectation,

$$e_M(y) \equiv \int_0^\infty x \; dM(x \mid y). \tag{1}$$

By identifying opportunities with expected incomes in a two–period context, we are clearly abstracting from agents' aversion to risk and intertemporal fluctuations, as well as their rate of time preference.[7] But, as explained in Section 2.3, this basic framework is readily extended to the more general case by redefining "opportunities" as permanent incomes or intertemporal utilities.

The monotonicity and continuity of $M(\cdot \mid y)$ imply that $e_M : X \to \mathbb{R}_+$ is a strictly increasing and continuous function, and therefore invertible on its range $e_M(X)$. *The distribution of conditional expected incomes* (or opportunities) induced by (X, F, M) is then given by the cdf

$$\Lambda_{F,M}(x) = F(e_M^{-1}(x)) \qquad \text{for all } x \in e_M(X), \tag{2}$$

with support $e_M(X)$.

We now consider the first question posed above. To build up intuition, let us start with an extreme case where everyone has the same expected future income, regardless of their current situation. It is natural to regard this situation as involving a perfect equality of opportunities: the *realized* income distribution next period may not be perfectly equal, or even be vastly unequal, but those differences would be due only to *unpredictable* shocks, as opposed to the persistence of initial disparities. Generalizing this intuition, one might say that mobility in an economy (X, F, M) equalizes opportunities, relative to social origins, whenever

$$\Lambda_{F,M} \succsim_L F. \tag{3}$$

This is, however, an inequality which is *conditional* on the current income distribution, and therefore provides only a local evaluation of mobility. To see why this is problematic, suppose that for some initial income distribution F we have $\Lambda_{F,M} \succ_L F$, but for the income distribution G which will prevail (with probability one) in the next period as a result of the mobility process, we have

[7] These aspects are the main topic of Gottschalk and Spolaore (2000), who show how to disentangle their effects on intertemporal social welfare.

$G \succ_L \Lambda_{G,M}$. In other words, M tends to equalize opportunities (relative to initial conditions or social origins) in period 1, but to disequalize them in period 2. Similarly, it could be that $\Lambda_{F,M} \succ_L F$, but that the ranking is reversed when we restrict attention to some subgroup of the population (e.g., the middle class, women, etc.), even though the evolution of their incomes is governed by the same dynamic process M as all other agents (they just have initial incomes distributed according to a different F'). Both of these examples show that declaring (X, F, M) as an intertemporally egalitarian economy just because (3) holds would not really be justified.

There is a sense in which the notion of equality of opportunity is inherently dynamic, and hence all the relevant information is contained in the *law of motion* that dictates the evolution of incomes, as opposed to the initial conditions. To provide such a global answer to [Q1] we need to rank *mobility processes,* as opposed to economies. Consequently, we shall declare a process $M \in \mathcal{M}(X)$ to be *equalizing* (or *progressive*) when

$$\Lambda_{F,M} \succsim_L F \quad \text{for } \textit{all } F \in \mathcal{F}(X). \tag{4}$$

In words, a mobility process is equalizing when it leads to ex-ante income prospects that are more evenly distributed than initial incomes or endowments, regardless of this initial distribution.

Let us now turn to the second question of interest. Once again, [Q2] can be approached either locally or globally. For the same reasons as before, we deem the global approach to be preferable. We therefore declare a mobility process M on X *more equalizing (* or *more progressive)* than another $N \in \mathcal{M}(X)$, and write $M \succcurlyeq_{\text{eq}} N$, when

$$\Lambda_{F,M} \succsim_L \Lambda_{F,N} \quad \text{for } \textit{all } F \in \mathcal{F}(X). \tag{5}$$

The expected incomes of agents starting from different social positions are then more equally distributed under M than under N, no matter what the profile of initial inequality looks like. Suppose for instance that some country B, whose economic structure and institutions (education system, labor market regulations, etc.) result in a mobility process M^B, were to adopt those of another country A, resulting in the process $M^A \succcurlyeq_{\text{eq}} M^B$. Such a reform would reduce inequalities of opportunity in B's population, even if its initial income distribution was very different from that of A; for instance, each country could initially be in its own steady–state. Conversely, if A were to adopt B's mobility structure, its opportunities would become more unequal. It is in this sense that $\succcurlyeq_{\text{eq}}$ qualifies society A as unambiguously more mobile than society B. As a special case, a mobility process M is equalizing, in the sense of (4), if and only if $M \succcurlyeq_{\text{eq}} I$, where $I(x \mid y) \equiv 1_{\{x \geq y\}}$ for all $(x, y) \in \mathbb{R}_+ \times X$, corresponds to the preservation of the status quo.

Finally, one should note that $\succcurlyeq$ is a (partial) ordering defined over $\mathcal{M}(X)$, and therefore depends on the set of feasible incomes X. The importance of this point will become more apparent when we focus on discrete processes modeled in terms of transition matrices.

2.2 A Characterization Theorem

While the ordering $\succcurlyeq_{\text{eq}}$ appears useful with regard to the measurement of inequality of opportunity, its definition makes it hard to determine when two random processes can actually be ranked in this way. The following theorem and corollaries, which draw on the literature on inequality measurement and progressive taxation (e.g., Jakobsson (1976), Fellman (1976)) provides several operational characterizations of $\succcurlyeq_{\text{eq}}$.

Theorem 1 *Let $X \subseteq \mathbb{R}_+$ and $M, N \in \mathcal{M}(X)$. The following statements are equivalent:*

(i) $M \succcurlyeq_{eq} N$;

(ii) e_M/e_N is decreasing on $X \cap \mathbb{R}_{++}$;

(iii) There exists a strictly increasing mapping $\xi : e_N(X) \to \mathbb{R}_+$ such that the mapping $x \mapsto \xi(x)/x$ is decreasing on $e_N(X) \cap \mathbb{R}_{++}$, and $e_M = \xi \circ e_N$.

These results are quite intuitive. Condition (ii) states that expected incomes ("opportunities") are equalized at a faster rate under the mobility process M than under the process N. This is equivalent, as stated in condition (iii), to the fact that expected income under M can be obtained from expected income under N via a *progressive* redistribution scheme. Theorem 1 thus reflects the main idea of this paper: a mobility processes is nothing else than a *redistribution scheme* –albeit a stochastic one. It is therefore equalizing, in terms of expected incomes, to the extent that it is progressive, in the formal sense of having an increasing *average tax rate*, $t(y) \equiv (y - e_M(y))/y$. It is disequalizing to the extent that it is regressive. The following corollaries provide simple characterizations in terms of residual elasticities and expected growth rates.

Corollary 1. *Let M and N be two mobility processes on X such that e_M and e_N are differentiable on $X \cap \mathbb{R}_{++}$. Then:*

$$M \succcurlyeq_{\text{eq}} N \quad \text{if and only if} \quad \eta_M(y) \equiv \frac{y e'_M(y)}{e_M(y)} \leq \frac{y e'_N(y)}{e_N(y)} \equiv \eta_N(y), \text{ for all } y \in X \cap \mathbb{R}_{++}.$$

Corollary 2. *A mobility process $M \in \mathcal{M}(X)$ is equalizing, $M \succcurlyeq_{\text{eq}} I$, if and only if poorer agents have higher expected income growth than richer ones: the mapping $y \to e_M(y)/y$ is decreasing on $X \cap \mathbb{R}_{++}$. When e_M is differentiable, this means that $\eta_M \leq 1$ everywhere.*

We can also exploit the connection with Lorenz dominance to provide a welfaristic support for the mobility ordering $\succcurlyeq_{\text{eq}}$. Indeed, it is easy to see from (5) that if M is more progressive than N and leads to higher average income next period, it yields higher social welfare in terms of any utilitarian social welfare function defined over individual's expected incomes (opportunities).

Corollary 3. *If $M \succcurlyeq_{\text{eq}} N$, then for any $F \in \mathcal{F}(X)$ such that $\int_0^\infty e_M \, dF \geq \int_0^\infty e_N \, dF$,*

$$\int_0^\infty u \, d\Lambda_{F,M} \geq \int_0^\infty u \, d\Lambda_{F,N}$$

for all concave and continuous utility functions u defined on $\mathbb{R}_+$.

2.3 Remarks and Extensions

2.3.1 Perfect Immobility and Perfect Mobility

In evaluating mobility orderings, it is often considered intuitive that the identity process ($I(x \mid y) \equiv \mathbf{1}_{\{x \geq y\}}$) should correspond to the smallest element, and be viewed –implicitly or explicitly – as the worst case scenario (e.g. Shorrocks (1978a), Dardanoni (1993)). More generally, according to what Kanbur and Stiglitz (1986) term "the diagonals view," *any* increase in relative income movement (in a transition matrix context, shifting probability weight from diagonal to off-diagonal elements) should imply a higher ranking in the mobility ordering.

Our ordering is quite different in that respect, because it recognizes that relative income movements can be disequalizing as well as equalizing, and it is only the latter type that will count positively as "mobility." Thus, in general, there does not exist a *smallest* element in $\mathcal{M}(X)$ with respect to $\succcurlyeq_{\text{eq}}$, just as there generally does not exist a most regressive tax scheme. In particular, the identity process (no movement) does *not* have his property. The next section provides a discrete example, but this most clearly seen for $X = \mathbb{R}_+$. Let M be such that future income is any convex function of current income, plus mean–zero noise (but ensuring that realizations remain positive with probability one). Clearly, the distribution of $e_M(y)$ is more unequal than that of y, under any initial conditions.

When the income support is bounded above and below, on the other hand, one can show that I is a *minimal* element in $\mathcal{M}(X)$ with respect to $\succcurlyeq_{\text{eq}}$, i.e. there is no $M \in \mathcal{M}(X)$ such that $I \succ_{\text{eq}} M$. Indeed, by Theorem 1 this would imply $e_M(y_2)/y_2 \geq e_M(y_1)/y_1$ for all pairs $y_1, y_2 \in X$ with $y_2 \geq y_2 > 0$. But we must also have $e_M(\min X) \geq \min X$ and $e_M(\max X) \leq \max X$. Both conditions are compatible only if $e_M(y)/y = 1$ for all y, which contradicts $I \succ_{\text{eq}} M$. Note that this minimal property of I is entirely due to the assumption of fixed upper and lower bounds on feasible incomes, which seems rather arbitrary. Moreover, I is still not a smallest element of $\preccurlyeq_{\text{eq}}$, since one can again find mobility processes which are more regressive than immobility on some subset of X. The next section will provide specific examples in the context of transition matrices.

Is there a *greatest* element in $\mathcal{M}(X)$ with respect to our mobility ordering? Because of the requirement of strict monotonicity, there does not exist one either but –in contrast to the absence of a smallest element– this is only a minor technical wrinkle. If we extend the ordering $\succcurlyeq_{\text{eq}}$ to the set $\mathcal{M}_*(X)$ which includes processes with $M(x \mid \cdot)$ that are only weakly increasing, we immediately observe that any process M such that $M(\cdot \mid y)$ is independent of y is a greatest element in $\mathcal{M}_*(X)$ with respect to $\succcurlyeq_{\text{eq}}$. This is because under such a process, all agents have the same conditional distribution of future incomes, and in particular the same expected income.

2.3.2 Permanent Income and Intertemporal Utility

When introducing our ordering as one based on the equalization of opportunities, we defined the latter as conditional expected incomes. In terms of welfare, this corresponds to agents who live for two periods, care only about the second one (e.g., their children's expected lifetime income), and are risk–neutral. It is, however, straightforward to extend the analysis to a multiperiod setting (say, infinite horizon) where agents care about permanent incomes, or more generally about some additively separable intertemporal utility. Given the one–period transition M, let $M^{(t)} \in \mathcal{M}(X)$ denote the t–period–ahead mobility process, which is defined recursively by $M^{(1)} \equiv M$ and

$$M^{(t+1)}(x|y) \equiv \int_X M(x|z) dM^{(t)}(z|y) \tag{6}$$

for all $t \in \mathbb{N}$. When agents are risk–neutral and have discount factor ρ, we need only replace $e_M(y)$ in Theorem 1 and its corollaries by

$$e_M(y;\rho) \equiv (1-\rho)\sum_{t=0}^{\infty} \rho^t e_{M^{(t)}}(y), \tag{7}$$

provided course the series converges. All the results hold unchanged, for what is now a mobility ranking relating to the reduction of *lifetime inequality.* Similarly, when agents' instantaneous utility function is $u(\cdot)$, $e_M(y)$ is replaced by

$$e_M(y;u,\rho) \equiv (1-\rho)\sum_{t=0}^{\infty} \rho^t \int_X u(x) dM^{(t)}(x|y), \tag{8}$$

yielding parallel results for a mobility ordering based on the equalization of the *lifetime utilities.*[8] Of course, computing $e(y;\rho)$ or $e_M(y;u,\rho)$ from M may be easy or difficult. In the case of discrete mobility processes represented by transition matrices (see Section 3), it will be extremely simple.

2.3.3 Strongly Equalizing Mobility Processes

In the previous subsection we showed how to rank mobility processes according to equalization of permanent incomes, for a given discount factor ρ. It may also be of interest to know whether a mobility process equalize opportunities, relative to initial conditions, *over any horizon t* (say, for grandchildren as well as children), or *for any discount factor ρ*. Formally, the question is whether M being equalizing guarantees that $M^{(t)}$ is equalizing for all $t \in \mathbb{N}$. Unfortunately, the answer is negative, as one can show by means of simple examples. We can, however, identify a stronger condition on M which ensures that this appealing "horizon–independence" property holds.

Let $M \in \mathcal{M}(X)$, and define

$$E_M(y,\theta) \equiv \int_0^\theta x \; dM(x \mid y), \quad y \in X, \; \theta \in (0,\infty].$$

Clearly, E_M is a nonnegative-valued mapping on $X \times (0,\infty]$. We say that M is *strongly equalizing* if, *for each* $\theta \in (0,\infty]$, the mapping

$$y \mapsto \frac{1}{y} E_M(y,\theta) \text{ is decreasing in } y \text{ on } X \cap \mathbb{R}_{++}.$$

[8] An informal but perceptive antecedent to this idea can be found in Loury (1981), who wrote (about his model): *"the graph of the ... indirect utility function will be "flatter" when society is more mobile.... This makes the cross–sectional distribution of welfare... less unequal than would be the case with little or no mobility".* Our results formalize and generalize this intuition, showing in particular that what matters is the elasticity, rather than the slope, of the mapping between initial conditions and lifetime utilities.

Note that we have $E_M(y, \infty) = e_M(y)$ for all y, so by Theorem 1 a strongly equalizing process is equalizing. More interesting is the following key property.

Proposition 1 *Let X be a closed subset of $\mathbb{R}_+$. If $M \in \mathcal{M}(X)$ is strongly equalizing, so is $M^{(t)}$ for any $t \in \mathbb{N}$.*

If mobility in an economy is governed by the same strongly equalizing process over t-periods, then expected incomes are equalized, relative to initial conditions, over all relevant horizons. Clearly, expected present values of incomes then have the same property, for any discount factor.[9] Unfortunately, it does not seem possible to extend this result so as to define a "more strongly equalizing" ordering between arbitrary processes M and N which would be similarly preserved through iteration.

3 Discrete Markov Processes

3.1 Equalizing Transition Matrices: Properties and Characterization

In empirical applications, it is common to focus on discrete income distributions and represent mobility by transition matrices between n income states. In this section we elaborate on the properties of our mobility ordering $\succcurlyeq_{\text{eq}}$ in this case.

A transition matrix $P \equiv [p_{ij}]$ is any $n \times n$ stochastic matrix (i.e., $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$ for all i and j). To interpret it as modeling the evolution of individuals' *incomes*, we attach an income level y_i to income state i. We assume that $0 < y_1 < \cdots < y_n$ and denote the *income state vector* as $\mathbf{y} \equiv (y_1, ..., y_n)'$, where a prime signifies transposition. An income distribution in this setting corresponds to a probability row vector $\boldsymbol{\pi} = (\pi_1, ..., \pi_n)$, where π_i is interpreted as the proportion of individuals who are in income state i. Or, consistently with our general definition, one may think of an income distribution as any cdf with support contained in $X = \{y_1, ..., y_n\}$.

A transition matrix P is said to be *monotone* if an individual in income state $i+1$ faces a better lottery over her future income than an individual in income state i :

$$\sum_{j=1}^{k} p_{i+1,j} \geq \sum_{j=1}^{k} p_{ij} \quad \text{for all } i,\, k \text{ in } \{1, ..., n-1\}, \tag{9}$$

with strict inequality for some k.[10] A transition matrix P induces a mobility process M_P on $\{y_1, ..., y_n\}$ in a natural way:[11]

$$M_P(y_k \mid y_i) \equiv \sum_{j=1}^{k} p_{ij}, \quad k = 1, ..., n-1. \tag{10}$$

[9] As an example, the familiar loglinear–lognormal process, where $\ln y_{t+1} = \alpha \ln y_t + \varepsilon_{t+1}$, and the ε_t's are i.i.d. and normal, is strongly equalizing for all $\alpha < 1$.

[10] Monotone transition matrices were introduced to the mathematical literature by Keilson and Ketser (1977), and are now widely used in modeling income mobility; see Conlisk (1989), (1990) and Dardanoni (1993), (1995).

[11] $M_P(\cdot \mid y_i)$ is easily extended to a cdf on $\mathbb{R}_+$ by making it a step function that is constant on any $[y_j,\, y_{j+1})$.

Consequently, monotone transition matrices will be ranked by ordering with respect to $\succcurlyeq_{\text{eq}}$ the mobility processes induced over the income states in $\mathbf{y}$. We shall thus write $P \succcurlyeq_{\text{eq}}^{\mathbf{y}} Q$, and say that P is more *equalizing* (or *progressive*) than Q over the state space X, whenever $M_P \succcurlyeq_{\text{eq}} M_Q$.[12] In words, $P \succcurlyeq_{\text{eq}}^{\mathbf{y}} Q$ means that the distribution of conditional expected incomes (opportunities) induced by P, which for a fraction π_i of the population are equal to

$$e_P(y_i) \equiv \sum_{j=1}^{n} p_{ij}\, y_j, \tag{11}$$

is more equal than that similarly induced by Q, *for all* initial income distributions $\boldsymbol{\pi}$ defined on $\{y_1, ..., y_n\}$. Similarly, given an income state vector $\mathbf{y}$, we say that P is equalizing (or progressive) if $P \succcurlyeq_{\text{eq}}^{\mathbf{y}} I$, where I is the $n \times n$ identity matrix.

The following theorem provides easy-to-apply methods for checking whether or not two transition matrices can be ranked on the basis of $\succcurlyeq_{\text{eq}}^{\mathbf{y}}$, given an income state vector $\mathbf{y}$. In its statement, we denote by $D^*[A]$ the first superdiagonal of any square matrix A, and by $D_*[A]$ its first subdiagonal.

Theorem 2 *Let $\mathbf{y} \in \mathbb{R}^n_{++}$ (with $y_1 < \cdots < y_n$) and let P and Q be two $n \times n$ monotone transition matrices. The following statements are equivalent.*

(i) $P \succcurlyeq_{eq}^{\mathbf{y}} Q$;

(ii) $\dfrac{e_P(y_1)}{e_Q(y_1)} \geq \cdots \geq \dfrac{e_P(y_n)}{e_Q(y_n)}$;

(iii) $(D^* - D_*)[P\mathbf{y}\mathbf{y}'Q'] \geq 0$.

The interpretation of conditions (i) and (ii) in terms of progressivity is similar to that of Theorem 1. Condition (iii) is the new, "operational" one, being immediate to check.[13] Turning next to social welfare, we note again that, provided P does not yield a lower mean income than Q, the statement $P \succcurlyeq_{\text{eq}}^{\mathbf{y}} Q$ has a fairly strong utilitarian implications.

Corollary 1 *Let $\mathbf{y} \in \mathbb{R}^n_{++}$ (with $0 < y_1 < \cdots < y_n$) and let P and Q be two $n \times n$ monotone transition matrices. If $P \succcurlyeq_{eq}^{\mathbf{y}} Q$, then for all probability vectors $\boldsymbol{\pi}$ such that $\boldsymbol{\pi} P\mathbf{y} \geq \boldsymbol{\pi} Q\mathbf{y}$, we have*

$$\sum_{i=1}^{n} \pi_i u(e_P(y_i)) \geq \sum_{i=1}^{n} \pi_i u(e_Q(y_i))$$

for all concave and increasing utility functions u defined on $\mathbb{R}_+$.

Finally, the defining properties of an equalizing (progressive) transition matrix P naturally correspond to the particular case where $Q = I$ in the above results.

[12] The reason why we make explicit in our notation the dependence of this ordering on the income state space $\mathbf{y}$ will be discussed below.

[13] Note that, for any square matrix A, $(D^* - D_*)[A] = D^*[A - A']$.

3.2 Permanent Income and Intertemporal Utility

We now extend the analysis to multiperiod settings where agents care about their permanent income (lifetime or dynastic), or more generally about some intertemporal utility, as in (8). Following Dardanoni (1993), let us denote $P(\rho) \equiv (1-\rho)(I-\rho P)^{-1}$, for any transition matrix P and discount factor $\rho \in (0,1)$; given that P is monotone, so is $P(\rho)$. Next, for any increasing utility function u on $\mathbb{R}_+$, let (by a slight abuse of notation), $u(\mathbf{y}) \equiv (u(y_1), ...u(y_n))'$ be the utility state vector. Since $P(\rho)u(\mathbf{y})$ is the vector of conditional lifetime utilities, we have by Theorem 2:

Proposition 2 *Let $\mathbf{y} \in \mathbb{R}^n_{++}$ (with $y_1 < \cdots < y_n$) and let P and Q be any two $n \times n$ monotone transition matrices. The following statements are equivalent:*

(i) $P(\rho) \succcurlyeq^{u(\mathbf{y})}_{eq} Q(\rho)$: starting from any initial distribution $\boldsymbol{\pi}$, the Lorenz curve for agents' intertemporal utilities which obtains under the mobility process P is everywhere below that which obtains under Q.

(ii) $(D^* - D_*)[P(\rho)u(\mathbf{y})u(\mathbf{y})'Q(\rho)'] \geq 0$.

Note that this ranking (and the associated test) are conditional on the value chosen for the discount factor ρ –as in Kanbur and Stiglitz (1986), and Formby, Smith and Zheng (1995). By contrast, Dardanoni (1993) provides sufficient conditions which ensure that, if P and Q can be ranked according to his ordering (to start with, they must have the same steady–state), then $P(\rho)$ and $Q(\rho)$ have the same ranking for all values of ρ. It does not seem that any ρ–independent criterion can be provided for our ordering, except in special cases.

One such case which is of interest relates to the question of whether a mobility process is equalizing, relative to initial conditions, not just for next period's expected incomes, but also for lifetime incomes or welfare levels. For this, we shall make use again of strongly equalizing mobility processes, for which progressivity is actually a ρ–independent property. Consistent with the general definition in Section 2.3.3, a monotone matrix P is said to be *strongly equalizing* over the income space $\mathbf{y}$ if, for each $i = 1, ..., n-1$,

$$\frac{1}{y_i}\sum_{j=1}^{k} p_{ij}\, y_j \geq \frac{1}{y_{i+1}}\sum_{j=1}^{k} p_{i+1,j}\, y_j, \quad k = 1, ..., n.$$

Proposition 3 *Let P be an $n \times n$ monotone transition matrix. Given a utility state vector $u(\mathbf{y}) \in \mathbb{R}^n$ (with $u(y_1) < \cdots < u(y_n)$), if P is strongly equalizing over $u(\mathbf{y})$, then so is P^t for all t. Furthermore, $P(\rho)$ is then equalizing for all $\rho \in (0,1)$.*

3.3 Discussion and Relation to Other Orderings

Remark 1. A discrete mobility process was defined by a transition matrix P *and* an income state vector $\mathbf{y}$ over which it operates. In contrast to the view implicitly taken in much of the mobility literature, *income* mobility cannot, we believe, be adequately studied or even defined independently of the values taken by income, and this for several reasons.

The first issue goes back to the empirical meaning of a transition matrix. When we represent a country's mobility process by a transition matrix P, these coefficients correspond to the frequencies of transitions which were observed to occur between certain well-defined income levels –more precisely, income intervals– such as (say): 0–25,000, 25,000–50,000, 50,000–75,000, 75,000–100,000, 100,000–∞. *Nothing allows us to pretend* that the same transition probabilities obtain between any arbitrary five income levels or intervals, such as 0-1,000, 1,000–60,000, 60,000–61,000, 61,000-1,000,000, 1,000,000–∞. In fact, the data would surely contradict this notion.[14]

One may hope to avoid this problem by using interfractile transition matrices, where the ith state always corresponds to the ith fractile of the income distribution. This route leads to another difficulty, however, because it does not take into account the magnitude of the (absolute or relative) income changes associated to movements between fractiles. Clearly, the same transition probabilities P between some income states (y_1, y_2, y_3) on the one hand, and between $(y_1, y_1 + \varepsilon, y_2 \times 10^6)$ on the other (where ε is small), represent very different mobility processes in any economically meaningful sense of the term. Both the prospects faced by individuals (expected incomes, risks) and the implied magnitude and persistence of inequality are radically different. This problem arises very concretely with intefractile matrices, due in particular to the skewedness of empirical income distributions. For instance, in the 1979 US income distribution (estimated from PSID data, see Hungerford (1993)), a move from the 1^{st} to the 2^{nd} decile would correspond on average to a near doubling of family income, while rising from the 5^{th} to the 6^{th} or the 6^{th} to the 7^{th} would bring an average gain of only about 15%. The increase would become again much more significant for a move from the 9^{th} to the 10^{th}, which raises average income by 55%. What is more, such ratios typically differ from one country to another, thereby rendering intercountry mobility comparisons based solely on interfractile transition matrices conceptually problematic.

To summarize, an ordering or index purely based on changes in ranks (independently of their income implications), or more generally on the properties of P alone, as most of those found in the literature are, can in general not fully account for economic mobility, whether in the sense of equalization of opportunities or in terms of intertemporal welfare consequences. This leads us to the view that the measurement of income mobility should be based on $(P, \mathbf{y})$, as opposed to P alone.

Remark 2. It is interesting to note the kind of "duality" which exist between our ordering and those of Dardanoni (1993), or Formby, Smith and Zheng (1995). In both cases one compares the distributions of expected future incomes (or their present values) under P and Q. In these two papers this comparison is conditional on a particular value of $\boldsymbol{\pi}$ (common steady–state vector in the first case; arbitrary fixed $\boldsymbol{\pi}$ in the second), but required to hold for all $\mathbf{y}$. In our ordering it is conditional on $\mathbf{y}$, but required to hold for all $\boldsymbol{\pi}$.

Could one insist on an ordering which was independent of the income state vector $\mathbf{y}$, as well as of the initial distribution $\boldsymbol{\pi}$? A simple example will make clear why no such "global–global"

[14]This is of course a problem with any analysis based on transition matrices. The fundamental difficulty is that a discrete transition process provides only a very partial representation of the actual law of motion for incomes, which in reality operates on a large subset of $\mathbb{R}_+$.

ordering exists, as soon as $n \geq 3$. Suppose that one could find some 3×3 monotone transition matrix $P = [p_{ij}]$ which implied a progressive or equalizing mobility process over *all* possible income supports, i.e. $P \succ_{\text{eq}}^{\mathbf{y}} I$, for all $y_1 < y_2 < y_3$. This would mean that

$$\frac{p_{11}y_1 + p_{12}y_2 + p_{13}y_3}{y_1} > \frac{p_{21}y_1 + p_{22}y_2 + p_{23}y_3}{y_2} > \frac{p_{31}y_1 + p_{32}y_2 + p_{33}y_3}{y_3},$$

For the first inequality to remain valid as y_2 tends to y_1 it must be that $p_{23} \leq p_{13}$, which can be rewritten as $p_{21} + p_{22} \geq p_{11} + p_{12}$. Similarly, by letting y_2 tend to y_3 in the second inequality we get $p_{21} \leq p_{31}$. But both of these conditions imply that future income is (weakly) stochastically *de*creasing in current income. The intuition is simple: as y_2 and y_1 become very close, for instance, transition probabilities between these two states become almost irrelevant. Thus, starting from a current income of y_1 versus y_2, having higher expected income growth $e_M(y)/y$ becomes equivalent to having a higher probability of rising to y_3. Progressivity is thus incompatible with strict monotonicity, and consistent with weak monotonicity only when $p_{13} = p_{23}$ and $p_{21} = p_{31}$. But then, when y_3 becomes very large compared to y_2 the second inequality clearly becomes violated, even in its weak form. In summary, for $n \geq 3$ there is no transition matrix which is more equalizing (progressive) than the identity over all income supports.

Remark 3. Because it makes mobility a "signed" concept, the ordering we define on monotone matrices also differs from previous ones in the sense that the identity mapping is not the smallest element –because it is not the worst one from the point of view of inequality of opportunities. This was shown earlier with continuous processes defined on all of $\mathbb{R}_+$ but, as a simpler example, consider the transitions

$$J \equiv \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

defined over $y_1 < y_2 < y_3$. In this scenario, which could be made stochastic, the middle class "falls through the floor" and joins the ranks of the poor –an obvious oversimplification of a real policy concern. Clearly, for any initial distribution such that $\pi_1 = 0$ (this could describe the whole population, or only some subgroup), the distribution of conditional expected incomes associated to J is more unequal than the one associated to I. (Moreover, the restriction $\pi_1 = 0$ is entirely due to the fact that a fixed lower bound $y_1 > 0$ is imposed on all incomes). The converse holds when $\pi_3 = 0$, so neither $I \succ_{eq} J$ nor $J \succ_{eq} I$ holds.[15]

That "immobility" is not the worst–case scenario is a natural property of *any* mobility criterion which cares about relative income movement not as an end in itself, but because of its impact on individual and aggregate welfare. Dardanoni (1993) also argued that mobility processes should be evaluated on the basis of their (ex–ante) equalizing properties, but restricted his ordering to mobility matrices which share the same steady–state, and his Lorenz comparisons to initial situations where

[15] This example shows that $\succcurlyeq_{\text{eq}}^{\mathbf{y}}$ satisfies neither the *monotonicity* assumption of Shorrocks (1978b) nor the axiom of *diagonalizing switches* of Atkinson (1983). Of course, this is not surprising since these properties really concern the measurement of relative income movement (progressive or regressive), as opposed to equality of opportunity.

the economy finds itself in this steady–state. In terms of our example, this means comparing I and J's welfare implications only when $\pi_2 = 0$, i.e. when no one is located in the range where the "tax scheme" imposed by J is regressive. In that case, and only then, $\Lambda_{\pi,J,} = \Lambda_{\pi,I}$. It is therefore the local nature of the inequality comparisons underlying Dardanoni's ordering (justified by a focus on "exchange mobility" within the invariant distribution) which prevents it from registering the fact that some forms of mobility may aggravate existing inequalities, hence result in a Lorenz curve for expected or permanent incomes which is strictly outside that which prevails under "immobility".[16]

A related point can be made in relation to non–welfare based approaches, which make no distinction between equalizing and disequalizing income movements. We shall illustrate it in relation to Shorrocks' (1978a) immobility criterion, but it applies equally to the "diagonals view" of mobility matrices (e.g., Atkinson (1983)), including eigenvalue–based criteria. For simplicity, let there be two periods, and let inequality be measured as the variance of relative incomes. Shorrock's index of immobility is then the ratio of the variance in two–period (undiscounted) relative incomes to the sum of the one-period variances. This ratio, call it R, is shown to be always less than one, and equal to one only in the case of perfect immobility. In discussing its economic interpretation, Shorrocks states that: *"mobility is regarded as the degree to which equalization occurs as the period is extended. This view captures the prime importance of mobility for economists".* We agree with these statements (taken in an ex–ante sense), but not with the claim that R necessarily captures this notion of mobility as an equalizing force: what R records is just relative movement, *whether equalizing or disequalizing.*[17] Conversely, if mobility is defined according to R, it is not the case that "mobility is unambiguously good," as stated by Shorrocks, even staying within the context of risk–neutral agents.

Consider, for instance, three individuals with period–zero incomes 2, 3 and 4, and period one incomes $2+\varepsilon$, 3 and $4-\varepsilon$, respectively. The two-period average income vector is thus $(4+\varepsilon, 6, 8-\varepsilon)$. Since this mobility process, call it $M(\varepsilon)$, simply transfers ε dollars from rich to poor, it is regressive for $\varepsilon < 0$, and progressive for $\varepsilon > 0$. Yet we know that for *any* value of ε, $M(\varepsilon)$ ranks lower according to Shorrock's (1978a) criterion than $M(0) = I$, even though when $\varepsilon < 0$ all that "mobility" does is to aggravate existing inequalities. Moreover, it is easily observed that: $R(2/3) = R(-2) = .80$, even though the first process increases the poor's income by a third at the expense of the rich, while the latter takes away all the poor's income to give it to the rich! Similarly, $R(.30) > R(-.5) > R(-1) > R(-2)$: as we move from the first process, which reduces initial disparities, to the next three, which increasingly accentuate them, the index records rising mobility.[18] **Figure 1**, which

[16] Formby, Smith and Zheng (1995) relax Dardanoni's (1993) joint–steady state requirement, and compare mobility processes starting from any *given* distribution π. This is still a local comparison, although no longer restricted to a common steady-state. They do not examine the issue of a minimal or smallest element.

[17] Indeed, Shorrocks proves that $R \leq 1$ but not that inequality in m–period incomes must necessarily decline with m (as pictured on his graph), nor that such declines are directly related to decreases in R. The example he provides does have these properties, however, as the process is not only progressive but even non-monotonic (incomes have a serial correlation of -1).

[18] Note that $M(\varepsilon)$ is deterministic, for simplicity. One could obviously introduce some uncertainty without changing any of what follows, or interpret the table's entries as conditional expected incomes rather than probability–one

plots $R(\varepsilon)$, makes clear more generally that the kind of mobility measured by Shorrock's index need not always be of the type such that *"equalization is more pronounced in a very mobile society".* By contrast, our ordering clearly conclude that $M(\varepsilon) \succ_{\text{eq}} M(\delta)$ whenever $\varepsilon > \delta$.[19]

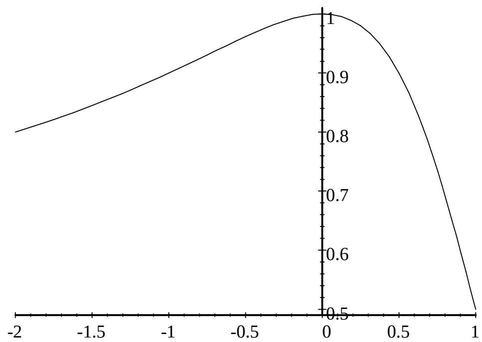

Figure 1: Shorrocks's immobility criterion R, versus the degree of equalization ε.

4 Summary Indices of Equalizing Mobility

Since our mobility ordering is the direct translation of inequality and progressivity measurement to a dynamic context, the same practical issues arise as in those literatures. When $P \succ_{\text{eq}} Q$ the first process is unambiguously better from the point of view of equalizing opportunities, but one would still like to quantify this difference. Even more importantly, when P and Q are not rankable according to $\succ_{\text{eq}}$, one would still like to compare them according to some unidimensional criterion, consistent with this ordering.

Fortunately, we do not need to devise and defend a new mobility index. According to our view, *mobility is progressivity,* so one should simply use existing and familiar measures of the latter –more specifically, of residual progressivity. Thus, given an economy (X, FM), one can simply:

(1) Compare inequality of initial incomes and inequality of conditional expected incomes (more generally, permanent incomes or intertemporal utilities), for example by taking the difference in the corresponding Gini coefficients:

$$\rho_M^{RS} \equiv Gini(F) - Gini(F \circ e_M^{-1}). \tag{12}$$

This gap, equal to the area between the two Lorenz curves, increases as M rises in the mobility ordering $\succcurlyeq_{\text{eq}}$. And indeed, ρ_M^{RS} is nothing but the familiar Reynolds–Smolensky (1977) *index of residual progressivity*, applied to the "redistributive scheme" e_M.[20]

realizations. In any case, we impose $\varepsilon \leq 1$ to maintain monotonicity.

[19] Of course, a mobility process which is globally regressive is very unrealistic. But, more generally, there might be equalization of opportunities over a certain range and disequalization over another, so it is important for a mobility index to be sensitive to the difference between the two. This is especially true when comparing a process M not to the identity but to some other process N, since what matters then is *relative progressivity,* which could well be decreasing even when both processes are progressive.

[20] For an extensive discussion of progressivity indices, see Lambert (1993).

(2) Alternatively, compute the average *residual elasticity* if the process is differentiable,

$$\bar{\eta} \equiv \int_X \frac{y e_M'(y)}{e_M(y)} \, dF(y), \tag{13}$$

or a discrete analogue (also consistent with $\succcurlyeq_{\text{eq}}$), if it is not:

$$\bar{\eta} \equiv \sum_{i=1}^{n-1} \tilde{\pi}_i \left(\frac{\ln(e_M(y_{i+1})/e_M(y_i))}{\ln(y_{i+1}/y_i)} \right), \tag{14}$$

where the weights are for instance set to $\tilde{\pi}_i \equiv \pi_i \Big/ \sum_{i=1}^{n-1} \pi_j$, $i = 1, ...n-1$.[21,22]

Given the parallel with taxation, it may also be interesting to characterize the mobility process in terms of implicit *average and marginal tax rates.* This metric is both familiar and intuitive, and readily allows comparisons between the reshu- ing of incomes due to the workings of the economy and those due to public policy. Normalizing both the initial and the expected distribution by their means, the expected average tax rate at any income level $y \in X$ is thus $t(y) \equiv 1 - (e_M(y)/y) \big/ \left(\mu_{\Lambda_{F,M}} / \mu_F \right)$; that is, it equals *one minus the expected growth rate* of relative income. The average marginal tax rate over the population is then $\bar{\tau} \equiv \int_X t'(y) \, dF(y)$ in the differentiable case.[23] The drawback is that this aggregate index is not always consistent with true progressivity.

5 Empirical Applications

5.1 Medium Term Earnings and Income Mobility in the US

We shall first illustrate our general methodology using PSID data from two sources, which correspond to different horizons and economic units. The first one is Gottschalk (1997), who provides the interquintile transition matrix M_{74}^{91} for individual male labor earnings over the 17 year period between 1974 and 1991. To (re)construct the income state vector, $\mathbf{y}$, we assign to each quintile its mean income level, as observed either in 1974 ($\mathbf{y}_{74}$) or in 1991 ($\mathbf{y}_{91}$). The distribution over $\mathbf{y}$ is $\boldsymbol{\pi} = (.2, .2, .2, .2, .2)$.

Table 1 here

[21] Alternatively, $\tilde{\pi}_i \equiv \pi_{i+1} \big/ \sum_{i=1}^{n-1} \pi_{j+1}$, or some combination of the two values. In our empirical applications we use the midpoint, but this choice makes almost no difference. Finally, one can similarly define the *income–weighted* elasticity, by replacing $dF(y)$ with $(y/\mu_F)dF(y)$ in (13), or the π_j's by $\pi_j \, y_j$'s in the definition of the weights $\tilde{\pi}_i$ entering (14).

[22] In the discrete case the $e_M(y_i)$'s are the components of Py, making $\bar{\eta}$ easy to compute. In the continuous case it is clear that when the mobility process is loglinear, $\ln x = \alpha + \beta \ln y + \varepsilon$, $\bar{\eta}$ is just equal to β, so the standard regression on individual data estimates the "right" measure of mobility.

[23] In the discrete case, the implicit marginal tax rate between y_i and y_{i+1} is given by

$$1 - \tau_i \equiv \left(\frac{e_M(y_{i+1}) - e_M(y_i)}{y_{i+1} - y_i} \right) \Bigg/ \left(\frac{\Sigma_{i=1}^n \pi_j e_M(y_j)}{\Sigma_{i=1}^n \pi_j y_j} \right),$$

for all $i = 1, ...n-1$. The average marginal tax rate is then $\bar{\tau} \equiv \sum_{i=1}^{n-1} \tilde{\pi}_i \tau_i$,with the weights $\tilde{\pi}_i$'s defined as before, or as in footnote 21.

As shown in **Table 1,** mobility prospects over the 17–year period reduce the Gini coefficient from .415 for initial incomes in 1974 to .226 or .255 for 1991 expected incomes, depending on whether it is assumed that the general rise in inequality which occurred in the late 70's and 80's was initially unexpected (column 1), or fully anticipated (column 2).[24] The Reynolds–Smolensky (1977) index of progressivity is thus . 189 or .160, respectively. This represents a substantial degree of equalization: by comparison, the same index for the US income tax system was only .031 in 1979 and .025 in 1988 (see Bishop, Chow and Formby (1997)).[25] In fact, the mobility process was globally *progressive* ($M_{74}^{91} \succ_{\text{eq}} I$), with average tax rates $t_i \equiv 1 - e_M(y_i)/y_i$ equal to $-282\%, -57.9\%, -16.9\%, +10.0\%$ and $+41.4\%$ (column 1). This means that the average person in the bottom quintile in 1974 could expect their earnings to grow 3.82 times as fast as the population mean over the following 17 years. Conversely, someone starting in the top quintile had expected relative losses of 10%. Note also that while inequality of outcomes had risen to .466 by 1991, at most .255/.466 = . 55% of this total reflected ex–ante unequal opportunities as of 1974. In other words, only 55% could have been predicted on the basis of initial earnings disparities.[26] Finally, column (3) describes mobility prospects starting in 1991, assuming that the transition matrix and quintile shares remain unchanged. The numbers obtained are comparable to those of the other columns.

Our second source is Hungerford (1993), who describes shorter transitions over the 7 year intervals 1969 − 1976 and 1979 − 1986, at the decile level. This allows for a finer description of states and transitions, but at the cost of higher sampling errors. There are also differences in the nature of the data; for instance Gottschalk's transitions relate to individual male earnings, whereas Hungerford's is for family incomes.[27] The income state vector is constructed as explained above, either from $\mathbf{y}_{69}$ or from $\mathbf{y}_{76}$ (used in place of $\mathbf{y}_{79}$, which is not reported by Hungerford); $\boldsymbol{\pi}$ is now ten–dimensional, with all entries equal to .1.

Table 2 here

The results are presented in **Table 2.** Mobility reduces the Gini coefficient from . 363 for 1969 initial conditions to between .217 and .235 for 1976 income prospects, depending on the terminal income state vector which is used (columns 1 and 2, respectively). The Reynolds–Smolensky index of progressivity is thus .146 or .128, which is still large compared to fiscal redistributions, but clearly less than in the earlier data. As intuition suggests, longer horizons allow more equalization

[24] Formally, the difference is whether conditional probabilities over quintiles in 1991 (computed using M_{74}^{91}) are translated into expected incomes using the initial distribution of relative incomes $(\boldsymbol{\pi}, \mathbf{y}_{74})$, or the ex–post realized one, $(\boldsymbol{\pi}, \mathbf{y}_{91})$. The fist assumption seems more realistic, since no one in the early 70's foretold the tide of rising inequality. Clearly, it makes little difference whether column 1 or column 2 is used.

[25] These numbers reflect only the progressivity of income taxes. If the incidence of public transfers and in–kind benefits was taken into account, the overall degree of fiscal progressivity would undoubtedly be higher.

[26] The Lorenz curve for 1991 income realizations is everywhere below the curve for 1994 incomes, which itself is below the curve for expected 1991 incomes (conditional on 1974 levels). This makes clear the fact that mobility equalizes (ex–ante) *opportunities,* but not (ex–post) *outcomes.*

[27] We refer the reader to these two sources for a full description of their data.

of opportunities –provided the mobility process is indeed progressive (rather than regressive, as in the last example of Section 3.3).

The numbers for the 1979–1986 period, given in the bottom panel of the table, are virtually identical to those of the earlier one.[28] This is consistent with Hungerford's conclusion, based on traditional tests, that there was no measurable change in mobility between the two periods. Indeed, the *"superdiagonals test"* of Proposition 3 shows that neither M_{69}^{76} nor M_{89}^{86} dominates the other in the sense of $\succcurlyeq_{\text{eq}}$. On the other hand, progressivity is satisfied by both matrices at almost every decile: the sequence $\mathbf{t} \equiv (t_i)_{i=1}^{10}$ of average tax rates for the ten deciles is decreasing, except at one or two points showing a slight increase, probably due to measurement error.[29] For instance, the 1979–1986 transition (as measured in column 3) yields

$$\mathbf{t} = (-136.4\%, -61.2\%, -41.0\%, -15.7\%, -16.8\%, -5.8\%, -4.2\%, -7.9\%, 15.0\%, 33.9\%).$$

Note that all but those families who start in the top two deciles have expected gains in their relative incomes.

5.2 Intergenerational Mobility: the U.S. versus Italy

Finally, we turn to intergenerational mobility, which is probably where equality of opportunity matters most. For this purpose we use the data of Rustichini, Ichino and Checchi (1999) for the United States and Italy. This consists of father–to–son transition probabilities between four "occupational income" classes, whose boundaries correspond to equiproportional increases in income.[30] This presentation of the data is particularly well suited to our purpose, as the transition matrices M_{US} and M_{IT} operate on (nearly) the same income state vector $\mathbf{y}$, up to a constant of proportionality (see the discussion in Section 3.3). By contrast, the distribution of incomes π_{US} and π_{IT} are very different, so that orderings which require a common (steady–state) distribution would not be applicable.

Based on standard indicators, Rustichini et al. find greater social mobility in the United States than in Italy. In addition to revisiting the issue with indices of progressive mobility, we shall apply our more stringent test of dominance (according to $\succ_{\text{eq}}$), asking which process better equalizes children's opportunities, for any arbitrary distribution of parental backgrounds. The results are summarized in **Table 3.**[31]

[28] The most relevant comparison is between numbers in the bottom and top panels of the *same column,* which compare how M_{69}^{76} and M_{79}^{86} operate on the same income distribution.

[29] These points often coincide with those where the transition matrices reported by Hungerford (1993) show slight non–monotonicities

[30] Fathers and sons are described by their occupations, and to each occupation is assigned its median income, as an indicator of long–term economic status. Finally, these "occupational incomes" are grouped into four intervals, whose boundaries differ by the same growth factor: denoting as $\underline{y}$ and $\bar{y}$ the minimum and maximum levels of income, class $k \in \{1, ...4\}$ corresponds $y \in [\underline{y}(1+g)^{k-1}, \underline{y}(1+g)^k]$, where $1 + g \equiv (\bar{y}/\underline{y})^{1/4}$. In the data, $g_{US} = 1.476$ and $g_{IT} = 1.467$. Since these numbers are quite close to each other, we define the (normalized) income state vector for both countries as $\mathbf{y} = (1, 1+\bar{g}, (1+\bar{g})^2, (1+\bar{g})^3)'$, where $\bar{g} \equiv \sqrt{g_{US} g_{IT}} = 1.472$.

[31] In particular, each country's mobility process, M_{US} or M_{IT}, is evaluated on both the US and the Italian income distributions, that is, on π_{US} (first column) as well as on π_{IT} (second column).

Table 3 here

The first observation is that there is much more cross-sectional inequality in the US than Italy. As illustrated on **Figure 2**, the US Lorenz curve for fathers' occupational incomes is everywhere below its Italian counterpart, with respective Gini coefficients of $G_{US} = .200$ and $G_{IT} = .160$. When we look at the extent to which these differences in social origins determine the next generation's opportunities, however, the picture is very different. The two Lorenz curves for sons' conditional expected incomes are virtually indistinguishable, with Ginis of $\hat{G}_{US} = .063$ and $\hat{G}_{IT} = .056$. The corresponding indices of progressivity are $\rho_{US}^{RS} = .137$ and $\rho_{IT}^{RS} = .104$. This comparison is not really fair, however, because there is less to equalize in Italy in the first place. As explained in the earlier sections, the appropriate comparison is between the effects of the two mobility processes on a *common* initial income distribution. This means comparing the entries in the top and bottom panels of **Table 3** within the *same column,* as opposed to across columns. For instance, had the Italian mobility process M_{IT} operated on the US distribution of fathers' incomes, the Gini in sons' opportunities would have been reduced by $\rho^{RS} = .121$; this is still less than the US number of $\rho_{US}^{RS} = .137$, but the gap is much smaller than before. Conversely, had Italy "imported" the US mobility process, inequality of opportunities would have fallen by only $\rho^{RS} = .116$.

Note that whether looking at Ginis, elasticities or average marginal tax rates, the ranking is the same for both initial distributions. And indeed, the last row of **Table 4** reveals that the US mobility process is in fact *unambiguously more egalitarian* than the Italian one: the superdiagonals test of Theorem 2 yields

$$M_{US} \succ_{\text{eq}} M_{IT} \succ_{\text{eq}} I.$$

This is rather remarkable, given the stringency of the requirements, even if the magnitude of the differences is fairly moderate. Thus, according to this data, the American intergenerational mobility process is a greater equalizer of *opportunities* than the Italian one. This contrasts sharply with *outcomes*, which are not at all equalized: in both countries, sons' ex–post income realizations exhibit the same degree of inequality as fathers' incomes. Finally, it is interesting to compute the actual profile of Italy–to–US relative net–of tax rates, or *relative growth rates in relative incomes,* $(1 - t_{IT}(y_i)) \,/\, (1 - t_{US}(y_i)) \;=\; e_{M_{IT}}(y_i)/e_{M_{US}}(y_i)$. The values are .90, . 91, . 95, . and .97, which shows that most of the mobility difference between the two countries occurs when the father rises above the second income class. The advantage thereby conferred to the son is markedly stronger in Italy than in the United States.

Appendix

Proof of Theorem 1. The equivalence of (ii) and (iii) is easy to establish, so we focus here on the equivalence of (i) and (ii).[32] For this purpose, we need the following two claims.

Claim 1. For any integrable functions $f, g : [0,1] \to \mathbb{R}_{++}$ such that f/g is decreasing, we have

$$\left(\int_0^p f(q)\ dq\right) \Big/ \left(\int_0^1 f(r)\ dr\right) \geq \left(\int_0^p g(q)\ dq\right) \Big/ \left(\int_0^1 g(r)\ dr\right) \text{ for all } p \in [0,1]. \qquad \text{(A.1)}$$

Proof of Claim 1. Fix any $p \in (0,1]$, and note that

$$\begin{aligned} & \left(\int_0^p f(q)\ dq\right)\left(\int_0^1 g(r)\ dr\right) - \left(\int_0^p g(q)\ dq\right)\left(\int_0^1 f(r)\ dr\right) \\ = & \int_0^p \int_p^1 \left(f(q)g(r) - g(q)f(r)\right)\ drdq = \int_0^p \int_p^1 g(r)g(q)\left(\frac{f(q)}{g(q)} - \frac{f(r)}{g(r)}\right)\ drdq. \end{aligned}$$

That the last expression is nonnegative follows immediately from the fact that f/g is decreasing. ‖

Claim 2. For all $p \in (0,1]$ and $F \in \mathcal{F}(X)$, $\Lambda_{F,M}^{-1}(p) = e_M(F^{-1}(p))$ and $\Lambda_{F,N}^{-1}(p) = e_N(F^{-1}(p))$.

Proof of Claim 2. Take any $p \in (0,1]$, and note that

$$\begin{aligned} \Lambda_{F,M}^{-1}(p) &= \inf\{x \in supp(\Lambda_{F,M}) : \Lambda_{F,M}(x) \geq p\} = \inf\{x \in e_M(X) : F(e_M^{-1}(x)) \geq p\} \\ &= \inf\{e_M(y) \in X : F(y) \geq p\} = e_M\left(\inf\{y \in X : F(y) \geq p\}\right) = e_M(F^{-1}(p)), \end{aligned}$$

where the fourth equality follows from the continuity and strict monotonicity of e_M on X. ‖

Now, to show that (ii) implies (i), first extend e_M and e_N to $X \cup \{0\}$ by setting $e_M(0) = \inf e_M(X)$ and $e_N(0) = \inf e_N(X)$. Next, let $f \equiv e_M \circ F^{-1}$ and $g \equiv e_N \circ F^{-1}$. Since $F^{-1}(0) = 0$ and $F^{-1}(0,1] \subseteq supp(F) \subseteq X$, the functions f and g are well-defined. It is easy to see that they are also monotonic, and satisfy all the assumptions of Claim 1. Therefore, by Claims 1 and 2:

$$L_{\Lambda_{F,M}}(p) = \int_0^p f(q)\ dq \Big/ \left(\int_0^1 f(r)\ dr\right) \geq \int_0^p g(q)\ dq \Big/ \left(\int_0^1 g(r)\ dr\right) = L_{\Lambda_{F,N}}(p)$$

for all $p \in (0,1]$; that is, $M \succcurlyeq_{\text{eq}} N$.

Conversely, suppose that there exist $a, b \in X$ with $b > a > 0$ such that $e_M(b)/e_N(b) > e_M(a)/e_N(a)$. We need to show that $\Lambda_{F,M} \succsim_L \Lambda_{F,N}$ does not hold for some $F \in \mathcal{F}(X)$. Indeed, let $F(q) = \frac{1}{2}\mathbf{1}_{[a,b)} + \frac{1}{2}\mathbf{1}_{[b,\infty)}$, and observe that

$$L_{\Lambda_{F,M}}(1/2) = \frac{e_M(a)}{e_M(a) + e_M(b)} < \frac{e_N(a)}{e_N(a) + e_N(b)} = L_{\Lambda_{F,N}}(1/2),$$

which completes the proof. ■

[32] Various versions of this result in the *tax context* have been proved, for instance, in Jakobsson (1976), Fellman (1976), Lambert (1993) or Le Breton et al. (1996). We were, however, unable to find in the literature the more general version needed for Theorem 1, which covers both the continuous and discrete cases (as well as mixtures of the two). At any rate, our proof is short and self-contained.

Proof of Proposition 1. Fix any $t \geq 1$ and any $\theta \in (0, \infty]$, and observe first that:

$$\begin{aligned} E_{M^{(t+1)}}(y,\theta) &= \int_0^\theta x\ dM^{(t+1)}(x \mid y) = \int_0^\theta x\ d\left(\int_X M(x \mid z)\, dM^{(t)}(z \mid y)\right) \\ &= \int_0^\theta \int_X x\ dM(x \mid z)\, dM^{(t)}(z \mid y) = \int_X E_M(z,\theta)\ dM^{(t)}(z \mid y). \end{aligned}$$

Now, define $\Psi(z) \equiv E_M(z,\theta)/z$, for all $z \in X \cap \mathbb{R}_{++}$. Since X is closed, $\mathbb{R}_{++} \backslash X$ is open (in $\mathbb{R}_{++}$) and hence there exists a countable collection of disjoint open intervals $\{(a_n, b_n)\}_{n=1}^\infty$ in $\mathbb{R}_{++}$ such that $\mathbb{R}_{++} \backslash X = \cup (a_n, b_n)$. By linear interpolation we can then extend Ψ to a continuous, decreasing function on all of $\mathbb{R}_{++}$; we denote this extension again by Ψ for simplicity. We then have:

$$\frac{1}{y} E_{M^{(t+1)}}(y,\theta) = \int_X \Psi(z) \frac{z}{y}\ dM^{(t)}(z \mid y) = \int_0^\infty \Psi(z) \frac{z}{y}\ dM^{(t)}(z \mid y). \tag{A.2}$$

Integrating by parts with respect to z, then, we find

$$\begin{aligned} \frac{1}{y} E_{M^{(t+1)}}(y,\theta) &= \left[\Psi(z) \int_0^z \frac{u}{y}\ dM^{(t)}(u \mid y)\right]_0^\infty - \int_0^\infty \left(\int_0^z \frac{u}{y}\, dM^{(t)}(u \mid y)\right) d\Psi(z) \\ &= \left[\Psi(z) \frac{E_{M^{(t)}}(y,z)}{y}\right]_0^\infty - \int_0^\infty \frac{E_{M^{(t)}}(y,z)}{y}\ d\Psi(z) \\ &= \lim_{z \to \infty} \Psi(z) \left(\frac{e_{M^{(t)}}(y)}{y}\right) - \int_0^\infty \frac{E_{M^{(t)}}(y,z)}{y} d\Psi(z). \end{aligned} \tag{A.3}$$

Since Ψ is positive and decreasing in z, it is clear from this expression that $E_{M^{(t+1)}}(y,\theta)/y$ is decreasing in y whenever $M^{(t)}$ is strongly equalizing. The claimed result follows by induction. ■

Proof of Theorem 2. The equivalence of (i) and (ii) follows from Theorem 1. To show the equivalence with (iii), note that (i) holds if and only if

$$\frac{\mathbf{e}_i' P\mathbf{y}}{\mathbf{e}_i' Q\mathbf{y}} \geq \frac{\mathbf{e}_{i+1}' P\mathbf{y}}{\mathbf{e}_{i+1}' Q\mathbf{y}}, \quad i = 1, ..., n-1, \tag{A.4}$$

where $\mathbf{e}_i$ is the ith unit vector. But his can be rewritten as $\mathbf{e}_i' P\mathbf{y}\mathbf{y}' Q' \mathbf{e}_{i+1} \geq \mathbf{e}_i' Q\mathbf{y}\mathbf{y}' P' \mathbf{e}_{i+1}$, or

$$\mathbf{e}_i' P\mathbf{y}\mathbf{y}' Q' \mathbf{e}_{i+1} - \mathbf{e}_{i+1}' P\mathbf{y}\mathbf{y}' Q' \mathbf{e}_i \geq 0, \quad i = 1, ..., n-1,$$

which is equivalent to (iii). ■

Proof of Proposition 3 The first statement follows directly from Proposition 1. For the second one, take any $\rho \in (0,1)$ and observe that since $P(\rho) = (I - \rho P)^{-1} = \sum_{t=0}^\infty \rho^t P^t$, we have:

$$\begin{aligned} (D^* - D_*)[P(\rho) u(\mathbf{y}) u(\mathbf{y})' I)] &= (1-\rho)(D^* - D_*) \left[\sum_{t=0}^\infty \rho^t \left(P^t u(\mathbf{y}) u(\mathbf{y})'\right)\right] \\ &= (1-\rho) \sum_{t=0}^\infty \rho^t (D^* - D_*)[P^t u(\mathbf{y}) u(\mathbf{y})']. \end{aligned}$$

Since P is strongly equalizing (given $u(\mathbf{y})$) we have $P^t \succcurlyeq_{\text{eq}}^{u(\mathbf{y})} I$ for each $t \in \mathbb{N}$, hence $(D^* - D_*)[P^t u(\mathbf{y}) u(\mathbf{y})'] \geq 0$ by Theorem 2. Thus $(D^* - D_*)[P(\rho) u(\mathbf{y}) u(\mathbf{y})'] \geq 0$, implying $P(\rho) \succcurlyeq_{\text{eq}}^{u(\mathbf{y})} I$ by Theorem 2. ■

References

[1] Atkinson, A. B. (1970). "On the Measurement of Inequality," *Journal of Economic Theory* 2, 244-263.

[2] Atkinson, A. B. (1983). "The Measurement of Economic Mobility," in *Social Justice and Public Policy*, ed. by A. B. Atkinson, Cambridge: MIT Press.

[3] Bartholomew, D.J (1982) "Stochastic Models for Social Processes," 3^d ed., Wiley: New York.

[4] Bénabou, R. and Ok, E. A. (2000). "Social Mobility and the Demand for Redistribution: the POUM Hypothesis," *Quarterly Journal of Economics,* forthcoming.

[5] Bishop. J., Chow, V. and Formby J. (1995). "The Redistributive Effect of Direct Taxes: A Comparison of Six Luxembourg Income Study-Countries," *Journal of Income Distribution*, 5, 65-90.

[6] Conlisk, J. (1989) "Ranking Mobility Matrices," *Economics Letters,* 29, 231-235.

[7] Conlisk, J. (1990). "Monotone Mobility Matrices." *Journal of Mathematical Sociology,* 15, 173-191.

[8] Dardanoni, V., (1993). "Measuring Social Mobility." *Journal of Economic Theory*, 61, 372-394.

[9] Dardanoni, V. (1995). "Income Distribution Dynamics: Monotone Markov Chains Make Light Work," *Social Choice and Welfare* 12, 181-192.

[10] Dasgupta, P., Sen, A. and Starett, D. (1973) "Notes on the Measurement of Inequality," *Journal of Economic Theory,* 6(2), 180–187.

[11] Fellman, J. (1976). "The Effect of Transformation on Lorenz Curves," Econometrica, 44, 823-4.

[12] Fields, G. and Ok, E. A. (1999a). "The Measurement of Income Mobility: An Introduction to the Literature," in *Handbook of Inequality Measurement,* ed. by J. Silber, Dordrecht, Germany: Kluwer Academic Publishers, 557–596.

[13] Fields, G. and Ok, E. A. (1999b). "Measuring Movement of Incomes," *Economica*, 66, 455–471.

[14] Formby, J., Smith, J. and Zheng B. (1995). "Economic Growth, Welfare and the Measurement of Social Mobility". University of Alabama–Tuscaloosa, mimeo, November.

[15] Foster, J. E. (1985). "Inequality Measurement," in *Fair Allocation*, ed. by H. P. Young. Providence: American Mathematical Society Proceedings in Applied Mathematics, Vol. 33.

[16] Gottschalk, P. (1997) "Inequality, Income Growth and Mobility: The Basic Facts," *Journal of Economic Perspectives,* 11(2), 21–40.

[17] Gottschalk, P. (1997) and Danziger, S. "Family Income Mobility: How Much is There and Has It Changed?", Boston College mimeo.

[18] Gottschalk, P. and Spolaore E. (2000) "On the Evaluation of Economic Mobility," Boston College mimeo, March.

[19] Jakobsson, U. (1976). "On the Measurement of the Degree of Progressivity," *Journal of Public Economics,* 5, 161–168.

[20] Hungerford, T.L., (1993). "U.S. Income Mobility in the Seventies and Eighties." *Review of Income and Wealth*, 31(4), 403–417.

[21] Kanbur, R. and Stromberg, J-O. (1988). "Income Transitions and Income Distribution Dominance," *Journal of Economic Theory* 45, 408-416.

[22] Kanbur, R. and Stiglitz, J. (1986). "Intergenerational Mobility and Dynastic Inequality," Princeton University mimeo.

[23] Keilson, J. and A. Ketser, (1977). "Monotone Matrices and Monotone Markov Chains." *Stochastic Processes and their Applications,* 5, 231-241.

[24] Kolm, S.C. (1960) "The Optimal Production of Justice," in *Public Economics*, J. Margolis and H. Guitton eds., MacMillan: London.

[25] Lambert, P. (1993) "The Distribution and Redistribution of Income: A Mathematical Analysis." 2nd. ed. Manchester: Manchester University Press.

[26] Le Breton, M., Moyes, P. and Trannoy A. (1996). "Inequality Reducing Properties of Composite Taxation," *Journal of Economic Theory* 69, 71-103.

[27] Loury, G. (1981) "Intergenerational Transfers and the Distribution of Earnings," *Econometrica,* 49, 843–867.

[28] Reynolds, M. and Smolensky, E.E. (1977) *"Public Expenditures, Taxes and the Distribution of Income: The United States 1950, 1961, 1970"*. New York: Academic Press.

[29] Rustichini, A., Ichino, A. and Checchi, D. (1999) "More Equal But Less Mobile? Education Financing and Intergenerational Mobility in Italy and the US," *Journal of Public Economics,* December, 70(3), pp. 399.424.

[30] Sen, A. K. (1997). *On Economic Inequality.* Expanded Edition, Oxford: Clarendon Press.

[31] Shorrocks, A. (1978a). "Income Inequality and Income Mobility," *Journal of Economic Theory,* 19, 376–393.

[32] Shorrocks, A. (1978b). The Measurement of Mobility," *Econometrica* 45, 1013-1024.

[33] Stokey, N. (1998), "Shirtsleeves to Shirtsleeves: The Economics of Social Mobility," in *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures 1983-1997*, N. L. Schwartz, D. Jacobs and E. Kalai eds., Cambridge: Econometric Society Monographs No. 23.

[34] United States Bureau of the Census, *"Current Population Reports"* and data available at http://www.census.gov/ftp/pub/hhes/www/income.html. U.S. Department of Commerce, Income Statistics Branch, HHES Division, Washington, D.C. 20233-8500.

.

Initial and Expected Incomes	$(\pi, y_{74}) \rightarrow (\pi, M_{74}^{91} \cdot y_{74})$	$(\pi, y_{74}) \rightarrow (\pi, M_{74}^{91} \cdot y_{91})$	$(\pi, y_{91}) \rightarrow (\pi, M_{74}^{91} \cdot y_{91})$
Gini	.415 $\rightarrow$ .226	.415$\rightarrow$. 255	.466 $\rightarrow$.255
ΔGini = ρ^{RS}	.137	.160	.211
1-Average Residual Elasticity	.442	.391	.450
Average Marginal Tax Rate	.458	.628	.463
Dominance Tests: $M_{74}^{91} \succ_{\text{eq}}^{\mathbf{y}} I$ and $M_{74}^{91} \succ_{\text{eq}}^{\mathbf{y}} I$, for $y \in \{y_{74}, y_{91}\}$			

Table 1: Male earnings mobility in the US

Transitions: M_{69}^{76}			
Initial and Expected Incomes	$(\pi, y_{69}) \rightarrow (\pi, M_{69}^{76} \cdot y_{69})$	$(\pi, y_{69}) \rightarrow (\pi, M_{69}^{76} \cdot y_{79})$	$(\pi, y_{79}) \rightarrow (\pi, M_{69}^{76} \cdot y_{79})$
Gini	.362 $\rightarrow$ .217	.362 $\rightarrow$. 235	.393 $\rightarrow$. 235
ΔGini = ρ^{RS}	.137	.127	.157
1-Average Residual Elasticity	.432	.378	.466
Average Marginal Tax Rate	.349	.304	.340
Transitions: M_{79}^{86}			
Initial and Expected Incomes	$(\pi, y_{69}) \rightarrow (\pi, M_{79}^{86} \cdot y_{79})$	$(\pi, y_{69}) \rightarrow (\pi, M_{79.}^{86} \cdot y_{79})$	$(\pi, y_{79}) \rightarrow (\pi, M_{79}^{86} \cdot y_{79})$
Gini	.362 $\rightarrow$. 219	.362 $\rightarrow$. .238	.393 $\rightarrow$. .238
ΔGini = ρ^{RS}	.144	. 15 9	.155
1-Average Residual Elasticity	.444	.392	.430
Average Marginal Tax Rate	.353	.382	.318
Dominance Tests: see the text.			

Table 2: Famility income mobility in the US

United States Mobility		
Initial and Expected Incomes	$(\pi_{US}, y) \rightarrow (\pi_{US}, M_{US} \cdot y)$	$(\pi_{IT}, y) \rightarrow (\pi_{IT}, M_{US} \cdot y)$
Gini	.200 → .063	.160→.044
ΔGini $= \rho^{RS}$	.137	.116
1-Average Residual Elasticity	.727	.733
Average Marginal Tax Rate	.707	.752
Italian Mobility		
Initial and Expected Incomes	$(\pi_{US}, y) \rightarrow (\pi_{US}, M_{IT} \cdot y)$	$(\pi_{IT}, y) \rightarrow (\pi_{IT}, M_{IT} \cdot y)$
Gini	.200→.078	.160 → . 056
ΔGini $= \rho^{RS}$	.121	.104
1-Average Residual Elasticity	..640	.669
Average Marginal Tax Rate	.640	.688
Dominance Tests: $M_{US} \succ^{\mathbf{y}}_{\text{eq}} M_{IT} \succ^{\mathbf{y}}_{\text{eq}} I$		

Table 3: Intergenerational mobility in the US and Italy

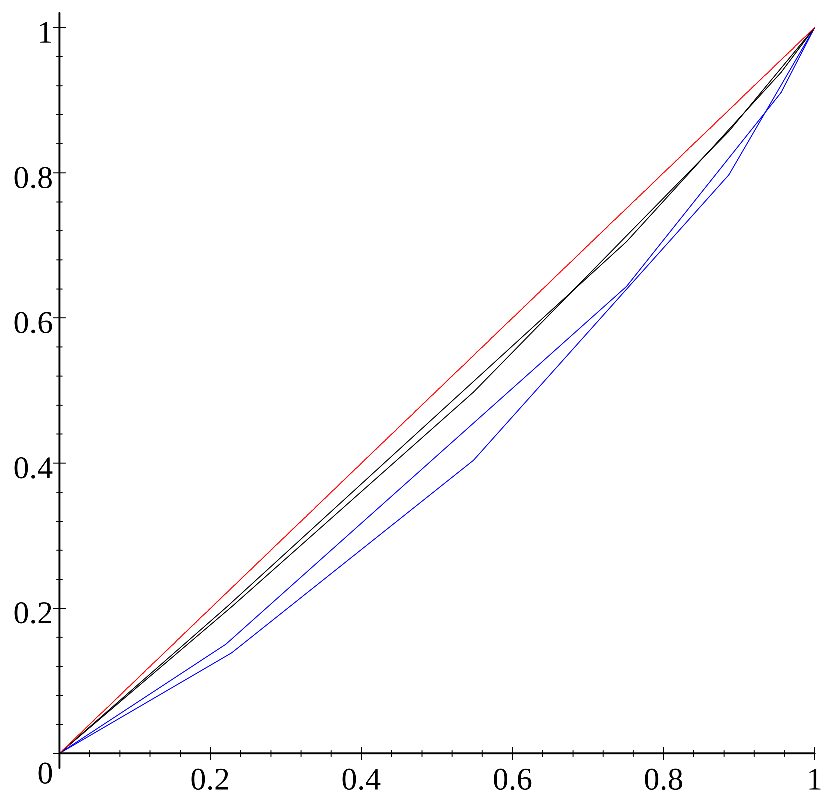

Figure 2: Lorenz curves for fathers' incomes and son's predicted incomes in the US and Italy: —— $= L_{US}$, – – – $= L_{IT}$, ,······· $= \hat{L}_{US}$, -·-·-·- $= \hat{L}_{IT}$