NBER WORKING PAPER SERIES

SOCIAL MOBILITY AND THE DEMAND FOR REDISTRIBUTION: THE POUM HYPOTHESIS

Roland Bénabou
Efe A. Ok

Working Paper 6795
http://www.nber.org/papers/w6795

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 1998

We thank Abhijit Banerjee and Ignacio Ortuno-Ortin for their helpful comments. Financial support from the National Science Foundation, the MacArthur Foundation and the C.V. Starr Center is gratefully acknowledged. The views expressed here are those of the author and do not reflect those of the National Bureau of Economic Research.

Social Mobility and the Demand for
Redistribution: The POUM Hypothesis
Roland Bénabou and Efe A. Ok
NBER Working Paper No. 6795
November 1998
JEL No. D31, D72, P16, H20

ABSTRACT

Even relatively poor people oppose high rates of redistribution because of the anticipation that they or their children may move up the income ladder. This "prospect of upward mobility" (POUM) hypothesis commonly advanced as an explanation of why most democracies do not engage in large-scale expropriation and highly progressive redistribution. But is it compatible with everyone -- especially the poor -- holding rational expectations, given that not everyone can simultaneously expect to end up richer than average? This paper establishes the formal basis for the POUM hypothesis. 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 laissez-faire coalition is larger, the more concave the transition function and the longer the policy horizon. We illustrate the general analysis with an example (calibrated to the U.S.) 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. We also analyze empirical mobility matrices from the PSID and find that the POUM effect is indeed a significant feature of the data.

Roland Bénabou
IDEI
Université des Sciences Sociales
Place Anatole France
F-31042 Toulouse Cedex
France
and NBER
benabou@cict.fr

Efe A. Ok
Department of Economics
New York University
269 Mercer Street
New York, NY 10011

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

Introduction

The following argument is 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] The question we ask in this paper is simple: does this story make sense?

To the best of our knowledge this "prospect of upward mobility" hypothesis has never been formalized, which seems rather surprising for such a recurrent theme in the political economy of redistribution. A moment's reflection may explain why: this "intuitive" argument is flawed, or more precisely severely incomplete, if it is to be compatible with everyone holding rational expectations over their income dynamics.

There are three implicit premises. The first is that redistributive 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, or that future income is fairly predictable. Otherwise, they must realize that it may go down as well as up, in which case redistribution provides valuable insurance. 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. Finally, even if a positive fraction of agents below the mean today can rationally expect to be above it tomorrow, the expected incomes 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?

The contribution of this paper is to formally examine the "prospect of upward mobility" (POUM) 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.

[1] See for example Roemer (1998) or Putterman (1996). There are of course other complementary explanations, such as the deadweight loss from taxation, or bias against the poor in the political system; Putterman (1996) provides a review. The prospect of upward mobility hypothesis is also related to Hirschman's (1973) famous "tunnel effect," although his argument is more about how individuals or groups use observations on the mobility experience of others to update their beliefs concerning their own prospects.

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 77% of the population below mean income, so that they would support purely contemporary redistributions. Yet when voters look ahead to the next period, 67% 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.

There are two intuitive ways to understand the key role of concavity –a requirement which is stronger than simple mean reversion or convergence of incomes.[2] 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, $y' = f(y)$. Without loss of generality, normalize f so that someone with income equal to the average, $y = \mu$, 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 poorer, can thus *rationally* expect to be richer than average in the next period, and will therefore oppose future redistributions. Alternatively, let us now normalize the transition function so that tomorrow's and today's mean incomes coincide, $\mu' = \mu$. To say that f is concave is then equivalent to saying that y' is obtained from y through a progressive, balanced budget, redistributive scheme, which shifts the Lorenz curve upwards. 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.

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. 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 equalizing 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.

[2]The latter could occur, for instance, with a (globally or locally) linear or even *convex* transition process, as long as its slope was less than one in the relevant range.

The paper will formalize these intuitions and examine their robustness to the presence of aggregate and idiosyncratic uncertainty, discounting over longer horizons, risk–aversion, and endogenous mobility. It will also provide two important analytical examples. The first one is the Markov process mentioned earlier, 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. A calibrated version of this simple model fits the main features of the US income distribution and intergenerational persistence rather well. The second example is a log–linear, log-normal process where complete closed–form solutions are obtained. This autoregressive specification is common in econometric studies of income persistence, and implies a strictly concave transition function between income levels. Finally, the paper offers a direct 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.

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 rather different. Piketty's main concern is to explain persistent differences in attitudes towards redistribution as resulting from divergent beliefs about the determinants of social mobility. He therefore studies the inference problem of agents who share the same redistributive goals but have conflicting priors over the contributions of family background and individual effort to personal success. Because their main source of information is personal or dynastic experimentation through costly effort they may never completely learn the true mobility process, and thus end up with different long–run posterior beliefs over the incentive costs of taxation. We focus instead on agents who have complete knowledge of the true (stochastic) mobility process, and whose primary concern when voting is to maximize the present value of their after–tax incomes, or that of their progeny, rather than a common social objective function. 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 is organized as follows. Section 1 introduces basic concepts and notations. Sections 2 and 3 develop the main theoretical analysis, first in a deterministic, then in a stochastic context. Section 4 discusses some extensions, while Section 5 works out the lognormal example. Finally, Section 6 conducts the empirical exercise. All the proofs are gathered in the appendix.

1 Preliminaries

We consider an endowment economy, populated by a continuum of individuals whose initial levels of income lie in some interval $X \equiv [0, \bar{y}]$, $0 < \bar{y} \leq \infty$. We shall often simplify the notation by identifying each individual with her initial income $y \in X$.

An **income distribution** is defined as a strictly increasing function $F : X \to [0,1]$ such that $F(0) = 0$, $\lim_{y \to \bar{y}} F(y) = 1$ and $\mu_F \equiv \int_X y\, dF < \infty$. The class of all such distributions will be denoted by $\mathcal{F}$. We shall be particularly interested in income distributions which are positively skewed, or more generally whose mean is at least as great as their median, denoted $m_F \equiv F^{-1}(1/2)$. This subset of $\mathcal{F}$ will be denoted $\mathcal{F}_+$.

A **redistribution scheme** is defined as any function $\tau : \mathcal{F} \to \mathcal{F}$ which preserves mean income: $\mu_{\tau(F)} \equiv \int_X y\, d\tau(F) = \mu_F$, for all $F \in \mathcal{F}$. We thus abstract from any deadweight losses which such a scheme might realistically entail, so as to better highlight the different mechanism which is our focus. Both represent complementary forces which reduce the demand for redistribution, and could 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 β and the collected revenue is redistributed in a lump-sum manner.[3] Given a pre-tax income distribution $F(y)$, the post-tax distribution is then $\tau_\beta(F) \equiv F \circ r_\beta^{-1}$, where $r(y) \equiv (1-\beta)y + \beta\mu_F$ for all $y \in X$. We shall mostly work with just the two extreme members of the set $\mathcal{P} \equiv \{\tau_\beta \,|\, 0 \leq \beta \leq 1\}$, namely, τ_0 and τ_1. Clearly, τ_0 corresponds to the "*laissez-faire*" policy, $r_0(y) \equiv y$, whereas τ_1 corresponds to "*complete equalization,*" $r_1(y) \equiv \mu_F$.

Our focus on these two polar cases is not as restrictive as it might initially appear. First, the analysis directly extends to the comparison between an arbitrary pair of proportional redistribution schemes, say τ_β and $\tau_{\beta'}$, with $0 \leq \beta < \beta' \leq 1$. Second, τ_0 and τ_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, they are the only candidates in this class that can be majority rule (Condorcet) winners. Thus, for any distribution with median income below the mean, τ_1 beats every other linear scheme under pairwise majority voting if individuals' preferences are defined only in terms of their *present* disposable incomes. We shall see that this conclusion may be dramatically altered when individuals' voting behavior also incorporates concerns about their *future* incomes. Finally, we do provide extensions of the analysis to certain non-linear (progressive or regressive) schemes in Section 5.

The third key feature of the economy is the *mobility process.* We shall initially focus on deterministic income dynamics, then incorporate random shocks, starting in Section 3. While the stochastic case is obviously of primary interest, the deterministic one provides more transparent intuitions, as well as useful intermediate results. All proofs are gathered in the appendix.

[3]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. More fundamentally, when unrestricted non-linear redistributive schemes are allowed there is no voting equilibrium (in pure strategies): the core of the voting game is empty.

2 Income Dynamics and Voting under Certainty

It will be assumed for now that individual (pre-tax) incomes or endowments evolve through time according to a **transition function** $f : X \to X$ which is continuous and strictly increasing.[4] The resulting income stream of an individual with initial endowment $y \in X$ is then $y, f(y), f^2(y), \ldots, f^t(y), \ldots$, 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 about voting and income mobility to its most elementary form, we focus at first on a two-period 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 the voting behavior of the constituent agents could be the well-being of their offspring, who will be subject to the tax policy designed by the current generation.[5] Accordingly, agent $y \in X$ votes for τ_1 over τ_0 if she expects her period one earnings to be below the per capita average:

$$f(y) < \int_X f \, dF_0 = \mu_{F_1}. \tag{1}$$

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

$$f(\mu_{F_0}) = f\left(\int_X y \, dF_0\right) > \int_X f \, dF_0 = \mu_{F_1}, \tag{2}$$

so the agent with mean income at date zero will oppose date one redistributions.[6] 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

$$f(y_f^*) = \mu_{F_1}. \tag{3}$$

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– characterizes concavity, the latter is both neces-

[4] More generally, f describes the transitions which agents expect to occur, and therefore base their votes upon.

[5] One can also think of the case where the tax rate is set for two periods, and agents care about some present value of income, as a mixture between two polar situations: that where only current income matters (as usually assumed), and that where only future income matters (as here). This perhaps more realistic intermediate setup is covered by the multi–period model studied in the next subsection.

[6] Note that Jensen's inequality does not presuppose, but rather establishes, the fact that F_1 has a finite mean. Indeed, denoting the right derivative of f (which exists everywhere by concavity) by f'_+, we have $f(y) \le f(\mu_{F_0}) + f'_+(\mu_{F_0})(y - \mu_{F_0})$ for all $y \in X$. Thus: $\mu_{F_1} = \int_X f(y)\,dF_0(y) \le \int_X f(\mu_{F_0})\,dF_0(y) + f'_+(\mu_{F_0}) \int_X (y - \mu_{F_0})\,dF_0(y) = f(\mu_{F_0}) < \infty$. This result will be used repeatedly in the paper.

sary and sufficient for the "prospect of upward mobility" hypothesis to be valid, under any linear redistribution scheme.[7]

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

(a) 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 τ_1 over τ_0, while all those in $(y_f^*, \bar{y}]$ vote for τ_0 over τ_1.

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

Compared to the standard case where individuals base their votes solely on the effect of taxation on their current incomes, we see that popular support for redistribution falls by a measure $F_0(\mu_{F_0}) - F_0(y_f^*) > 0$. Moreover, the underlying intuition suggests that the more concave is the transition function, the fewer people should vote for redistribution. This simple result, shown below, will turn out to be extremely 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 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 $f \circ g^{-1} \in \mathcal{T}$. Clearly, the relation $\succ$ is irreflexive, asymmetric and transitive, hence it is a strict partial ordering on $\mathcal{T}$.[8] The following elementary observation makes clear why it is relevant in our context.

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

The interpretation is straightforward. If two societies start from an identical pre-tax income distribution, the demand for redistribution will be lower in the one whose transition function already equalizes incomes at a faster (i.e., more marginally progressive) rate.

Can the prospect of upward mobility be strong enough for τ_0 to beat τ_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.[9] The construction of such a transition function is discussed below, together with a caveat. When combined with Proposition 2 and a continuity property, it allows us to show the following, more general result.

---

[7] For any τ_β and $\tau_{\beta'}$ in $\mathcal{P}$ such that $0 \leqslant \beta < \beta' \leqslant 1$, agent $y \in X$ votes for $\tau_{\beta'}$ over τ_β iff $(1-\beta)f(y) + \beta\mu_{F_1} < (1-\beta')f(y) + \beta'\mu_{F_1}$, which in turn holds iff (1) holds. Thus, as noted earlier, nothing is lost by focusing only on the two extreme schemes in $\mathcal{P}$, namely τ_0 and τ_1.

[8] Analogues of this partial ordering are widely used in the theory of risk aversion, especially when individual preferences are defined over more than one good (see Kihlstrom and Mirman (1974)). A number of remarkable properties of this ordering are obtained by Debreu (1976) and Kannai (1977).

[9] In this case, τ_0 is the unique Condorcet winner in $\mathcal{P}$. We shall assume throughout that indifferent agents abstain from voting.

Theorem 3 *For any $F_0 \in \mathcal{F}_+$, there exists an $f \in \mathcal{T}$ such that τ_0 beats τ_1 under pairwise majority voting for all transition functions that are more concave than f, and τ_1 beats τ_0 for all transition functions that are less concave than f.*

The caveat mentioned above is that for a majority of individuals to vote for "laissez-faire", 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 $\mu_{F_1} < f(m_{F_0}) = m_{F_1}$.[10] There are two reasons why this is far less problematic, from an empirical point of view, 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 present a stochastic version of Theorem 3 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 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 to be defeated.[11]

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 $t = 0, \ldots, T$, 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 the basic findings from the two-period analysis carry over to a great extent to this setting.

[10] The simplest type of transition function which achieves this outcome is $f(y) \equiv \min\{y, m_{F_0} + \alpha(y - m_{F_0})\} + k$, where α is small enough and k is any constant, which could for instance be chosen so as to ensure that $\mu_{F_1} = \mu_{F_0}$. For details see the proof in the appendix, which also shows that Theorem 3 –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}.

[11] In the deterministic case the minimum bias required varies monotonically with the skewness of F_1, as seen above. In the stochastic case they need not be related, so that the "laissez-faire" policy can be the perfectly democratic outcome of the election even with $m_{F_1} < \mu_{F_1}$, i.e. with $F_1 \in \mathcal{F}_+$. The political process could even be biased towards the poor, rather than the rich, without affecting the result; see the proof of Theorem 5.

First, there again exists a unique tipping point $y_f^*(T)$ such that all agents with initial income less than $y_f^*(T)$ vote for τ_1, while all those richer than $y_f^*(T)$ vote for τ_0. When voters only consider current incomes, or when the policy has no lasting effects, it coincides with the mean: $y_f^*(0) = \mu_{F_0}$. When future incomes are factored in, however, 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 a strictly decreasing function of T. 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 $f^2, \ldots, f^T$, etc., and each of these functions is more concave than its predecessor.

Second, whether or not the increase in the vote for τ_0 is enough to ensure its victory over τ_1 in the election depends on the particular forms of f, F, on the degree of forward–looking, and on the specifics of the political system (e.g., relative propensities to vote of the different income classes). With standard majority voting, for instance, a generalization of Theorem 3 can be established, provided of course that agents care enough about future incomes.

Theorem 4 *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 τ_0.*

(b) *For all δ and T large enough, there exists an $f \in \mathcal{T}$ such that τ_0 ties with τ_1 under pairwise majority voting. Moreover, τ_0 beats τ_1 if the duration of the redistribution scheme is extended beyond T, and is beaten by τ_1 if this duration is reduced below T.*

Simply put, longer horizons magnify the strength of the "prospect of upward mobility" effect.[12] Finally, note that relaxing the assumption of constant marginal utility for money does not alter the results. If individuals have access to perfect credit markets, their lifetime utility remains an increasing function of the present discounted value of their net incomes, so nothing changes. Even when there are no such borrowing and lending opportunities, so that agents must consume their disposable income in each period, the results remain: Theorem 4 easily extends to the case of agents who vote for τ_0 over τ_1 whenever $\sum_{t=0}^{T} \delta^t U(f^t(y)) > \sum_{t=0}^{T} \delta^t U(\mu_{F_t})$, given any continuous and strictly increasing utility function U on X. In the presence of uninsurable uncertainty over future incomes, however, we shall see that risk aversion does complicate matters by creating a demand for redistribution for insurance purposes. This second effect works in the opposite direction of the POUM hypothesis.

---

[12] The reason why δ and T must be large enough in part (b) of Theorem 4 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 $T \geq 1$. In either case, the same caveat discussed following Theorem 3 now applies to the skewness of the terminal distribution F_T, in the absence of idiosyncratic shocks.

3 Income Dynamics and Voting under Uncertainty

3.1 Stochastic Income Processes

The assumption that individuals know their future incomes with certainty is obviously unrealistic. Moreover, in the absence of idiosyncratic shocks the cross-sectional distribution converges over time 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.

We shall now identify each individual by an index $i \in [0,1]$, and denote her endowment at date t by y_t^i. The evolution of y_t^i is determined by a **stochastic transition function** f, whose properties are discussed below, and a random shock Θ_{t+1}^i whose realization is denoted θ_{t+1}^i:

$$y_{t+1}^i = f(y_t^i, \theta_{t+1}^i), \quad t = 0, ..., T-1. \tag{5}$$

We require that the random variables Θ_t^i, $(i,t) \in [0,1] \times \{1, \ldots, T\}$, all have a common probability distribution function P, with support Ω. This means that everyone faces the same uncertain environment, which is stationary across periods.[13] It is important to note, however, that we put no restriction on the correlation of shocks across individuals. We thus allow for purely *aggregate shocks* ($\Theta_t^i = \Theta_t^j$ for all i, j in $[0,1]$), purely *idiosyncratic shocks* (the Θ_t^i's are independent across agents and sum to zero), and all cases in between.[14]

In the deterministic case the transition function was taken to be continuous, strictly increasing, and concave (but not affine). The most strict extension of these requirements to the stochastic case is that they should hold with probability one. Let us therefore denote as $\mathcal{T}_P$ the set of (P, F_0)-measurable functions $f : X \times \Omega \rightarrow X$ such that $\text{Prob}[\{\theta \,|\, f(\cdot\,;\theta) \in \mathcal{T}\}] = 1$. It is clear that any f in $\mathcal{T}_P$ has the following properties:

(i) The expectation $\mathsf{E}_\Theta[f(y\,;\Theta)]$ is well-defined on X, and $\mathsf{E}_\Theta[f(\cdot\,;\Theta)]$ belongs to $\mathcal{T}$.

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

For some of our purposes, the requirement that $f \in \mathcal{T}_P$ is too strong, so we shall develop our analysis for the larger set of ((P, F_0)-measurable) functions from $X \times \Omega$ to X which simply satisfy properties (i) and (ii). We denote it as $\mathcal{T}_P^*$.

---

[13] The first requirement is one of anonymity: two individuals with the same income history have the same probability distribution over future endowments. The second is made only to simplify the notation; one could easily allow for time-variable or serially correlated shocks.

[14] 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 all Θ_t^i's as jointly measurable, for any t. Consequently, the law of large numbers and Fubini's theorem (switching the order of double integrals) are applied as usual.

3.2 Two-Period Analysis

We first return to the case where risk-neutral agents vote in period 0 over distributing period 1 incomes. Agent $y^i \in X$ then prefers τ_0 to τ_1 if and only if

$$\mathsf{E}_{\Theta^i}\left[f(y^i;\Theta^i)\right] > \mathsf{E}\left[\mu_{F_1}\right], \tag{6}$$

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 uncertainty it remains a random variable. 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 :

$$\int_X \mathsf{E}_{\Theta^i}\left[f(y;\Theta^i)\right] dF_0(y) < \mathsf{E}_{\Theta^i}\left[f(\mu_{F_0};\Theta^i)\right]. \tag{7}$$

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

$$\mathsf{E}_{\Theta^i}\left[f(y_f^*;\Theta^i)\right] = \mathsf{E}\left[\mu_{F_1}\right]. \tag{8}$$

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 τ_0 to beat τ_1 under majority voting, observe that the inequality in (7) 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.[15] 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. Moreover, 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. For instance, a sufficiently skewed distribution of shocks will ensure that $F_1 \in \mathcal{F}_+$ without affecting the

[15] 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 $f \succ_P g$.

cutoff y_f^*. Combining a formal proof of this claim with the monotonicity of the cutoff, we establish a stochastic generalization of Theorem 3.

Theorem 5 *For any $F_0 \in \mathcal{F}_+$ and any $\sigma \in (0,1)$, there exists a probability distribution P and an $f \in \mathcal{T}_P^*$ such that $F_1(\mu_{F_1}) \geq \sigma$ and, under pairwise majority voting, τ_0 beats τ_1 for all transition functions in $\mathcal{T}_P^*$ that are more concave than f in expectation, while τ_1 beats τ_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 τ_0 can even be the preferred policy of a majority of voters.

3.3 Steady–State Distributions

The presence of idiosyncratic uncertainty is not only realistic, but 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

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

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 $y_{f,F}^* < \mu_F$.[16] 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, $\mathsf{E}\left[y_t^i \,|\, y_0^i\right]$. In the long run everyone's expected income converges to the population mean μ_F, but

[16]While concavity of $\mathsf{E}_\Theta[f(\cdot;\Theta)]$ is still a sufficient condition, it is no longer a necessary one if the inequality $y_{f,F}^* < \mu_F$ is required to hold *only* for the steady–state distribution(s) F induced by f and P, rather than for *all* initial distributions. But even then, 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 $y_{f,F}^* \geq \mu_F$ for all distributions, including stationary ones.

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 the Markov matrix:

$$M = \begin{bmatrix} 1-r & r & 0 \\ ps & 1-s & (1-p)s \\ 0 & q & 1-q \end{bmatrix}, \tag{9}$$

where $(p, q, r, s) \in (0,1)^4$.[17] We require that the conditional distribution of next period's income y_{t+1}^i be stochastically increasing in current income y_t^i.[18] It is easily checked that this holds if and only if

$$s < \min\left\{\frac{1-r}{p}, \frac{1-q}{1-p}\right\}. \tag{10}$$

Expected income $\mathsf{E}\left[y_{t+1}^i \mid y_t^i\right]$ is then strictly increasing in current income y_t^i; its concavity will follow from a stronger requirement imposed below. We now turn to the steady–state properties of this economy. The invariant income distribution induced by M over $\{a_1, a_2, a_3\}$ is the unique probability vector π that solves $\pi M = \pi$, namely,

$$\pi_1 = \frac{pqs}{r(1-p)s + q(r+ps)}, \quad \pi_2 = \frac{qr}{r(1-p)s + q(r+ps)} \quad \text{and} \quad \pi_3 = 1 - \pi_1 - \pi_2. \tag{11}$$

The corresponding mean income is $\mu = \pi_1 a_1 + \pi_2 a_2 + (1 - \pi_1 - \pi_2)a_3$. The median income is a_2 provided that $\pi_1 < 1/2 < \pi_1 + \pi_2$, that is,

$$s < \frac{rq}{|qp - r(1-p)|}. \tag{12}$$

Agents with $y_t^i = a_2$ (hence, a fortiori, those with $y_t^i = a_3$) will oppose redistribution of time

[17] In the formalism of our general model, the random variable Θ is three-dimensional: $\Theta \equiv (\Theta_1, \Theta_2, \Theta_3)$, where the probability distribution of Θ_i over $\{a_1, a_2, a_3\}$ is given by the i-th row of M, $i = 1, \ldots, 3$. The transition function is then, simply: $f(y, \Theta) = \mathbf{1}_{\{y=a_1\}}\Theta_1 + \mathbf{1}_{\{y=a_2\}}\Theta_2 + \mathbf{1}_{\{y=a_3\}}\Theta_3$. The only (minor) difference in this formulation is the restriction to a discrete support.

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

$t+1$ incomes if and only if $\mathsf{E}\left[y^i_{t+1} \mid y^i_t = a_2\right] > \mu$. One can show that this condition holds for all $a_1 < a_2 < a_3$ when [19]

$$\frac{r(1-q)}{r(1-p)+qp} \leq s < \frac{q(1-r)}{r(1-p)+qp}. \tag{13}$$

The entire set of requirements on s can thus be summarized as:

$$\underline{s} \equiv \frac{r(1-q)}{r(1-p)+qp} < s < \min\left\{\frac{1-r}{p}, \frac{1-q}{1-p}, \frac{q(1-r)}{r(1-p)+qp}, \frac{rq}{|qp-r(1-p)|}\right\} \equiv \bar{s}. \tag{14}$$

One easily verifies that $\underline{s} < \bar{s}$ if and only if $r < q$ and

$$1-2q < \min\left\{\frac{r(1-p)}{pq}, \frac{pq}{r(1-p)}\right\}, \tag{15}$$

which holds for all $q > \max\{r, 1/2\}$. It then suffices to chose any s in $(\underline{s}, \bar{s})$ for all the desired conditions to hold, independently of (a_1, a_2, a_3). The final requirement is that the stationary income distribution be positively skewed: $\mu > a_2$, or equivalently

$$\frac{a_3-a_2}{a_2-a_1} > \frac{pq}{r(1-p)}. \tag{16}$$

In conclusion, there is a wide set of parameters $(p, q, r, s; a_1, a_2, a_3)$ for which the steady–state of this economy has all the desired properties.[20] In particular, *over half* of the population is always poorer than average, but *over half* of the population always has expected income above average. Put differently, the distribution of expected incomes is *negatively* skewed even though the distribution of actual incomes is *positively* skewed and every one has rational expectations. But, one might ask, is such a process empirically plausible?

It is actually not difficult to find a specification which matches the broad facts of the US income distribution –say, in the 1990 Census– and its intergenerational persistence. Let $p = .55$, $q = .6$, $r = .5$, and $s = .7$, leading to the stationary distribution $(\pi_1, \pi_2, \pi_3) = (.33, .44, .23)$. Hence, while 67% of the population has *expected* income above the mean, in each period only 23% actually end up with *realized* incomes above the mean.[21] Choosing $(a_1, a_2, a_3) = (16000, 36000, 91000)$, we obtain a rather remarkable fit with the data, especially in light of the model's simplicity.

[19] To see this, note that $\lambda a + \lambda' b > 0$ holds for all $0 < a < b$ if and only if $\lambda' \geq 0$ and $\lambda + \lambda' > 0$. Since $\mathsf{E}\left[y^i_{t+1} \mid y^i_t = a_2\right] - \mu = (1-s-\pi_2)(a_2-a_1) + ((1-p)s - (1-\pi_1-\pi_2))(a_3-a_1)$, the inequality $\mathsf{E}\left[y^i_{t+1} \mid y^i_t = a_2\right] > \mu$ therefore holds for all $a_1 < a_2 < a_3$ whenever $(1-p)s \geq 1-\pi_1-\pi_2$ and $1-ps > 1-\pi_1$. Condition (13) follows from this observation.

[20] The requirement that $\mathsf{E}\left[y^i_{t+1} \mid y^i_t = a_2\right] > \mu > a_2$ automatically implies that the conditional expectation $g(y) \equiv \mathsf{E}\left[y^i_{t+1} \mid y^i_t = y\right] = \mathsf{E}_\Theta\left[f(y;\Theta)\right]$ is strictly concave in y (hence $f \in \mathcal{T}_P^{\prime}$). This can be verified by direct computation, or more simply by observing that the strictly increasing, piece–wise linear function defined on $[a_1, a_3]$ by linearly interpolating g between $(a_1, g(a_1))$, $(a_2, g(a_2))$ and $(a_3, g(a_3))$ is, necessarily, either concave or convex. If it were convex this would imply: $g(a_2) < g(\mu) = g(\Sigma_{j=1}^3 \pi_j\, a_j) \leq \Sigma_{j=1}^3 \pi_j\, g(a_j) = \mu$, a contradiction.

[21] In addition we also verify, numerically, that the middle class has expected income above the mean not just in the next period, but in all future periods.

		Data (1990)	Model
Median family income	\$	35,353	36,000
Mean family income	\$	42,652	41,872
Standard deviation of family incomes	\$	29,203	28,138
Share of bottom 1/3	%	11.62	12.61
Share of bottom 3/4 [77%]	%	52.23	50.44
Share of top 1/4 [23%]	%	47.77	49.56
Intergenerational correlation of log–incomes		0.4 to 0.5	.45.

Table 1: Distribution and Persistence of Income in the United States [22]

The resulting 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:

(a) no deadweight loss concern enters into voters' calculations;

(b) *three quarters* of the population is always *poorer than average;*

(c) most of the members of the "laissez–faire" coalition ($\pi_2/(\pi_2+\pi_3) = 66\%$ of them) are below mean income, and know that their children are four times more likely than not to end up in that same situation, where they would benefit from redistribution.

3.4 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). Thus, agents 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:

$$y_t^i = f(..., f(f(y_0^i;\theta_1^i);\theta_2^i);...;\theta_t^i) \equiv f^t(y_0^i;\underline{\theta}_t^i), \quad t = 1, ..., T, \tag{17}$$

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:

[22] 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). Cooper, Durlauf and Johnson (1994) allow for non–linearities in the transmission process, and find persistence to be a U–shaped function of initial income.

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

where we suppressed the index i on the random variables $\underline{\Theta}_t^i$ since they all have the same probability distribution $P^t(\underline{\theta}_t) = \prod_{k=1}^{t} P(\theta_k)$ on Ω^t. Under the policy τ_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 τ_1 over τ_0 if and only if

$$V^T(y_0^i) > \int_X V^T(y)\, dF_0(y). \tag{18}$$

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 6 *Let $F_0 \in \mathcal{F}$, $\delta \in (0,1]$, $T \geq 1$. For any probability distribution P and any transition function $f \in \mathcal{T}_P$, there exists a unique $y_f^*(T) < \mu_{F_0}$ such that all agents in $[0, y_f^*(T))$ vote for τ_1 over τ_0, while all those in $(y_f^*(T), 1]$ vote for τ_0 over τ_1.*

Note that Proposition 6 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. For $f \in \mathcal{T}_P$, can one obtain a stronger result, similar to that of the deterministic case, namely that the tipping point decreases as the time horizon lengthens? While this seems quite intuitive, and will indeed occur in the "natural" example of Section 5, it may in fact not hold without relatively strong additional assumptions. Technically speaking, this is because the expectation operator does not, in general, preserve the "more concave than" relation. One interesting sufficient condition that insures this result is that the $t+1$–step transition function be more concave in expectation than the t–step transition function.

Proposition 7 *Let $F_0 \in \mathcal{F}$, $\delta \in (0,1]$, $T \geq 1$, and let P be a probability distribution on Ω. If, for all t, $f^{t+1}(\cdot\,;\underline{\Theta}_{t+1}) \succ_{P^{t+1}} f^t(\cdot\,;\underline{\Theta}_t)$, that is,*

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

then the larger the political horizon T, the larger the share of the votes that go to τ_0.

When the condition holds, the proof is analogous to Theorem 4(a), and so is the interpretation: the more forward-looking voters are, or the more long-lived the tax scheme is, the lower is political support for redistribution. An immediate and useful corollary is that this monotonicity holds when the transition function is of the form $f(y,\theta) = y^{\alpha}\phi(\theta)$, where α is any number in $(0,1)$ and ϕ can be an arbitrary function.[23] Such will be the case in the log-linear example of Section 5.

4 Extending the Basic Framework

4.1 The Effect of 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 τ_0 from those who prefer τ_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, and will be made explicit in the next section's example, 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 E_{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.

4.2 Endogenous Mobility

While we have focused on pure endowment economies, the POUM mechanism remains operative when agents make effort and savings decisions. One complication that arises with accumulation is that the transition function now depends on the chosen redistributive policies. Future tax rates matter through their disincentive effects on savings and labor supply, while current taxes and transfers matters when agents undertake investments subject to borrowing constraints (since disposable resources then determine the level of investment, hence future earnings). Bénabou (1996, 1997) develops such a model, allowing also for a continuous policy choice but using very specific functional and distributional assumptions, as in Section 5 below. The main difference with the endowment

[23] Note that the mere multiplicative separability of f into $f(y;\theta) = \gamma(y)\phi(\theta)$ is not sufficient to ensure that the requirement of Proposition 7 is satisfied.

economy is that social mobility is now endogenous: it increases with the progressivity of redistribution, which relaxes the liquidity constraints hindering investment by the poor. The demand for redistribution may be higher or lower, but its comparative statics properties with respect to the concavity of the transition function (as reflecting the production and investment technologies), the length of the political horizon, and agents' risk–aversion, are all very similar to those derived here.[24]

5 The Log–Normal Case

In this section we consider a particular specification which yields simple, explicit solutions, even in the cases of risk–aversion and non–linear taxes where no general results exist.

5.1 Dynamics and Distribution of Incomes

Let the transition function be log-linear: $f(y;\theta) = \theta y^{\alpha}$ for all $y \geq 0$, with $\alpha \in (0,1)$ ensuring strict concavity in the first argument. Individual endowments thus evolve according to the stochastic process:

$$\ln y_{t+1}^i = \alpha \ln y_t^i + \ln \theta_{t+1}^i, \quad t = 0, 1, \dots . \tag{19}$$

We assume that both the initial income levels and the shocks are log-normally distributed:

$$\ln y_0^i \sim \mathcal{N}(m_0, \Delta_0^2) \quad \text{and} \quad \ln \Theta_t^i \sim \mathcal{N}(-s^2/2, s^2). \tag{20}$$

Notice that $\mathsf{E}[\Theta_t^i]$ is normalized to one. The log–linear specification is very common in the empirical literature on income or wage dynamics, whether intra- or inter-generational. Moreover, with log-normal shocks the cross-sectional distribution also remains log–normal over time, and this is quite a good approximation to actual income distributions. It is clear from the above assumptions that $\ln y_t^i \sim \mathcal{N}(m_t, \Delta_t^2)$, with the mean and variance given by the following recursion equations:

$$m_{t+1} = \alpha m_t - s^2/2 \quad \text{and} \quad \Delta_{t+1}^2 = \alpha^2 \Delta_t^2 + s^2. \tag{21}$$

Note that m_t is the logarithm of median income ($m_t = \ln m_{F_t}$), whereas per capita income is given by $\ln \mu_{F_t} = m_t + \Delta_t^2/2$. From (19) one can easily compute agent i's log–income in period t as

$$\ln y_t^i = \alpha^t \ln y_0^i + \sum_{k=1}^{t} \alpha^{t-k} \ln \theta_k^i. \tag{22}$$

Therefore, *conditionally* on its initial level, y_t^i is also log–normally distributed. Taking expectations and variances in (22), and using the recursion on m_t in (21) to simplify, we have:

[24] Intuitively, the demand for current redistribution tends to increase (because it allows a reallocation of investment resource towards those who have a higher marginal product, due to tighter liquidity constraints), while the demand for future redistribution tends to decrease (because savings distortions may now compound those in labor supply). The first mechanism tends to increase the growth of aggregate income, the second one to reduce it.

$$\mathsf{E}_{\Theta_1^i \dots \Theta_t^i}\left[\ln y_t^i \,|y_0^i\right] = m_t + \alpha^t \left(\ln y_0^i - m_0\right), \tag{23}$$

$$\mathsf{Var}_{\Theta_1^i \dots \Theta_t^i}\left[\ln y_t^i \,|y_0^i\right] = s^2 \left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right). \tag{24}$$

From these moments one can easily compute $\mathsf{E}_{\Theta_1^i \dots \Theta_t^i}\left[(y_t^i)^\lambda \,|y_0^i\right]$, for any value of λ. We shall make repeated use of this result in what follows.

5.2 Risk–Neutral Agents

We begin once again with the case where risk–neutral agents vote in period zero over the linear tax rate to be implemented *at date t*. Agent i thus supports redistribution when $\mathsf{E}_{\Theta_1^i \dots \Theta_t^i}\left[y_t^i \,|y_0^i\right] < \mu_{F_t}$.

Proposition 8 *The tipping point for risk-neutral agents who vote over linear tax schedules to be implemented in period t is:*

$$\ln y_{ft}^* = \ln \mu_{F_0} - (1-\alpha^t)\frac{\Delta_0^2}{2}.$$

This result clearly shows how political support for redistribution declines as the horizon lengthens, from the *mean* of initial income μ_{F_0} in the standard case of voting over current taxes, to the *median* in the limiting case where t tends to infinity. For a long enough horizon, or a high enough discount factor δ when agents care about expected present values, redistribution can therefore be blocked by arbitrarily small deviations from perfect democracy.[25] In contrast to the Markov example of Section 3.3, however, the cutoff never falls below the median.

In the long run the economy settles to its asymptotic income distribution, denoted F_∞, with variance $\Delta_\infty^2 = s^2/(1-\alpha^2)$ by (21). Replacing Δ_0^2 with Δ_∞^2 in the expression for y_{ft}^*, we obtain the value of the cutoff y_{ft}^{**} corresponding to F_∞:

$$\ln y_{ft}^{**} = \ln \mu_{F_\infty} - \left(\frac{1-\alpha^t}{1-\alpha^2}\right)\frac{s^2}{2}.$$

Our characterization of political support for redistribution (or lack thereof) thus remains unchanged in steady–state.

5.3 Risk–Aversion

The log-linear/log–normal specification also allows us demonstrate explicitly how the size of the coalition for redistributive policies is shaped by two opposing forces: (a) the concavity of the transition function, which tends to lower y_{ft}^*; (b) the concavity of the utility function, which tends to

[25] Note that lengthening the horizon, which raises α to a higher power, is a special case of taking a strictly concave transformation $h(y^\alpha) = y^{\alpha\gamma}$, $0 < \gamma < 1$.

increase it. For simplicity we continue to focus on the case where agents in period zero vote on the tax rate for period t.[26] Their preferences over random levels of disposable income at that date are given by $\mathsf{E}_{\Theta_1^i \dots \Theta_t^i}\left[U(y_t^i) \mid y_0^i\right]$, where

$$U(c) \equiv \frac{c^\beta - 1}{\beta}, \quad \beta \leq 1. \tag{25}$$

Agent i compares her expected utility under "laissez-faire" to the sure level to be received under (complete) redistribution at date t, and votes for τ_0 over τ_1 when $\mathsf{E}_{\Theta_1^i \dots \Theta_t^i}\left[(y_t^i)^\beta \mid y_0^i\right] > (\mu_{F_t})^\beta$.

Proposition 9 *The tipping point for risk–averse agents who vote over whether τ_0 or τ_1 should be implemented in period t is:*

$$\ln y_{ft}^* = \ln \mu_{F_0} - (1-\alpha^t)\frac{\Delta_0^2}{2} + \alpha^{-t}(1-\beta)\left(\frac{s^2}{2}\right)\left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right).$$

Under risk–neutrality ($\beta = 1$), this expression reduces of course to the result derived earlier, where the cutoff y_{ft}^* declines from the mean μ_{F_0} to the median as t becomes larger. The role of (relative) risk–aversion $1-\beta$ is to increase everyone's demand for redistribution, thereby raising the cutoff y_{ft}^*. With one-period ahead decisions, for instance, the cutoff remains below the mean if and only if

$$\alpha(1-\alpha)\frac{\Delta_0^2}{2} > (1-\beta)\left(\frac{s^2}{2}\right). \tag{26}$$

Note how this turns on the comparison between the *concavity* of the transition function, $\alpha(1-\alpha)$, and the degree of relative *risk–aversion*, $1-\beta$.[27] As the horizon lengthens, however, the cutoff y_{ft}^* inevitably rises above the mean, and ultimately tends to infinity. This is because all agents face the same long–run distribution of income F_∞. Thus when making decisions over redistribution in the very far future, they essentially have an ex–ante perspective, and (in the absence of offsetting distortions) their desire for insurance becomes the dominant factor. When mobility prospects and this insurance motive are combined, finally, the size of the "laissez–faire" coalition may have an inverse U–shape with respect to the horizon or duration of the redistributive scheme.

5.4 Non–Linear Taxation

To illustrate the fact that the paper's insights are not limited to the set of linear schemes $\mathcal{P}$, we shall now extend the preceding results to a one–dimensional family on non–linear tax schemes, denoted $\mathcal{P}'$. This family also has the advantage of yielding simple, explicit results when combined with the log–normal specification and CRRA utility functions. For each $\tau \in (-\infty, 1]$, consider the

[26] The more standard case where agents care about an expected present value of utilities is, qualitatively speaking, a weighted average over $t = 0, \dots, T$, of the "pure" t–periods–ahead problems considered here, and therefore leads to similar results.

[27] In steady–state, condition (26) simplifies to $\alpha/(1+\alpha) > 1-\beta$.

following redistributive policy, to be implemented at date t. Agent i pays net taxes $\varphi_{t,\tau}(y^i)$, leaving her with a disposable income equal to

$$\hat{y}^i_{t,\tau} = y^i_t - \varphi_{t,\tau}(y^i) \equiv (y^i_t)^{1-\tau}(\tilde{y}_{t,\tau})^\tau, \tag{27}$$

where the break–even level $\tilde{y}_{t,\tau}$ is defined by the government budget constraint $\int_X \varphi_{t,\tau}\, dF_t = 0$, or

$$\int_X (y)^{1-\tau}(\tilde{y}_{t,\tau})^\tau\, dF_t(y) = \mu_{F_t}. \tag{28}$$

Note that the scheme is progressive for $\tau > 0$ and regressive for $\tau < 0 : \varphi_{t,\tau}(\cdot)$ is convex in the first case, concave in the second. We shall refer to τ as the progressivity rate. Clearly, "laissez–faire" and "complete redistribution" correspond to $\tau = 0$ and $\tau = 1$, respectively.[28]

To compute an agent's welfare under a policy $\tau \in \mathcal{T}'$, let us first derive the distribution of her future post–tax income $\hat{y}^i_{t,\tau}$, conditional on her current pre-tax income y^i_0. From (23), (24) and (27) it follows that

$$\ln \hat{y}^i_{t,\tau} \sim \mathcal{N}\left(\alpha^t(1-\tau)(\ln y^i_0 - m_0) + (1-\tau)m_t + \tau \ln \tilde{y}_{t,\tau}\,,\ (1-\tau)^2 s^2 \left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right)\right), \tag{29}$$

where $\tilde{y}_{t,\tau}$ can be obtained as a function of τ, m_t, and Δ^2_t by solving (28).

We can now examine the outcome of pairwise voting over two possible schemes in $\mathcal{P}'$ with arbitrary progressivity rates $\underline{\tau}$ and $\bar{\tau}$, $-\infty < \underline{\tau} < \bar{\tau} \leq 1$. Given a relative risk-aversion of $1-\beta$, agent i will prefer the latter when

$$\mathsf{E}[(\hat{y}^i_{t,\bar{\tau}})^\beta\, |y^i_0] > \mathsf{E}[(\hat{y}^i_{t,\underline{\tau}})^\beta\, |y^i_0]. \tag{30}$$

Proposition 10 *The tipping point $y^*_{f^t}(\underline{\tau}, \bar{\tau})$ for risk-averse agents choosing between two non–linear redistributive schemes $\bar{\tau}$ and $\underline{\tau}$ in $\mathcal{P}'$ is:*

$$\ln y^*_{f^t}(\underline{\tau}, \bar{\tau}) = \ln \mu_{F_0} - [1 - \alpha^t(2 - \bar{\tau} - \underline{\tau})]\left(\frac{\Delta^2_0}{2}\right) + \alpha^{-t}(1-\beta)(2 - \bar{\tau} - \underline{\tau})\left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right)\left(\frac{s^2}{2}\right).$$

In the case of an all-or-nothing policy decision ($\bar{\tau} = 1$, $\underline{\tau} = 0$) this coincides of course with the expressions derived earlier. More generally, risk–aversion has the same effect of raising the cutoff as before; the longer the horizon t, the more so. To isolate the new effects which arise from non–linear taxation, let us now focus on the case of risk–neutrality, $\beta = 1$. The cutoff is then below the mean

[28] In the terminology introduced by Musgrave and Thin (1948), the elasticity $1-\tau$ of post–tax to pretax income is the rate of "residual progressivity". The scheme (27) takes it to be the same at all income levels, in which case τ is also the income–weighted average marginal tax rate paid by agents, as shown in Bénabou (1997). This isoelastic or "constant residual progression" specification has been used to study insurance or risk–taking in a static context by Feldstein (1969), Kanbur (1979) and Persson (1983), and in models with accumulation by Bénabou (1996, 1997).

μ_{F_0} if and only if

$$\alpha^t(2 - (\bar{\tau} + \underline{\tau})) < 1. \tag{31}$$

In the usual case of contemporaneous redistribution ($t = 0$), this becomes $\bar{\tau} + \underline{\tau} > 1$. If the first scheme is progressive while the second is regressive ($\underline{\tau} \leq 0 < \bar{\tau} \leq 1$, with at least one strict inequality) the indifference point is always above the mean, in conformity with the main result of Marhuenda and Ortuño–Ortin (1995).[29] This is intuitive, since a progressive scheme redistributes income disproportionately from the very rich to everyone else, including those with average resources. When mobility prospects are taken into account, however, this result does not hold any more: $y^*_{ft}(\underline{\tau}, \bar{\tau})$ falls below μ_{F_0} for t large enough. More generally and more importantly, the cutoff declines with the time between voting and the implementation of the policy, converging again (in the absence of an insurance motive) to the median m_0 as t tends to infinity. Finally, note that even for a contemporaneous redistribution ($t = 0$) the cutoff can be below the mean if we are comparing –perhaps more realistically– two progressive policies with $0 < \underline{\tau} < \bar{\tau} < 1 < \bar{\tau} + \underline{\tau}$, rather than a regressive and a progressive one.[30]

6 Uncovering 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. Furthermore, the Markovian example of Section of Section 3.3 showed that this effect can be strong enough to swing a majority, or even a supermajority, while still maintaining empirically reasonable values for income inequality and the average degree of serial persistence. The final question which naturally arises is whether this effect is at all 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, but to show that the POUM effect can be measured quite simply from income mobility and inequality data –with rather interesting results. In doing so we shall continue to abstract, as in most of the paper, from the other forces which contribute to shaping the political equilibrium (e.g., tax distortions and demand for insurance). The question we ask, therefore, is 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%?*[31]

As a first pass at the numbers, let us continue to work within the loglinear–lognormal specification. Given an autoregressive coefficient α and an initial variance of log–incomes Δ_0^2, Proposition

[29] See also Mitra, Ok and Koçkesen (1997) for a generalization.

[30] In contrast to the other results, which are in closely related the general propositions in our paper, this one is more dependent on the particular form of progressivity which we have assumed.

[31] 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 in no way general features of stationary processes. Rather than simply "mean–reverting", the income dynamics need to be "mean–crossing from below" (in expectation), over some range.

8 shows that the proportion of agents with conditional expected incomes below the mean at any horizon t is $\Phi(\alpha^t\Delta_0/2)$, where Φ is the c.d.f. of a standard normal. According to the Census data used earlier in Table 1, the standard deviation of families' log–incomes in the United States in 1990 was about .64. Under log–normality this means that $\Phi(.32) = 62.5\,\%$ of the families are poorer than average. Yet with a typical estimate of the intergenerational persistence coefficient, say $\alpha = .4$, only $\Phi(.4 \times .32) = 55.9\%$ of the children have expected incomes below the mean. As the issue debated shifts from current to future redistribution, the POUM effect thus moves about 7% of the population towards "laissez–faire". This is by no means negligible, especially since the differentials rates of political participation according to socioeconomic class which are observed in the U.S. imply that the pivotal agent is almost surely located above the 56th percentile.[32]

However, the loglinear example is not really suited to this empirical exercise, because it imposes concavity from the start, and because the cutoff can never fall below the median.[33] In what follows we shall therefore use the much more detailed and flexible description of the mobility process provided by empirical mobility matrices. These are often estimated in terms of income quintiles, which is too coarse a grid for our purposes, especially given the importance of what happens near the median. We shall therefore use the more disaggregated data compiled by Hungerford (1993) from the PSID (Panel Study on Income Dynamics), 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 provide less noisy measures of "permanent income".

(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}.

Let us start by examining these two income distributions:

$$a_{69} = (.211\ \ .410\ \ .566\ \ .696\ \ .822\ \ .947\ \ 1.104\ \ 1.302\ \ 1.549\ \ 2.393)'$$
$$a_{79} = (.179\ \ .358\ \ .523\ \ .669\ \ .801\ \ .933\ \ 1.084\ \ 1.289\ \ 1.588\ \ 2.576)'.$$

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%

[32]Bénabou (1996, revised version) uses 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 to compute the resulting bias with respect to the median. It is found to vary between 6% (when only voting propensities are taken into account) and 24% (when only propensities to contribute to campaigns are taken into account), with most values being above 10%. Depending on the (unknown) relative efficacy of the different forms of activity at influencing actual policies, the pivot is thus located somewhere between 55% and 75%, and probably occurs around 60%.

[33]More generally, a single persistence coefficient captures only a small part of the relevant information about income transitions. For instance, the Markovian example of Section 3.3 shows that the same degree of serial correlation in log–incomes can be generated by a very different process, where the POUM effect is so strong as to turn a supermajority from pro–redistribution to pro–"laissez-faire".

in 1969 and 64.4% in 1979.[34]

Next, we apply the appropriate empirical transition matrix to compute the vector of conditionally expected relative incomes $t \times 7$ years ahead, namely $(M_{69}^{76})^t \cdot a_{69}$ or $(M_{79}^{86})^t \cdot a_{79}$, for $t = 1, ...4$.[35] 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. The results are presented in **Table 2.**

	Horizon (years)				
Mobility Matrix	0	7	14	21	28
1969–1976	%				
Annual income	63.39	61.83	54.22	48.77	47.25
"Permanent" income	63.39	60.77	56.36	52.91	50.38
1979–1986	%				
Annual income	64.42	60.90	51.31	48.11	46.54
"Permanent" income	64.42	58.80	54.29	51.43	49.10
1969–1976 × 1979–1986	%				
Annual income	63.39		52.53		46.83
"Permanent" income	63.39		55.08		50.08

Table 2: Income Percentile of the Political Cutoff
Source: authors' calculations using PSID data from Hungerford (1993)

The message delivered is consistent across all specifications: the POUM effect is present and significant in the data –even at relatively short horizons, but especially over longer ones. It affects approximately 3.5% of the population over 7 years, and 10% over 14 years. Since the patterns of voting and political participation by the different socioeconomic classes imply that the "upper fifties" represent a lower bound on the rank of the pivotal group, a horizon of a decade or so could already suffice to swing the political outcome against redistribution.[36] In any case, over a horizon of approximately 20 years mobility prospects wipe out the entire 13-15% point interval between mean and median incomes, bringing a strict majority to the "laissez–faire" side. Thus, in both 1969 and 1979, 64% of the population was poorer than average in terms of current income and yet

[34]In passing, it is interesting to note that the Lorenz curve for 1979 is everywhere below its 1969 counterpart, meaning that income inequality increased unambiguously –albeit slightly– between these two dates.

[35]There are two implicit assumptions in this procedure. First, by iterating a 7–year transition matrix to compute mobility over 14, 21 and 28 years we are treating the transition process as stationary over time. Similarly, by applying these matrices to the income distribution vector at the beginning of the transition period we are abstracting from any changes in the deciles' relative incomes during that time. These are obviously simplifying approximations, imposed by the availability of the data. To check the robustness of the results, however, we also used the composite matrix $M_{69}^{76} \cdot M_{79}^{86}$ to recompute the 14 and 28 year transitions; see the bottom part of Table 2. Similarly, we applied the transition matrices M_{69}^{76} and $M_{69}^{76} \cdot M_{79}^{86}$, and their iterates, to the income distribution a_{79} instead of a_{69} (the two were seen in footnote 34 to be somewhat different). In all cases the results remained essentially unchanged.

[36]See footnote 32.

51% could rationally see themselves as richer than average in terms of expected income two decades down the road.

We conclude from this empirical exercise that the POUM hypothesis is not only a theoretical possibility but also a significant feature of the actual process of socioeconomic mobility.[37] Naturally, it does not explain all by itself (especially with discounting) why democracies with skewed distributions of incomes maintain relatively low rates of redistribution. But the empirical evidence brought to light by **Table 2** shows that it represents –alongside with deadweight losses, political bias, and risk aversion– an important entry in the balance of forces which determine the equilibrium rate of redistribution.

7 Conclusion

This paper has established the formal basis for the "prospect of upward mobility" hypothesis in regard to the political economy of redistribution. Voters poorer than average will nonetheless opt for a zero or 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. We provided an example where, in every period, three quarters of the population have less than mean income, yet a two–thirds majority supports a zero tax rate for their children's and all future generations. This is in spite of the fact that there are no deadweight loss concerns, and that voters in the pivotal middle class know that their children have no more than a 20% chance of rising above the mean income level. A calibrated version of this simple model was shown to match the main features of the US income distribution and the average degree of intergenerational persistence.

Using income mobility and inequality data from the PSID, we also provided direct empirical evidence of the POUM effect. As the horizon over which incomes are forecasted increases from 0 to 7, then 14 years, the proportion of agents with expected income above the mean rises from 36% to 39% and 47%, respectively. Over a 20 year horizon, mobility prospects bring the size of this "laissez–faire" coalition up to 51%, thus "erasing" the entire interval between mean and median incomes.

At the same time, the prospect of upward mobility effect is subject to limitations, which we have also analyzed. In particular, there must be sufficient inertia or commitment power in the choice of fiscal policy or institutions, and voters' risk–aversion must not be too large compared to the curvature of the transition function.

[37]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.

Appendix

Proof of Proposition 2

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

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

The proposition follows from the fact that f is strictly increasing. ||

Proof of Theorem 3

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

$$g_{\eta,\alpha}(y) \equiv \begin{cases} y & \text{if } 0 \leq y \leq \eta \\ \\ \eta + \alpha(y - \eta), & \text{if } \eta < y \leq \bar{y} \end{cases}, \tag{A.1}$$

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 monotonicity there exists a unique (and easily computable) $\alpha(\eta) \in (0,1)$ 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 τ_0 to τ_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 4

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

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

for any $T \geq 1$.

Since $V^T(f)(\cdot)$ is clearly continuous and increasing, there must therefore exist a unique $y^*_f(T) \in (y_{fT}, \mu_{F_0})$ 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 $y^*_{fT+1} < y^*_{fT}$, we have $y^*_{fT+1} < y^*_f(T)$. This implies that $\mu_{F_{T+1}} = f^{T+1}(y^*_{fT+1}) < f^{T+1}(y^*_f(T))$, 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 $g_{\eta,\alpha}$. Recall that, for all $y \in X$,

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

where $\eta < \mu_{F_0}$ and $\alpha \in [0, 1]$. Let us first observe that the iterates of such a function are simply:

$$(g_{\eta,\alpha})^t(y) \equiv \min\{y, \eta + \alpha^t(y - \eta)\} = g_{\eta,\alpha^t}(y). \tag{A.7}$$

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 τ_0 (under which she receives η in each period) to τ_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

$$\frac{\mu_{F_0}-\eta}{\eta-\int_X \min\{y,\eta\}dF_0} < \sum_{t=1}^{T}\delta^t = \frac{\delta(1-\delta^T)}{1-\delta}. \tag{A.8}$$

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)–(A.6) 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 the 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" each agent with initial $y < \eta$ would receive y in every period, while each agent with $y > \eta$ would receive $\eta + \alpha^t(y-\eta) > \eta$. Therefore, η is the cutoff $y_f^*(T)$ separating those who support τ_0 from those who support τ_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 5

As in the proof Theorem 3, 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 $f : X \times \Omega \to X$ as follows:

$$\begin{array}{lllll}
\bullet\ \text{if } 0 \le y \le \eta-\varepsilon, & f(y;\theta) \equiv y & & \text{for all } \theta & \\
\bullet\ \text{if } \eta-\varepsilon < y \le \eta, & f(y;\theta) \equiv \begin{cases} \eta-\varepsilon & \text{if } \theta=0 \\ \frac{y-p(\eta-\varepsilon)}{1-p} & \text{if } \theta=1 \end{cases} & & \begin{array}{l}\text{(probability } p) \\ \text{(probability } 1-p)\end{array} & \text{(A.9)} \\
\bullet\ \text{if } \eta \le y \le \bar{y}, & f(y;\theta) \equiv \begin{cases} \eta-\varepsilon & \text{if } \theta=0 \\ \eta+\frac{\alpha(y-\eta)+p\varepsilon}{1-p} & \text{if } \theta=1 \end{cases} & & \begin{array}{l}\text{(probability } p) \\ \text{(probability } 1-p)\end{array} &
\end{array}$$

By construction, $\mathsf{E}_\Theta[f(y;\Theta)] = g(y)$ for all $y \in X$, therefore, $\mathsf{E}_\Theta[f(\cdot\,;\Theta)] = g \in \mathcal{T}$. It remains to be checked that $f(y;\Theta)$ is strictly stochastically increasing in y : for any $(y,x) \in X^2$. That is, the conditional distribution $M(x|y) \equiv P(\{\theta \in \Omega \,|\, f(y;\theta) \le x\})$ 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 $h : X \to R$; this latter form of the property is easily verified by using (A.9).

Because $\mathsf{E}_\Theta\left[f(\cdot\,;\Theta)\right] = g$, so that $\mu_{F_1} = \eta$ by (A.2), it is clear that the cutoff between the agents who prefer τ_0 and those who prefer τ_1 is $y_f^* = \eta$. This tipping point can be set to any value below μ_{F_0} (such as the median m_{F_0}), and nonetheless the date one income distribution F_1 will remain positively skewed, as long as p is high enough. Indeed,

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

Thus, for any η, the fraction of agents who end up with income below the mean, $1 - F_1(\mu_1)$, can be made arbitrarily close to 1, by choosing p close to 1. At the same time, choosing η small ensures that the size of the coalition supporting redistribution, $1 - F_0(\eta)$, remains arbitrarily small.

To conclude the proof of the theorem, 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 7

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 4, with h_t playing the role of f^t. ||

Proof of Propositions 8, 9 and 10

We shall prove Proposition 10 directly, since it includes the other two as special cases. Our task is thus to compare the expected utility achieved by an agent with relative risk–aversion $1 - \beta \geq 0$ under two arbitrary policies $\bar{\tau}$ and $\underline{\tau} < \bar{\tau}$ in the family $\mathcal{P}'$ of non-linear redistributive schemes defined in Section 5.4. Recall in particular that $\underline{\tau} = 0$ and $\bar{\tau} = 1$ coincide with τ_0 and τ_1 in $\mathcal{P}$, namely with "laissez-faire" and complete redistribution, and that under risk–neutrality the cutoff between these two extreme policies also applies to any pair of linear schemes in $\mathcal{P}$.

Let τ be any policy in $\mathcal{P}'$. Given the distribution $\ln y_t^i \sim \mathcal{N}(m_t, \Delta_t^2)$ we can compute the integrals in (28) and obtain:

$$\tau \ln \tilde{y}_{t,\tau} = \ln \mu_{F_t} - \ln\left(\int_X (y)^{1-\tau}\, dF_t(y)\right) = m_t + \Delta_t^2/2 - (1-\tau)m_t - (1-\tau)^2\Delta_t^2/2$$

that is,

$$\ln \tilde{y}_{t,\tau} = m_t + (2-\tau)\Delta_t^2/2 = \ln \mu_{F_t} + (1-\tau)\Delta_t^2/2. \tag{A.10}$$

Consider next the conditional distribution of agent i's post-tax income which will result from τ. Clearly, $\ln \hat{y}^i_{t,\tau} = (1-\tau)\ln y^i_t + \tau \ln \tilde{y}_{t,\tau} = (1-\tau)(\ln y^i_t - m_t) + m_t + \tau(2-\tau)\Delta^2_t/2$ is also normally distributed:

$$\ln \hat{y}^i_{t,\tau} \sim \mathcal{N}\left((1-\tau)\alpha^t(\ln y^i_0 - m_0)_t + m_t + \tau(2-\tau)\Delta^2_t/2\,,\,(1-\tau)^2 s^2\left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right)\right). \tag{A.11}$$

Finally, let $V^i_{t,\tau}$ denote agent i's expected utility under the policy τ. With the utility function (25), equation (A.11) implies:

$$\begin{aligned}
\ln\left(1+\beta V^i_{t,\tau}\right) &= \ln \mathsf{E}_{\Theta^i_1 \ldots \Theta^i_t}\left[(\hat{y}^i_t)^\beta \mid y^i_0\right] \\
&= \mathsf{E}_{\Theta^i_1 \ldots \Theta^i_t}\left[\ln(\hat{y}^i_t)^\beta \mid y^i_0\right] + \frac{\beta^2}{2}\mathsf{Var}_{\Theta^i_1 \ldots \Theta^i_t}\left[\ln \hat{y}^i_t \mid y^i_0\right] \\
&= \beta m_t + \beta(1-\tau)\alpha^t(\ln y^i_0 - m_0) + \beta\tau(2-\tau)\left(\frac{\Delta^2_t}{2}\right) + \beta^2(1-\tau)^2\left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right)\left(\frac{s^2}{2}\right) \\
&= \beta \ln \mu_{F_t} + \beta(1-\tau)\alpha^t(\ln y^i_0 - m_0)(1-\tau)^2 + \left[-\beta\left(\frac{\Delta^2_t}{2}\right) + \beta^2\left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right)\left(\frac{s^2}{2}\right)\right].
\end{aligned}$$

But by (21), $\Delta^2_t = \alpha^{2t}\Delta^2_0 + s^2(1-\alpha^{2t})/(1-\alpha^2)$, and hence,

$$\begin{aligned}
\ln\left(1+\beta V^i_{t,\tau}\right) &= \beta \ln \mu_{F_t} + \beta(1-\tau)\alpha^t(\ln y^i_0 - m_0) \\
&\quad -\beta(1-\tau)^2\alpha^{2t}\left(\frac{\Delta^2_0}{2}\right) - \beta(1-\beta)(1-\tau)^2\left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right)\left(\frac{s^2}{2}\right).
\end{aligned} \tag{A.12}$$

The difference in agent i's expected welfare between two policies $\bar{\tau}$ and $\underline{\tau}$ in $\mathcal{P}'$ is thus

$$\begin{aligned}
\ln\left(1+\beta V^i_{t,\bar{\tau}}\right) - \ln\left(1+\beta V^i_{t,\underline{\tau}}\right) &= (\bar{\tau}-\underline{\tau})\left[\beta\alpha^t(m_0 - \ln y^i_0) + \beta(2-\bar{\tau}-\underline{\tau})\alpha^{2t}\left(\frac{\Delta^2_0}{2}\right)\right. \\
&\quad \left. +\beta(1-\beta)(2-\bar{\tau}-\underline{\tau})\left(\frac{1-\alpha^{2t}}{1-\alpha^2}\right)\left(\frac{s^2}{2}\right)\right].
\end{aligned} \tag{A.13}$$

Setting this expression to zero yields the indifference point claimed in Proposition 10. Choosing $\bar{\tau} = 1$ and $\underline{\tau} = 0$ then yields Propositions 9, and Proposition 8 corresponds to the subcase where $\beta = 1$. ||

References

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

[2] Bénabou, R., (1996), "Unequal Societies." NBER Working Paper 5583, May. Revised, March 1997.

[3] Bénabou, R., (1997) "What Levels of Redistribution Maximize Income and Efficiency?" Mimeo, New York University, March .

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

[5] 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.

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

[7] Debreu, G., (1976), "Least Concave Utility Functions." *Journal of Mathematical Economics,* 3, 121-129.

[8] Feldstein, M. (1969) "The Effects of Taxation on Risk Taking," *Journal of Political Economy,* 77, 755–764.

[9] 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, 4, 544–566.

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

[11] Kannai, Y., (1977), "Concavifiability and Constructions of Concave Utility Functions." *Journal of Mathematical Economics,* 4, 1-56.

[12] Kanbur, S.M. (1979) "Of Risk Taking and the Personal Distribution of Income," *Journal of Political Economy* 87(4), 769–797.

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

[14] Kihlstrom R. E. and L. J. Mirman, (1974), "Risk Aversion with Many Commodities." *Journal of Economic Theory,* 8, 361-388.

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

[16] Mitra, T., Ok, E. A. and L. Koçkesen, (1998), "Popular Support for Progressive Taxation and the Relative Income Hypothesis." *Economics Letters*, 58, 69-76.

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

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

[19] Musgrave, R., and Thin, T. (1948) "Income Tax Progression 1929–48," *Journal of Political Economy,* 56, 498–514.

[20] Persson, M. (1983) "The Distribution of Abilities and the Progressive Income Tax," *Journal of Public Economics,* 22, 73–88.

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

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

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

[24] Putterman, L., (1996), "Why Have the Rabble Not Redistributed the Wealth? On the Stability of Democracy and Unequal Property." Mimeo, Brown University.

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

[26] Solon, Gary (1992) "Intergenerational Income Mobility in the United States." *American Economic Review*, 82, 3, 393-408.

[27] 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.

[28] Zimmerman, D. (1992) "Regression Toward Mediocrity in Economic Stature," *American Economic Review,* 82, 3, 409-429.

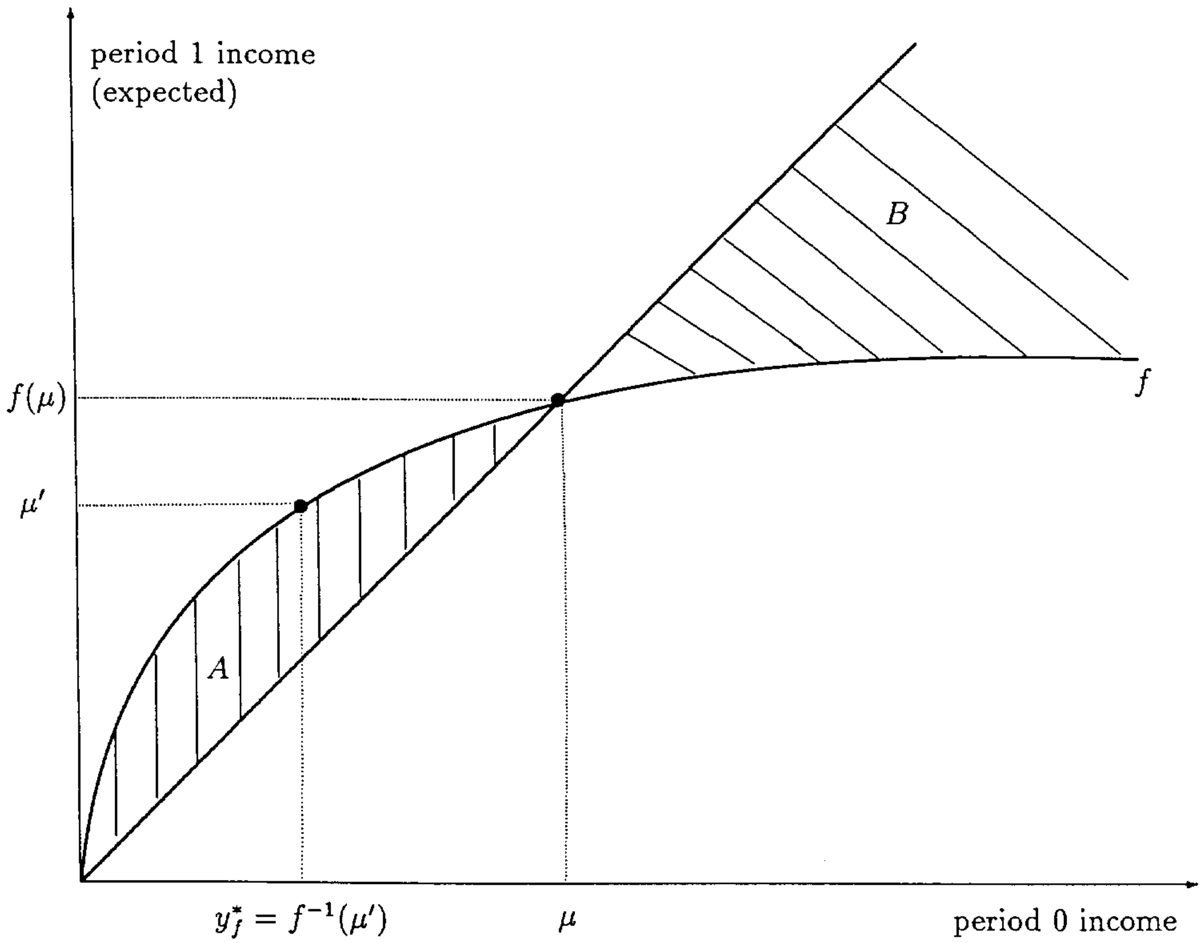

Figure 1: Concave transition function

Note that:

- $A < B$, therefore $\mu' < f(\mu)$
- The figure also applies to the stochastic case, with f replaced by Ef everywhere

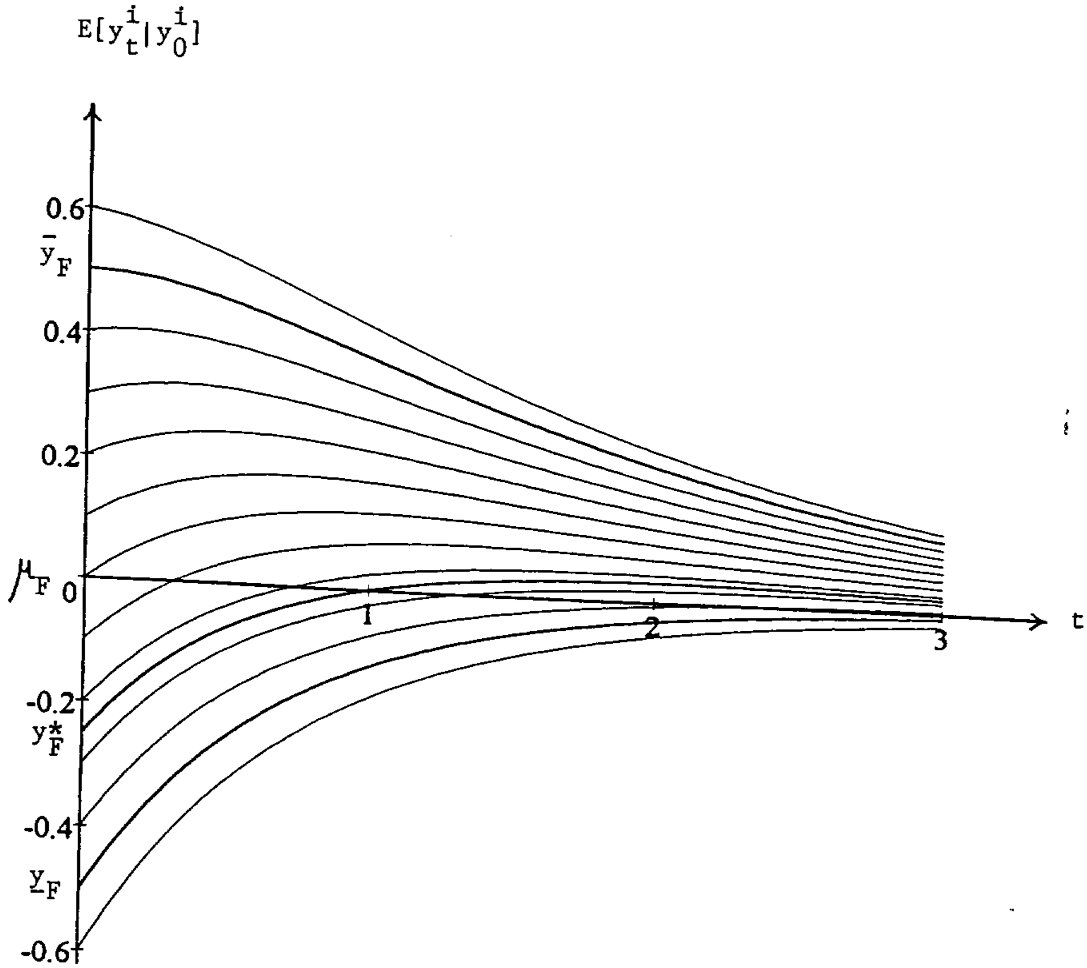

Figure 2: Expected future income under a concave transition function (semi-logarithmic scale)