1. Conference Name: EPCS2006

2. Title of the paper: Saving the POUM Hypothesis from its Saviors

3. Name(s) of Author(s) with the name of the person presenting the paper followed by an asterisk (*):

Philipp Harms* and Stefan Zink

4. Institutional affiliation of author(s), with complete mailing address and e-mail of the presenter.

Philipp Harms*
RWTH Aachen University, Faculty of Business and Economics
Templergraben 64/III, 52062 Aachen, Germany.
Email: harms@rwth-aachen.de.
Phone: +49(0)241 80 96203.

Stefan Zink
University of Konstanz, Dept. of Economics,

5. Abstract of paper.

The two main explanations for the moderate levels of redistribution observed in most democracies refer to deadweight costs of taxation and to people's desire to protect their prospect-of-upward-mobility (POUM). In this paper we demonstrate that these two arguments, which are usually presented in isolation, are mutually reinforcing: if a society is characterized by a sufficient degree of social mobility, losses from redistribution are likely to outweigh the immediate gains, even if current median income is below average. Our result thus suggests that social mobility is much more important in limiting the extent of redistribution than generally assumed.

Saving the POUM Hypothesis from its Saviors

Philipp Harms*
RWTH Aachen University
Study Center Gerzensee

Stefan Zink†
University of Konstanz

16th December 2005

Abstract

The two main explanations for the moderate levels of redistribution observed in most democracies refer to deadweight costs of taxation and to people's desire to protect their prospect–of–upward–mobility (POUM). In this paper we demonstrate that these two arguments, which are usually presented in isolation, are mutually reinforcing: if a society is characterized by a sufficient degree of social mobility, losses from redistribution are likely to outweigh the immediate gains, even if current median income is below average. Our result thus suggests that social mobility is much more important in limiting the extent of redistribution than generally assumed.

*Address: Templergraben 64/III, 52062 Aachen, Germany. Email: harms@rwth-aachen.de. Phone: +49(0)241 80 96203.

†Address: Dept. of Economics, Box D 146, D-78457 Konstanz, Germany. Email: stefan.zink@uni-konstanz.de. Phone: +49(0)7531 88 2605. We are grateful to seminar participants at the Silvaplana Workshop 2004 and at RWTH Aachen University for helpful comments.

1 Introduction

It is not surprising that we observe redistributive taxation: given that most countries' income distributions are markedly skewed to the right, the poor majority clearly benefits from redirecting a part of the rich minority's income into their own pockets – be it via direct transfers or the financing of public goods. The greater puzzle is the rather *moderate* level of observed tax rates. In a static environment without redistribution costs one would expect the median voter to prefer a 100 percent tax over all lower tax rates, and complete redistribution should thus be a natural outcome of the political process.

The two key arguments that have been proposed to explain this apparent "paradox of redistribution" depart from the assumptions expressed in the previous sentence, emphasizing that redistribution is *costly*, and that the environment is *not static*. Thus, a relatively poor median voter may oppose complete redistribution if redistribution comes at a cost – either because the tax base responds to changes in the tax system, or because collecting taxes requires resources. This is the argument brought forward, most prominently, by Meltzer and Richard (1981). Conversely, Benabou and Ok (2001) focus on the role of intertemporal tradeoffs in shaping agents' appetite for redistribution: if taxation is persistent and if individuals have a prospect of upward mobility (POUM) – i.e., if they expect their future income to rise above the population mean – they may oppose excessive tax rates although complete redistribution would maximize their current after–tax income.[1]

The goal of this paper is to demonstrate that these two arguments, which are usually treated separately, are *mutually reinforcing*. We claim that simultaneously accounting for costs of taxation and the prospect of upward mobility provides an explanation of the above–mentioned puzzle which is much more powerful than the two approaches in isolation. To support our claim, we model an economy in which agents' income is determined both by their past income and by chance. There is social mobility in the sense that a sequence of lucky draws enables agents whose initial income is relatively low to overtake those agents whose initial income is relatively high. Redistribution takes place in the form of a linear tax-cum-transfer scheme and is costly – i.e., only a part of collected tax revenues

[1]Harms and Zink (2004) offers a broader survey of the various approaches to explain the paradox of redistribution.

is available for transfers and the provision of public goods. Moreover, redistribution is persistent – that is, a tax rate that is determined at time t prevails for the next n periods. We show that in this setup the tax rate preferred by the median voter may be close to zero even if his current income is below the economy–wide average.

This is not a trivial result. Benabou and Ok (2001) emphasize that the transition function linking current and expected future income has to be *concave* to support the POUM hypothesis.[2]. However, if credit constraints are the source of income persistence, the notion of a concave relationship between current and expected future incomes is problematic, since indivisible investments are likely to result in convex parts of the transition function. If such threshold effects are an important real–world phenomenon, the POUM argument, as presented by Benabou and Ok (2001), does not bite.[3] We show that taking into account costs of redistribution allows for a much less restrictive concept of social mobility: in our setup, the median voter need not expect his income to rise *above* the mean in the future. All we need is a plausible mechanism that ensures convergence of expected incomes towards the mean.[4]

Second, if there is social mobility, discincentives play a much more powerful role than in Meltzer and Richard (1981). Although concerns about other agents' reaction to taxation may already attenuate the redistributive ambitions of poor voters in a static framework, these concerns are unlikely to dominate the immediate gains from redistribution if the median voter's income is much lower than the mean – especially if agents' propensity to adjust their factor supply is low.[5] We show that, by moving agents' expected incomes ever closer to the mean, social mobility considerably reinforces the relative importance of redistribution costs for agents' preferred tax rate.

[2] "There exists a range of incomes below the mean where agents oppose lasting redistribution if (and, in a sense, only if) tomorrow's expected income is an increasing and *concave* function of today's income" (Benabou and Ok, 2001:449)

[3] Experimental support for the POUM-hypothesis provided by Checchi and Filippin (2003) does not solve this problem since concavity of transition functions is a component of their experimental design.

[4] In fact, we will claim that social mobility is *equivalent* to convergence of income prospects.

[5] The original argument by Meltzer and Richard (1981) requires a large labor supply elasticity. A related argument discussed by, among others, Epple and Romer (1991), Persson and Tabellini (1992) and Cremer et al. (1996) requires a high degree of international factor mobility.

To summarize: if taxation does not distort economic decisions, one has to impose restrictive (and often implausible) assumptions on the dynamics of the income distribution to support the prospect–of–upward-mobility hypothesis. Conversely, in a very unequal society without social mobility, the tax-base effects of redistribution are likely to pale relative to the immediate benefits from redistributive taxation. The goal of this paper is to demonstrate that combining the two aspects provides a promising argument to explain the paradox of redistribution.

The rest of the paper is structured as follows: Section 2 introduces our assumptions on the political process and on agents' income dynamics, and derives our key result on the relationship between social mobility, the persistence of redistribution and the equilibrium tax rate. Section 3 offers an approach to explain persistence of redistribution as an institution that emerges as the outcome of a majority vote. Section 4 summarizes and concludes. Most proofs are shifted to the Appendix.

2 Persistent redistribution, social mobility, and the equilibrium tax rate

2.1 The political process

We consider an economy populated by a large number of agents whose mass is normalized to 1. Following Benabou and Ok (2001), we consider an endowment economy where the income dynamics is exogenous. Agents are heterogeneous with respect to their current and future incomes y_t^i. While period 0 incomes are known, there is some uncertainty about the future.

In period 0, there is a majority vote over the extent of income taxation. Knowing y_0^i, agents choose the tax rate $\tau \in [0, 1]$ of a linear tax–cum–transfer scheme. In periods 0 to n, incomes are taxed at rate τ, and the proceeds are redistributed equally among the population. We assume that there are convex costs of collecting taxes. More specifically, for given average income $\overline{y}_t$, tax revenues amount to $\tau\overline{y}_t$, but only $(\tau - \tau^2/2)\overline{y}_t$ can be redistributed in a lump–sum manner. This specification is meant to capture the "iceberg costs" of redistribution, but also the disincentive effect associated with distortionary taxation, i.e. the reduction in the supply of labor, capital and effort.[6]

Utility is assumed to be linear in consumption and there is no discounting. Of course, in an uncertain environment, an income tax provides ex–ante insurance against idiosyncratic risks, and agents' preferred tax rates are likely to increase in the volatility of shocks and in their degree of risk aversion. By working with a linear utility function, we abstract from this effect. Moreover, by assuming zero discounting we take the assumption that future utility levels matter for current distributional preferences to the extreme. If the discount factor is not too small, our results remain valid in a more general setting where agents discount future utility.

Given these assumptions, we can derive agent i's preferred tax rate as the value of τ which maximizes the sum of current and future expected net incomes:

$$\mathrm{E}\left\{ \sum_{t=0}^{n} \left[(1-\tau) y_t^i + (\tau - \tau^2/2)\overline{y}_t \right] \middle| \, y_0^i \right\}. \tag{1}$$

Utility as a function of τ is single–peaked and the maximum is reached for the following tax rate:

$$\tau^{i,opt} = \max\left\{ 0, 1 - \frac{\sum_{t=0}^{n} \mathrm{E}(y_t^i)}{\sum_{t=0}^{n} \mathrm{E}(\overline{y}_t)} \right\} \tag{2}$$

Equation (2) shows that an agent's preferences depend on the average distance between his expected income and the population mean. Thus, the attitude towards taxation is determined not by the *current* but by the *average future* income position. For $\sum_{t=0}^{n} \mathrm{E}(y_t^i) \geq \sum_{t=0}^{n} \mathrm{E}(\overline{y}_t)$ the agent wants no taxation at all. Benabou and Ok (2001) demonstrate that this can be the case if the income transition process is characterized by decreasing returns. Conversely, if $\sum_{t=0}^{n} \mathrm{E}(y_t^i) - \sum_{t=0}^{n} \mathrm{E}(\overline{y}_t) < 0$, the preferred tax rate depends negatively on the distance between expected personal and expected average income, and moving closer to the mean reduces the demand for redistribution. Whether such a convergence actually takes place depends on agents' income dynamics to which we turn in the following section.

[6]The alternative approach to explicitly model gross–income as decreasing in τ would complicate the analysis without providing any further insights.

2.2 Income dynamics

We assume that agents' incomes evolve in the following fashion:

$$y_{t+1}^i = f(y_t^i, a_{t+1}^i), \tag{3}$$

where a_{t+1}^i represents a stochastic shock. The function f is defined on a subset $Y \times A$ of $\mathbb{R}_+^2$ with $Y = [0, y^{max}]$ and $A = [a^{min}, a^{max}]$. Both Y and A are equipped with the corresponding Borel σ–algebras $\mathcal{Y}$ and $\mathcal{A}$. The transition function is continuous and strictly increasing in both arguments: starting with a high value of y improves future income prospects, but agents with modest initial conditions have a chance of becoming rich if they draw a particularly high realization of a.

There are various ways to motivate the persistence of incomes as expressed by equation (3): one interpretation would be that individuals' income paths are relatively stable because earnings are largely determined by some invariant factors like talent or skill. It may also be due to capital market imperfections which make investment opportunities depend on current resources. Alternatively, (3) could be read as formalizing the well–documented persistence of income across generations (see Bowles and Gintis, 2002). This persistence, in turn, may reflect the genetic transmission of talent, or once again borrowing limitations.[7] The latter interpretation suggests to cast the model in terms of *dynasties* instead of individuals. However, since we do not explicitly model people's bequest behavior, we prefer to stick to the concept of individual agents with a planning horizon that stretches over several periods. Apart from continuity and monotonicity, there are no other restrictions on the the shape of f, i.e. we admit both concave or convex segments.

Regarding the distribution of shocks, we impose that a_{t+1}^i is independent of y_t^i and drawn according to some probability law π on $(A, \mathcal{A})$ which is absolutely continuous with respect to the Lebesgue–measure. The corresponding density is positive in the interior of A so that the distribution function, say F, is strictly increasing. The realizations of a_t^i are independent across individuals.

Given our assumptions regarding f and the distribution of a, the income dynamics for one individual can be described by some time–invariant mapping

[7] For an influential paper that formalizes this relationship see Galor and Zeira (1993).

$P : Y \times \mathcal{Y} \to [0,1]$ with[8]

$$P(y, C) = \pi\left(f_y^{-1}(C)\right) \quad (y \in Y, C \in \mathcal{Y}).$$

If current income is y, then $P(y, C)$ denotes the probability that income in the following period lies in the set C. It is straightforward to show that P is a transition kernel, i.e. for every $y \in Y$, $P(y, \cdot)$ is a probability measure on $(Y, \mathcal{Y})$, and for every $C \in \mathcal{Y}$, $P(\cdot, C)$ is $\mathcal{Y}$–measurable. For future reference, we denote the space of all probability measures on the measurable space $(Y, \mathcal{Y})$ by $\mathcal{P}(Y)$.

Since f is increasing in y and the realization of a is independent of the past, we can infer how the probability measures $P(y, \cdot)$ relate for different values of y.

Lemma 1 *The transition function P is monotone in the sense that the probability measure $P(y, \cdot)$ dominates $P(y', \cdot)$ in the first–order sense (written $P(y, \cdot) \succeq P(y', \cdot)$) if $y \geq y'$.*

Intuitively, rich individuals have better income prospects before the resolution of uncertainty. Nonetheless, the presence of random shocks guarantees that people's position in the income hierarchy may change over time. In particular, initially poor agents may overtake richer ones if they are sufficiently lucky. Our central assumption is that this can occur even if the initial income difference is quite large. Hence, we assume that our economy allows for *social mobility* in the following sense:

Assumption 1

Social Mobility (SM) The sequences y_t and y'_t with

$$\begin{aligned} y_0 = y^{max} \quad &\textit{and} \quad y_{t+1} = f(y_t, a^{min}) \\ y'_0 = 0 \quad &\textit{and} \quad y'_{t+1} = f(y'_t, a^{max}) \end{aligned}$$

satisfy $\lim_{t\to+\infty} y_t < \lim_{t\to+\infty} y'_t$.

The intuition behind this assumption is quite simple: equation (3) states that an agent's future incomes depend on his current income and on a stochastic shock. Despite a positive correlation between current and future incomes, the position in the income hierarchy may thus change. Under condition SM, the random variables

[8] Here f_x denotes the x–section of f.

a_t^i are sufficiently important such that a series of favorable shocks allows an agent with low initial income to eventually overtake an agent who started out from a much higher level but kept drawing the lowest realization of a. This may occur in finite time: under SM, there exists some natural number $k > 0$ and some $\hat{y}$ in the interior of Y such that $y_k < \hat{y} < y'_k$. Since the density of π with respect to the Lebesgue measure is strictly positive in the interior of A, both the probability that an individual with a high initial level of y ends up below $\hat{y}$ and the probability that a dynasty with a low initial level of y reaches the interval $[\hat{y}, \overline{y}]$ is strictly positive. Let $\epsilon > 0$ denote the minimum of these probabilities.[9] For this kind of social mobility to play a role within our planning horizon, we take for granted that k is smaller than n. In section 3, we will argue that societies' constitutional design may guarantee that this relation holds.

For a given income level y^i, we can denote the distribution of agent i's income in the following period as $P(y^i, \cdot)$. For any probability measure μ on $(Y, \mathcal{Y})$, we define $T\mu$ by:

$$T\mu(C) := \int P(s, C)\, \mu(s) ds \quad (C \in \mathcal{Y}).$$

We can interpret $T\mu(C)$ as the probability that an agent's income in period $t+1$ lies in C, given that his income in period t is drawn according to μ. In other words, if income is distributed according to μ in period t, it is distributed according to $T\mu$ in $t+1$. Monotonicity of the transition function P translates into the following property of the operator T:

Lemma 2 *The operator T is increasing with respect to $(\mathcal{P}(Y), \succeq)$, i.e. $\mu \succeq \mu'$ $(\mu, \mu' \in \mathcal{P}(X))$ implies $T\mu \succeq T\mu'$.*

The introduction of T allows us to characterize the stochastic evolution of an individual's income in a concise way: Given y_0^i, the distribution of agent i's income evolves as follows:

$$\delta_{y_0^i}, T\delta_{y_0^i} = P(y_0^i, \cdot), T^2\delta_{y_0^i}, \ldots, T^{(n-1)}\delta_{y_0^i}$$

where T^k is the kth iteration of T and δ_x is the Dirac–measure with unit mass at x.

[9] Conversely, the existence of some $\hat{y} \in (0, \overline{y})$ such that the poor end up above $\hat{y}$ and the rich below it with positive probability also implies SM. We make use of this property in the appendix.

In what follows, we want to argue that SM is equivalent to a convergence of income prospects. To establish such a result, we need to quantify the distance between any two income distributions. We therefore endow $\mathcal{P}(Y)$ with the Kolmogorov metric which gives the distance between two probability measures μ and ν as the vertical distance between the corresponding distribution functions F_μ and F_ν: $d_K(\mu, \nu) := \sup_{x \in X} |F_\mu(x) - F_\nu(x)|$. Using this topology, we can show that the sequence $T^k\mu$ converges to some probability measure $\tilde{\mu} \in \mathcal{P}(X)$, regardless of the initial distribution μ, if, and only if, SM holds:

Theorem 1 *The following statements are equivalent:*

1) SM

2) There exists some probability measure $\tilde{\mu} \in \mathcal{P}(Y)$ which is the unique fixed point of T. For all $\mu \in \mathcal{P}(X)$ the sequence $T^j\mu$ converges to $\tilde{\mu}$ exponentially fast in the Kolmogorov metric.

If SM is satisfied, initially poor individuals have a chance to outdo the initially rich after k periods. This guarantees that, while T itself is a weak contraction, the kth iteration T^k is a strict contraction. After j steps, the income of an agent who started with y_0^i is distributed according to $T^j\delta_{y_0^i}$. The distance between this income path and the fixed point of T therefore never increases in terms of the Kolmogorov metric, but strictly declines after k steps. Hence, for j large enough, $T^j\delta_{y_0^i}$ is arbitrarily close to $\tilde{\mu}$. In fact, the proof of Theorem 1 also gives us some idea about the *speed of convergence*. In particular, $d(T^j\mu, \tilde{\mu}) \leq (1-\epsilon)^{[j/k]} d(\mu, \tilde{\mu})$, where ϵ is the minimum of $\mathrm{P}(y_k \leq \hat{y})$ and $\mathrm{P}(y'_k \geq \hat{y})$.[10] Obviously, the expression on the right hand side depends on the relative importance of initial conditions and luck in determining agents' incomes: if, for instance, the economic past does not play an important role, individual income expectations converge fast, which would be reflected by a low value of k and, a high value of ϵ. This, in turn, would imply a quick convergence of $T^j\mu$ and $\tilde{\mu}$.

Conversely, the convergence of income distributions implies that the law of motion which governs individual income dynamics allows for social mobility in the sense of SM: if the rich could never be overtaken by the poor, long–run income prospects would crucially depend on agents' starting position, and the

[10] As usual, $[x]$ denotes the largest natural number below x.

convergence of income distributions would never take place. Theorem 1 thus states that social mobility and convergence of income prospects represent two sides of the same coin.

If SM holds, individuals experience all income levels over time. In particular, the rich will become poor and conversely the poor will eventually be rich. Since individual income dynamics follows this process independently and since there are a very large number of agents, the long–run distribution of individual incomes $\tilde{\mu}$ can be interpreted as the cross–sectional income distribution at every point in time. That is, while individuals keep changing places within the income hierarchy, the economy as a whole remains stationary:

Assumption 2 *The cross–sectional income distribution is given by $\tilde{\mu}$ in every period. The corresponding mean income is constant and denoted by $\overline{y}$.*

Assumption 2 combined with Theorem 1 implies that, even if an individual is not yet close to the economy–wide average, his distance from the equilibrium distribution, as measured by the Kolmogorov metric, never increases but eventually declines. Since, for every individual, the distribution function of future incomes becomes similar to the economy's cross–sectional distribution function, expected income in the future approaches average income:

Corollary 1 *Under SM, for all $y_0^i \in Y$, $\lim_{m\to+\infty} E(y_m|y_0^i) = \overline{y}$.*

If some sequence b_m of real numbers converges to $\tilde{b}$, then the sequence $1/(m+1)\sum_{i=0}^{m} b_i$ is also convergent to $\tilde{b}$. Corollary 1 therefore has the following consequence:

Corollary 2 *Under SM, for all $y_0^i \in Y$,*

$$\lim_{m\to+\infty} E\left[1/(m+1)\sum_{t=0}^{m} y_t^i \middle| y_0^i\right] = \lim_{m\to+\infty} 1/(m+1)\sum_{t=0}^{m} E\left(y_t^i \middle| y_0^i\right) = \overline{y}\ . \qquad (4)$$

Hence, if the time horizon is large enough, SM implies that, regardless of initial conditions, the mean of an agent's expected future incomes equals the population mean.

2.3 The political equilibrium

Section 2.1 showed that people's attitude towards redistribution depends crucially on the average distance between expected incomes and the population mean. Under our assumption that the economy as a whole is stationary, the individual's preferred tax rate (see equation (2)) can be rewritten as:

$$\tau^{opt,i} = \max\left\{0, 1 - \frac{1/(n+1)\sum_{t=0}^{n} \mathrm{E}(y_t^i)}{\overline{y}}\right\}, \tag{5}$$

where n denotes the number of periods during which a tax chosen at $t = 0$ is in place.

Section 2.2 described the evolution of agents' income prospects. On the one hand, the monotonicity properties stated by Lemma 1 and Lemma 2 imply that, for any two agents a and b with $y_0^a \geq y_0^b$ and any finite value of n, we have $T^n\delta_{y_0^a} \succeq T^n\delta_{y_0^b}$ and thus $1/(n+1)\sum_{t=0}^{n} \mathrm{E}(y_t^a|y_0^a) \geq 1/(n+1)\sum_{t=0}^{n} \mathrm{E}(y_t^b|y_0^b)$. Hence, agents' expected incomes – averaged over $n+1$ periods – depend on their initial position. Since preferences are moreover single–peaked, we can conclude that the agent with median income in period 0 (y_0^M) is politically decisive.

On the other hand, we know from Theorem 1 that, for all agents, the distribution function of future incomes approaches the cross–sectional distribution function, and that expected income – averaged over all future time periods – converges to the population mean (Corollary 2). Hence, the distance between $1/(m+1)\sum_{t=0}^{m} \mathrm{E}(y_t^M|y_0^M)$ and $\overline{y}$ shrinks as m increases. We are thus ready to state our central result:

Theorem 2 *If SM holds, the majority vote leads to an arbitrarily low tax rate, given that taxation is sufficiently persistent.*

Note that our argument does not rest on the assumption that n is infinite: since T^k is a strong contraction, a relatively short time horizon may be sufficient to substantially tame agents' appetite for redistribution. Moreover, the cross–sectional distribution of incomes need not be characterized by declining inequality.[11] What matters is how agents assess the prospect of upward mobility, i.e. the feasibility of "rags–to–riches" careers. To support the perception of a

[11] Recall that we assumed that the distribution of income and thus the extent of inequality is stationary.

high degree of social mobility, social and political institutions have to be such that chance and effort are at least as important in determining future incomes as initial conditions.[12]

3 Endogenizing the persistence of redistribution

Our results so far showed that, given SM, the support for redistribution is declining in the persistence of the proposed tax scheme. We assumed that taxes are persistent enough for this kind of convergence to play a role. In what follows, we will take the effect of persistence on the level of taxes as granted and show that limits to tax changes may be deliberately anchored in a society's constitutional design.

To make our point as simple as possible, we consider the following variant of the framework described above: agents live for three periods. In the first period, there is a constitutional election on how persistent future taxation will be. More specifically, a majority vote decides whether the tax rate chosen at the beginning of the second period (τ_2) is applied in the third period as well, or if a vote on taxation takes place in every period. As before, we assume a simple tax–cum–transfer scheme, the proceeds of which, $\tau\overline{y}$, less the costs of taxation, $\tau^2/2\overline{y}$, are distributed in a lump–sum manner.

For simplicity, we impose an extreme form of social mobility: in every period, agents' incomes are drawn according to some fixed probability law χ. Hence, regardless of the current income realization, everybody has the same income prospects for the future, and the individual distribution of future income coincides with the cross-sectional income distribution.[13] The probability law χ is continuous and its distribution function is denoted by F. We assume that median income $y^M = F^{-1}(0.5)$ lies below the population average $\overline{y}$.

Before we turn to the constitutional election, we look at the tax–setting stage under the two different scenarios of persistence. In both cases, the person with median income at the time of the vote is politically decisive. If taxation is not

[12]Piketty (1995) develops a model in which agents learn about the relative importance of effort and initial conditions in determining economic success, and in which their distributional preferences are shaped by the experienced degree of social mobility.

[13]Our example thus corresponds to the case where f does not depend on its first argument y.

persistent, a majority vote determines τ in every period. The median voter maximizes his net income (including transfers), which is $(1-\tau)y^M + (\tau - \tau^2/2)\overline{y}$. This implies a tax rate equal to

$$\tau_2^{M,opt} = \tau_3^{M,opt} = 1 - y^M/\overline{y} \tag{6}$$

in both periods.

If the tax rate chosen in period 2 is applied in both the second and the third period, the second period median voter has the following expected net income:

$$\begin{aligned}(1-\tau)y^M + (\tau - \tau^2/2)\overline{y} + \mathrm{E}\left[(1-\tau)y_3 + (\tau - \tau^2/2)\overline{y}\right] \\ = y^M + \overline{y} + \tau(\overline{y} - y^M) - \tau^2\overline{y}.\end{aligned}$$

Maximizing this expression with respect to τ yields:

$$\tau^{M,opt} = 1 - \frac{0.5(y^M + \overline{y})}{\overline{y}} = 0.5(1 - y^M/\overline{y}). \tag{7}$$

Obviously, the persistence of redistribution leads to a lower tax rate chosen in period 2. This is an example of the general result derived above: if the period-2 median voter cannot be sure to gain from redistribution in period 3, his incentives to implement high taxes in the second period decrease.

Consider now the constitutional election in period 1: since we assumed an extreme form of social mobility, with all agents expecting to earn mean income in the future, there is unanimity in period 1 that future tax rates should be zero. If an outright prohibition of taxation is not feasible – for example, because the government needs to remain flexible in the face of aggregate uncertainty (see Casamatta et al., 2000, as well as Boyer and Laffont, 1998) – it is in agents' interest to establish rules which limit the future extent of taxation. Persistence of redistribution is such a rule, since agents rationally anticipate that the prospect of upward mobility reduces the median voter's preferred tax rate in period 2.

Due to our assumption of an extreme form of social mobility – namely that all agents share the same expectations about their future incomes – our result mirrors the case where the institutional choice takes place behind a perfect "veil of ignorance". This correspondence would break down if we dropped the assumption that incomes are independent across periods. The conclusions would, however, be similar: if convergence is not immediate, agents' income prospects depend on their current income level. However, under SM, the median voter at the time of the constitutional election can still expect to move closer to the mean in the future, which reduces his interest in redistribution. Knowing that persistence will lead to less redistribution, he will implement a constitutional system which excludes frequent tax changes so that a tax scheme, once chosen, prevails for a longer time span.

4 Summary and conclusions

Our goal was to demonstrate that, rather than providing two separate explanations for the "paradox of redistribution", the deadweight–cost–based argument of Meltzer and Richard (1981) and the prospect–of–upward–mobility hypothesis of Benabou and Ok (2000) are mutually reinforcing: if a society is characterized by a sufficient degree of social mobility in the sense that adverse initial conditions can be overcome and beneficial initial situations may deteriorate, today's median voter identifies with the long–run average of the cross–sectional income distribution. This makes him reluctant to tolerate the distortions associated with redistributive taxation.

While we presented our key arguments using a highly stylized model, we believe that the mechanisms we emphasized go a long way in explaining the "paradox of redistribution". More specifically, the empirical finding of Rodriguez (1999) that measured inequality does not affect the extent of redistribution can be explained by acknowledging the role of *income prospects* in shaping agents' distributional preferences. In fact, there is ample survey evidence that agents desire less redistribution if they expect higher incomes in the future (Ravaillon and Lokshin, 2000; Alesina and La Ferrara, 2005). However, the earnings dynamics that Benabou and Ok (2001) claim to be necessary for such an effect to make theoretical sense – namely concavity of the transition function over the whole range of income values – may not be plausible. We believe that, by focusing on a less restrictive definition of social mobility, it is possible to return to the POUM hypothesis the theoretical and empirical weight it deserves.

5 References

Alesina, A., La Ferrara, E., 2005. Preferences for redistribution in the land of opportunities. Journal of Public Economics 89, 897–931.

Bhattacharya and Majumdar, 2001. On a class of stable random dynamical systems: Theory and applications, Journal of Economic Theory 96, 208–229.

Benabou, R., Ok, E.A., 2001. Social mobility and the demand for redistribution: the POUM hypothesis. Quarterly Journal of Economics 116, 447–487.

Bowles, S. and H. Gintis, 2002. The inheritance of inequality. Journal of Economic Perspectives 16, 3–30.

Boyer, M., Laffont, J.J., 1998. Toward a political theory of the emergence of environmental incentive regulation, Rand Journal of Economics 30, 137–157.

Casamatta, G., Cremer, H., Pestieau, P., 2000. Political sustainability and the design of social insurance, Journal of Public Economics 75, 341–364.

Checchi, D., Filippin, A., 2003. An experimental study of the POUM hypothesis. IZA Discussion Paper No. 912, Bonn.

Cremer, H., Fourgeaud, V., Monteiro, M.L., Marchand, M.,Pestieau, P., 1996. Mobility and redistribution: a survey. Public Finance / Finances Publiques 51, 325–52.

Epple, D., Romer, T., 1991. Mobility and redistribution. Journal of Political Economy 99, 828–858.

Galor, O.,Zeira, J., 1993. Income distribution and macroeconomics. Review of Economic Studies 60, 35–52.

Harms, P., Zink, S., 2003. Limits to redistribution in a democracy: a survey. European Journal of Political Economy 19, 651–668.

Meltzer, A.H., Richard, S.F., 1981. A rational theory of the size of government. Journal of Political Economy 89, 914–927.

Persson, T., Tabellini, G., 1992. The politics of 1992: fiscal policy and European integration. Review of Economic Studies 59, 689–701.

Piketty, T., 1995. Social mobility and redistributive politics. Quarterly Journal of Economics 110, 551–584.

Putterman, L., 1997. Why have the rabble not redistributed the wealth? On the stability of democracy and unequal property. In: Roemer, J.E. (Ed.), Property Relations, Incentives and Welfare. IEA Conference 115, pp. 359–389.

Ravallion, M., Lokshin, M., 2000. Who wants to redistribute? The tunnel effect in 1990s Russia. Journal of Public Economics 76, 87–104.

Rodriguez, F.C., 1999. Does distributional skewness lead to redistribution? Evidence from the United States. Economics and Politics 11, 171–199.

Roemer, J.E., 1998. Why the poor do not expropriate the rich: An old argument in new garb. Journal of Public Economics 70, 399–424.

6 Appendix

Proof of Theorem 1:
Before we start the proof, we need a little bit of notation. In the main text we characterized the evolution of income by f. This analysis can easily be extended to the k–period transition. We recursively define the following family of mappings $f^k : Y \times A^k \to X$ via

$$f^1(y_0, a_1) := f(y_0, a_1)$$

and

$$f^k(y_0, a_1, \ldots, a_k) := f\left[f^{k-1}(y_0, a_1, \ldots, a_{k-1}), a_k\right].$$

Thus, if an individual starts from y_0 and realizes the vector $\mathbf{a} = (a_1, \ldots, a_k)$ in the subsequent periods, his endowment in period k is given by $f^k(y_0, \mathbf{a})$.[14] It is easy to prove by induction that f^k is strictly increasing in its first argument. Just as in the main text, we can show that f^k implies well–defined k–step transition functions P^k and n–step transition operators T^k:

$$T^k\mu(C) = \int \mu\left[f^k(\cdot, \mathbf{a})^{-1}(C)\right] \pi^k(d\mathbf{a}) \tag{8}$$

and

$$P^k(x, C) = \pi^k\left[f^k(x, \cdot)^{-1}(C)\right] \tag{9}$$

where π^k denotes the product measure $\pi \otimes \ldots \otimes \pi$ on the product space $(A^k, \mathcal{A} \otimes \ldots \otimes \mathcal{A})$. A simple argument shows that P^k and T^k also result from a k–time iteration of P and T, respectively. Consequently, T^k is increasing with respect to $(\mathcal{P}(X), \succeq)$ for all k. We can prove the following equivalence:

[14]Bold face letters indicate vectors.

Lemma 3 *The following statements are equivalent:*

i) SM

ii) There exists some $k > 0$ and some $\hat{y} \in (0, y^{max})$ such that $\pi^k\{\mathbf{a} \in A^k | f^k(0, \mathbf{a}) \geq \hat{y}\} > 0$ and $\pi^k\{\mathbf{a} \in A^k | f^k(y^{max}, \mathbf{a}) \leq \hat{y}\} > 0$

Proof. i) $\Rightarrow$ ii): There exists some $k > 0$ such that $f^k(0, a^{max}, \ldots, a^{max}) > f^k(y^{max}, a^{min} \ldots, a^{min})$. Choose $\hat{y}$ between these real numbers. Due to continuity of f, the sets $\{\mathbf{a} \in A^k | f^k(0, \mathbf{a}) \geq \hat{y}\}$ and $\{\mathbf{a} \in A^k | f^k(y^{max}, \mathbf{a}) \leq \hat{y}\}$ have positive Lebesgue measure. Absolute continuity of π with respect to the Lebesgue–measure proves the claim.
ii) $\Rightarrow$ i): Since f is strictly increasing in its first argument, both sequences $f^k(0, a^{max}, \ldots, a^{max})$ and $f^k(y^{max}, a^{min} \ldots, a^{min})$ are monotone and convergent. Since f is strictly increasing in a and π is continuous, $\pi^k\{\mathbf{a} \in A^k | f^k(0, \mathbf{a}) \geq \hat{y}\} > 0$ and $\pi^k\{\mathbf{a} \in A^k | f^k(y^{max}, \mathbf{a}) \leq \hat{y}\} > 0$ imply $f^k(0, a^{max}, \ldots, a^{max}) > \hat{y} > f^k(y^{max}, a^{min}, \ldots, a^{min})$ which proves the claim. □

We are now ready to prove Theorem 1. 1) $\Rightarrow$ 2): Fix $k > 0$ and $\hat{y} \in (0, y^{max})$ such that $\pi^k\{\mathbf{a} \in A^k | f^k(0, \mathbf{a}) \geq \hat{y}\} > 0$ and $\pi^k\{\mathbf{a} \in A^k | f^k(y^{max}, \mathbf{a}) \leq \hat{y}\} > 0$ and let $\epsilon > 0$ denote the minimum of these probabilities (see Lemma 3). Following the reasoning in the proof of Theorem 2.1 in Bhattacharya and Majumdar (2001), we will first show that for all pairs (μ, ν)

$$d_K(T^k\mu, T^k\nu) \leq (1 - \epsilon) d_K(\mu, \nu)$$

holds, i.e. T^k is a contraction.

For this purpose, fix a pair (μ, ν). It follows from (8) that, for $y \in Y$,

$$\left|T^k\mu[0, y] - T^k\nu[0, y]\right| \\ = \left| \int \mu \left[f^k(\cdot, \mathbf{a})^{-1}[0, y]\right] - \nu \left[f^k(\cdot, \mathbf{a})^{-1}[0, y]\right] \pi^k(d\mathbf{a}) \right|. \quad (10)$$

Suppose first that $y \geq \hat{y}$ and define

$$\Gamma_y := \left\{\mathbf{a} \in A^k | f^k(0, \mathbf{a}) \leq y\right\}.$$

The mapping $f^k(\cdot, \mathbf{a})$ being strictly increasing and continuous has two useful implications:

i) $f^k(\cdot, \mathbf{a})^{-1}[0, y]$ is either empty or has the form $[0, z]$ for some well–defined $z \in Y$,

ii) $f^k(\cdot, \mathbf{a})^{-1}[0, y] = Y$ for $\mathbf{a} \in \Gamma_y$.

Using ii), we obtain, for $\mathbf{a} \in \Gamma_y$,

$$\mu \left[f^k(\cdot, \mathbf{a})^{-1}[0, y] \right] - \nu \left[f^k(\cdot, \mathbf{a})^{-1}[0, y] \right] = \nu(X) - \mu(X) = 0$$

so that (10) equals

$$\left| \int_{A^k \backslash \Gamma_y} \mu \left[f^k(\cdot, \mathbf{a})^{-1}[0, y] \right] - \mu \left[f^k(\cdot, \mathbf{a})^{-1}[0, y] \right] \, \pi^k(d\mathbf{a}) \right| . \tag{11}$$

Application of i) guarantees that, for all $\mathbf{a} \in A^k \backslash \Gamma_y$, the integrand in (11) is either 0 or equal to $|\mu[0, z] - \nu[0, z]|$ for some $z \in Y$. In both cases, it is obviously not larger than $d_K(\mu, \nu)$.

Now, $\pi^k(A^k \backslash \Gamma_y) \leq (1 - \epsilon)$ is implied by $\Gamma_y \supseteq \left\{ \mathbf{a} \in A^k | f^k(y^{max}, \mathbf{a}) \leq \hat{y} \right\}$. We therefore obtain

$$\left| T^k(\mu)[0, y] - T^k(\nu)[0, y] \right| \leq (1 - \epsilon) d_K(\mu, \nu). \tag{12}$$

In case of $y < \hat{y}$, let $\Gamma_y = \left\{ \mathbf{a} \in A^k | f^k(0, \mathbf{a}) > y \right\} \supseteq \left\{ \mathbf{a} \in A^k | f^k(0, \mathbf{a}) \geq \hat{y} \right\}$. Since $f^k(\cdot, \mathbf{a})^{-1}[0, y] = \emptyset$ for $\mathbf{a} \in \Gamma_y$, we find

$$\begin{aligned} &\left| T^k(\mu)[0, y] - T^k(\nu)[0, y] \right| \\ &\qquad = \left| \int_{A^k \backslash \Gamma_y} \mu \left[f^k(\cdot, \mathbf{a})^{-1}[0, y] \right] - \mu \left[f^k(\cdot, \mathbf{a})^{-1}[0, y] \right] \, \pi^k(d\mathbf{a}) \right| . \end{aligned}$$

By $\pi^k(A^k \backslash \Gamma_y) \leq (1 - \epsilon)$ and i), the above reasoning shows that

$$\left| T^k(\mu)[0, y] - T^k(\nu)[0, y] \right| \leq (1 - \epsilon) d_K(\mu, \nu). \tag{13}$$

Combining (12) and (13) yields

$$d_K(T^k(\mu), T^k(\nu)) \leq (1 - \epsilon) d_K(\mu, \nu).$$

Moreover, T is nonexpansive since

$$\begin{aligned} d_K(T(\mu), T(\nu)) &= \sup_{y \in Y} \left| \int \mu(f(\cdot, a)^{-1}[0, y]) - \nu(f(\cdot, a)^{-1}[0, y]) \, \pi(da) \right| \\ &\leq d_K(\mu, \nu). \end{aligned}$$

Combining that T is nonexpansive and T^k is a contraction yields, that for arbitrary $n \in \mathbb{N}$, $d_K(T^n(\mu), T^n(\nu)) \leq (1-\epsilon)^{[n/k]}$ where $[n/k]$ denotes the largest natural number which is smaller than or equal to n/k.

Since T^k is a contraction and $\mathcal{P}(X)$ endowed with the Kolmogorov distance is a complete metric space (see Bhattacharya and Majumdar, 2001), the Banach fixed point theorem guarantees that T^k has a unique fixed point, say $\tilde{\mu}$. It is easy to see that $\tilde{\mu}$ is also a fixed point of T: $T(\tilde{\mu}) = \tilde{\mu}$. Due to $T^k(T(\tilde{\mu})) = T(T^k(\tilde{\mu})) = T(\tilde{\mu})$ the probability measure $T(\tilde{\mu})$ is also a fixed point of T^k. Uniqueness of the fixed point of T^k guarantees the rest. Hence, $d_K(T^n(\mu), \tilde{\mu}) = d_K(T^n(\mu), T^n(\tilde{\mu})) \leq (1-\epsilon)^{[n/k]} d_K(\mu, \tilde{\mu})$ which proves our claim. □

2) $\Rightarrow$ 1): Suppose all sequences $T^n\mu$ converge to $\tilde{\mu}$. Then $\tilde{\mu}$ must be the unique fixed point of T. Since the topology induced by the Kolmogorov metric is stronger than the weak*–topology, we moreover know that $T^n\mu$ converges weakly to $\tilde{\mu}$. It is easy to see that $\tilde{\mu}$ is non–degenerate: if $\tilde{\mu}$ was equal to some point measure δ_y for some $y \in Y$, $T\delta_y = \delta_y$ would require $P(y, \{y\}) = 1$; due to strict monotonicity of $f(y, \cdot)$, this would contradict non–degenerateness of π. The fact that $\tilde{\mu}$ is non–degenerate implies the existence of $\hat{y} \in Y$ with the following properties: i) $\tilde{\mu}(\{\hat{y}\}) = 0$, ii) $\tilde{\mu}([\hat{y}, y^{max}]) > 0$, and iii) $\tilde{\mu}([0, \hat{y}]) > 0$. It follows from i) that the boundaries of $[0, \hat{y}]$ and $[\hat{y}, y^{max}]$ in Y have $\tilde{\mu}$–measure 0. As a consequence, weak convergence of $T^n\delta_{\overline{x}}$ and $T^n\delta_0$ to $\tilde{\mu}$ implies $\lim_{n\to+\infty} T^n\delta_{y^{max}}([0, \hat{y}]) = \tilde{\mu}([0, \hat{y}]) > 0$ and $\lim_{n\to+\infty} T^n\delta_0([\hat{y}, y^{max}]) = \tilde{\mu}([\hat{y}, y^{max}]) > 0$. These considerations prove the existence of some natural number k for which $T^k\delta_{y^{max}}([0, \hat{y}]) > 0$ and $T^k\delta_0([\hat{y}, y^{max}]) > 0$. Since $T^k\delta_0([\hat{y}, y^{max}]) = P^k(0, [\hat{y}, y^{max}])$ and $T^k\delta_{\overline{x}}([0, \hat{y}]) = P^k(y^{max}, [0, \hat{y}])$, SM must hold according to Lemma 3. □