A Political Matthew Effect: Democratic Redistribution with Feedback Loops

Dylan T. Moore

September 2, 2021

Preliminary Draft

Abstract

In a democratic society where economic inequality and political inequality are mutually reinforcing, is plutocracy inevitable? I explore this question using a simple dynamic model of democratic redistribution. Two candidates iteratively compete in elections fought over redistributive policy. Campaign spending is financed by citizen political donations, creating a feedback loop through which the current distribution of income affects the future distribution. The impact of these "plutocratic feedback loops" depends on assumptions about citizen political and economic behavior, as well as campaign spending technology. Long run convergence to a plutocratic equilibrium can occur for arbitrarily small levels of initial economic inequality. However, the opposite scenario is also possible: a society which is initially extremely unequal may nonetheless be destined for egalitarianism. The long run distribution of income can also exhibit extreme sensitivity to initial conditions. Tiny differences in initial inequality may determine whether democratic redistribution induces or prevents plutocracy.

"For whosoever hath, to him shall be given, and he shall have more abundance: but whosoever hath not, from him shall be taken away even that he hath." –Matthew 13:12, King James Bible

Whether or not democracy is compatible with economic inequality is a fundamental question in democratic theory. Empirical evidence from the United States (Flavin, 2011; Gilens, 2012; Bartels, 2016; Gilens & Page, 2014), and other liberal democracies (Rosset, 2016; Lesschaeve, 2017) suggests that affluent citizens may have a disproportionate influence over policy outcomes in representative democracies. This kind of systematic inequality of political influence seems to violate intuitive notions of the ideal democracy as a polity featuring equality of political power amongst the citizenry [Dahl, 2006].

While there are various possible mechanisms that might generate a positive correlation between wealth/income and political influence in a democracy [Rosset, 2016, Shore, 2016], many of these share a key feature: they predict that greater economic inequality will induce greater inequality of political influence. Empirical evidence form cross-jurisdictional comparisons provides suggestive evidence supporting this hypothesis (Karabarbounis, 2011; Scruggs & Hayes, 2017). On the other hand, inequality of political influence may also induce economic inequality if it leads to redistributive policy that is more favorable to more influential citizens. If relatively affluent citizens have disproportionate political influence, then politicians catering to their policy preferences may adopt policies that further enhance the economic advantages of these citizens. Thus, the evolution of redistribution in a democratic society might be shaped by a feedback loop whereby political inequality begets economic inequality, which begets further political inequality, and so on.

Such feedback loops can, in principle, causes large changes in economic and political inequality over time. This substantially complicates the task of providing an empirically grounded answer to the question of the compatibility of democracy and economic inequality. A society which appears to exhibit a stable compromise between political equality and economic inequality at one point in time could nonetheless collapse into a plutocracy in the long run, purely as a result of the normal operation of democratic institutions.

This document presents preliminary findings from an investigation of this idea using a very simple model of iterated democratic redistribution. In section 1, I consider a model of pure redistribution i.e. (pie-splitting). I assume the political process takes a particular form which determines the evolution of the income distribution and explore the properties of this type of this process, with a particular focus on its implications for the long-run behavior of the income distribution. Section 2 provides a microfoundation for the political process assumed in section 1: an extension of probabilistic voting theory which incorporates campaign spending financed by citizen donations. Section 3 extends the model beyond the case of pure redistribution and considers the case where the only available tool for redistribution is a nonlinear income tax. A surprising connection is found between the conditions supporting a stable egalitarian outcome in the pure redistribution model and the condition permitting stable taxation of the rich in the nonlinear tax model. Section 4 concludes.

1 Simple Model of Pure Redistribution

Consider a model with a finite set of infinitely lived citizens each indexed by some $i \in \{1, 2, \ldots, n\}$. In each time period $t \in \{0, 1, 2, \ldots\}$, each citizen i has some income $I_{t,i} \geq 0$ from which they derive utility. For any time $t > 0$, the *income distribution* at time t, $\mathbf{I_t}$, is a vector of these individual incomes and is an element of the set $\mathcal{I} \equiv \left\{\mathbf{I} \in \mathbb{R}^n_+ : \bar{I} = \frac{1}{n}\sum_{i=1}^n I_i\right\}$ where $\bar{I}$ is the average endowment of a citizen. The *initial income distribution* $\mathbf{I_0}$ is a vector of citizen incomes at time 0, and is an element of the set $\mathcal{I}_0 \equiv \mathcal{I} \setminus \{\mathbf{I} \in \mathcal{I} : I_i = 0 \text{ for some } i\}$.

At each time period, $t \in \{0, 1, \ldots\}$ the citizens participate in a *political process* $\Gamma : \mathcal{I} \to \mathcal{I}$ which maps from the time t income distribution $\mathbf{I_t} \in \mathcal{I}$ to a time $t+1$ income distribution $\mathbf{I_{t+1}} \in \mathcal{I}$. In this section I consider a political process Γ which can be described as

$$\mathbf{I_{t+1}} = \Gamma\left(\mathbf{I_t}\right) \equiv \arg\max_{\tilde{\mathbf{I}} \in \mathcal{I}} \left\{\sum_{i=1}^n \gamma\left(I_{t,i}\right) v\left(\tilde{I}_i\right)\right\} \tag{1}$$

where $\gamma : \mathbb{R}_+ \to \mathbb{R}_{++}$ is a strictly increasing, strictly positive function, and $v : \mathbb{R}_+ \to \mathbb{R}$ is a strictly increasing, strictly concave indirect utility function. That is to say, the income distribution at time $t+1$ is the solution to a weighted utilitarian welfare maximization problem where citizen i's welfare is weighted by $\gamma\left(I_{t,i}\right)$, the output of some increasing function of their time t income. It is important to emphasize that the weights applied to citizen welfare in the objective function in equation (1) are not normatively meaningful. To avoid confusion, I refer to $\gamma(\cdot)$ as the *political influence function* and $\gamma\left(I_{t,i}\right)$ as citizen i's *political influence* at time t.

In this section, I leave the details of the political process unspecified, simply assuming that it's behavior can be described by equation (1). Section 2 introduces a spatial elections model which provides one way to microfound this assumption. Other models of redistributive politics might also be used to justify the type of reduced form description presented in equation (1).

Notice, equation (1) forms a recurrence relation on $\mathcal{I}$ which—together with the set of all possible initial income distributions ($\mathcal{I}_0$)—defines a nonlinear discrete dynamical system. We can thus employ standard methods for the analysis of such systems to answer questions about the dynamics of the distribution of income in this model.

Some properties of the model are immediately clear.

Lemma 1 (Political Process is Rank Preserving)**.** *The political process described by equation 1 is rank-preserving. That is to say for any citizens $i, j \in \{1, 2, \ldots, n\}$ and time period $t \geq 0$*

$$I_{t,i} > I_{t,j} \iff I_{t+1,i} > I_{t+1,j}.$$

Proof. Suppose that for some citizens i and j, $I_{t+1,i} > I_{t+1,j}$. By strict concavity of the objective function in equation 1 there is a unique solution to the maximization problem. Therefore it must be the case that

$$\gamma\left(I_{t,i}\right) v\left(I_{t+1,i}\right) + \gamma\left(I_{t,j}\right) v\left(I_{t+1,j}\right) > \gamma\left(I_{t,i}\right) v\left(I_{t+1,j}\right) + \gamma\left(I_{t,j}\right) v\left(I_{t+1,i}\right)$$

because otherwise $\mathbf{I_{t+1}}$is not the solution to the maximization problem in equation 1. Equivalently, it must be that

$$\left(\gamma\left(I_{t,i}\right) - \gamma\left(I_{t,j}\right)\right)\left(v\left(I_{t+1,i}\right) - v\left(I_{t+1,j}\right)\right) > 0$$

which implies $I_{t,i} > I_{t,j}$.[1] A similar argument demonstrates that $I_{t,i} > I_{t,j} \implies I_{t+1,i} \geq I_{t+1,j}$. Finally, by continuity and strict monotonicity of γ and v, when $I_{t,i} > I_{t,j}$ the solution to the maximization problem in 1 will not satisfy $I_{t+1,i} = I_{t+1,j}$.[2] □

First-Order Condition Approach

Assumption 1. *The indirect utility function v is continuously differentiable and* $\lim_{I \to 0} v'(I) = \infty$.

Under assumption 1, the maximization problem in equation (1) is strictly concave with a unique interior solution and, therefore, the political process Γ can equivalently be characterized by the first-order conditions of this maximization problem. That is to say, at any time t with income distribution $\mathbf{I_t}$, the time $t+1$ income distribution $\mathbf{I_{t+1}}$ is characterized by the first-order conditions

$$\gamma\left(I_{t,i}\right) v'\left(I_{t+1,i}\right) = \gamma\left(I_{t,j}\right) v'\left(I_{t+1,j}\right) \qquad \text{for all } i, j \in \{1, 2, \ldots, n\},\ i \neq j \tag{2}$$

and the feasibility constraint

$$\frac{1}{n}\sum_{i=1}^{n} I_{t+1,i} = \bar{I}. \tag{3}$$

[1]Since γ is strictly increasing.

[2]Because if $I_{t+1,i} = I_{t+1,j}$, there exists some sufficiently small transfer from j to i which would increase the value of the objective function.

1.1 Fixed Points and Stability

An income distribution $\mathbf{I} \in \mathcal{I}$ is a *fixed point* if $\mathbf{I} = \Gamma(\mathbf{I})$. In this section, I will discuss main two types of income distributions which are fixed points of the recurrence relation (1).

Egalitarian Fixed Point The most obvious fixed point income distribution is the *egalitarian distribution* satisfying $I_i = \bar{I}$ for all $i \in \{1, 2, \dots, n\}$. I use the terms *egalitarian distribution* and *egalitarian fixed point* interchangeably because the egalitarian distribution is always a fixed point of the political process (equation 1). To see why, note that if the income distribution at time t is egalitarian then all citizens types have the same social welfare weight at time t: $\gamma\left(\bar{I}\right)$. Since agents have homogeneous indirect utility functions which are strictly increasing and strictly concave, the unique solution to the weighted utilitarian maximization problem will therefore be the egalitarian distribution.

Plutocratic Fixed Points I shall also consider a class of fixed point income distributions where some subset of citizens are relatively "rich" and the remaining agents are relatively "poor".

Consider a non-empty, strict subset of citizens $S \subset \{1, 2, \dots, n\}$ with $s \equiv |S|$. Suppose that for some $T > 0$, the income distribution $\mathbf{I}$ satisfies:

- $I_j = \bar{I} + T$ for all $j \in S$; and,
- $I_i = \bar{I} - \frac{s}{n-s}T$ for all $i \in \{1, 2, \dots, n\} \setminus S$.

I call such an income distribution a *plutocratic distribution* with s *affluent citizens* and *plutocratic transfer* T, highlighting the fact that the s citizens in set S have the same level of above average income while all other citizens have the same level of below average income.[3] If a particular plutocratic distribution with s affluent citizens is a fixed point of the political process (equation 1) I refer to it as a *plutocratic fixed point* with s affluent citizens.[4]

Under assumption 1 the first-order condition approach can be applied to characterize a plutocratic fixed point. In particular, if the plutocratic distribution with s affluent citizens and transfer T is a fixed point, then (2) implies that

$$\gamma\left(\bar{I} + T\right) v'\left(\bar{I} + \mathrm{T}\right) = \gamma\left(\bar{I} - \frac{s}{n-s}T\right) v'\left(\bar{I} - \frac{s}{n-s}T\right)$$

[3]Note that a plutocratic distribution as defined here does not necessarily match common descriptions of plutocratic societies income inequality in such a distribution can be arbitrary small (but must be non-zero). As well, although a plutocratic coalition S could be very exclusive, to the point of containing only one citizen, it can also be quite large, containing as many as $n - 1$ citizens.

[4]Note, the assumption of homogeneous citizens implies that properties of the income distribution are invariant to the identity of citizens assigned.

It is important to note that, unlike the egalitarian fixed point, a plutocratic fixed point with s affluent citizens may not always exist: that is to say, there will not always be a distribution fitting the description above which is a fixed point of the political process.

In some cases, for simplicity of exposition, I will refer to the egalitarian distribution as being a special case of the plutocratic distribution with a transfer of $T = 0$.

General Fixed Points

It is frequently helpful to think about the evolution of the income distribution in terms of the evolution of ratios of marginal utilities between different citizens. To that end, I define the *evolution function*

$$f(I, I') \equiv \log\left(\frac{\gamma(I')}{\gamma(I)}\right) - \log\left(\frac{v'(I)}{v'(I')}\right). \tag{4}$$

Notice that the first-order conditions (2) imply that for any agents i, j and any time period t we have

$$f(I_{t,i}, I_{t,j}) = \log\left(\frac{v'(I_{t+1,i})}{v'(I_{t+1,j})}\right) - \log\left(\frac{v'(I_{t,i})}{v'(I_{t,j})}\right). \tag{5}$$

Thus, the output of the evolution function is (approximately) equal to the percentage change in the ratio of marginal utilities between any two citizens with some initial incomes $(I_{t,i}, I_{t,j})$ caused by one iteration of the political process.

The evolution function can be helpfully used to provide a convenient description of fixed point income distributions.

Lemma 2. *Under assumption (1),*

$$\mathbf{I}^{\star} \in \mathcal{I} \textit{ is a fixed point} \iff f\left(I_i^{\star}, I_j^{\star}\right) = 0 \; \forall i, j.$$

Proof. That $f\left(I_i^{\star}, I_j^{\star}\right) = 0 \; \forall i, j$ is a necessary condition for any fixed point is trivial given (5). To see that it is also a sufficient condition, note that the condition is equivalent to the set of first-order conditions (2).[5] $\square$

[5] The inclusion of $\lim_{I \to 0} v'(I) = \infty$ in assumption 1 is required for sufficiency, because otherwise there might be fixed points which were not interior solutions of the maximization problem (1).

Stability of Fixed Points

A key property of interest is how the income distribution evolves when it is "close to" a given fixed point. To be more precise, for any $\epsilon > 0$, define

$$\mathcal{N}_\epsilon (\mathbf{I}) := \left\{\mathbf{I'} \in \mathbb{R}^n_+ : \|\mathbf{I'} - \mathbf{I}\| < \epsilon\right\}$$

as the ϵ-neighborhood of some income distribution $\mathbf{I}$, where $\|\cdot\|$ is the Euclidean norm.

If the income distribution should be expected to converge towards a given fixed point whenever it is sufficiently close, then the fixed point is said to be "locally stable".

Definition 1 (Local Stability)**.** A fixed point income distribution $\mathbf{I}^\star$ is *locally stable* (or *stable*) if there exists some $\epsilon > 0$ such that for any initial income distribution $\mathbf{I_0} \in \mathcal{N}_\epsilon (\mathbf{I}^\star)$, the income distribution converges to $\mathbf{I}^\star$: $\lim_{t\to\infty} \mathbf{I_t} = \mathbf{I}^\star$.

Alternatively, if the income distribution should be expected to diverge away from a given fixed point whenever it is sufficiently close, the fixed point is said to be "locally unstable".

Definition 2 (Local Instability)**.** A fixed point income distribution $\mathbf{I}^\star$ is *locally unstable* (or *unstable*) if there exists some $\epsilon > 0$ such that for any initial income distribution $\mathbf{I_0} \in \mathcal{N}_\epsilon (\mathbf{I}^\star)$, the income distribution will eventually permanently exit the ϵ-neighborhood: that is, if there exists a $\tau > 0$ such that for all $t > \tau$, $\mathbf{I_t} \notin \mathcal{N}_\epsilon (\mathbf{I}^\star)$.

1.2 Evolution of Plutocratic Distributions

In general, the evolution of income distributions via the political process may be quite complex. However, when the initial income distribution is plutocratic, this phenomenon becomes more tractable because in this case, the otherwise multidimensional dynamical system is reducible to one with a single state variable. To see this, note that lemma 1 implies that if the time t income distribution is plutocratic then so is the time $t+1$ income distribution.[6] Thus, whenever the initial income distribution is plutocratic the resulting sequence of income distributions is fully described by a sequence of plutocratic transfers $\{T_t\}_{t=0}^{\infty}$.

The dynamic behavior of this sequence of transfers can be analyzed using a simplified version of the evolution

[6] With the same set of affluent citizens at t and $t+1$.

function which is a function of the plutocratic transfer alone:

$$\hat{f}(T;n,s) \equiv f\left(\bar{I} - \frac{s}{n-s}T, \bar{I} + T\right)$$
$$= \log\left(\frac{\gamma\left(\bar{I}+T_t\right)}{\gamma\left(\bar{I}-\frac{s}{n-s}T_t\right)}\right) - \log\left(\frac{v'\left(\bar{I}-\frac{s}{n-s}T_t\right)}{v'\left(\bar{I}+T_t\right)}\right)$$

When evaluated at T_t this function outputs the (approximate) percentage change in the ratio of marginal utilities between an affluent citizen and a non-affluent citizen caused by own iteration of the political process at time t:

$$\hat{f}(T_t;n,s) = \log\left(\frac{v'\left(\bar{I}-\frac{s}{n-s}T_{t+1}\right)}{v'\left(\bar{I}+T_{t+1}\right)}\right) - \log\left(\frac{v'\left(\bar{I}-\frac{s}{n-s}T_t\right)}{v'\left(\bar{I}+T_t\right)}\right). \tag{6}$$

Note, as per Lemma 2, the income distribution induced by the transfer T is a fixed point if and only if $\hat{f}(T;n,s) = 0$.

Lemma 3 (Monotonic Convergence Lemma)**.** *Let assumption 1 hold.*

(i) *Suppose that at some time $t \geq 0$ the income distribution is plutocratic with s affluent citizens and plutocratic transfer $T_t \in \left[0, \frac{n-s}{s}\bar{I}\right)$. Then the time $t+1$ income distribution is plutocratic with s affluent citizens and a plutocratic transfer $T_{t+1} \in \left[0, \frac{n-s}{s}\bar{I}\right)$. Furthermore,*

$$\hat{f}(T_t;n,s) > 0 \iff T_{t+1} > T_t \iff \hat{f}(T_{t+1};n,s) > 0.$$

(ii) *Suppose that the initial income distribution is plutocratic with s affluent citizens and plutocratic transfer $T_0 \in \left(0, \frac{n-s}{s}\bar{I}\right)$. The resulting sequence of income distributions are induced by a strictly monotonic sequence of transfers $\{T_t\}_{t=0}^{\infty}$ which converges to some limiting transfer $T^\star \in \left[0, \frac{n-s}{s}\bar{I}\right)$. The transfer $T^\star$ induces a fixed point income distribution.*

Proof. To prove part (i) first note that by lemma 1 there exists some $T_{t+1} > 0$ such that $I_{t+1,i} = \bar{I} + T_{t+1}$

for all $i \in S$ and $I_{t+1,j} = \bar{I} - \frac{s}{n-s} T_{t+1}$ for all $j \notin S$. Therefore,

$$\begin{aligned}
\hat{f}\left(T_t; n, s\right) > 0 &\iff \log\left(\frac{v'\left(\bar{I} - \frac{s}{n-s} T_{t+1}\right)}{v'\left(\bar{I} + T_{t+1}\right)}\right) - \log\left(\frac{v'\left(\bar{I} - \frac{s}{n-s} T_t\right)}{v'\left(\bar{I} + T_t\right)}\right) > 0 \\
&\iff T_{t+1} > T_t \\
&\iff \log\left(\frac{\gamma\left(\bar{I} + T_{t+1}\right)}{\gamma\left(\bar{I} - \frac{s}{n-s} T_{t+1}\right)}\right) - \log\left(\frac{\gamma\left(\bar{I} + T_t\right)}{\gamma\left(\bar{I} - \frac{s}{n-s} T_t\right)}\right) > 0 \\
&\iff \hat{f}\left(T_{t+1}; n, s\right) > 0.
\end{aligned}$$

The first and final lines above follow from the first-order conditions (2); the second line from the fact that $v'(\cdot)$ is strictly decreasing; the third from the fact that $\gamma(\cdot)$ is strictly increasing.

To prove part (ii), first note that the fact that the sequence $\{T_t\}_{t=0}^{\infty}$ is strictly monotonic follows immediately from part (i). And because the space of possible transfers $\left[0, \frac{n-s}{s}\bar{I}\right)$ is bounded, the sequence must converge to some point $T^\star \in \left[0, \frac{n-s}{s}\bar{I}\right]$. To see that the sequence cannot converge to $\frac{n-s}{s}\bar{I}$, first note that such convergence would imply that any local neighborhood of $\frac{n-s}{s}\bar{I}$ contains infinitely many points of the sequence. Note as well, under assumption 1 we have $\lim_{I \to 0} v'(I) = \infty$ which implies that

$$\lim_{T \to \frac{n-s}{s}\bar{I}} \hat{f}\left(T; n, s\right) = -\infty$$

because $\gamma\left(\frac{n}{s}\bar{I}\right)$, $\gamma(0)$, and $v'\left(\frac{n}{s}\bar{I}\right)$ are finite but $\lim_{T \to \frac{n-s}{s}\bar{I}} v'\left(\bar{I} - \frac{n-s}{s}\bar{I}\right) \to \infty$. Thus, for all T sufficiently close to $\bar{I}$ we have $\hat{f}\left(T; n, s\right) < 0$. But this implies that there are infinitely many points of the sequence $\{T_t\}_{t=0}^{\infty}$ where $\hat{f}\left(T_t; n, s\right) < 0$ so by the monotonic convergence lemma the sequence is strictly monotonically decreasing in time, and thus cannot converge to $\frac{n-s}{s}\bar{I}$. Thus, $\{T_t\}_{t=0}^{\infty}$ converges to some $T^\star \in \left[0, \frac{n-s}{s}\bar{I}\right)$. Finally, to see that $T^\star$ must be a fixed point recall that, as per equation (6), $\hat{f}\left(T_t; n, s\right)$ is equal to the approximate percentage change in the ratio of marginal utilities between affluent and non-affluent citizens as a result of the period t political process. Thus, if

$$\lim_{t \to \infty} \left|T_{t+1} - T_t\right| = 0 \iff \lim_{t \to \infty} \hat{f}\left(T_t; n, s\right) = 0.$$

By continuity of $\hat{f}\left(\cdot; n, s\right)$, $\lim_{t \to \infty} \hat{f}\left(T_t; n, s\right) = \hat{f}\left(T^\star; n, s\right)$, so it must be the case that $\hat{f}\left(T^\star; n, s\right) = 0$, and therefore $T^\star$ is a fixed point. $\square$

The monotonic convergence lemma provides the foundation for a simple characterization of the behavior of the sequence of transfers in the vicinity of a fixed point.

Theorem 1 (Necessary and Sufficient Conditions for Convergence with Plutocratic Distributions)**.** *Suppose assumption 1 holds, and that the initial income distribution is plutocratic with* s *affluent citizens and some plutocratic transfer* $T_0 \in \left[0, \frac{n-s}{s}\bar{I}\right)$*. As well, assume that the plutocratic income distribution with* s *affluent citizens and some transfer* $T^\star \in \left[0, \frac{n-s}{s}\bar{I}\right)$ *is a fixed point. The income distribution converges to this fixed point if and only if one of the following conditions holds:*

(i) $T_0 \in (0, T^\star)$ *and* $\hat{f}(T; n, s) > 0$ *for all* $T \in [T_0, T^\star)$*;*

(ii) $T_0 \in \left(T^\star, \frac{n-s}{s}\bar{I}\right)$ *and* $\hat{f}(T; n, s) < 0$ *for all* $T \in (T^\star, T_0]$*; or,*

(ii) $T_0 = T^\star$*.*

Proof. First, we show convergence to the fixed point must occur if (i) is satisfied. Since $\hat{f}(T_0; n, s) > 0$, the monotonic convergence lemma (Lemma 3) implies that the sequence $\{T_t\}_{t=0}^{\infty}$ is strictly monotonic and converges to some fixed point. To see that it must converge to $T^\star$ in particular, it suffices to show that $T_t < T^\star$ for all $t \geq 0$, since by assumption there are no fixed points in the interval $[T_0, T^\star)$.[7] This can be shown by induction. By assumption, $T_0 < T^\star$. Thus, if we can show that $T_t < T^\star$ implies $T_{t+1} < T^\star$, the proof will be complete. Applying the first-order condition (2), and the characterization of fixed points in lemma 2 we have

$$\begin{aligned} T_t < T^\star \iff & \frac{\gamma\left(\bar{I} + T_t\right)}{\gamma\left(\bar{I} - \frac{s}{n-s}T_t\right)} < \frac{\gamma\left(\bar{I} + T^\star\right)}{\gamma\left(\bar{I} - \frac{s}{n-s}T^\star\right)} \\ \iff & \frac{v'\left(\bar{I} - \frac{s}{n-s}T_{t+1}\right)}{v'\left(\bar{I} + T_{t+1}\right)} < \frac{v'\left(\bar{I} - \frac{s}{n-s}T^\star\right)}{v'\left(\bar{I} + T^\star\right)} \\ \iff & T_{t+1} < T^\star \end{aligned}$$

because $\gamma(\cdot)$ is strictly increasing and $v'(\cdot)$ is strictly decreasing. Thus, if $T_0 < T^\star$ then $T_t < T^\star$ for all $t \geq 0$, and consequently the sequence of transfers $\{T_t\}_{t=0}^{\infty}$ converges to $T^\star$.

A similar argument shows that monotonic convergence to the fixed point occurs if condition (ii) is satisfied. That convergence occurs when condition (iii) is satisfied is trivial.

It remains to show that a necessary condition for convergence of the sequence $\{T_t\}_{t=0}^{\infty}$ to $T^\star$ is that one of (i), (ii), or (iii) must hold. I proceed by contradiction. Suppose that the sequence converges but none of these assumptions holds. This implies that $T_0 \neq T^\star$ and that either:

[7] Because $\hat{f}(T; n, s) \neq 0$ over this interval.

- $T_0 \in (0, T^\star)$ and $\hat{f}(T; n, s) \leq 0$ for at least one $T \in [T_0, T^\star)$; or,
- $T_0 \in \left(T^\star, \frac{n-s}{s}\bar{I}\right)$ and $\hat{f}(T; n, s) \geq 0$ for at least one $T \in (T^\star, T_0]$.

In either case, this implies that either that $\hat{f}(T_0; n, s) = 0$ so that T_0 is itself a fixed point, or $\hat{f}(T_0; n, s) \neq 0$ and by the monotonic convergence lemma (Lemma 3) we have that the sequence $\{T_t\}_{t=0}^{\infty}$ is monotonically moving away from $T^\star$. $\square$

Note, Theorem 1 also applies to the special case where fixed point in question is the plutocratic distribution with a transfer of $T^\star = 0$: the egalitarian fixed point.[8]

1.3 The Two Citizen Case

To build intuition and aid in subsequent derivations, I begin with the simplest possible case: a society with two citizens ($n = 2$).

Throughout this section, without loss of generality, I assume that citizen 2 has weakly higher income than citizen 1: $I_{t,2} \geq I_{t,1}$ for all t.[9] Note, with two citizens the all income distributions can be described by a single parameter which describes how much additional income the higher income citizen relative to the average income. Specifically, the distribution of income at time t can be written as

$$I_{t,1} = \bar{I} - T_t, \quad \text{and} \quad I_{t,2} = \bar{I} + T_t \tag{7}$$

for some $T_t \in \left[0, \bar{I}\right]$.

For convenience, throughout this section I will suppress the dependence of the simplified evolution function from the preceding section on the parameters n and s: $\hat{f}(T_t) \equiv \hat{f}(T_t; 2, 1)$.

Note, in the two citizen case, all income distributions are either egalitarian or plutocratic. When $T_t = 0$, we have the egalitarian distribution. When $T_t > 0$, we have a plutocratic distribution with one affluent citizen and a plutocratic transfer of T_t. This means that we can apply the previously obtained characterization of the evolution of plutocratic distributions (Theorem 1) to fully characterize the evolution of the income distribution. The following corollary is an immediate application Theorem 1.

Corollary 1 (Necessary and Sufficient Conditions for (In)Stability (Two Citizens))**.** *Suppose assumption 1 holds and $n = 2$. Further suppose that the plutocratic income distribution with s affluent citizens and some*

[8] In particular, Theorem 1 says that the income distribution converges to the egalitarian fixed point if and only if either $T_0 = 0$, or $\hat{f}(T; n, s) < 0$ for all $T \in (0, T_0]$.

[9] The rank preservation lemma (1) ensures that if the inequality holds in one period, it holds for all periods.

transfer $T^\star \in \left[0, \frac{n-s}{s}\bar{I}\right)$ *is a fixed point. This fixed point is stable if and only if there exists some* $\epsilon > 0$ *such that*

$$\hat{f}(T) > 0 \textit{ for all } T \in (T^\star - \epsilon, T^\star) \cap \left[0, \frac{n-s}{s}\bar{I}\right)$$

and

$$\hat{f}(T) < 0 \textit{ for all } T \in (T^\star, T^\star + \epsilon).$$

Alternatively, the fixed point is unstable if and only if there exists some $\epsilon > 0$ *such that*

$$\hat{f}(T) < 0 \textit{ for all } T \in (T^\star - \epsilon, T^\star) \cap \left[0, \frac{n-s}{s}\bar{I}\right)$$

and

$$\hat{f}(T) > 0 \textit{ for all } T \in (T^\star, T^\star + \epsilon).$$

First, let us consider the circumstances under which the egalitarian distribution is locally stable. To obtain a more intuitive local stability result, we'll adopt the additional assumption of differentiability of the political influence and marginal utility functions.

Assumption 2. *The political influence function* γ *is differentiable. The indirect utility function* v *is twice differentiable.*

Under these conditions, the egalitarian distribution is stable if the responsiveness of citizen political influence to income is sufficiently small relative to the (local) concavity of agent utility functions at this distribution. To simplify presentation, let

$$\sigma(I) \equiv -\frac{Iv''(I)}{v'(I)}$$

be the *coefficient of relative risk aversion* for the indirect utility function of a citizen with income I. As well, let

$$\mathcal{E}^\gamma(I) \equiv \frac{I\gamma'(I)}{\gamma(I)}$$

be the *elasticity of political influence* for a citizen with income I.

Theorem 2 (Conditions for Stable (or Fragile) Egalitarianism)**.** *Suppose assumptions 1 and 2 hold, and* $n = 2$*. Then egalitarian distribution is stable if the coefficient of relative risk aversion at* $\bar{I}$ *is larger than the elasticity of the political influence function at* $\bar{I}$

$$\mathcal{E}^\gamma(\bar{I}) < \sigma(\bar{I}), \tag{8}$$

and is unstable if the coefficient of relative risk aversion at $\bar{I}$ is smaller than the elasticity of the political influence function at $\bar{I}$

$$\mathcal{E}^{\gamma}\left(\bar{I}\right) > \sigma\left(\bar{I}\right). \tag{9}$$

Proof. First note that

$$\hat{f}'(0) = \frac{2}{\bar{I}}\left(\mathcal{E}^{\gamma}\left(\bar{I}\right) - \sigma\left(\bar{I}\right)\right).$$

Thus,$\mathcal{E}^{\gamma}\left(\bar{I}\right) < \sigma\left(\bar{I}\right) \iff \hat{f}'(0) < 0$. And $\hat{f}'(0) < 0$ implies that there exists some $\epsilon > 0$ such that $\hat{f}(T) < 0$ for all $T \in (0, \epsilon)$, and stability follows from Corollary 1. Similarly, $\mathcal{E}^{\gamma}\left(\bar{I}\right) > \sigma\left(\bar{I}\right) \iff \hat{f}'(0) > 0$ and $\hat{f}'(0) > 0$ implies that there exists some $\epsilon > 0$ such that $\hat{f}(T) > 0$ for all $T \in (0, \epsilon)$, with instability following from Corollary 1. □

To understand this result, first note that the local concavity of the objective function in equation (1) is increasing in the concavity of the indirect utility functions. The degree of local concavity reduces the size of the income differences between agents that are generated by given differences in the welfare weights of each agent. Meanwhile, the size of the welfare weight differences induced by given income differences is locally proportion to the income elasticity of welfare weights. Thus Theorem 2 simply states that, on the one hand, if an arbitrarily small amount of income inequality induces sufficiently large welfare weight inequality, the egalitarian distribution is not locally stable. On the other hand, if an arbitrarily small amount of income inequality induces sufficiently small welfare weight inequality, the egalitarian distribution is locally stable. The local concavity of the objective function determines how large "sufficiently large" is and how small "sufficiently small" is.[10]

We can similarly characterize the stability of plutocratic fixed points in terms of the relative magnitudes of the elasticities of citizen political influence and the concavity of indirect utility functions.

Theorem 3 (Stability of a Plutocratic Fixed Point (Two Citizens))**.** *Suppose assumptions 1 and 2 hold, and $n = 2$. Further, suppose that the income distribution induced by the transfer $T^{\star}$ is a fixed point. This fixed point is stable if*

$$\frac{\mathcal{E}^{\gamma}\left(\bar{I}+T^{\star}\right)}{\bar{I}+T^{\star}} + \frac{\mathcal{E}^{\gamma}\left(\bar{I}-T^{\star}\right)}{\bar{I}-T^{\star}} < \frac{\sigma\left(\bar{I}+T^{\star}\right)}{\bar{I}+T^{\star}} + \frac{\sigma\left(\bar{I}-T^{\star}\right)}{\bar{I}-T^{\star}} \tag{10}$$

and is unstable if

$$\frac{\mathcal{E}^{\gamma}\left(\bar{I}+T^{\star}\right)}{\bar{I}+T^{\star}} + \frac{\mathcal{E}^{\gamma}\left(\bar{I}-T^{\star}\right)}{\bar{I}-T^{\star}} > \frac{\sigma\left(\bar{I}+T^{\star}\right)}{\bar{I}+T^{\star}} + \frac{\sigma\left(\bar{I}-T^{\star}\right)}{\bar{I}-T^{\star}}. \tag{11}$$

[10] In section 2, where I introduce an explicit model of the political process, I provide further discussion on the interpretation of the indirect utility function v and it's properties (such as its coefficient of relative risk aversion) in the context of this model.

Proof. As with theorem 2, the proof follows immediately from the application of Corollary 1, together with the fact that

$$\hat{f}'(T) = \frac{\mathcal{E}^{\gamma}\left(\bar{I}+T^{\star}\right)}{\bar{I}+T^{\star}} + \frac{\mathcal{E}^{\gamma}\left(\bar{I}-T^{\star}\right)}{\bar{I}-T^{\star}} - \frac{\sigma\left(\bar{I}+T^{\star}\right)}{\bar{I}+T^{\star}} + \frac{\sigma\left(\bar{I}-T^{\star}\right)}{\bar{I}-T^{\star}}$$

□

Note, in the case of a plutocratic fixed point what matter for stability is the comparison of the magnitudes of two particular linear combinations of political influence elasticities and coefficients of relative risk aversion across the two agents.

The next theorem states that whenever the egalitarian fixed point is unstable, at least one plutocratic fixed point must exist.

Theorem 4 (Existence of and Convergence to a Plutocratic Fixed Point (Two Citizens))**.** *Suppose assumption 1 holds, and $n = 2$. If the egalitarian fixed point is unstable, then at least one plutocratic fixed point must exist. Furthermore, for any initial income distribution which is plutocratic, we have*

$$\lim_{t\to\infty} \mathbf{I_t} \quad \textit{is a plutocratic fixed point.}$$

Proof. In the proof of Theorem 2 above, I showed that condition (9) is equivalent to $\hat{f}'(0) > 0$. This implies that for some sufficiently small initial transfer $T_0 > 0$ we have $\hat{f}(T_0) > 0$. Thus, for any such initial transfer Theorem 1 implies that the sequence of transfers $\{T_t\}_{t=0}^{\infty}$ must be (strictly) monotonically increasing with time and converges to a plutocratic fixed point. □

Note, if assumption 2 also holds then combining the above result with Theorem 2 we have that whenever $\mathcal{E}^{\gamma}\left(\bar{I}\right) > \sigma\left(\bar{I}\right)$ there must exist at least one plutocratic fixed point and that the income distribution converges to a plutocratic fixed point if it is not initially egalitarian.

Theorem 4 follows from two facts. First, that citizen welfare weights are strictly positive and bounded above, so no citizen's utility ever receives zero weight nor can the relative welfare weights of any two citizen ever converge to zero over time. Second, that the marginal utility of income in the neighborhood of zero income is sufficiently large that full expropriation of the income of the poor is not a politically feasible option.

Theorem 5 (Dual Stability and Existence of Additional Fixed Points (Two Citizens))**.** *Suppose that assumption 1 holds, and $n = 2$. Further suppose that both the egalitarian fixed point and some plutocratic fixed point with transfer $T^{\star}$ exist and are stable. Then there exists at least one additional plutocratic fixed point with transfer $T^{\star\star} \in (0, T^{\star})$ which is not stable.*

Proof. By Corollary 1, we have $\hat{f}(T) < 0$ for all sufficiently small $T > 0$ and for some $\epsilon > 0$ we also have $\hat{f}(T) > 0$ for all $T \in (T^\star - \epsilon, T^\star)$. Thus, by continuity of $\hat{f}(\cdot)$, there exists some $T^{\star\star} \in (0, T^\star)$ and $\delta > 0$ such that $\hat{f}(T^{\star\star}) = 0$ and $\hat{f}(T) > 0$ for all $T \in (T^{\star\star}, T^{\star\star} + \delta)$. By Corollary 1, this fixed point is not stable. □

1.4 The Three Citizen Case

Theoretical results are a work-in-progress...

Theorem 6 (No Additional Fixed Points (Three Citizens))**.** *Suppose* $\mathbf{I} \in \mathcal{I}$ *is a fixed point. Then* $\mathbf{I}$ *is either the egalitarian distribution or a plutocratic distribution.*

Proof. Work-in-progress. □

Simulation

It is possible to build more intuition about the dynamics of the political process in the three citizen case ($n = 3$) by conducting some simulations. The three citizen case can be easily depicted graphically, as I do below in figure 1. Each sub-figure contains a triangular plot, which displays the simulated dynamics of the political redistribution process over time for a variety of initial income distributions. Each point contained in these triangles represents a specific income distribution. The x-coordinate reflects the share of total income going to citizen 1, the y-coordinate reflects the share of total income going to citizen 2, and the remaining share of income goes to citizen 3. In this way, all possible income distributions can be represented as points in the triangle. The vertices of the triangle correspond to the extreme cases where one citizen has 100% of total income.

I generated these plots numerically under the assumption that $v(\cdot)$ is CRRA with a coefficient of relative risk aversion of $\sigma = 0.1$. I further assume that the political influence function is of the form

$$\gamma(I) = 1 + bI^\eta$$

where $b = 0.0001$ and η varies across sub-figures. I assume that average income is $\bar{I} = 10000$. In section 2 I show that this political influence function specification can be microfounded via a political donation mechanism under the assumption that there is a convergent equilibrium of the election model and there are constant returns to campaign spending.

Each sub-figure divides the triangular region into differently colored regions. Within a given sub-figure, all points within a region of a given color correspond to initial income distributions that converge to the same fixed point income distribution. Each figure also depicts the simulated dynamics of the political system for 18 different initial income distributions using lines of connected black arrows. Each arrow in one of these lines corresponds to the impact that a single iteration of the political process had on the income distribution.

Sub-figure 1a depicts the scenario where $\eta = 0.85$. In this case, the egalitarian fixed point is the only stable fixed point, and for any initial income distribution (i.e. any point in the interior of the triangle), the income distribution converges to the egalitarian distribution over time.

Sub-figure 1b depicts the scenario where $\eta = 0.9$. In this case, the egalitarian fixed point is unstable. Instead, there are three stable plutocratic fixed points. For almost all initial income distributions, the system converges to one of these plutocratic fixed points, where a single citizen has almost 90% of total income. The exceptions include the case where the initial income distribution is egalitarian (of course) as well as some edge cases where the income distribution converges to a saddle point.[11]

Sub-figure 1b depicts the scenario where $\eta = 0.875$. In this case, the egalitarian fixed point is stable, and there are also three stable plutocratic fixed points where one citizen has about 60% of total income. Any initial income distribution in the purple region results in convergence to egalitarianism, whereas (almost) all initial distributions outside this region converge to one of these stable plutocratic fixed points.[12] This figure highlights the possibility that long-run inequality in this model of democratic redistribution can exhibit sensitivity to initial conditions: a tiny difference in the initial income distribution can make the difference between convergence to egalitarianism and convergence to plutocracy.

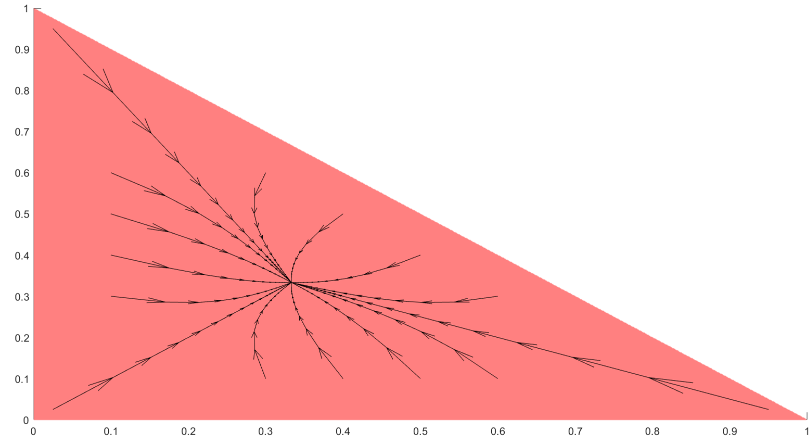
Together these figures also highlight the fact that small changes in parameter values can make a big difference to the behavior of the political system. If the model is microfounded via a campaign donation mechanism, then the parameter η can be interpreted as the (causal) income elasticity of political donations. This elasticity could reasonably be expected to change in response to campaign finance policy changes, changes in citizen tastes for political involvement, or changes in campaign technology. These simulations suggest that even relatively small such changes could have large consequences for the trajectory of economic inequality in a democracy.

[11] There are three saddle points which are (unstable) plutocratic fixed points where two citizens have the same level of above average income, and the remaining citizen has less than average.

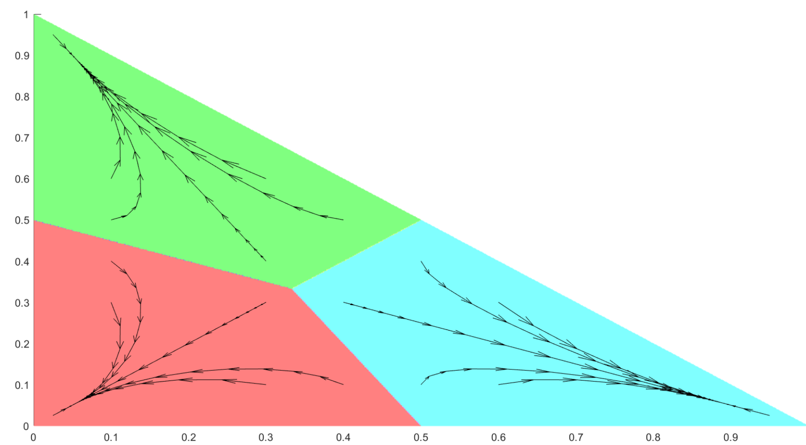
[12] As in the prior example, there are some edge cases where the income distribution converges to saddle points. These saddle points are plutocratic fixed points where one citizen gets a slightly smaller share of total income (around 55%). They are not stable.

Figure 1: Dynamics of Redistribution

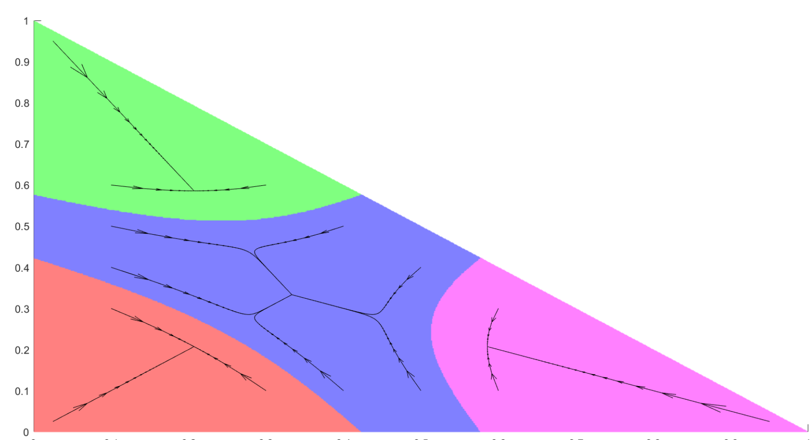
(a) Stable Egalitarianism ($\eta = 0.85$)



(b) Unstable Egalitarianism ($\eta = 0.9$)



(c) Dual Stability ($\eta = 0.875$)



2 Microfounding the Political Process

In the preceding section, I simply assume the political process can be described by equation (1). In this section, I justify this assumption using an extension of probabilistic voting theory which incorporates campaign finance. Note, this is not the only possible way to microfound equation (1). Bai & Lugunoff (2013) consider other models that imply political influence is increasing in relative income.

2.1 Setup

At each time period t, a two candidate election occurs. Each candidate $c \in \{A, B\}$ announces a policy platform consisting only of a proposed income distribution for the next period: $\mathbf{I^c_{t+1}}$. Each candidate c also engages in some amount of campaign spending D^c. For simplicity, in this section I focus on the case of pure redistribution.[13] That is to say, I assume each candidate c's policy is chosen from the set of all income distributions with the same total income as the initial distribution:

$$\mathbf{I^c_{t+1}} \in \left\{ \mathbf{I} \in \mathbb{R}^n_+ : \sum_{i=1}^{n} I_i = n\bar{I} \right\},$$

where $\bar{I} \equiv \frac{1}{n}\sum_{i=1}^{n} I_{0,i}$ is average income in the initial distribution.

Voting

Given a pair of policy announcements $\left(\mathbf{I^A_{t+1}}, \mathbf{I^B_{t+1}}\right)$ and campaign expenditures $\left(D^A_t, D^B_t\right)$, suppose that citizen i votes for candidate A if and only if

$$v\left(I^A_{t+1,i}\right) + \rho\left(D^A_t\right) + \epsilon^A_{t,i} > v\left(I^B_{t+1,i}\right) + \rho\left(D^B_t\right) + \epsilon^B_{t,i} \tag{12}$$

where $\rho(\cdot)$ captures the persuasive effect of campaign spending and $\left(\epsilon^A_{t,i}, \epsilon^B_{t,i}\right)$ are a pair of random utility shocks. Suppose that for each candidate c, time period t, and citizen i, $\epsilon^c_{t,i}$ is iid Type-II extreme value. This implies that citizen vote probabilities take the well-known binomial logit form:

$$P^c_{t,i}\left(\mathbf{I^A_{t+1}}, \mathbf{I^B_{t+1}}; D^A_t, D^B_t\right) \equiv \Pr\{i \text{ votes for } c \text{ at time } t\} = \frac{\exp\left(v\left(I^c_{t+1,i}\right) + \rho\left(D^c_t\right)\right)}{\exp\left(v\left(I^A_{t+1,i}\right) + \rho\left(D^A_t\right)\right) + \exp\left(v\left(I^B_{t+1,i}\right) + \rho\left(D^B_t\right)\right)}.$$

[13] That is, I assume there is no excess burden from redistribution.

Donations

In addition to voting, citizens may donate to the candidates. For simplicity, suppose in time period t, and citizen i will donate some amount $d\left(I_{t,i}\right)$ to candidate A if and only if condition (12) holds. I assume that $d\left(\cdot\right)$ is a strictly increasing function of after-tax income so that donations are a normal good. This implies that the expected donation citizen i makes to candidate c at time t is $P_{t,i}^{c}d\left(I_{t,i}\right)$. This donation behavior requires assuming that citizens have preferences over total donations and other consumption goods that are independent of their preferences over candidates. That is, a citizen's preference over candidates determines which candidate gets their money, but not the amount of money. Note, this is a very strong assumption, adopted here only for convenience of exposition. Later in the paper [work-in-progress] I discuss a version of the model where a citizens' donation size depends on the intensity of their preference for their preferred candidate.

Candidate's Problem

Each candidate c chooses their policy position $\mathbf{I^{c}_{t+1}}$ in order to maximize their expected vote share

$$\frac{1}{n}\sum_{i=1}^{n}P_{t,i}^{c}\left(\mathbf{I^{A}_{t+1}},\mathbf{I^{B}_{t+1}};D_{t}^{A}\left(\mathbf{I^{A}_{t+1}},\mathbf{I^{B}_{t+1}}\right),D_{t}^{B}\left(\mathbf{I^{A}_{t+1}},\mathbf{I^{B}_{t+1}}\right)\right) \tag{13}$$

where the pair of campaign expenditure functions $\left(D_{t}^{A}\left(\mathbf{I^{A}_{t+1}},\mathbf{I^{B}_{t+1}}\right),D_{t}^{B}\left(\mathbf{I^{A}_{t+1}},\mathbf{I^{B}_{t+1}}\right)\right)$ are implicitly defined as solutions to the system of equations

$$D_{t}^{c}\left(\mathbf{I^{A}_{t+1}},\mathbf{I^{B}_{t+1}}\right)=\sum_{i=1}^{n}P_{t,i}^{c}\left(\mathbf{I^{A}_{t+1}},\mathbf{I^{B}_{t+1}};D_{t}^{A},D_{t}^{B}\right)d\left(I_{t,i}\right)\qquad\forall c\in\left\{A,B\right\}. \tag{14}$$

This definition ensures that campaign expenditures by each candidate are fully financed by the donations they receive (in expectation).[14]

2.2 Electoral Equilibrium

An electoral equilibrium at time t is a pair of policy positions such that neither candidate can increase their expected vote share (13) by deviating, given the other candidate's policy. Such an equilibrium can be

[14] The requirement that a candidate's budget constraint need only be satisfied in expectation can be thought of as a close approximation to a state-contingent reality when n is large, or by assuming the candidates have access to some kind of actuarily fair insurance.

characterized via the candidates' first-order conditions:

$$\frac{\partial P_{t,i}^{c}}{\partial I_{t+1,i}^{c}}+\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\mathrm{d}P_{t,j}^{c}}{\mathrm{d}D_{t}^{c}}\right)d\left(I_{t,i}\right)\frac{\partial P_{t,i}^{c}}{\partial I_{t+1,i}^{c}}+\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\mathrm{d}P_{t,j}^{c}}{\mathrm{d}D_{t}^{-c}}\right)\left(d\left(I_{t,i}\right)\frac{\partial P_{t,i}^{-c}}{\partial I_{t+1,i}^{c}}\right)=\lambda^{c}\qquad\forall c\in\{A,B\},\forall i$$

where λ^c is the multiplier on government budget constraint requiring that total income is constant. Given my assumptions regarding citizen voting and donations preferences, as well as the distribution of citizens' random utility shocks, these conditions can be simplified to:[15]

$$\left[1+\left(\left(\rho'\left(D^{A}\right)+\rho'\left(D^{B}\right)\right)\frac{1}{n}\sum_{j=1}^{n}P_{t,j}^{A}\left(1-P_{t,j}^{A}\right)\right)d\left(I_{t,i}\right)\right]P_{t,i}^{A}\left(1-P_{t,i}^{A}\right)v'\left(I_{t+1,i}^{c}\right)=\lambda^{c}\qquad\forall c\in\{A,B\},\forall i. \tag{15}$$

Convergent Equilibrium

Note, by symmetry, whenever any electoral equilibrium exists, a convergent electoral equilibrium exists. In this case, the first-order conditions (15) can be even further simplified. For any convergent electoral equilibrium $(\mathbf{I_{t+1}},\mathbf{I_{t+1}})$, $\mathbf{I_{t+1}}$ must satisfy

$$\left[1+\left(\frac{\rho'\left(\frac{1}{2}\bar{D}\left(\mathbf{I_t}\right)\right)}{2}\right)d\left(I_{t,i}\right)\right]v'\left(I_{t+1,i}\right)=\lambda\qquad\forall i \tag{16}$$

where $\bar{D}\left(\mathbf{I_t}\right)\equiv\sum_{i=1}^{n}d\left(I_{t,i}\right)$ is total donations from all citizens to the two candidates. The representation above follows from the fact that in a convergent equilibrium both candidates receive an equal amount of donations and votes (in expectation).

For the remainder of this section, I assume that every election features a unique electoral equilibrium which is convergent.

Political Influence

Equation (16) provides a way to microfound the political process (equation 1). Note, we can also think of equation (16) as being the first-order condition of the following weighted utilitarian maximization problem:

$$\max_{\tilde{\mathbf{I}}\in\mathbb{R}_{+}^{n}}\left\{\sum_{i=1}^{n}\gamma\left(I_{t,i};\mathbf{I_t}\right)v\left(\tilde{I}_{i}\right):\frac{1}{n}\sum_{i=1}^{n}\tilde{I}_{i}=\bar{I}\right\}$$

[15]In particular, the result below follows primarily from two facts: first, that $P_{t,i}^{A}+P_{t,i}^{B}=1$; and, second, if $f\left(x,y\right)=\frac{\exp x}{\exp x+\exp y}$, then

$$\frac{\partial f}{\partial x}=f\left(1-f\right)=-\frac{\partial f}{\partial y}.$$

where

$$\gamma\left(I_{t,i};\mathbf{I_t}\right)\equiv 1+\left(\frac{\rho'\left(\frac{1}{2}\bar{D}\left(\mathbf{I_t}\right)\right)}{2}\right)d\left(I_{t,i}\right) \tag{17}$$

is the political influence of citizen i at time t.

Recall, in the political process assumed in section 1 the political influence function depended only on one individual's income. Equation (17) does not satisfy this definition, since the political influence function also depends on aggregate donations, $\bar{D}\left(\mathbf{I_t}\right)$, which is itself influenced by the distribution of income.

To justify the political influence function of section 1 thus requires a further restriction. Two possible options are:

1. Assume that citizen donation preferences are homothetic. That is to say, they spend a constant fraction of their income on donations:

$$d\left(I_{t,i}\right)\equiv \alpha I_{t,i} \qquad \text{for some } \alpha\in(0,1)\,.$$

   This formulation ensures that total donations are constant across all income distributions,

$$\bar{D}\left(\mathbf{I_t}\right)=\alpha\bar{I},$$

   thus eliminating any need to address the question of how changes in the income distribution affect the marginal value of campaign spending. Note, however, that this also implies that the political influence function is linear

$$\gamma\left(I_{t,i};\mathbf{I_t}\right)=1+\rho'\left(\frac{\alpha}{2}\bar{I}\right)\frac{\alpha}{2}I_{t,i}.$$

   Thus, this assumption substantially restricts the form of the political influence function.

2. Assume that the marginal value of campaign spending $\rho'\left(\frac{1}{2}\bar{D}\right)$ is constant in total doantions $\bar{D}$. This approach provides a rationalization for the type of functional form employed in my simulations presented in section 1.4, as it allows for a nonlinear donation function $d\left(I_{t,i}\right)$.

3 Redistribution via Nonlinear Income Taxation

In this section, I extend the model discussed above to the case where the only available technology for redistribution is a nonlinear income tax. Section 3.1 shows that the evolution of the nonlinear income tax in a political system like that described in section 2 can be described using the standard ABC-style formulas of optimal tax theory. However, it is very difficult to derive any analytical results about the full tax schedule, even using such characterizations. In section 3.2, I side step these analytical difficulties by focusing on the key research question motivating this paper: under what conditions will democratic redistribution lead to plutocracy? To answer this question, I focus on the predictions my model makes about top tax rates. Generalizing results from section 1, I show taxation of the richest members of society is only feasible if the coefficient of relative risk aversion among the rich exceeds the elasticity of political influence with respect to after-tax income. I also demonstrate that, under a range of assumptions, the elasticity of political influence is proportionate to the elasticity of political donations with respect to after-tax income. This provides a possible justification for donations caps: when there is uncertainty about the effects of other policies, only donation caps can ensure that taxation of the rich is politically feasible.

3.1 Setup

Citizen Economic Behavior

Suppose there is a continuum of citizens who differ in their wage rate w. Let $F(w)$ be the distribution of wages. A citizen with wage rate w chooses taxable income z, political donations d, and other consumption c to solve

$$\max_{c,d,z}\left\{\phi(c,d)-h\left(\frac{z}{w}\right):c+d=z-T(z)\right\}$$

where $T(z)$ is a nonlinear income tax.

The assumption of additive separability of the utility function implies that a wage w citizen's choice of taxable income can alternatively be thought of as solving the problem

$$\max_{z}\left\{u(I)-h\left(\frac{z}{w}\right):I=z-T(z)\right\}$$

where I is after-tax income and

$$u(I)\equiv\max_{c,d}\{\phi(c,d):c+d=I\}$$

is the value of the agent's consumption-donation problem when they have after-tax income of I. This also makes it clear that all citizens with the same after-tax income I make the same choice of political donations $d(I)$. I shall make extensive use of this alternative perspective below.

Let $z(w;T(\cdot))$ be the choice of taxable income a wage w agent makes at the tax schedule $T(\cdot)$, let $I(w;T(\cdot))$

be the corresponding after-tax income, and let

$$V\left(w;T\left(\cdot\right)\right)\equiv u\left(I\left(w;T\left(\cdot\right)\right)\right)-h\left(\frac{z\left(w;T\left(\cdot\right)\right)}{w}\right)$$

be the indirect utility function of an citizen with wage w for a given tax schedule $T\left(\cdot\right)$.

Citizen Political Behavior

Following the same approach as section 2, I assume that a wage w citizen with random utility shocks $\left(\epsilon_t^A,\epsilon_t^B\right)$ votes for candidate A in the election at time t if and only if

$$V\left(w;T_{t+1}^A\left(\cdot\right)\right)+\rho\left(D_t^A\right)+\epsilon_t^A>V\left(w;T_{t+1}^B\left(\cdot\right)\right)+\rho\left(D_t^B\right)+\epsilon_t^B$$

and that a wage w citizen who votes for candidate A also sends their political donations $d\left(I\left(w;T_t\left(\cdot\right)\right)\right)$ to candidate A. If we once again adopt the assumption that the random utility shocks are iid Extreme Value Type II, we get binomial logit expressions for vote probabilities as in section 2.

Politician Behavior

Our two political candidates now compete in the election at time t by proposing the next period's nonlinear income tax schedule. In particular, candidate c's policy in the time t election, $T_{t+1}^c\left(\cdot\right)$, may be any nonlinear function in the set $\left\{T\left(\cdot\right):\mathcal{R}\left(T\left(\cdot\right)\right)=0\right\}$, where

$$\mathcal{R}\left(T\left(\cdot\right)\right)\equiv\int T\left(z\left(w;T\left(\cdot\right)\right)\right)f\left(w\right)\mathrm{d}w$$

is the revenue from the tax schedule $T\left(\cdot\right)$. For tractability, I will also assume that any such schedule is twice continuously differentiable and ensures the the second-order condition $\frac{\mathrm{d}z\left(w;T\left(\cdot\right)\right)}{\mathrm{d}w}\geq 0$ for all w is satisfied.

If the politician's objective is to maximize their expected vote share (as in section 2), and we assume there is a convergent equilibrium, then we can once again show that the equilibrium policy at time $t+1$ solves the weighted utilitarian maximization problem

$$\max_{T\left(\cdot\right)}\left\{\int\gamma\left(I_t\left(w\right);\bar{D}_t\right)V\left(w;T\left(\cdot\right)\right)f\left(w\right)\mathrm{d}w:\mathcal{R}\left(T\left(\cdot\right)\right)=0\right\}\tag{18}$$

where $I_t\left(w\right)\equiv I\left(w;T_t\left(\cdot\right)\right)$ is the time after-tax income of a wage w citizen under the time t tax schedule,

$\bar{D}_t \equiv \int d\left(I\left(w;T\left(\cdot\right)\right)\right) f\left(w\right) \mathrm{d}w$ is aggregate donations under that tax schedule, and

$$\gamma\left(I;D\right) \equiv 1 + \left(\frac{\rho'\left(\frac{1}{2}D\right)}{2}\right) d\left(I\right) \tag{19}$$

is the political influence of citizen with after-tax income I when aggregate donations are D.

Equilibrium Tax Schedule

Because (18) is simply the familiar social planner's problem from the study of optimal nonlinear income taxation, we can characterize the equilibrium tax schedule using the standard ABC formula. Because I have assumed additively separable preferences, the marginal tax rates of the schedule at time $t+1$ will be satisfy

$$\frac{T'_{t+1}\left(z_{t+1}\left(w\right)\right)}{1-T'_{t+1}\left(z_{t+1}\left(w\right)\right)} = \frac{E_z^w}{E_z^{1-T'}} \left(\frac{1-F\left(w\right)}{wf\left(w\right)}\right) u'\left(c_{t+1}\left(w\right)\right) \int_w^\infty \left(1 - \frac{\gamma\left(I_t\left(m\right);\bar{D}_t\right) u'\left(I_{t+1}\left(m\right)\right)}{\lambda_{t+1}}\right) \left(\frac{1}{u'\left(I_{t+1}\left(m\right)\right)}\right) \frac{f\left(m\right)\mathrm{d}m}{1-F\left(w\right)} \tag{20}$$

where $E_z^w \equiv \frac{w}{z_{t+1}} \frac{\partial z_{t+1}}{\partial w}$ is the elasticity of taxable income with respect to the wage rate, $E_z^{1-T'} \equiv \frac{1-T'}{z_{t+1}} \frac{\partial z_{t+1}}{\partial(1-T')}$ is the elasticity of taxable income with respect to the net-of-tax rate, and

$$\lambda_{t+1} \equiv \int_0^\infty \gamma\left(I_t\left(m\right);\bar{D}_t\right) u'\left(I_{t+1}\left(m\right)\right).$$

3.2 When Is Taxation of the Rich Politically Feasible?

Equation (20) is useful, but the task of fully characterizing the evolution of the income tax schedule via the political remains quite difficult. But such a complete characterization may not be necessary to answer the fundamental question at the heart of this paper: when is democracy susceptible to a collapse into plutocracy? Put another way, under what conditions is taxation of the very richest members of society politically sustainable. To answer this question, we can examine the predictions of this model regarding top tax rates.

Suppose that as $w \to \infty$ citizen preferences become isolelastic,

$$u\left(I\left(w\right)\right) \to \frac{I\left(w\right)^{1-\sigma}}{1-\sigma} \qquad \text{and} \qquad h\left(\frac{z\left(w\right)}{w}\right) \to \frac{\left(\frac{z(w)}{w}\right)^{1+\frac{1}{e}}}{1+\frac{1}{e}},$$

and Pareto parameter converges to a constant

$$\frac{wf\left(w\right)}{1-F\left(w\right)} \to \alpha.$$

Then top income tax rates are characterized by

$$\frac{T'_{t+1}\left(z_{t+1}\left(w\right)\right)}{1-T'_{t+1}\left(z_{t+1}\left(w\right)\right)}=\frac{1+e}{e}\cdot\frac{1}{\alpha}\cdot I_{t+1}\left(w\right)^{-\sigma}\int_{w}^{\infty}\left(1-\frac{\gamma\left(I_{t}\left(m\right);\bar{D}_{t}\right)I_{t+1}\left(m\right)^{-\sigma}}{\lambda_{t}}\right)\left(\frac{1}{I_{t+1}\left(m\right)^{-\sigma}}\right)\frac{f\left(m\right)\mathrm{d}m}{1-F\left(w\right)}. \quad (21)$$

Taxation of the very rich is only occurs at time $t+1$ if the right-hand side of this expression is weakly increasing in the wage rate at the top of the wage distribution, because then (21) implies that

$$\frac{\mathrm{d}}{\mathrm{d}w}\left[\frac{T'_{t+1}\left(z_{t+1}\left(w\right)\right)}{1-T'_{t+1}\left(z_{t+1}\left(w\right)\right)}\right]=0. \quad (22)$$

If condition (22) is satisfied, then the marginal tax rate converges to some constant $T'_{t+1}\left(z_{t+1}\left(w\right)\right)\to\tau_{t+1}\leq 1$ at the top of the wage distribution. If condition (22) is violated on other hand, marginal tax rates diverge to an infinite marginal subsidy (i.e. $T'_{t+1}\left(z_{t+1}\left(w\right)\right)\to-\infty$), and positive taxation of the rich will not occur at time $t+1$.[16]

Note, one way to ensure condition (22) holds is to adopt an analogue of the standard optimal tax theory assumption that welfare weights converge to a constant at the top of the wage distribution. In particular, suppose that at time $t+1$ we have that as $w\to\infty$,

$$\frac{\gamma\left(I_{t}\left(w\right);\bar{D}_{t}\right)I_{t+1}\left(w\right)^{-\sigma}}{\lambda_{t}}\to\tilde{\gamma}_{t+1}. \quad (23)$$

Then, borrowing standard arguments from optimal tax theory with isoelastic preferences, equation (21) implies that the top marginal income tax rate at time $t+1$ will satisfy

$$\frac{\tau_{t+1}}{1-\tau_{t+1}}=\frac{1}{e}\frac{\left(1+e\right)\left(1-\tilde{\gamma}_{t+1}\right)}{\left(\alpha-\frac{(1+e)\sigma}{1-\sigma e}\right)} \quad (24)$$

or

$$\tau_{t+1}=\frac{\left(1+e\right)\left(1-\tilde{\gamma}_{t+1}\right)}{\left(1+e\right)\left(1-\tilde{\gamma}\right)+e\left(\alpha-\frac{(1+e)\sigma}{1-\sigma e}\right)}. \quad (25)$$

So when condition (23) holds, the political equilibrium at time $t+1$ will feature finite top tax rates characterized by (22).

To better understand condition (23), note that

$$\frac{\mathrm{d}}{\mathrm{d}w}\left[\gamma\left(I_{t}\left(w\right);\bar{D}_{t}\right)I_{t+1}\left(w\right)^{-\sigma}\right]<0\iff\frac{wI'_{t}\left(w\right)}{I_{t}\left(w\right)}\frac{I_{t}\left(w\right)\gamma'\left(I_{t}\left(w\right);\bar{D}_{t}\right)}{\gamma\left(I_{t}\left(w\right);\bar{D}_{t}\right)}-\frac{wI'_{t+1}\left(w\right)}{I_{t+1}\left(w\right)}\sigma<0.$$

[16] Note, the marginal tax rate cannot diverge to an infinite positive level because agents will not continue to work if marginal tax rates exceed one.

If $\frac{\mathrm{d}}{\mathrm{d}w}\left[\gamma\left(I_t\left(w\right);\bar{D}_t\right)c_{t+1}\left(w\right)^{-\sigma}\right]=0$ at the top, then $\tilde{\gamma}_{t+1}$ is some non-negative constant and the top tax rate is described by equation (24). If $\frac{\mathrm{d}}{\mathrm{d}w}\left[\gamma\left(I_t\left(w\right);\bar{D}_t\right)c_{t+1}\left(w\right)^{-\sigma}\right]<0$ at the top, then $\tilde{\gamma}=0$, and the top tax rate is still described by (24), and is the revenue-maximizing rate. If $\frac{\mathrm{d}}{\mathrm{d}w}\left[\gamma\left(I_t\left(w\right);\bar{D}_t\right)c_{t+1}\left(w\right)^{-\sigma}\right]>0$ at the top, then taxation of the rich is not politically feasible.

With isoelastic preferences and a constant marginal tax rate at the top it can be shown that $\lim_{w\to\infty}\frac{wI_t'(w)}{I_t(w)}=\lim_{w\to\infty}\frac{wI_{t+1}'(w)}{I_{t+1}(w)}$. Thus, we obtain the striking result that taxation of the rich is politically feasible only if the coefficient of relative risk aversion for the rich is smaller than the elasticity of the political influence function,

$$\sigma\geq\lim_{w\to\infty}\frac{I_t\left(w\right)\gamma'\left(I_t\left(w\right);\bar{D}_t\right)}{\gamma\left(I_t\left(w\right);\bar{D}_t\right)}. \tag{26}$$

This extends a key result from section 1 (see, for example, proposition **??**) to a much richer setting.

4 Conclusion

This paper introduces a simple model of democratic redistribution and shows that a political system which has some bias towards the preferences of citizens with higher incomes may, in the long-run, produces either egalitarian or plutocratic outcomes depending on two key factors: how responsive citizens' political behavior is to their receipt of additional income and, how sensitive their political influence is to their income. In the context of this model, small difference in initial conditions can lead to large difference in long-run economic inequality. Moreover, the level of economic inequality in a given time period may provide little information about the long-run level of inequality.

This paper remains a work in progress. In future drafts, I intend to expand on my existing results for both the pure redistribution model and the nonlinear taxation model. Additionally, I will explore how diminishing returns to campaign spending influence the model's long run behavior.

References

Robert A Dahl. *On political equality*. Yale University Press, 2006.

Jan Rosset. Economically based inequalities in political representation: Where do they come from? In *Understanding Inequality: Social Costs and Benefits*, pages 241–256. Springer, 2016.

Jennifer Shore. Political inequality: Origins, consequences, and ways ahead. In *Understanding inequality: Social costs and benefits*, pages 225–240. Springer, 2016.