Power Fluctuations and Political Economy

Citation: Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. "Power fluctuations and political economy." Journal of Economic Theory, Volume 146, Issue 3, May 2011, Pages 1009–1041.© 2010 Elsevier Inc.

As Published: http://dx.doi.org/10.1016/j.jet.2010.11.002

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/61784

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Power Fluctuations and Political Economy*

Daron Acemoglu
MIT and CIFAR

Michael Golosov
Yale and NES

Aleh Tsyvinski
Yale and NES

October 2010.

Abstract

We study (constrained) Pareto efficient allocations in a dynamic production economy where the group that holds political power decides the allocation of resources. For high discount factors, the economy converges to a first-best allocation where labor supply decisions are not distorted. For low discount factors, distortions do not disappear and fluctuate over time. Most importantly, the set of sustainable first-best allocations is larger when there is less persistence in the identity of the party in power (because this encourages political compromise). This result contradicts the common presumption that there will be fewer distortions when there is a "stable ruling group".

Keywords: commitment problem, dynamic political economy, Olson-McGuire hypothesis, political compromise, political economy, political power.

JEL Classification: P16, P48.

*The authors thank the associate editor, Ramon Marimon, and three anonymous referees for useful comments.We also thank participants at the Canadian Institute of Advanced Research, Econometric Society Meeting 2006, Erasmus, EUI, LAEF conference "Dynamic Political Economy and Optimal Taxation," Minneapolis Fed conference "Economics without Frontiers," MIT, New Economic School, University of Pennsylvania, University of Washington, Society of Economic Dynamics, Stanford, Stockholm, and Yale. Financial support from the AFOSR and the NSF is gratefully acknowledged. Golosov and Tsyvinski also thank Einaudi Institute of Economics and Finance for hospitality.

1 Introduction

In this paper, we investigate (constrained) Pareto efficient equilibria in an infinite-horizon production economy in which political power fluctuates between different social groups ("parties"). These groups may correspond to social classes with different incomes or to citizens living in different regions. The process for power fluctuation is taken as given. Our objective is to understand the implications of political economy frictions/constraints on the allocation of resources. The key to political economy friction in our model is *lack of commitment*: the group currently in power determines the allocation of resources (the allocation of total production across different groups in the society), and there are no means of making binding commitments to future allocations. This political economy friction leads to an additional sustainability constraint for the group in power, to ensure that it does not expropriate the available resources.

We characterize the (constrained) Pareto efficient allocations in this economy.[1] This focus enables us to understand the implications of political economy frictions on "the best possible" allocations, as clearly identifying the role of political economy in production and consumption distortions.[2] These allocations can be identified as the solution to an optimization problem subject to the participation and sustainability constraints, with different Pareto weights given to the utilities of different groups. We refer to allocations that involve full consumption smoothing and no distortions as "first best" (or as "sustainable first best" to emphasize that these are achieved despite the sustainability constraints). In these allocations, each individual supplies the same amount of labor and receives the same level of consumption at every date, irrespective of which group is in power. The sustainability constraints resulting from political economy imply that first-best allocations may not be supported because the group in power could prefer to deviate from a first-best allocation. In this case, Pareto efficient allocations will involve distortions (in the sense that marginal utility of consumption and disutility of labor are not equalized) and consumption and labor will fluctuate over time.[3]

[1] "Constrained" here means that all of these allocations are subject to the *sustainability constraints* resulting from the lack of commitment. To simplify the terminology, we simply refer to the "constrained Pareto efficient allocations" as "Pareto efficient allocations" unless additional emphasis is needed.

[2] An alternative, complementary strategy is to focus on Pareto dominated equilibria that may emerge either in our game or in some related institutional setting. Much of the political economy literature investigates the role of specific institutions and thus implicitly focuses on such allocations. Such Pareto dominated allocations will naturally induce further distortions relative to the allocations we characterize.

[3] Constrained Pareto efficient allocations have a quasi-Markovian structure and can be characterized recurs-

We present three sets of theoretical results. First, we characterize the structure of political economy frictions as a function of the preference and production structure, the identity of the group in power and the stochastic process regulating power switches. We show that as long as a first-best allocation is not sustainable at the current date, the labor supply (and production) of individuals who belong to groups that are not in power will be distorted downwards—i.e., "taxed". This downward distortion results from the sustainability constraints reflecting the political economy considerations. Intuitively, an increase in production raises the amount that the group in power can allocate to itself for consumption rather than allocating it among the entire population. Reducing aggregate production relaxes the political economy constraints and reduces the rents captured by the group in power. Since starting from an undistorted allocation, the gain to society from rents to the ruling group is first order, while the loss is second order, (non-first-best) constrained Pareto efficiency allocations involve distortions and underproduction (relative to the first best).

Second, we characterize the dynamics of the distortions caused by the political economy factors. When discount factors are low, no first-best allocation is sustainable. Consequently, distortions always remain, even asymptotically. In particular, we show that in this case all Pareto efficient allocations converge to an invariant *non-degenerate* distribution of consumption and labor supply across groups, whereby distortions as well as the levels of consumption and labor supply for each group fluctuate according to an invariant distribution. We then focus on the special case with two social groups (two parties) and quasi-linear preferences. In this environment we show that if there is any sustainable first-best allocations, then any Pareto efficient allocation path (meaning an efficient allocation starting with any Pareto weights) eventually reaches a first-best allocation, and both distortions and fluctuations in consumption and labor supply disappear.

Third and most importantly, we use our framework to discuss a central question in political economy—whether a more stable distribution of political power (as opposed to frequent power switches between groups) leads to "better public policies." That is, whether it leads to policies involving lower distortions and generating greater total output. A natural conjecture is that a stable distribution of political power should be preferable because it serves to increase the

ively conditional on the identity of the group that is in power and Pareto weights. Dynamics are determined by updating the Pareto weights recursively.

"effective discount factor" of the group in power, thus making "cooperation" easier. This conjecture receives support from a number of previous political economy analyses. For example, Olson [**?**] and McGuire and Olson [31] contrast an all-encompassing long-lived dictator to a "roving bandit" and conclude that the former will lead to better public policies than the latter. The standard principal-agent models of political economy, such as Barro [15], Ferejohn [21], Persson, Roland and Tabellini [34], [35], also reach the same conclusion, because it is easier to provide incentives to a politician who is more likely to remain in office.

Our analysis shows that this conjecture is generally not correct (in fact, its opposite is true). The conjecture is based on the presumption that incentives can be given to agents only when they remain in power. Once a politician or a social group leaves power, they can no longer be punished or rewarded for past actions. This naturally leads to the result that there is a direct link between the effective discount factor of a political agent and its likelihood of staying in power. This presumption is not necessarily warranted, however. Members of a social group or an individual can be rewarded not only when they are in power, but also *after they have left power*. Consequently, the main role of whether power persists or not is not to affect the effective discount factor of different parties in power, but to determine their deviation payoff. Greater persistence implies better deviation payoffs; in contrast, in the first best, there are no fluctuations in consumption and labor supply, thus along-the-equilibrium-path payoffs are independent of persistence. This reasoning leads to the opposite of the McGuire-Olson conjecture: more frequent power switches tend to reduce political economy distortions and expand the set of sustainable first-best allocations. Although our results stand in contrast to the McGuire-Olson conjecture, they are consistent with the line of argument going back to Aristotle that emphasizes the importance of power turnover in supporting democratic institutions (see, e.g., Przeworski [38]).[4]

Finally, we also illustrate the relationship between persistence of power and the structure of Pareto efficient allocations numerically. We verify the result that greater persistence reduces the set of sustainable first-best allocations. However, we also show that an increase in the frequency of power switches does not necessarily benefit all parties. Interestingly, greater persistence might harm—rather than benefit—the party in power. This is because with greater

[4] For example, in a famous passage, Aristotle, emphasizes: "A basic principle of the democratic constitution is liberty.... and one aspect of liberty is being ruled and ruling in turn." (quoted in Hansen [24], p. 74).

persistence, when power finally switches away from the current incumbent, the sustainability constraint of the new government will be more binding, and this will necessitate a bigger transfer away from the current incumbent in the future.

Our paper is related to the large and growing political economy literature.[5] Several recent papers also study dynamic political economy issues which is the focus of our paper. These include, among others, Acemoglu and Robinson [7], [8], Acemoglu, Egorov and Sonin [2], Battaglini and Coate [13], Hassler et al. [25], Krusell and Rios-Rull [27], Lagunoff [28], [29], Roberts [39] and Sonin [41]. The major difference of our paper from this literature is our focus on Pareto efficient allocations rather than Markov perfect equilibria. Almost all of the results in the paper are the result of this focus.

In this respect, our work is closely related to and builds on previous analyses of constrained efficient allocations in political economy models or in models with limited commitment. These include, among others, the limited-commitment risk sharing models of Thomas and Worrall [43] and Kocherlakota [26] and the political economy models of Dixit, Grossman and Gul [19] and Amador [11], [12]. The main difference between our paper and these previous studies is our focus on the production economy. Several of our key results are derived from the explicit presence of production (labor supply) decisions.[6] In addition, to the best of our knowledge, no existing work has systematically analyzed the impact of the Markov process for power switches on the set of Pareto efficient allocations.[7]

The paper most closely related to our work is a recent and independent contribution by Aguiar and Amador [10], who consider an international political economy model in which a party that comes to power derives greater utility from current consumption then groups not in power. Similar to our environment, there is also no commitment and the identity of the power fluctuates over time. Aguiar and Amador characterize a class of tractable equilibria, which lead to fluctuations in taxes on investment (expropriation), slow convergence to steady state

[5] See, among others, Persson and Tabellini [33], Persson, Roland and Tabellini [34], [35], Besley and Coate [16], [17], Baron [14], Grossman and Helpman [22], [23], Dixit and Londergan [20] and Bueno de Mesquita et al. [18].

[6] Acemoglu, Golosov and Tsyvinski [4], [5], [6] and Yared [44] also consider dynamic political economy models with production, but their models do not feature power switches between different social groups. Battaglini and Coate [13] study the Markov perfect equilibria of the model of debt policy with power switches.

[7] Acemoglu and Robinson [9] and Robinson [40] also question the insight that long-lived all-encompassing regimes are growth-promoting. They emphasize the possibility that such regimes may block beneficial technological or institutional changes in order to maintain their political power.

due to commitment problems, and potential differential responses to openness depending on the degree of "political economy frictions" parameterized by the difference in the differential utility from consumption for the group in power. In contrast to our model political economy distortions disappear in the long run. In their model the backloading argument similar to Acemoglu, Golosov and Tsyvinski [4], [6] applies as despite the current impatience the parties agree on the long term allocations. Battaglini and Coate [13] is also closely related, since they investigate the implications of dynamic political economy frictions in a model with changes in the identity of the group in power, though focusing on Markovian equilibria and implications for debt and government expenditure.

The rest of the paper is organized as follows. Section 2 introduces the basic environment and characterizes the first-best allocations. Section 3 describes the political economy game. Section 4 analyzes the structure of (constrained) Pareto efficient allocations, characterizes the level and dynamics of distortions, and also provides a complete characterization of the dynamics of distortions in the case with two parties. Section 5 studies the effect of frequency (persistence) of power switches on political economy distortions. Section 6 provides a numerical illustration. Section 7 concludes, while the Appendix contains a number of technical details and the proofs omitted from the text.

2 Environment and Benchmark

In this section, we introduce the model and describe efficient allocations without political economy constraints.

2.1 Demographics, Preferences and Technology

We consider an infinite horizon economy in discrete time with a unique final good. The economy consists of N parties (groups). Each party j has utility at time $t = 0$ given by

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u_j(c_{j,t}, l_{j,t}), \tag{1}$$

where $c_{j,t}$ is consumption, $l_{j,t}$ is labor supply (or other types of productive effort), and $\mathbb{E}_0$ denotes the expectations operator at time $t = 0$. To simplify the analysis, we assume, without loss of any economic insights, labor supply belongs to the closed interval $\left[0, \bar{l}\right]$ for each party. We also impose the following assumption on utility functions.

Assumption 1 ***(Preferences)*** *The instantaneous utility function*

$$u_j : \mathbb{R}_+ \times \left[0, \bar{l}\right] \to \mathbb{R},$$

for $j = 1, ..., N$ is uniformly continuous, twice continuously differentiable in the interior of its domain, strictly increasing in c, strictly decreasing in l and jointly strictly concave in c and l, with $u_j(0,0) = 0$ and satisfies the following Inada conditions:

$$\lim_{c \to 0} \frac{\partial u_j(c,l)}{\partial c} = \infty \text{ and } \lim_{c \to \infty} \frac{\partial u_j(c,l)}{\partial c} = 0 \text{ for all } l \in \left[0, \bar{l}\right],$$

$$\frac{\partial u_j(c,0)}{\partial l} = 0 \text{ and } \lim_{l \to \bar{l}} \frac{\partial u_j(c,l)}{\partial l} = -\infty \text{ for all } c \in \mathbb{R}_+.$$

The differentiability assumptions enable us to work with first-order conditions. The Inada conditions ensure that consumption and labor supply levels are not at corners. The concavity assumptions are also standard. Nevertheless, these will play an important role in our analysis, since they create a desire for consumption and labor supply smoothing over time.

The economy also has access to a linear aggregate production function given by

$$Y_t = \sum_{j=1}^{N} l_{j,t}. \tag{2}$$

2.2 Efficient Allocation without Political Economy

As a benchmark, we start with the efficient allocation without political economy constraints. This is an allocation that maximizes a weighted average of different groups' utilities, with *Pareto weights* vector denoted by $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)$, where $\alpha_j \geq 0$ for $j = 1, ..., N$ denotes the weight given to party j. We adopt the normalization $\sum_{j=1}^{N} \alpha_j = 1$. The program for the (unconstrained) efficient allocation can be written as:

$$\max_{\left\{[c_{j,t}, l_{j,t}]_{j=1}^{N}\right\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\sum_{j=1}^{N} \alpha_j u_j(c_{j,t}, l_{j,t})\right] \tag{3}$$

subject to the resource constraint

$$\sum_{j=1}^{N} c_{j,t} \leq \sum_{j=1}^{N} l_{j,t} \text{ for all } t. \tag{4}$$

Standard arguments imply that the *first-best allocation*, $\left\{\left[c_{j,t}^{fb}, l_{j,t}^{fb}\right]_{j=1}^{N}\right\}_{t=0}^{\infty}$, which is a solution to the program (3), satisfies the following conditions:

$$\text{no distortions: } \frac{\partial u_j(c_{j,t}^{fb}, l_{j,t}^{fb})}{\partial c} = -\frac{\partial u_j(c_{j,t}^{fb}, l_{j,t}^{fb})}{\partial l} \text{ for } j = 1, ..., N \text{ and all } t, \tag{5}$$

$$\text{perfect smoothing: } c_{j,t}^{fb} = c_j^{fb} \text{ and } l_{j,t}^{fb} = l_j^{fb} \text{ for } j = 1, ..., N \text{ and all } t. \tag{6}$$

The structure of the first-best allocations is standard. Efficiency requires the marginal benefit from additional consumption to be equal to the marginal cost of labor supply for each individual, and also requires perfect consumption and labor supply smoothing. Note that different parties can be treated differently in the first-best allocation depending on the Pareto weight vector $\boldsymbol{\alpha}$, i.e., receive different consumption and labor allocations.

3 Political Economy

We now consider a political environment in which political power fluctuates between the N parties $j \in \mathcal{N} \equiv \{1, ..., N\}$. The game form in this political environment is as follows.

1. In each period t, we start with one party, j', in power.
2. All parties simultaneously make their labor supply decisions $l_{j,t}$. Output $Y_t = \sum_{j=1}^{N} l_{j,t}$ is produced.
3. Party j' chooses consumption allocations $c_{j,t}$ for each party subject to the feasibility constraint
$$\sum_{j=1}^{N} c_{j,t} \leq \sum_{j=1}^{N} l_{j,t}. \tag{7}$$
4. A first-order Markov process m determines who will be in power in the next period. The probability of party j being in power following party j' is $m(j \mid j')$, with $\sum_{j=1}^{N} m(j \mid j') = 1$ for all $j' \in \mathcal{N}$.

A number of features is worth noting about this setup. First, this game form captures the notion that political power fluctuates between groups. Second, it builds in the assumption that

the allocation of resources is decided by the group in power (without any prior commitment to what the allocation will be). The assumption of no commitment is standard in political economy models (e.g., Persson and Tabellini [33], Acemoglu and Robinson [8]), while the presence of power switches is crucial for our focus (see also Dixit, Grossman and Gul [19], and Amador, [11], [12]). We have simplified the analysis by assuming that there are no constraints on the allocation decisions of the group in power and by assuming no capital.

In addition, we impose the following assumption on the Markov process for power switches.

Assumption 2 ***(Power Switches)*** *The first-order Markov chain $m(j \mid j')$ is irreducible, aperiodic and ergodic.*

We are interested in *subgame perfect equilibria* of this infinitely-repeated game. More specifically, as discussed in the Introduction, we will look at subgame perfect equilibria that correspond to constrained Pareto efficient allocations, which we refer to as *Pareto efficient perfect equilibria*.

To define these equilibria, we now introduce additional notation. Let $h^t = (h_0, ..., h_t)$, with $h_s \in \mathcal{N}$ be the history of power holdings. Let H^∞ denote the set of all such possible histories of power holding. Let $\mathbf{L}^t = \left(\{l_{j,0}\}_{j=1}^N, ..., \{l_{j,t}\}_{j=1}^N\right)$ be the history of labor supplies, and let $\mathbf{C}^t = \left(\{c_{j,0}\}_{j=1}^N, ..., \{c_{j,t}\}_{j=1}^N\right)$ be the history of allocation rules. A *(complete) history* of this game ("history" for short) at time t is $\omega^t = \left(h^t, \mathbf{C}^{t-1}, \mathbf{L}^{t-1}\right)$, which describes the history of power holdings, all labor supply decisions, and all allocation rules chosen by groups in power. Let the set of all potential date t histories be denoted by Ω^t. In addition, denote an intermediate-stage (complete) history by $\hat{\omega}^t = \left(h^t, \mathbf{C}^{t-1}, \mathbf{L}^t\right)$, and denote the set of intermediate-stage full histories by $\hat{\Omega}^t$. The difference between ω and $\hat{\omega}$ lies in the fact that the former does not contain information on labor supplies at time t, while the latter does. The latter history will be relevant at the intermediate stage where the individual in power chooses the allocation rule.

We can now define strategies as follows. First define the following sequence of mappings $\hat{l} = \left(\hat{l}^0, \hat{l}^1, ..., \hat{l}^t, ...\right)$ and $\hat{C} = \left(\hat{C}^0, \hat{C}^1, ..., \hat{C}^t, ...\right)$, where $\hat{l}^t : \Omega^t \to \left[0, \bar{l}\right]$ determines the level of labor a party will supply for every given history $\omega^t \in \Omega^t$, and $\hat{C}^t : \hat{\Omega}^t \to \mathbb{R}_+^N$ determines a sequence of allocation rules, which a party would choose, if it were in power, for every given intermediate-stage history $\hat{\omega}^t \in \hat{\Omega}^t$, such that $\hat{C}$ satisfies the feasibility constraint (7). A date t strategy for party j is $\sigma_j^t = \left(\hat{l}^t, \hat{C}^t\right)$. Denote the set of date t strategies by Σ^t. A strategy

for party j is $\sigma_j = \left(\left\{\sigma_j^t\right\} : t = 0, 1, ...\right)$ and the set of strategies is denoted by Σ. Denote the expected utility of party j at time t as a function of its own and others' strategies given intermediate-stage history $\hat{\omega}^t$ (which subsumes history ω^t) by $U_j\left(\sigma_j, \sigma_{-j} \mid \hat{\omega}^t\right)$.

We next define various concepts of equilibria which we use throughout the paper.

Definition 1 *A* subgame perfect equilibrium (SPE) *is a collection of strategies* $\sigma^* = \left(\left\{\sigma_j^t\right\} : j = 1, ..., N,\ t = 0, 1, ...\right)$ *such that* σ_j^* *is best response to* σ_{-j}^* *for all* $\hat{\omega}^t \in \hat{\Omega}^t$ *(and* $\omega^t \in \Omega^t$*) and for all* j*, i.e.,* $U_j\left(\sigma_j^*, \sigma_{-j}^* \mid \hat{\omega}^t\right) \geq U_j\left(\sigma_j, \sigma_{-j}^* \mid \hat{\omega}^t\right)$ *for all* $\sigma_j \in \Sigma$*, for all* $\hat{\omega}^t \in \hat{\Omega}^t$ *(and* $\omega^t \in \Omega^t$*), for all* $t = 0, 1, ...$ *and for all* $j \in \mathcal{N}$*.*

Definition 2 *A (constrained)* Pareto efficient perfect equilibrium*,* σ^{**}*, is a collection of strategies that form an SPE such that there does not exist another SPE* σ*, whereby* $U_j\left(\sigma_j, \sigma_{-j} \mid \hat{\omega}^0\right) \geq U_j\left(\sigma_j^{**}, \sigma_{-j}^{**} \mid \hat{\omega}^0\right)$ *for all* $j \in \mathcal{N}$*, with at least one strict inequality.*

In light of this definition, by *Pareto efficient allocations* we refer to the equilibrium-path allocations that result from a Pareto efficient perfect equilibrium. To characterize Pareto efficient allocations, we will first determine the worst subgame perfect equilibrium, which will be used as a threat against deviations from equilibrium strategies. These are defined next. We write $j = \mathbf{j}\left(h^t\right)$ if party j is in power at time t according to history (of power holdings) h^t. We also use the notation $h^t \in H_j^t$ whenever $j = \mathbf{j}\left(h^t\right)$. A *worst SPE for party* j at time t following history ω^t is a collection of strategies σ^W that form a SPE such that there does not exist another SPE σ^{***} such that $U_j\left(\sigma_j^{***}, \sigma_{-j}^{***} \mid \hat{\omega}^t\right) < U_j\left(\sigma_j^W, \sigma_{-j}^W \mid \hat{\omega}^t\right)$ for all $\hat{\omega}^t \in \hat{\Omega}^t, \forall t$.

4 Characterization of (Constrained) Pareto Efficient Allocations

4.1 Preliminary Results

Let $V_j^W(j')$ denote the expected payoff of party j in the worst subgame perfect equilibrium for party j from period $t+1$ on, conditional on party j' being in power at time t, i.e. $j' = \mathbf{j}(h^t)$. Lemma 1 in the Appendix clarifies why, given Assumption 2, $V_j^W(j')$ only depends only on the identify of party in power at time t. The next proposition provides a constraint maximization problem characterizing the Pareto efficient allocations subject to the political economy constraints (represented here by the sustainability constraints).

Proposition 1 *Suppose Assumptions 1 and 2 hold. Then, an outcome of any Pareto efficient subgame perfect equilibrium is a solution to the following maximization problem for all h^t:*

$$\max_{\{c_j(h^t),l_j(h^t)\}_{j=1,..,N;\ h^t}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\sum_{j=1}^{N} \alpha_j u_j(c_j\left(h^t\right), l_j\left(h^t\right)) \right] \tag{8}$$

subject to, for all h^t,

$$\sum_{j=1}^{N} c_j\left(h^t\right) \leq \sum_{j=1}^{N} l_j\left(h^t\right), \tag{9}$$

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s u_j\left(c_j\left(h^{t+s}\right), l_j\left(h^{t+s}\right)\right) \geq \beta V_j^W(\mathbf{j}(h^t)) \text{ for all } j \neq \mathbf{j}(h^t), \tag{10}$$

and

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s u_{\mathbf{j}(h^t)}\left(c_{\mathbf{j}(h^t)}\left(h^{t+s}\right), l_{\mathbf{j}(h^t)}\left(h^{t+s}\right)\right) \geq v_{\mathbf{j}(h^t)}\left(\sum_{j\neq\mathbf{j}(h^t)} l_j\left(h^t\right)\right) + \beta V^W_{\mathbf{j}(h^t)}(\mathbf{j}(h^t)), \tag{11}$$

for some Pareto weights vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)$, where

$$v_j(Y) \equiv \max_{\tilde{l}\in[0,\bar{l}]} u_j\left(Y + \tilde{l}\ , \tilde{l}\right). \tag{12}$$

Proof. See the Appendix. ■

In this constrained maximization problem, (10) represents the participation constraint for each group that is not in power.[8] Group j can never receive less than $\beta V_j^W(\mathbf{j}(h^t))$, which it can guarantee by supplying and consuming zero today and then receiving its payoff in the worst subgame perfect equilibrium from tomorrow on starting with group $\mathbf{j}(h^t)$ in power today. In addition, (11) is the participation or the *sustainability constraint* of group $\mathbf{j}(h^t)$ that is in power. This group can always deviate today and consume all output, giving itself current utility $v_{\mathbf{j}(h^t)}\left(\sum_{j\neq\mathbf{j}(h^t)} l_j\left(h^t\right)\right)$, and from tomorrow on receive its payoff in the worst subgame perfect equilibrium, $\beta V^W_{\mathbf{j}(h^t)}(\mathbf{j}(h^t))$ starting with itself in power today. To simplify the notation, we define

$$V_j(h^{t-1}) \equiv \mathbb{E}\left\{ \sum_{s=0}^{\infty} \beta^s u_j\left(c_j\left(h^{t+s}\right), l_j\left(h^{t+s}\right)\right) | h^{t-1} \right\}$$

[8] The maximization (8) subject to (9), (10), and (11) is a potentially non-convex optimization problem, because (11) defines a non-convex constraint set. This implies that randomizations may improve the value of the program (see, for example, Prescott and Townsend [36], [37]). The working paper version of our paper establishes the analogs of our results in the presence of explicit randomization on public signals. We omit these details here to economize on space and notation.

$$V_j[h^{t-1}, i] \equiv \mathbb{E}\left\{\sum_{s=0}^{\infty} \beta^s u_j\left(c_j\left(h^{t+s}\right), l_j\left(h^{t+s}\right)\right) | h^{t-1}, \mathbf{j}(h^t) = i\right\}$$

The difference between $V_j(h^{t-1})$ and $V_j[h^{t-1}, i]$ is that the former denotes expected lifetime utility of party j in period t before the uncertainty which party is in power in that period is realized, while the latter denotes the expected lifetime utility after realization of this uncertainty. From the above definition and Assumption 2,

$$V_j(h^{t-1}) = \sum_{j'=1}^{N} m(j'|h_{t-1}) V_j[h^{t-1}, j'].$$

Proposition 1 implies that in order to characterize the entire set of Pareto efficient perfect equilibria, we can restrict attention to strategies that follow a particular prescribed equilibrium play, with the punishment phase given by σ^W.[9]

4.2 Political Economy Distortions

We next characterize the structure of distortions arising from political economy. Our first result shows that as long as sustainability/political economy constraints are binding, the labor supply of parties that are not in power is distorted *downwards*. There is a positive wedge between their marginal utility of consumption and marginal disutility of labor. In contrast, there is no wedge for the party in power. Recall also that without political economy constraints, in the first-best allocations, the distortions are equal to zero.

Proposition 2 *Suppose that Assumptions 1 and 2 hold. Let the (normalized) Lagrange multiplier on the sustainability constraint (11) given history h^t, be denoted by $\lambda_{\mathbf{j}(h^t)}\left(h^t\right)$. Then as long as $\lambda_{\mathbf{j}(h^t)}\left(h^t\right) > 0$, the labor supply of all groups that are not in power, i.e., $j \neq \mathbf{j}\left(h^t\right)$, is distorted* downwards*, in the sense that*

$$\frac{\partial u_j\left(c_j\left(h^t\right), l_j\left(h^t\right)\right)}{\partial c} > -\frac{\partial u_j\left(c_j\left(h^t\right), l_j\left(h^t\right)\right)}{\partial l}.$$

The labor supply of a party in power, $j' = \mathbf{j}\left(h^t\right)$, is undistorted, i.e.,

$$\frac{\partial u_{j'}\left(c_{j'}\left(h^t\right), l_{j'}\left(h^t\right)\right)}{\partial c} = -\frac{\partial u_{j'}\left(c_{j'}\left(h^t\right), l_{j'}\left(h^t\right)\right)}{\partial l}.$$

[9] Notice, however, that this proposition applies to Pareto efficient outcomes, *not* to the strategies that individuals use in order to support these outcomes. These strategies *must be* conditioned on information that is not contained in the history of power holdings, h^t, since individuals need to switch to the worst subgame perfect equilibrium in case there is any deviation from the implicitly-agreed action profile. This information is naturally contained in ω^t. Therefore, to describe the subgame perfect equilibrium strategies we need to condition on the full histories ω^t.

Proof. See the Appendix ■

The intuition for why there will be downward distortions in the labor supply of groups that are not in power is similar to that in Acemoglu, Golosov and Tsyvinski [4], [6]. Positive distortions, which are the equivalent of "taxes," discourage labor supply, reducing the amount of output that the group in power can "expropriate" (i.e., allocate to itself as consumption following a deviation). This relaxes the sustainability constraint (11). In fact, starting from an allocation with no distortions, a small distortion in labor supply creates a second-order loss. In contrast, as long as the multiplier on the sustainability constraint is positive, this small distortion creates a first-order gain in the objective function, because it enables a reduction in the rents captured by the group in power. We show in the appendix that the problem can be re-written recursively using the updated Pareto weights α_j following Marcet and Marimon [30]. This intuition also highlights that the extent of distortions will be closely linked to the Pareto weights given to the group in power. In particular, when α_j is close to 1 and group j is in power, there will be little gain in relaxing the sustainability constraint (11). In contrast, the Pareto efficient allocation will attempt to provide fewer rents to group j when α_j is low, and this is only possible by reducing the labor supply of all other groups, thus distorting their labor supplies. Note that the Lagrange multiplier $\lambda_{\mathbf{j}(h^t)}\left(h^t\right)$ is a measure of distortions. This follows immediately from Proposition 2, and more explicitly its proof in the Appendix, which shows that the wedges between the marginal utility of consumption and the marginal disutility of labor are directly related to $\lambda_{\mathbf{j}(h^t)}\left(h^t\right)$. This is useful as it will enable us to link the level and behavior of distortions to the behavior of the Lagrange multiplier $\lambda_{\mathbf{j}(h^t)}\left(h^t\right)$.

The (constrained) efficient allocations will be "first-best" if and only if the Lagrange multipliers associated with all constraints (10) and (11) are equal to zero (so that there are no distortions in a first-best allocation). That is, a first-best allocation starting at history h^t involves $\lambda_j\left(h^{t+s}\right)=0$ for all j and all subsequent h^{t+s}.

We also call any allocation $\{c_j^*, l_j^*\}_{j=1,...,N}$ a *sustainable first-best allocation* if $\{c_j^*, l_j^*\}$ is a first-best allocation that satisfies for each $j=1,...,N$

$$\frac{1}{1-\beta}u_j\left(c_j^*, l_j^*\right) \geq v_j\left(\sum_{i\neq j} l_i^*\right) + \beta V_j^W\left(j\right), \tag{13}$$

where recall that v_j is defined in (12), and

$$\frac{1}{1-\beta}u_j\left(c_j^*, l_j^*\right) \geq \beta V_j^W\left(j'\right) \text{ for each } j' \neq j. \tag{14}$$

The next theorem studies the dynamics of the distortions.

Theorem 1 *Suppose that Assumptions 1 and 2 hold. Then there exists* $\bar{\beta}$*, with* $0 < \bar{\beta} < 1$ *such that*

1. *For all* $\beta \geq \bar{\beta}$*, there is some first-best allocation that is sustainable;*
2. *If no first-best allocation is sustainable then* $\{c_j(h^t), l_j(h^t)\}_{j=1,...,N}$ *converges to an invariant non-degenerate distribution* F*.*

Proof. See the Appendix. ■

This theorem shows that for high discount factors, i.e., $\beta \geq \bar{\beta}$, there are first-best allocations that are sustainable. In contrast, when β is low, then there are no sustainable first best allocations and consumption and labor supply levels will fluctuate permanently as political power changes hands between different parties. The invariant distribution can be quite complex in general, and involve the consumption and labor supply levels of each group depending on the entire history of power holdings.

This theorem does not answer the question of whether first-best allocations, when they are sustainable, will be ultimately reached. We address this question for the case of two parties in the next section.

4.3 The Case of Two Parties

In this subsection, we focus on an economy with two parties (rather than N parties as we have done so far). We also specialize utility function to be quasi-linear. Under these conditions, we show that when there exists a sustainable first-best allocation (i.e., an undistorted allocations for some Pareto weights), the equilibrium will necessarily converge to a point in the set of first-best allocations. More specifically, starting with any Pareto weights, the allocations ultimately converge to undistorted allocations.

For the rest of this section, we impose the following assumption on the preferences.

Assumption 3 ***(Quasi-Linear Preferences)*** *The instantaneous utility of each party j satisfies $u_j\left(c_j-\eta_j\left(l_j\right)\right)$ with the normalization*

$$\eta_j'\left(1\right)=1. \tag{15}$$

Assumption 3 implies that there are no income effects in labor supply. Consequently, when there are no distortions, the level of labor supply by each group will be constant, and given the normalization in (15), this labor supply level will be equal to 1 (which is without loss of any generality). The absence of income effects also simplifies the analysis and dynamics, which is our main focus in this section.

Now we are ready to prove the result on the convergence to the first-best allocations.

Theorem 2 *Suppose that Assumptions 1 and 3 hold. If there exists a sustainable first-best, then*

$$\left\{\left(c_j\left(h^t\right),l_j\left(h^t\right)\right)\right\}_{j=1}^{2}\rightarrow\left\{\left(c_j^*,l_j^*\right)\right\}_{j=1}^{2}$$

where $\left\{\left(c_j^*,l_j^*\right)\right\}_{j=1}^{2}$ *is some first-best sustainable allocation.*

Proof. See the Appendix. ■

This theorem establishes that if there exist first-best allocations that are sustainable they will be ultimately reached. This implies that the political economy frictions in this situation will disappear in the long run. The resulting long-run allocations will not feature distortions and fluctuations in consumption and labor supply. Note, however, that the theorem does not imply that such first-best allocations will be reached immediately. Sustainability constraints may bind for a while, because the sustainable first-best allocations may involve too high a level of utility for one of the groups. In this case, a first-best allocation will be reached only after a specific path of power switches increases the Pareto weight of this group to a level consistent with a first-best allocation. After this point, sustainability constraints do not bind for either party, and thus Pareto weights are no longer updated and the same allocation is repeated in every period thereafter. Interestingly, however, this first-best allocation may still involve transfer from one group to another.

The analysis in this subsection relies on two assumptions: (a) preferences are quasi-linear and (b) there are only two parties. Relaxing any of this assumptions significantly complicates

the analysis, and we conjecture one can construct counterexamples to Theorem 2 when they are violated. The assumption of quasi-linearity is crucial to establish a key lemma in the Appendix. It ensures that an updated Pareto weight $\alpha_j\left(h^t\right)$ when party j is in power does not "jump" from below the first-best Pareto weight α_j^* to above it. The higher updated Pareto weight $\alpha_j(h^t)$ ensures higher expected utility from period $t+1$ for party j, but also introduces additional distortions in period t. With quasi-linearity, it is possible to show that the first effect dominates for all $\alpha_j(h^t)$ as long as $\alpha_j(h^t) \geq \alpha_j^*$, but it does not have to be true more generally.[10]

5 Political Stability and Efficiency

Our framework enables an investigation of the implications of persistence of power on the sustainability of first-best applications. In particular, the "stability" or persistence of power is captured by the underlying Markov process for power switches. If the Markov process $m\left(j \mid j'\right)$ makes it very likely that one of the groups, say group 1, will be in power all the time, we can think of this as a very "stable distribution of political power".

Such an investigation is important partly because a common conjecture in the political economy literature is that such stable distributions of political power are conducive to better policies. For example Olson [**?**] and McGuire and Olson [31] reach this conclusion by contrasting an all-encompassing long-lived dictator to a "roving bandit". They argue that a dictator with stable political power is superior to a roving bandit and will generate better public policies. This conjecture at first appears plausible, even compelling: what matters for better policies are high "effective discount factors," and frequent switches in the identity of powerholders would reduce these effective discount factors. Hence, stability (persistence) of power should be conducive to better policies and allocations. Similar insights emerge from the standard principal-agent models of political economy, such as Barro [15], Ferejohn [21], Persson, Roland and Tabellini [34], [35], because, in these models, it is easier to provide incentives to a politician who is more likely to remain in power. We next investigate whether a similar result applies in

[10] The assumption that $N = 2$ can be relaxed for the case where all N parties have the same preferences and the Markov process for power switches m is symmetric. In that case we conjecture that it can be shown that the party that starts with a highest initial Pareto weight has a sequence of updated Pareto weights which is monotonically decreasing and is bounded from below by $1/N$. In that case, the proof of Theorem 2 goes without changes. With asymmetric parties this condition is difficult to ensure.

our context. In particular, we ask which types of Markov processes make it more likely that a large set of first-best allocations are sustainable.

Our main result in this section is that this common conjecture is not generally correct. In fact, perhaps at first surprisingly, in our framework, essentially the opposite of this conjecture holds. In particular, we show that the opposite of this conjectures is true for the set of sustainable first-best allocations; greater persistence of power encourages deviations and leads to a smaller set of sustainable first-best allocations. In the next section, we complement this result by showing, numerically, how changes in persistence of power influences the utility of different players.

The next theorem shows our main result that higher persistence of power makes distortions more likely, in the sense that it leads to a smaller set of sustainable first-best allocations. This result is stated in the next theorem for the general case in which there are N parties and general rather than quasi-linear preferences (whereas, recall that, Theorem 2 was for two parties with quasi-linear preferences).

Theorem 3 *Consider an economy consisting of N groups, with group j having utility functions $u_j(c_j, l_j)$ satisfying Assumption 1. Suppose that $m(j \mid j) = \rho$ and $m(j' \mid j) = (1-\rho)/(N-1)$ for any $j' \neq j$. Then the set of sustainable first-best allocations is decreasing in ρ (i.e., $\bar{\beta}$ defined in Theorem 1 is increasing in ρ).*

Proof. Recall that a first-best allocation satisfies (13) and (14). First, we show that under the conditions of the theorem, (13) implies (14). From the specification of the power switching process, we have that group j will remain in power next period with probability ρ, and hence $V_j^W(j)$ satisfies

$$V_j^W(j) = \rho V_j^P + (1-\rho) V_j^{NP}, \tag{16}$$

where V_j^P and V_j^{NP} are respectively the utility of being in power and not in power after a deviation. These are given by

$$V_j^P = u_j\left(\tilde{l}_j, \tilde{l}_j\right) + \beta\rho V_j^P + \beta(1-\rho) V_j^{NP}, \tag{17}$$

and

$$V_j^{NP} = \beta\left(1 - \frac{1-\rho}{N-1}\right) V_j^{NP} + \beta\left(\frac{1-\rho}{N-1}\right) V_j^P, \tag{18}$$

where $\tilde{l}_j$ is a solution to $\partial u_j\left(\tilde{l}_j,\tilde{l}_j\right)/\partial c = -\partial u_j\left(\tilde{l}_j,\tilde{l}_j\right)/\partial l$ and the symmetry of m implies that V_j^{NP} is independent of which group j' succeeds j in power. Subtracting (17) from (18), we obtain

$$V_j^P - V_j^{NP} = \frac{u_j\left(\tilde{l}_j,\tilde{l}_j\right)}{1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)}. \tag{19}$$

Similarly for any $j' \neq j$,

$$V_j^W\left(j'\right) = \left(1-\frac{1-\rho}{N-1}\right)V_j^{NP} + \left(\frac{1-\rho}{N-1}\right)V_j^P,$$

which implies that

$$\begin{aligned} V_j^W(j) - V_j^W\left(j'\right) &= \left(\rho-\frac{1-\rho}{N-1}\right)\left(V_j^P - V_j^{NP}\right) \\ &= \left(\rho-\frac{1-\rho}{N-1}\right)\frac{u_j\left(\tilde{l}_j,\tilde{l}_j\right)}{1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)}. \end{aligned}$$

Now compare the right hand sides of (13) and (14)

$$\begin{aligned} v_j\left(\sum_{i\neq j} l_i^*\right) + \beta V_j^W(j) - \beta V_j^W(j') &= v_j\left(\sum_{i\neq j} l_i^*\right) + \beta\left(\rho-\frac{1-\rho}{N-1}\right)\frac{u_j\left(\tilde{l}_j,\tilde{l}_j\right)}{1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)} \\ &\geq u_j\left(\tilde{l}_j,\tilde{l}_j\right)\left(1+\frac{\beta\left(\rho-\frac{1-\rho}{N-1}\right)}{1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)}\right) \\ &= \frac{u_j\left(\tilde{l}_j,\tilde{l}_j\right)}{1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)} \\ &> 0. \end{aligned}$$

The second line of this equation follows from the fact that $v_j\left(\sum_{i\neq j} l_i^*\right) \geq u_j\left(\tilde{l}_j,\tilde{l}_j\right)$ and the last line follows because $1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)$ is bounded from below by $1-\beta > 0$. Therefore if any $\left(c_j^*, l_j^*\right)$ satisfies (13), it also satisfies (14).

To prove the theorem it is sufficient to show that if any $\left\{c_j^*, l_j^*\right\}_j$ satisfies (13) for some ρ and all j, it also satisfies it for any other $\rho' \leq \rho$. The left-hand side of this expression is independent of ρ, so is the first term on the right-hand side. Therefore, the desired result follows if the second term on the right-hand side, $V_j^W(j)$, is increasing in ρ. Substitute (19)

into (17) to obtain

$$V_j^P = \frac{1-\beta+\beta\left(\frac{1-\rho}{N-1}\right)}{(1-\beta)\left(1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)\right)} u_j\left(\tilde{l}_j, \tilde{l}_j\right),$$

and substituting this into (16), we obtain

$$V_j^W(j) = \frac{\beta\left(\frac{1-\rho}{N-1}\right)+(1-\beta)\rho}{(1-\beta)\left(1-\beta\rho+\beta\left(\frac{1-\rho}{N-1}\right)\right)} u_j\left(\tilde{l}_j, \tilde{l}_j\right),$$

which is increasing in ρ, establishing the desired result. ■

This theorem implies the converse of the Olson conjecture discussed above: the set of sustainable first-best allocations is maximized when there are *frequent power switches* between different groups. The Olson conjecture is based on the idea that "effective discount factors" are lower with frequent power switches, and this should make "cooperation" more difficult. Yet, "effective discount factors" would be the key factor in shaping cooperation (the willingness of the party in power to refrain from deviating) only if those in power can only be rewarded when in power. This is not necessarily the case, however, in reality or in our model. In particular, in our model deviation incentives are countered by increasing current utility and the Pareto weight of the party in power, and, all else equal, groups with higher Pareto weights will receive greater utility in all future dates. This reasoning demonstrates why "effective discount factor" is not necessarily the appropriate notion in this context. Instead, Theorem 3 has a simple intuition: the value of deviation for a group in power is determined by the persistence of power; when power is highly persistent, deviation becomes more attractive, since the group in power can still obtain relatively high returns following a deviation as it is likely to remain in power. In contrast, with more frequent power switches, the group in power is likely to be out of power tomorrow, effectively reducing the value of a deviation. Since first-best allocations, and thus first-best utilities, are independent of the persistence of power, this implies that greater persistence makes deviation more attractive relative to candidate first-best allocations, and thus first-best allocations become less likely to be sustainable.

6 The Form of the Pareto Frontier: Numerical Results

In this section, we numerically investigate the effect of persistence and frequency of power switches on the structure of Pareto efficient allocations. In particular, we study how both the

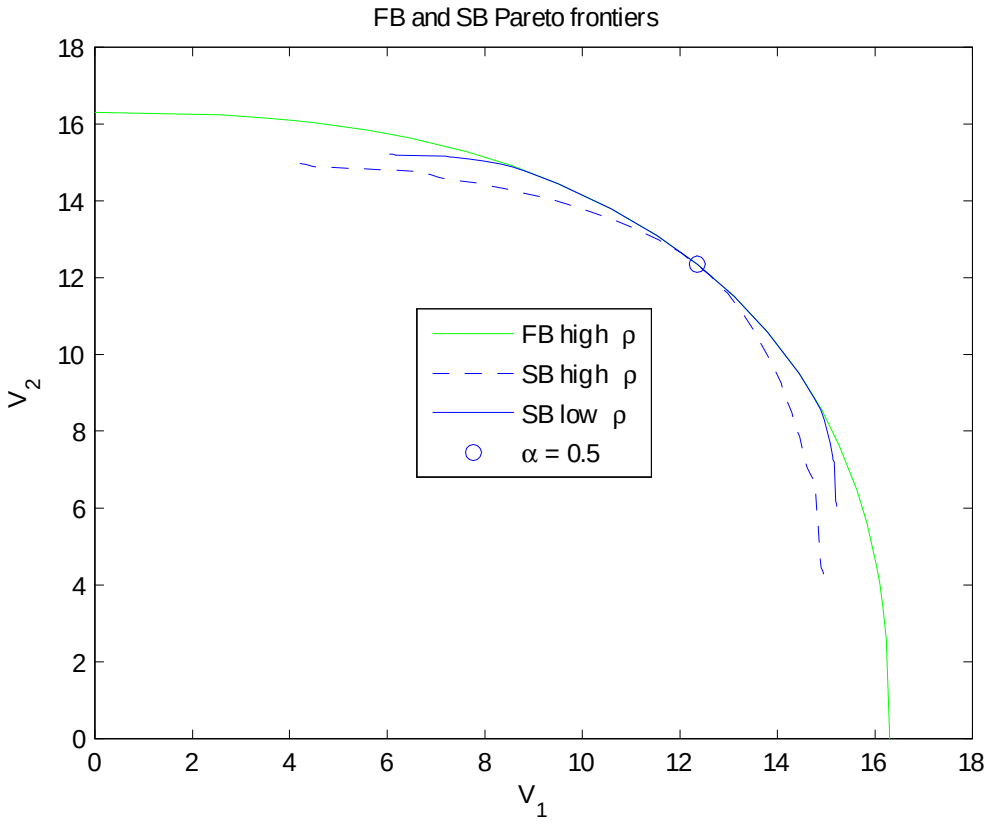
ex-ante Pareto frontier, which applies before the identity of the party in power is revealed, and the ex-post Pareto frontier, conditional on the identity of the party in power, vary with the degree of persistence. Our purpose is not to undertake a detailed calibration, but to provide illustrative numerical computations. We focus on an economy with two groups, $j = 1, 2$, and further simplify the discussion by assuming quasi-linear and identical utilities, given by

$$u_j\left(c - h\left(l\right)\right) = \frac{1}{1-\sigma}\left(c - \frac{1}{1+\gamma}l^{1+\gamma}\right)^{1-\sigma}.$$

We set $\gamma = 1$, $\sigma = 0.6$ and choose a symmetric Markov process for power switches with $m(1|1) = m(2|2) = \rho$, so that ρ is the persistence parameter (higher ρ corresponds to greater persistence). In Figure 1, we focus on the ex-ante Pareto frontier. For any given Pareto weight α we define ex-ante utility of party i by

$$V_i^{ea}\left[\alpha\right] = \frac{1}{2}V_i\left[\alpha, 1\right] + \frac{1}{2}V_i\left[\alpha, 2\right].$$

The figure plots $V_2^{ea}\left[\alpha\right]$ and $V_1^{ea}\left[\alpha\right]$ for different values of α and for two different values of levels of the persistence, ρ, $\rho = 0.9$ represented by the inner solid line, and $\rho = 0.6$ shown as the dashed line. We also show the first-best Pareto frontier for comparison (the outer solid line). We chose a discount factor β so that only one first-best allocation (that corresponding to the Pareto weights $\alpha_1 = \alpha_2 = 0.5$ is sustainable) when $\rho = 0.9$.



Consistent with Theorem 3, a larger set of first-best allocations is sustainable when persistence is lower. This can be seen by observing the common part of the first-best frontier and two other frontiers. For $\rho = 0.6$ this common part is larger than for $\rho = 0.9$ (which is just one point corresponding to $\alpha_1 = \alpha_2 = 0.5$). Also, the whole ex-ante Pareto frontier for low persistence lies above the Pareto frontier for high persistence, which implies that, before uncertainty about the identity of the party in power is realized, both parties are better off, and would prefer to be, in a regime with frequent power switches.

If the institutional characteristics of the society determining the frequency of power switches were chosen "behind the veil of ignorance," then this result would imply that both parties would prefer lower persistence. However, most institutional characteristics in practice are not determined behind a veil of ignorance. Different groups would typically have different amounts of political power, and in the context of our model, one would be "in power". In this case, what would be relevant is the *ex-post, not the ex-ante*, Pareto frontier. We next turn to the ex-post Pareto frontier. This is shown in Figure 2, assuming that party 1 is currently in power. As with

Figure 1, this figure also plots $V_2[\alpha, 1]$ against $V_1[\alpha, 1]$ for different values of α, and for high and low levels of persistence ($\rho = 0.9$ and $\rho = 0.6$). Figure 2 first shows that higher persistence imposes a greater "lower bound" on the possible payoff of party 1, which is in power. This can be seen from the fact that the beginning of the right solid line ($\rho = 0.6$) starts lower than the beginning of the left line ($\rho = 0.9$). This implies that greater persistence decreases the highest payoff that party 2 can get. The more surprising pattern in Figure 2 is that for high values of α, the value of party 1 is *lower* with higher persistence. This appears paradoxical at first, since higher persistence improves the deviation value of party 1. We investigate the reason for this pattern in greater detail in Figure 3.

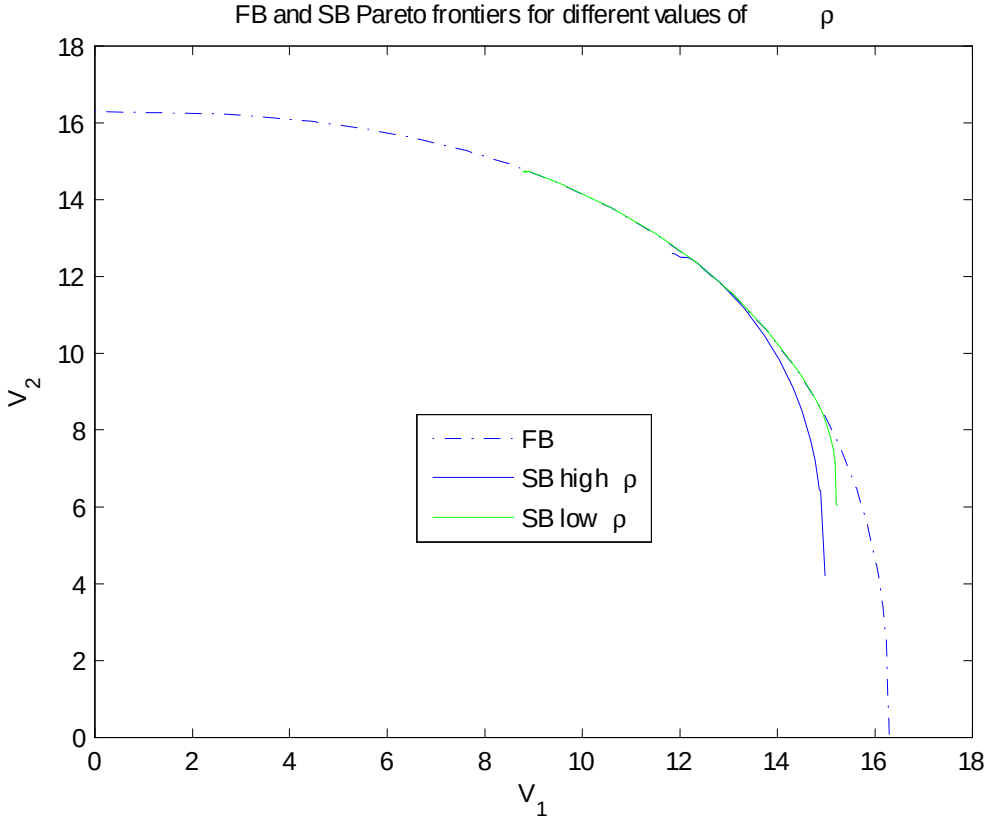
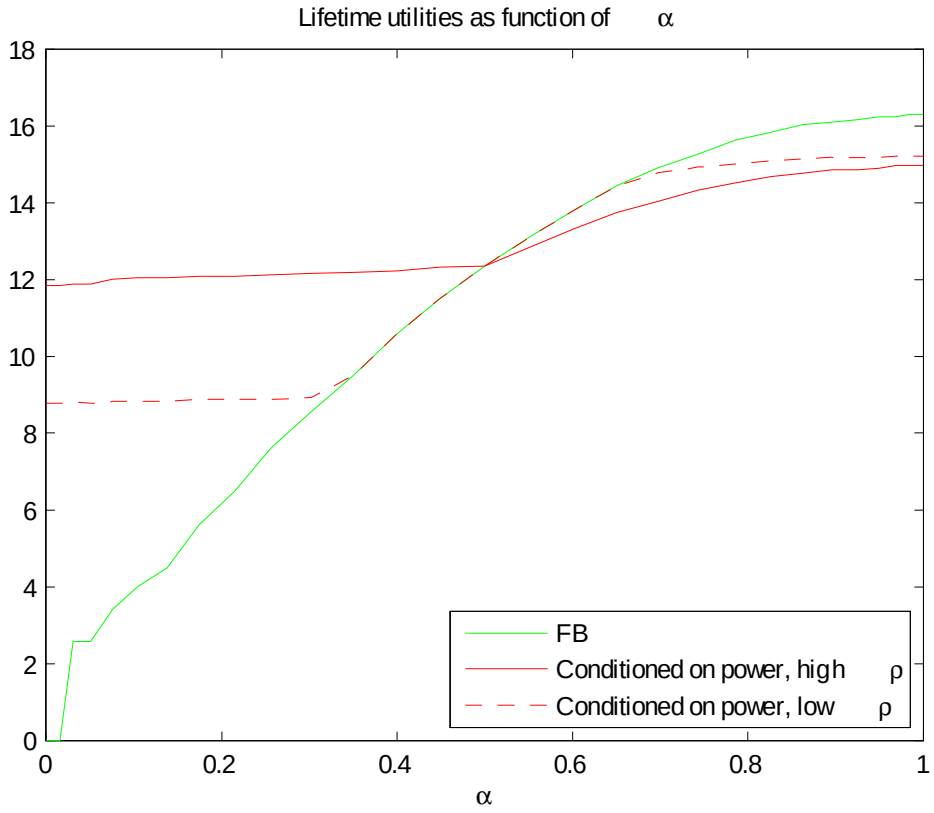


Figure 3 plots the payoff of party 1, $V_1[\alpha, 1]$, for different values values of α_1. The line representing the first-best allocation is monotonically increasing with the Pareto weight assigned to party 1. Figure 3 also shows that for α_1 sufficiently high, party 1 obtains higher value with lower persistence (for $\alpha_1 > 1/2$, the line representing $\rho = 0.9$ is below the dashed line representing $\rho = 0.6$). The reason for this is as follows. When α_1 is sufficiently high, the sustainability constraint of party 1 is slack. Thus greater persistence does not necessitate an

increase in current consumption or Pareto weight to satisfy its sustainability constraint. But it implies that the deviation value of party 2 will also be higher when it comes to power. When party 2 comes to power, its Pareto weight will be low and thus its sustainability constraint will be binding. A greater deviation value for party 2 at this point therefore translates into higher utility for it and lower utility for party 1. The anticipation of this lower utility in the future is the reason why the value of party 1 is decreasing in the degree of persistence in power switches for α_1 sufficiently high (greater than $1/2$ in the figure). The analogue of this argument holds for weights below $1/2$. Consistent with Figure 2, for low initial Pareto weights the utility of party 1 is increasing in the degree of persistence ρ (for $\alpha_1 < 1/2$, the line representing $\rho = 0.9$ is above the dashed line representing $\rho = 0.6$). Here, higher persistence of power increases the value of deviation and requires the planner to allocate more utility for this party with a low initial Pareto weight. This reasoning explains why the "lower bound" on the equilibrium payoffs for party 1 is higher with high persistence.

Another important implication of Figure 3 (already visible from Figure 2) is that changes in persistence do not necessarily correspond to Pareto improvements once the identity of the party in power is known. This highlights that even though the set of sustainable first-best allocations expands when the degree of persistence declines, along-the-equilibrium-path utility of both parties (conditional on the identity of the party in power) need not increase. This suggests that we should not necessarily expect a strong tendency for societies to gravitate towards institutional settings that increase the frequency of power switches.



7 Concluding Remarks

In this paper, we studied the (constrained) Pareto efficient allocations in a dynamic production economy in which the group in power decides the allocation of resources. The environment is a simple model of political economy. In our model, different groups have conflicting preferences and, at any given point in time, one of the groups has the political power to decide (or to influence) the allocation of resources. We made relatively few assumptions on the interactions between the groups; the process of power switches between groups is modeled in a reduced-form way with an exogenous Markov process. Our focus has been on the allocations that can be achieved given the distribution and fluctuations of political power in this society—rather than potential institutional failures leading to Pareto dominated equilibria given the underlying process of power switches. This focus motivated our characterization of Pareto efficient equilibria. In the constrained Pareto efficient equilibria, there are well-defined political economy distortions that change over time.

The distortions in Pareto efficient equilibria are a direct consequence of the sustainability constraints, which reflect the political economy interactions in this economy. If these sustainability constraints are not satisfied, the group in power would allocate all production to itself. The results here are driven by the location and shape of the Pareto frontier and by the "power" of different groups, which corresponds to what point the society is located along the Pareto frontier.

We showed how the analysis in the paper is simplified by the fact that these Pareto efficient allocations take a quasi-Markovian structure and can be represented recursively as a function of the identity of the group in power and updated Pareto weights. This recursive formulation allows us to provide a characterization of the level and dynamics of taxes and transfers in the economy.

In the two party case with quasilinear preferences we demonstrated that for high discount factors the economy converges to a first-best allocation in which there may be transfers between groups, but labor supply decisions are not distorted and the levels of labor supply and consumption do not fluctuate over time. When discount factors are low, the economy converges to an invariant stochastic distribution in which distortions do not disappear and labor supply and consumption levels fluctuate over time, even asymptotically.

Most importantly, we showed that the set of sustainable first-best allocations is "decreasing" in the degree of persistence of the Markov process for power change. This result directly contradicts a common conjecture that there will be fewer distortions when the political system creates a stable ruling group (see, e.g., Olson [32], or McGuire and Olson [31], as well as the standard principle-agent models of political economy such as Barro [15], Ferejohn [21], Persson, Roland and Tabellini [34], [35]). The reason why this conjecture is incorrect illustrates an important insight of our approach. In an economy where the key distributional conflict is between different social groups, these groups can be rewarded not only when they hold power, but also when they are out of power (and they engage in consumption and production). Consequently, the probability of power switches does not directly affect "effective discount factors," potentially invalidating the insight on which this conjecture is based. Because the persistence of the Markov process for power switches reduces deviation payoffs (while first-best payoffs are independent of persistence), greater persistence makes first-best allocations less likely to be sustainable.

While our analysis focused on the distortions introduced by the political economy friction, it is straightforward to derive implications of these results for tax policy. If the group in power sets taxes and transfers rather than directly deciding allocations, then the Pareto efficient allocation can be decentralized as a competitive equilibrium, but this would necessarily involve the use of distortionary taxes. This observation implies that the fluctuations of distortions, consumption and labor supply levels derived as part of the Pareto efficient allocations in this paper also correspond to fluctuations in taxes—not simply to the presence of and fluctuations in "wedges" between the marginal utility of consumption and the marginal disutility of labor. The result that distortionary taxes must be used to decentralize the Pareto efficient allocation has a simple intuition, further clarifying the source of distortions in our economy: distortionary taxes must be used in order to discourage labor supply, because greater labor supply would increase the amount of output at the group in power can expropriate, tightening its sustainability constraint. Starting from an undistorted allocation, a small increase in taxes (distortions) would have a second-order cost in terms of lost net output, while having a first-order benefit in terms of relaxing the sustainability constraint when the latter is binding (see also Acemoglu, Golosov and Tsyvinski [4]).

We believe that the framework studied here is attractive both because we can analyze the effect of political economy distortions without specifying all of the details of interactions between groups and the process of decision-making. Undoubtedly, these institutional details are important in practice, and may lead to the emergence of outcomes inside the constrained efficient Pareto frontier. A natural next step is then to investigate what types of institutional structures can support ("implement") the constrained Pareto efficient allocations. This would give a different perspective on the role of specific institutions, as potential tools regulating the allocation of political power in society and placing constraints on the exercise of such power so as to achieve constrained Pareto efficient allocations. Nevertheless, our results indicating that changes in the frequency of power switches that improve ex-ante welfare do not necessarily improve ex-post welfare for all groups suggest that even when such specific institutions implementing constrained Pareto efficient allocations exist, whether they will emerge in equilibrium needs to be studied in the context of well-specified models. We leave an investigation of these issues to future work.

Another important area for future research is to endogenize the Markov process for power

switches. In modern societies, fluctuations of political power between different groups arise because of electoral competition, possible political coalitions between different groups lending their support to a specific party or group, or in extreme circumstances, because different groups can use their *de facto* power, such as in revolutions or in civil wars, to gain *de jure* power (see, e.g., Acemoglu and Robinson, [8]).

8 Appendix

In this Appendix, we provide some of the technical details, results, and proofs omitted from the text.

8.1 Proof of Proposition 1

The next lemma describes the worst subgame perfect equilibrium. In that equilibrium, all parties that are not in power in any given period supply zero labor and receive zero consumption, while the party in power supplies labor and consumes all output to maximize its per period utility in such a way that marginal utility from consumption is equated with marginal disutility of labor.

Lemma 1 *Suppose Assumption 1 holds. The worst SPE is given by the collection of strategies* σ^W *such that for all* $j \neq \mathbf{j}\left(h^t\right)$*:* $l_j\left(\omega^t\right) = 0$ *for all* $\omega^t \in \Omega^t$*, and for* $j' = \mathbf{j}\left(\omega^t\right)$*:* $l_{j'}\left(h^t\right) = \tilde{l}_{j'}$ *for all* $\omega^t \in \Omega^t$ *where* $\tilde{l}_{j'}$ *is a solution to*

$$\frac{\partial u_{j'}(\tilde{l}_{j'}, \tilde{l}_{j'})}{\partial c} = -\frac{\partial u_{j'}(\tilde{l}_{j'}, \tilde{l}_{j'})}{\partial l} \tag{20}$$

and $c_j^t\left(\hat{\omega}^t\right) = 0$ *for* $j \neq j', c_{j'}^t\left(\hat{\omega}^t\right) = \sum_{i=1}^N l_i^t\left(\hat{\omega}^t\right)$ *for all* $\hat{\omega}^t \in \hat{\Omega}^t$*.*

Proof. We first show that σ^W is a best response for each party in all subgames when other parties are playing σ^W. Consider first party j that is not in power (i.e., suppose that party $j' \neq j$ is in power) at history ω^t. Consider strategy σ_j for party j that deviates from σ_j^W at time t, and then coincides with σ_j^W at all subsequent dates (following all histories). By the one step ahead deviation principle, if σ_j^W is not a best response for party j, then there exists such a strategy σ_j that will give higher utility to this party. Note, first that given σ_{-j}^W, for any σ_j, party j will always receive zero consumption, and moreover under σ^W, this has no effect on the continuation value of party j. Therefore, at such a history, we have

$$\begin{aligned} U_j\left(\sigma_j, \sigma_{-j}^W \mid \omega^t\right) &= u_j(0, l_{j,t}) + \beta \mathbb{E}\left[U_j\left(\sigma_j, \sigma_{-j}^W \mid \omega^{t+1}\right) \mid \omega^t\right] \\ &\leq \beta \mathbb{E}\left[U_j\left(\sigma_j^W, \sigma_{-j}^W \mid \omega^{t+1}\right) \mid \omega^t\right], \end{aligned}$$

for any such σ_j, where $l_{j,t}$ is the labor supply of party j at time t implied by the alternative strategy σ_j, and $\mathbb{E}\left[U_j\left(\sigma_j, \sigma_{-j}^W \mid \omega^{t+1}\right) \mid \omega^t\right]$ is the continuation value of this party from date $t+$

1 onwards, with the expectation taken over histories determining power switches given current history ω^t (where conditioning is on history ω^t which is available to party j when making decisions as this party is not in power). The second line follows in view of the fact that since $u_j(0,0)=0$, we have $u_j(0,l_{j,t})\leq 0$, and since under σ^W any change in behavior at t has no effect on future play and σ_j coincides with σ_j^W from time $t+1$ onwards, $\mathbb{E}\left[U_j\left(\sigma_j,\sigma_{-j}^W\mid\omega^{t+1}\right)\mid\omega^t\right]=\mathbb{E}\left[U_j\left(\sigma_j^W,\sigma_{-j}^W\mid\hat{\omega}^{t+1}\right)\mid\omega^t\right]$. This establishes that there is no profitable deviations from σ_j^W for any j not in power.

Next consider party j in power at history $\hat{\omega}^t$ (where we condition on intermediate-state history $\hat{\omega}^t$ as the party in power makes decisions after observing his history). Under σ^W, $l_{j',t}=c_{j',t}=0$ for all $j'\neq j$, and thus $l_{j,t}=c_{j,t}$. Consider again strategy σ_j for party j that deviates from σ_j^W at time t, and then coincides with σ_j^W at all subsequent dates. Then, using similar notation, we have

$$\begin{aligned}U_j\left(\sigma_j,\sigma_{-j}^W\mid\hat{\omega}^t\right) &= u_j(c_{j,t},l_{j,t})+\beta\mathbb{E}\left[U_j\left(\sigma_j,\sigma_{-j}^W\mid\hat{\omega}^{t+1}\right)\mid\hat{\omega}^t\right]\\ &\leq u_j(c_{j,t}^W,l_{j,t}^W)+\beta\mathbb{E}\left[U_j\left(\sigma_j^W,\sigma_{-j}^W\mid\hat{\omega}^{t+1}\right)\mid\hat{\omega}^t\right],\end{aligned}$$

for any such σ_j, where $l_{j,t}$ and $c_{j,t}$ are the labor supply and consumption of party j at this history under strategy σ_j, and $l_{j,t}^W$ and $c_{j,t}^W$ or the labor supply and consumption implied by σ_j^W. The second line follows since σ_j^W satisfies (20), and thus $u_j(c_{j,t}^W,l_{j,t}^W)=u_j(\tilde{l},\tilde{l})\geq u_j(c_{j,t},l_{j,t})$ for any σ_j, and again because $\mathbb{E}\left[U_j\left(\sigma_j,\sigma_{-j}^W\mid\hat{\omega}^{t+1}\right)\mid\hat{\omega}^t\right]=\mathbb{E}\left[U_j\left(\sigma_j^W,\sigma_{-j}^W\mid\hat{\omega}^{t+1}\right)\mid\hat{\omega}^t\right]$ (from the fact that under σ^W the current deviation by party j has no effect on future play and σ_j coincides with σ_j^W from time $t+1$ onwards). This establishes that there is no profitable deviations from σ_j^W for the party in power. Therefore, σ^W is a SPE. The proof is completed by showing that σ^W is also the worst SPE for any party j. To see this, suppose that all $j'\neq j$ choose strategy σ_{-j}^M to minimize the payoff of j. Since power switches are exogenous, party j can guarantee itself $u_j(\tilde{l},\tilde{l})$ whenever it is in power and $u_j(c_{j,t}^M,l_{j,t}^M)=0$ whenever it is not in power. Therefore,

$$U_j\left(\sigma_j^W,\sigma_{-j}^M\mid\hat{\omega}^t\right)\geq U_j\left(\sigma_j^W,\sigma_{-j}^W\mid\hat{\omega}^t\right)$$

for any σ_{-j}^M, and thus σ^W is the worst SPE (i.e., it involves all groups other than the one in power supplying zero labor and thus minimizing the utility of the group in power). Moreover, by the same argument σ^W is also the worst equilibrium for all parties, completing the proof.

■

To prove the proposition, we next show that (9)-(11) are necessary and sufficient conditions for any allocation $\{c_j(h), l_j(h)\}_{j=1}^N$ that is an outcome of some SPE. First, we show that any allocation $\{c_j(h), l_j(h)\}_{j=1}^N$ that satisfies (9)-(11) is an outcome of some SPE. For any history ω^t with $h^t \in \omega^t$ let $\sigma^*(\omega^t) = \left\{l_j\left(h^t\right)\right\}_{j=1}^N$ if $\omega^t = \left(h^t, \left(\left\{c_j\left(h^{t-s}\right), l_j\left(h^{t-s}\right)\right\}_{j=1}^N\right)_{s=1}^t\right)$, and $\sigma^*\left(\omega^t\right) = \sigma^W$ otherwise, and, analogously, $\sigma^*(\hat{\omega}^t) = \left\{c_j\left(h^t\right)\right\}_{j=1}^N$ if $\hat{\omega}^t = \left(h^t, \left(\left\{c_j\left(h^{t-s}\right), l_j\left(h^{t-s}\right)\right\}_{j=1}^N\right)_{s=1}^t\right)$, and $\sigma^*\left(\hat{\omega}^t\right) = \sigma^W$ otherwise. For any $j \neq \mathbf{j}(h^t)$ if $\sigma_j(\omega^t) \neq \sigma_j^*(\omega^t)$,

$$\begin{aligned} U_j(\sigma_j, \sigma_{-j}^* | \hat{\omega}^t) &\leq u_j(0,0) + \beta V_j^W(\mathbf{j}(h^t)) \\ &\leq U_j(\sigma_j^*, \sigma_{-j}^* | \hat{\omega}^t), \end{aligned}$$

where the last inequality follows from (10) and from the fact that $u_j(0,0) = 0$. Moreover, for $j' = \mathbf{j}(h^t)$, any $\sigma_{j'} \neq \sigma_{j'}^*$ implies

$$\begin{aligned} U'_{j'}(\sigma_{j'}, \sigma_{-j'}^* | \hat{\omega}^t) &\leq \max_{\tilde{l} \geq 0} u_{j'}\left(\sum_{j \neq j'} l_j\left(h^t\right) + \tilde{l}\,, \tilde{l}\right) + \beta V_{j'}^W(j') \\ &\leq U_{j'}(\sigma_{j'}^*, \sigma_{-j'}^* | \hat{\omega}^t). \end{aligned}$$

Therefore, σ^* is an equilibrium.

The necessity of (9)-(11) is straightforward. Condition (9) is feasibility constraint. In any equilibrium σ^*, we have, for $j \neq \mathbf{j}(h^t)$, that

$$U_j(\sigma_j^*, \sigma_{-j}^* | \omega^t) \geq U_j(\sigma_j, \sigma_{-j}^* | \omega^t)$$

for all $\sigma_j \neq \sigma_j^*$ (where conditioning is on history ω^t which is available to $j \neq \mathbf{j}(h^t)$ when making decisions). This implies:

$$\begin{aligned} U_j(\sigma_j^*, \sigma_{-j}^* | \omega^t) &\geq u_j(c_j, l_j) + \beta \mathbb{E}_t U_j(\sigma_j, \sigma_{-j}^* | \omega^{t+1}) \\ &\geq u_j(0, l_j) + \beta V_j^W(\mathbf{j}(h^t)) \\ &= u_j(0,0) + \beta V_j^W(\mathbf{j}(h^t)). \end{aligned}$$

Since $u_j(0,0)=0$, this string of inequalities yields (10). Similarly, for $j'=\mathbf{j}(h^t)$, we have that

$$U_{j'}(\sigma^*_{j'},\sigma^*_{-j'}|\hat{\omega}^t) \geq \max_{\tilde{l}\in[0,\bar{l}]} u_{j'}\left(\sum_{j\neq j'} l_j\left(h^t\right)+\tilde{l}\ ,\tilde{l}\right)+\beta\mathbb{E}_t U_j(\sigma_{j'},\sigma^*_{-j'}|\hat{\omega}^{t+1})$$
$$\geq \max_{\tilde{l}\in[0,\bar{l}]} u_{j'}\left(\sum_{j\neq j'} l_j\left(h^t\right)+\tilde{l}\ ,\tilde{l}\right)+\beta V^W_{j'}(j')$$

for all $\sigma_j\neq\sigma^*_j$, which gives condition (11). (Here conditioning is on $\hat{\omega}^t$ which is the intermediate-state history available when $j'=\mathbf{j}(h^t)$ makes decisions; in what follows we condition on $\hat{\omega}^t$ since it subsumes ω^t).

To see that σ^* is a (constrained) Pareto efficient equilibrium, suppose there is any other equilibrium σ^{**} that Pareto dominates σ^*. Since σ^{**} is a SPE, the outcome of σ^{**} must satisfy (9)-(11). But then the value of (8) would be higher under the outcome of σ^{**} than under σ^*, yielding a contradiction and completing the proof. ■

Recursive Characterization

Here we present a characterization result, which shows that the solution to (8)-(11), the maximization problem in Proposition 1, can be represented recursively. This characterization is used in the rest of the Appendix.

Let us first define $M\left(h^{t+s}\mid h^t\right)$ to be the (conditional) probability of history h^{t+s} at time $t+s$ given history h^t at time t according to the Markov process $m(j\mid j')$. We will show that after history h^{t-1}, (8)-(11) can be represented by the following Lagrangian:

$$\max_{\{c_j(h^s),l_j(h^s)\}_{j=1,..,N;\ h^s,s\geq t}} \mathcal{L}=\sum_{s=t}^{\infty}\sum_{h^s}\beta^s M\left(h^s\mid h^t\right)\left[\sum_{j=1}^{N}\left(\alpha_j+\mu_j\left(h^{t-1}\right)\right)u_j\left(c_j\left(h^s\right),l_j\left(h^s\right)\right)\right] \tag{21}$$
$$+\sum_{s=t}^{\infty}\sum_{h^s}\beta^s M\left(h^s\mid h^t\right)\sum_{j=1}^{N}\lambda_j\left(h^s\right)\times$$
$$\left(\sum_{s'=s}^{\infty}\sum_{h^{s'}}\beta^{s'-s}M\left(h^{s'}\mid h^s\right)u_j\left(c_j\left(h^{s'}\right),l_j\left(h^{s'}\right)\right)-\mathbf{I}_{j=\mathbf{j}(h^s)}v_{\mathbf{j}(h^s)}\left(\sum_{j'\neq\mathbf{j}(h^s)}l_{j'}\left(h^s\right)\right)-\beta V^W_j(\mathbf{j}\left(h^s\right))\right),$$

subject to (9), with μ_j's defined recursively as:

$$\mu_j\left(h^t\right)=\mu_j\left(h^{t-1}\right)+\lambda_j\left(h^t\right)$$

with the normalization $\mu_j\left(h^{\varnothing}\right)=0$ for $j=1,...,N$. The most important implication of the formulation in (21) is that for any h^t following h^{t-1}, the numbers

$$\alpha_j\left(h^{t-1}\right)\equiv\frac{\alpha_j+\mu_j\left(h^{t-1}\right)}{\sum_{j'=1}^{N}\left(\alpha_{j'}+\mu_{j'}\left(h^{t-1}\right)\right)}\text{ (for each } j=1,...,N) \tag{22}$$

can be interpreted as *updated Pareto weights*. Therefore, after history h^{t-1}, the problem is equivalent to maximizing the sum of utilities with these weights (subject to the relevant constraints). The problem of maximizing (21) is equivalent to choosing current consumption and labor supply levels for each group and also updated Pareto weights $\{\alpha_j\}_{j=1}^{N}$. This analysis establishes the following characterization result:

Lemma 2 *Suppose Assumptions 1 and 2 hold. Then the efficient allocation has a quasi-Markovian structure whereby consumption and labor allocations* $\left\{c_j\left(h^t\right),l_j\left(h^t\right)\right\}_{j=1,..,N;\ h^t}$ *depend only on* $\mathbf{s}\equiv\left(\left\{\alpha_j\left(h^{t-1}\right)\right\}_{j=1}^{N},\mathbf{j}\left(h^t\right)\right)$*, i.e., only on updated weights and the identity of the group in power.*

This recursive characterization implies that we can express $V_j(h^{t-1})$ and $V_j[h^{t-1},i]$ as $V_j(\boldsymbol{\alpha})$ and $V_j[\boldsymbol{\alpha},i]$.

Proof. The proof of this proposition builds on the representation suggested by Marcet and Marimon [30]. First note that the maximization problem (8)-(11), given in Proposition 1, can be written in Lagrangian form as follows (recalling that $\mathbf{j}\left(h^t\right)$ denotes the party in power at time t according to history h^t):

$$\begin{aligned}\max_{\{c_j(h^t),l_j(h^t)\}_{j=1,..,N;\ h^t}}\mathcal{L}&=\sum_{t=0}^{\infty}\sum_{h^t}\beta^t M\left(h^t\mid h^{\varnothing}\right)\left[\sum_{j=1}^{N}\alpha_j u_j\left(c_j\left(h^t\right),l_j\left(h^t\right)\right)\right]\\&+\sum_{t=0}^{\infty}\sum_{h^t}\beta^t M\left(h^t\mid h^{\varnothing}\right)\sum_{j=1}^{N}\lambda_j\left(h^t\right)\\&\times\left[\sum_{s=t}^{\infty}\beta^{s-t}\sum_{h^s}M\left(h^s\mid h^t\right)u_j\left(c_j\left(h^s\right),l_j\left(h^s\right)\right)-\mathbf{I}_{j=\mathbf{j}(h^t)}v_{\mathbf{j}(h^t)}\left(\sum_{j'\neq\mathbf{j}(h^t)}l_{j'}\left(h^t\right)\right)-\beta V_j^W(\mathbf{j}\left(h^t\right))\right]\end{aligned} \tag{23}$$

subject to (9). Here for $j=\mathbf{j}\left(h^t\right)$, $\beta^t M\left(h^t\mid h^{\varnothing}\right)\lambda_j\left(h^t\right)$ is the Lagrange multiplier on the sustainability constraint, (11) (party $\mathbf{j}\left(h^t\right)$ is the one in power at time t following history h^t) and for $j'\neq\mathbf{j}\left(h^t\right)$, $\beta^t M\left(h^t\mid h^{\varnothing}\right)\lambda_{j'}\left(h^t\right)$ is the Lagrange multiplier on the participation constraints of the parties that are not in power in history h^t, (10). $\mathbf{I}_{j=\mathbf{j}(h^t)}$ is an indicator

variable that takes the value 1 if $j = \mathbf{j}\left(h^t\right)$ and 0 otherwise, thus ensuring that the term $v_{\mathbf{j}(h^t)}\left(\sum_{j\neq\mathbf{j}(h^t)} l_j\left(h^t\right)\right)$ is only present when we consider group $\mathbf{j}\left(h^t\right)$. For any $T \geq 0$, we have

$$\begin{aligned} & \sum_{s=0}^{T}\sum_{h^s}\beta^s M\left(h^s \mid h^{\varnothing}\right)\lambda_j\left(h^s\right)\sum_{s'=s}^{T}\sum_{h^{s'}}\beta^{s'-s}M\left(h^{s'} \mid h^s\right)u_j\left(c_j\left(h^{s'}\right), l_j\left(h^{s'}\right)\right) \quad (24) \\ = & \sum_{s=0}^{T}\sum_{h^s}\beta^s M\left(h^s \mid h^{\varnothing}\right)\mu_j\left(h^s\right)u_j\left(c_j\left(h^s\right), l_j\left(h^s\right)\right), \end{aligned}$$

where $\mu_j\left(h^s\right) = \mu_j\left(h^{s-1}\right) + \lambda_j\left(h^s\right)$ for $h^s \in P\left(h^{s-1}\right)$ with the initial $\mu_j\left(h^{\varnothing}\right) = 0$ for all j. Substituting (24) in $\mathcal{L}$ in (23) and noting that after history h^{t-1} has elapsed, all terms preceding this history are given, we obtain (21). The result that optimal allocations only depend on $\left\{\alpha_j\left(h^{t-1}\right)\right\}_{j=1}^{N}$ and $\mathbf{j}\left(h^t\right)$ then follows immediately. ■

8.2 Proof of Proposition 2

Let $\zeta\left(h^t\right)$ be the Lagrange multiplier on (9), let $\mathbf{I}_{h^s\in h^t}$ be the indicator variable that take the value 1 if $h^t \in P(h^s)$ and 0 otherwise, and recall that the multiplier on (11) following history h^t with party $\mathbf{j}\left(h^t\right)$ in power is $\beta^t M\left(h^t \mid h^{\varnothing}\right)\lambda_{\mathbf{j}(h^t)}\left(h^t\right)$. Then under Assumptions 1 and 2, the (constrained) Pareto efficient allocation satisfies the following first-order conditions for any h^t:

$$\beta^t M\left(h^t \mid h^{\varnothing}\right)\left[\alpha_j + \sum_{s=0}^{t}\sum_{h^s}\mathbf{I}_{h^s\in h^t}\lambda_j\left(h^s\right)\right]\frac{\partial u_j\left(c_j\left(h^t\right), l_j\left(h^t\right)\right)}{\partial c} = \zeta\left(h^t\right) \text{ for all } j,$$

$$\begin{aligned} & \beta^t M\left(h^t \mid h^{\varnothing}\right)\left[\alpha_j + \sum_{s=0}^{t}\sum_{h^s}\mathbf{I}_{h^s\in h^t}\lambda_j\left(h^s\right)\right]\frac{\partial u_j\left(c_j\left(h^t\right), l_j\left(h^t\right)\right)}{\partial l} \\ = & -\zeta\left(h^t\right) + \beta^t M\left(h^t \mid h^{\varnothing}\right)\lambda_{\mathbf{j}(h^t)}\left(h^t\right)v'_{\mathbf{j}(h^t)}\left(\sum_{j'\neq\mathbf{j}(h^t)}^{N} l_{j'}\left(h^t\right)\right) \quad \text{for } j \neq \mathbf{j}\left(h^t\right), \end{aligned}$$

and

$$\beta^t M\left(h^t \mid h^{\varnothing}\right)\left[\alpha_j + \sum_{s=0}^{t}\sum_{h^s}\mathbf{I}_{h^s\in h^t}\lambda_j\left(h^s\right)\right]\frac{\partial u_j\left(c_j\left(h^t\right), l_j\left(h^t\right)\right)}{\partial l} = -\zeta\left(h^t\right) \text{ for } j = \mathbf{j}\left(h^t\right).$$

Combining these conditions establishes the proposition.■

8.3 Proof of Theorem 1

Part (a): A sustainable first-best allocation exists if there exists a feasible vector of consumption and labor supply levels, $\left\{c_j^*, l_j^*\right\}_j$, that satisfies (13) and (14) for each $j = 1, ..., N$. Clearly, as $\beta \to 0$, such sustainable first-best allocations exist. Conversely, as $\beta \to 0$, they do not, as the group in power will necessarily benefit by deviating. Since (13) and (14) define a closed set, this implies that there exists $\bar{\beta} \in (0, 1)$ such that for $\beta \geq \bar{\beta}$, some first-best allocations are sustainable and for $\beta < \bar{\beta}$, there are no sustainable first-best allocations.

Part (b): Recall that $\mathbf{s} \equiv (\boldsymbol{\alpha}, j)$, where $\boldsymbol{\alpha} \in \Delta^{N-1}$ are ex ante weights and $j \in \mathcal{N} \equiv \{1, ..., N\}$. Let $S = \Delta^{N-1} \times \mathcal{N}$ and $\mathcal{S}$ be the Borel σ-algebra over S.

Our model implies that each s is deterministically mapped into $\alpha' = h(s)$. Thus the stochastic process for $\mathbf{s}'$ is determined uniquely by the Markov process $m(j' \mid j)$. Therefore we can define

$$p\left((\boldsymbol{\alpha}, j), (\boldsymbol{\alpha}', j')\right) = \begin{cases} 0 & \text{if } \boldsymbol{\alpha}' \neq h(\boldsymbol{\alpha}, j) \\ m(j' \mid j) & \text{if } \boldsymbol{\alpha}' = h(\boldsymbol{\alpha}, j) \end{cases}$$

Note that $p((\boldsymbol{\alpha}, j), (\boldsymbol{\alpha}', j'))$ is uniformly bounded above by $\max_{j,j' \in \mathcal{N}} m(j' \mid j) < 1$ (the latter inequality by Assumption 2). Therefore, for all $s \in S$ and $A \in \mathcal{S}$, we have

$$P((\boldsymbol{\alpha}, j), A) = \int_A p\left((\boldsymbol{\alpha}, j), (\boldsymbol{\alpha}', j')\right) \mu\left(d(\boldsymbol{\alpha}', j')\right).$$

Stokey, Lucas and Prescott [42], Exercise 11.4f (which is straightforward to see) shows that if there exists a bounded above function p and and a finite measure μ such that

$$P(s, A) = \int_A p(s, s') \mu(ds')$$

for all $s \in S$ and $A \in \mathcal{S}$, then Doeblin's condition is satisfied. Almost sure convergence to an invariant limiting distribution then follows.

Since $s\left(h^t\right)$ converges to an invariant distribution, so do $\left\{c_j\left(h^t\right), l_j\left(h^t\right)\right\}_{j=1}^N$. It remains to show that invariant distribution over $\left\{c_j\left(h^t\right), l_j\left(h^t\right)\right\}_{j=1}^N$ is non-degenerate. Suppose, to obtain a contradiction, that it is degenerate, say given by $\left\{c_j^*, l_j^*\right\}_{j=1}^N$. First, suppose that (13) does not bind in the invariant distribution. From Proposition 2, $\left\{c_j^*, l_j^*\right\}_{j=1}^N$ must satisfy

$$\frac{\partial u_j\left(c_j^*, l_j^*\right)}{\partial c} = -\frac{\partial u_j\left(c_j^*, l_j^*\right)}{\partial l} \text{ for all } j.$$

Since in any Pareto efficient perfect equibrium feasibility constraint (9) must hold with equality and $\left\{c_j^*, l_j^*\right\}_{j=1}^N$ are time invariant, the allocation $\left\{c_j^*, l_j^*\right\}_{j=1}^N$ must be a first-best allocation, which is a contradiction. Therefore (13) must bind for some j, and thus from Proposition 2

$$\frac{\partial u_j\left(c_j^*, l_j^*\right)}{\partial c} > -\frac{\partial u_j\left(c_j^*, l_j^*\right)}{\partial l}$$

when j is in not power. However when j is in power, Proposition 2 implies

$$\frac{\partial u_j\left(c_j^*, l_j^*\right)}{\partial c} = -\frac{\partial u_j\left(c_j^*, l_j^*\right)}{\partial l}$$

which is impossibility. Therefore, $\left\{c_j\left(h^t\right), l_j\left(h^t\right)\right\}_{j=1}^N$ converges to nondegenerate distribution. ■

Proof of Theorem 2

We now state and prove three lemmas, which together will enable us to establish our main result in this section Theorem 2.

We first show that the party with a higher Pareto weight will receive higher value.

Lemma 3 *For any two vectors of Pareto weights $\boldsymbol{\alpha}, \boldsymbol{\alpha}'$, if $\alpha_i > \alpha_i'$, then $V_i[\boldsymbol{\alpha}, j] \geq V_i[\boldsymbol{\alpha}', j]$ for $j \in \{1, 2\}$.*

Proof. Without loss of generality, let $i = 1$. Then constrained Pareto efficiency implies

$$\alpha_1 V_1[\boldsymbol{\alpha}, j] + \alpha_2 V_2[\boldsymbol{\alpha}, j] \geq \alpha_1 V_1\left[\boldsymbol{\alpha}', j\right] + \alpha_2 V_2\left[\boldsymbol{\alpha}', j\right]$$

and

$$\alpha_1' V_1\left[\boldsymbol{\alpha}', j\right] + \alpha_2' V_2\left[\boldsymbol{\alpha}', j\right] \geq \alpha_1' V_1[\boldsymbol{\alpha}, j] + \alpha_2' V_2[\boldsymbol{\alpha}, j].$$

These conditions then imply

$$\left(\alpha_1 - \alpha_1'\right)\left(V_1[\boldsymbol{\alpha}, j] - V_2[\boldsymbol{\alpha}, j]\right) \geq \left(\alpha_1 - \alpha_1'\right)\left(V_1\left[\boldsymbol{\alpha}', j\right] - V_2\left[\boldsymbol{\alpha}', j\right]\right)$$

or

$$V_1[\boldsymbol{\alpha}, j] - V_1\left[\boldsymbol{\alpha}', j\right] \geq V_2[\boldsymbol{\alpha}, j] - V_2\left[\boldsymbol{\alpha}', j\right]. \tag{25}$$

Suppose that $V_1[\boldsymbol{\alpha}, j] < V_1[\boldsymbol{\alpha}', j]$. Then from (25), $V_2[\boldsymbol{\alpha}, j] < V_2[\boldsymbol{\alpha}', j]$, which is impossible since otherwise $(V_1[\boldsymbol{\alpha}', j], V_2[\boldsymbol{\alpha}', j])$ would Pareto dominate $(V_1[\boldsymbol{\alpha}, j], V_2[\boldsymbol{\alpha}, j])$, thus establishing the lemma. ■

Let $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*)$ be a vector of the Pareto weights for which first-best allocation is sustainable. Consider any other initial vector $\boldsymbol{\alpha}_0 \neq \boldsymbol{\alpha}^*$and suppose that the first-best allocation that corresponds to that vector is not sustainable. This implies that at least for one party the participation constraint (10) or the sustainability constraint (11) binds. The next lemma shows that sustainability constraint (11) does not bind if any party has a Pareto weight higher than α_j^* and that there exists $\bar{\alpha}_j \in (\alpha_j^*, 1]$ such that if $\alpha_j \leq \bar{\alpha}_j$, then the participation constraint, (10), of the other party also does not bind, so when $\alpha_j \in [\alpha_j^*, \bar{\alpha}_j]$, Pareto weights to not change.

Lemma 4 *Suppose Assumptions 1 and 3 hold. If* $\alpha_j\left(h^{t-1}\right) \geq \alpha_j^*$ *for some* j, h^{t-1} *and* $j = \mathbf{j}\left(h^t\right)$, *then* $\lambda_j\left(h^t\right) = 0$. *Moreover, under these hypotheses there exists* $\bar{\alpha}_j \in (\alpha_j^*, 1]$ *such that if* $\alpha_j\left(h^{t-1}\right) > \bar{\alpha}_j$, *then* $\alpha_j\left(h^t\right) = \bar{\alpha}_j$ *and if* $\alpha_j\left(h^{t-1}\right) \in \left[\alpha_j^*, \bar{\alpha}_j\right]$, *then* $\alpha_j\left(h^t\right) = \alpha_j\left(h^{t-1}\right)$.

Proof. Without loss of generality assume that $j = 1$. By Lemma 3, $V_2\left[(\alpha_1, 1-\alpha_1), 1\right]$ is decreasing in α_1. If $V_2\left[(1,0),1\right] \geq \beta V_2^W(1)$, set $\bar{\alpha}_1 = 1$, otherwise let $\bar{\alpha}_1$ be defined by $V_2\left[(\bar{\alpha}_1, 1-\bar{\alpha}_1), 1\right] = \beta V_2^W(1)$. Since $V_2\left[(\alpha_1, 1-\alpha_1), 1\right]$ is continuous and monotone in α_1 and $V_2\left[(\alpha_1^*, 1-\alpha_1^*), 1\right] > \beta V_2^W(1)$, such $\bar{\alpha}_1$ exists and $\bar{\alpha}_1 \in (\alpha_1^*, 1]$. Since for any α_1, (10) implies that $V_2\left[(\alpha_1, 1-\alpha_1), 1\right] \geq \beta V_2^W(1)$ and $V_2\left[(\alpha_1, 1-\alpha_1), 1\right]$ is decreasing in α_1, this construction implies that

$$V_2\left[(\alpha_1, 1-\alpha_1), 1\right] = V_2\left[(\bar{\alpha}_1, 1-\bar{\alpha}_1), 1\right] \text{ for all } \alpha_1 \geq \bar{\alpha}_1. \tag{26}$$

Since the equilibrium is (constrained) Pareto efficient, (26) implies that if $\alpha_1(h^{t-1}) \geq \bar{\alpha}_1$, then

$$V_1\left[\left(\alpha_1(h^{t-1}), 1-\alpha_1(h^{t-1})\right), 1\right] = V_1\left[(\bar{\alpha}_1, 1-\bar{\alpha}_1), 1\right]$$

and therefore the equilibrium for state $\mathbf{s} = \left(\left(\alpha_1(h^{t-1}), 1-\alpha_1(h^{t-1})\right), 1\right)$ coincides with the equilibrium for state $\bar{\mathbf{s}}_1 = \left((\bar{\alpha}_1, 1-\bar{\alpha}_1), 1\right)$.

Suppose $\alpha_1\left(h^{t-1}\right) \in [\alpha_1^*, \bar{\alpha}_1]$. Let us consider the relaxed problem of maximizing (8) without the constraints (10) and (11) following history h^t. We will characterize the solution to this relaxed problem and then show that the solution in fact satisfies (10) and (11) establishing that $\lambda_i\left(h^t\right) = 0$ for $i \in \{1,2\}$ and $\alpha_j\left(h^t\right) = \alpha_j\left(h^{t-1}\right)$.

The expected utility of party 1 in history h^t in the relaxed problem is

$$u_1\left(c_1\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right), l_1\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right)\right) + \beta\left(m(1|1)V_1\left[\boldsymbol{\alpha}\left(h^{t-1}\right), 1\right] + m(2|1)V_1\left[\boldsymbol{\alpha}\left(h^{t-1}\right), 2\right]\right)$$

where $\left(c_i\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right), l_i\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right)\right)$ is a solution to the maximization problem

$$\max_{\{c_i,l_i\}_{i=1,2}} \alpha_1\left(h^{t-1}\right) u_1\left(c_1, l_1\right) + \alpha_2\left(h^{t-1}\right) u_2\left(c_2, l_2\right)$$

subject to

$$c_1 + c_2 \leq l_1 + l_2.$$

Since there is no sustainability constraint, Assumption 3 immediately implies that $l_j\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right) = 1$ for all j, and moreover, $u_1\left(c_1\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right), l_1\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right)\right)$ is increasing in $\alpha_1\left(h^{t-1}\right)$.

Since Pareto weights $\boldsymbol{\alpha}^*$ correspond to the sustainable allocation,

$$\begin{aligned} u_1\left(c_1\left(\boldsymbol{\alpha}^*\right), l_1\left(\boldsymbol{\alpha}^*\right)\right) + \beta\left(m(1|1)V_1\left[\boldsymbol{\alpha}^*, 1\right] + m(2|1)V_1\left[\boldsymbol{\alpha}^*, 2\right]\right) \\ \geq v_1\left(l_2\left(\boldsymbol{\alpha}^*\right)\right) + \beta V_1^W(1) \end{aligned} \tag{27}$$

Once again, Assumption 3 implies that $l_j\left(\boldsymbol{\alpha}^*\right) = 1$ for all j. From Lemma 3, $V_1\left[\boldsymbol{\alpha}\left(h^{t-1}\right), j\right] \geq V_1\left[\boldsymbol{\alpha}^*, j\right]$ for all j. Therefore the solution to the relaxed problem satisfies (11) if

$$u_1\left(c_1\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right), 1\right) - v_1(1) \geq u_1\left(c_1\left(\boldsymbol{\alpha}^*\right), 1\right) - v_1(1)$$

Since $u_1\left(c_1\left(\boldsymbol{\alpha}\left(h^{t-1}\right)\right), 1\right)$ is increasing in $\alpha_1\left(h^{t-1}\right)$ and $\alpha_1\left(h^{t-1}\right) \geq \alpha_1^*$, this inequality is satisfied. The solution to the relaxed problem also satisfies (10) because $\alpha_1\left(h^{t-1}\right) \leq \bar{\alpha}_1$. ■

The previous lemma established that if party j is in power and has an updated Pareto weight above α_j^*, its next period updated Pareto weight remains above α_j^*. The next key step in our argument is to show that if a party has Pareto weight is below α_j^*, its next period updated Pareto weight is also below α_j^* (even if its current sustainability constraint is binding).

The next lemma is the key to the main result in this section. It shows that if the sustainability constraint does not hold for group j that is in power even though its Pareto weight is below α_j^*, then for all subsequent histories its Pareto weight will not exceed α_j^*. The proof utilizes quasi-linearity of preferences to put structure on the behavior of updated Pareto weights and the corresponding allocations.

Lemma 5 *Suppose Assumptions 1 and 3 hold. Suppose that $\alpha_j\left(h^{t-1}\right) < \alpha_j^*$ for some j, h^{t-1} and $j = \mathbf{j}\left(h^t\right)$. Then $\alpha_j\left(h^t\right) \leq \alpha_j^*$ for all subsequent h^t.*

Proof: Without loss of generality assume that $j = 1$. First note that if $\alpha_1(h^{t-1}) \leq \alpha_1^*$, then constraint (10) cannot bind for party 2. This is true because from Lemma 3 and the sustainability of (c_2^*, l_2^*), we have

$$V_2\left[\boldsymbol{\alpha}\left(h^{t-1}\right), 1\right] \geq V_2\left[\boldsymbol{\alpha}^*, 1\right] \geq \beta V_2^W(1).$$

If constraint (11) also does not bind, the result follows immediately. Suppose therefore that (11) binds. Then the Lagrange multiplier $\lambda_1\left(h^t\right) > 0$ and from (22), $\alpha_1\left(h^t\right) = \left(\alpha_1\left(h^{t-1}\right) + \lambda_1\left(h^t\right)\right) / \left(1 + \lambda_1\left(h^t\right)\right)$. Suppose, to obtain a contradiction, that $\alpha_1\left(h^t\right) > \alpha_1^*$. This implies:

$$\begin{aligned} &u_1\left(c_1\left(\boldsymbol{\alpha}\left(h^{t-1}\right), \lambda_1\left(h^t\right)\right), l_1\left(\boldsymbol{\alpha}\left(h^{t-1}\right), \lambda_1\left(h^t\right)\right)\right) - v_1\left(l_2\left(\boldsymbol{\alpha}\left(h^{t-1}\right), \lambda_1\left(h^t\right)\right)\right) \\ &+ \beta\left(m(1|1)V_1\left[\boldsymbol{\alpha}\left(h^t\right), 1\right] + m(2|1)V_1\left[\boldsymbol{\alpha}\left(h^t\right), 2\right]\right) = \beta V_1^W(1), \end{aligned} \tag{28}$$

where $\left\{c_i\left(\boldsymbol{\alpha}, \lambda\right), l_i\left(\boldsymbol{\alpha}, \lambda\right)\right\}_{i=1}^2$ is a solution to the maximization problem

$$\max_{\{c_i, l_i\}_{i=1,2}} \left(\alpha_1 + \lambda\right) u_1\left(c_1, l_1\right) + \alpha_2 u_2\left(c_2, l_2\right) - \lambda v_1\left(l_2\right)$$

subject to

$$c_1 + c_2 \leq l_1 + l_2.$$

Lemma 3 establishes that $V_1\left[\boldsymbol{\alpha}\left(h^t\right), j\right] \geq V_1\left[\boldsymbol{\alpha}^*, j\right]$ for all j. If for any λ such that $\left(\alpha_1 + \lambda\right) / \left(1 + \lambda\right) > \alpha_1^*$, we have

$$u_1(c_1\left(\boldsymbol{\alpha}, \lambda\right), l_1\left(\boldsymbol{\alpha}, \lambda\right)) - v_1(l_2\left(\boldsymbol{\alpha}, \lambda\right)) > u_1\left(c_1^*, l_1^*\right) - v_1\left(l_2^*\right), \tag{29}$$

then (27) would immediately imply that (28) cannot hold with equality, thus leading to a contradiction and establishing the desired result. Thus to complete the proof of lemma, we will establish that for any Δ such that $\left(\alpha_1 + \Delta\right) / \left(1 + \Delta\right) > \alpha_1^*$, (29) holds, which will thus complete the proof.

We will first prove some intermediate results. Let us adopt the following simpler notation for the rest of the proof:

$$u_{iC}\left(c_i, l_i\right) = \frac{\partial u_i(c_i, l_i)}{\partial c_i} \text{ and } u_{iL}\left(c_i, l_i\right) = \frac{\partial u_i(c_i, l_i)}{\partial l_i},$$

and similarly denote second order derivatives. Sometimes we will also drop (c_i, l_i) and use notation u_{iC}, u_{iL}, etc., whenever there is no confusion.

Claim 1 *Let* $c(\Delta;\boldsymbol{\alpha})$ *and* $l(\Delta,\boldsymbol{\alpha})$ *be the solution to the problem*

$$W(\Delta;\boldsymbol{\alpha}) = \max_{c\geq 0, l\in[0,1]} (\alpha_1+\Delta)\, u_1(c_1,l_1) + \alpha_2 u_2(c_2,l_2) - \Delta v_1(l_2)$$

subject to

$$c_1 + c_2 \leq l_1 + l_2.$$

Then

$$u_1(c_1(\Delta,\boldsymbol{\alpha}), l_1(\Delta,\boldsymbol{\alpha})) - v_1(l_2(\Delta,\boldsymbol{\alpha})) \tag{30}$$

is increasing in Δ.

Proof. Consider Δ' and $\Delta'' \neq \Delta'$, and denote the corresponding solutions to the above problem by $\{c_i', l_i'\}_{i=1,2}$ and $\{c_i'', l_i''\}_{i=1,2}$. By definition, this implies

$$\begin{aligned}
\left(\alpha_1+\Delta'\right) u_1(c_1',l_1') + \alpha_2 u_2(c_2',l_2') - \Delta' v_1(l_2') &\geq \left(\alpha_1+\Delta'\right) u_1(c_1'',l_1'') + \alpha_2 u_2(c_2'',l_2'') - \Delta' v_1(l_2'') \\
\left(\alpha_1+\Delta''\right) u_1(c_1'',l_1'') + \alpha_2 u_2(c_2'',l_2'') - \Delta'' v_1(l_2'') &\geq \left(\alpha_1+\Delta''\right) u_1(c_1',l_1') + \alpha_2 u_2(c_2',l_2') - \Delta'' v_1(l_2')
\end{aligned}$$

Summing these two inequalities in rearranging, we obtain

$$\begin{aligned}
\left(\Delta'-\Delta''\right)\left(u_1(c_1',l_1') - v_1(l_2')\right) &\geq \left(\Delta'-\Delta''\right)\left(u_1(c_1'',l_1'') - v_1(l_2'')\right) \\
\left(\Delta'-\Delta''\right)\left[\left(u_1(c_1',l_1') - v(l_2')\right) - \left(u_1(c_1'',l_1'') - v_1(l_2'')\right)\right] &\geq 0
\end{aligned}$$

Therefore, if $\Delta' > \Delta''$, then $(u_1(c_1',l_1') - v_1(l_2')) \geq (u_1(c_1'',l_1'') - v_1(l_2''))$, which establishes (30). ■

Claim 2 *Suppose Assumptions 1 and 3 hold. Then for any Pareto weight* α^* *and for any* $\Delta \in [0, \alpha_1^*/\alpha_2^*]$, *define* $\alpha_{1\Delta}$ *by*

$$\frac{\alpha_{1\Delta}+\Delta}{1+\Delta} = \alpha_1^*$$

and $\alpha_{2\Delta} = 1 - \alpha_{1\Delta}$. *Consider the maximization problem*

$$\Omega(\Delta) = \max_{c\geq 0, l\in[0,1]} (\alpha_{1\Delta}+\Delta)\, u_1(c_1,l_1) + \alpha_{2\Delta} u_2(c_2,l_2) - \Delta v_1(l_2) \tag{31}$$

subject to

$$c_1 + c_2 \leq l_1 + l_2.$$

Denote the solution to this problem by $\{c_i(\Delta), l_i(\Delta)\}_{i=1,2}$. *Then*

$$F(\Delta) \equiv u_1(c_1(\Delta), l_1(\Delta)) - v_1(l_2(\Delta)) \tag{32}$$

is increasing in Δ. *Moreover, for any* $\Delta > 0$ $F(\Delta) > F(0)$.

Proof. Consider the first-order conditions to the maximization problem in (31). Assumption 3 implies that $l_1(\Delta) = 1$. The other first-order conditions are

$$(\alpha_{1\Delta} + \Delta) u_{1C}(c_1(\Delta) - \eta_1(1)) = \alpha_{2\Delta} u_{2C}(c_2(\Delta) - \eta_2(l_2(\Delta))) \tag{33}$$

$$\frac{\Delta}{\alpha_{2\Delta}} \frac{v_1'(l_2(\Delta))}{u_{2C}(c_2(\Delta) - \eta_2(l_2(\Delta)))} = 1 - \eta_2'(l_2(\Delta)) \tag{34}$$

and

$$c_1(\Delta) + c_2(\Delta) = 1 + l_2(\Delta). \tag{35}$$

First, note that

$$\frac{\alpha_{1\Delta} + \Delta}{\alpha_{2\Delta}} = \frac{\alpha_1^*}{\alpha_2^*} \tag{36}$$

Combining (33) and (36), we obtain

$$\alpha_1^* u_{1C}(c_1(\Delta) - \eta_1(1)) = \alpha_2^* u_{2C}(c_2(\Delta) - \eta_2(l_2(\Delta))) \tag{37}$$

Now differentiate F (as defined in (32)), which yields

$$\begin{aligned} F'(\Delta) &= u_{1C} \frac{\partial c_1}{\partial \Delta} - v_1' \frac{\partial l_2}{\partial \Delta} \\ &= \frac{\partial l_2}{\partial \Delta} \left(u_{1C} \frac{\partial c_1}{\partial l_2} - v_1' \right) \end{aligned} \tag{38}$$

Substituting (35) into (37) and differentiating, we obtain

$$\alpha_1^* u_{1CC} \frac{\partial c_1(\Delta)}{\partial \Delta} = \alpha_2^* u_{2CC} \left(\left(1 - \eta_2'\right) \frac{\partial l_2}{\partial \Delta} - \frac{\partial c_1(\Delta)}{\partial \Delta} \right),$$

which implies

$$\begin{aligned} \frac{\partial c_1}{\partial l_2} &= \frac{\alpha_2^* u_{2CC}(1 - \eta_2')}{\alpha_1^* u_{1CC} + \alpha_2^* u_{2CC}} \\ &= \frac{\alpha_2^* u_{2CC}}{\alpha_1^* u_{1CC} + \alpha_2^* u_{2CC}} \times \frac{\Delta}{\alpha_{2\Delta}} \frac{v_1'}{u_{2C}} \\ &= \frac{\alpha_2^* u_{2CC}}{\alpha_1^* u_{1CC} + \alpha_2^* u_{2CC}} \times \frac{\Delta}{\alpha_{1\Delta} + \Delta} \frac{v_1'}{u_{1C}} \end{aligned}$$

where we used (34) in the second line and (33) in the third. Substituting this into (38), we obtain

$$\begin{aligned} F'(\Delta) &= v_1' \frac{\partial l_2}{\partial \Delta} \left(\frac{\alpha_2^* u_{2CC}}{\alpha_1^* u_{1CC} + \alpha_2^* u_{2CC}} \times \frac{\Delta}{\alpha_{1\Delta} + \Delta} - 1 \right) \\ &= v_1' \frac{\partial l_2}{\partial \Delta} \left(\frac{1}{\frac{\alpha_1^* u_{1CC}}{\alpha_2^* u_{2CC}} + 1} \times \frac{1}{1 + \frac{\alpha_{1\Delta}}{\Delta}} - 1 \right) \end{aligned}$$

The expression in the brackets is negative, therefore $F'(\Delta)$ has the opposite sign of $\frac{\partial l_2}{\partial \Delta}$. The desired result follows from the fact that $\frac{\partial l_2}{\partial \Delta} \leq 0$, which we establish next.

Consider two different Δ', Δ'' and denote the corresponding solutions to (31) by $\{c_i', l_i'\}_{i=1,2}$ and $\{c_i'', l_i''\}_{i=1,2}$. By definition, we have

$$\begin{aligned}
\left(\alpha_{1\Delta'} + \Delta'\right) u_1(c_1', l_1') + \alpha_{2\Delta'} u_2(c_2', l_2') - \Delta' v_1(l_2') &\geq \left(\alpha_{1\Delta'} + \Delta'\right) u_1(c_1'', l_1'') + \alpha_{2\Delta'} u_2(c_2'', l_2'') - \Delta' v_1(l_2'') \\
\left(\alpha_{1\Delta''} + \Delta''\right) u_1(c_1'', l_1'') + \alpha_{2\Delta''} u_2(c_2'', l_2'') - \Delta'' v_1(l_2'') &\geq \left(\alpha_{1\Delta''} + \Delta''\right) u_1(c_1', l_1') + \alpha_{2\Delta''} u_2(c_2', l_2') - \Delta'' v_1(l_2').
\end{aligned}$$

Now dividing these two inequalities by $\alpha_{2\Delta'}$ and $\alpha_{2\Delta''}$ respectively, we obtain

$$\begin{aligned}
\frac{\alpha_{1\Delta'} + \Delta'}{\alpha_{2\Delta'}} u_1(c_1', l_1') + u_2(c_2', l_2') - \frac{\Delta'}{\alpha_{2\Delta'}} v_1(l_2') &\geq \frac{\alpha_{1\Delta'} + \Delta'}{\alpha_{2\Delta'}} u_1(c_1'', l_1'') + u_2(c_2'', l_2'') - \frac{\Delta'}{\alpha_{2\Delta'}} v_1(l_2'') \\
\frac{\alpha_{1\Delta''} + \Delta''}{\alpha_{2\Delta''}} u_1(c_1'', l_1'') + u_2(c_2'', l_2'') - \frac{\Delta''}{\alpha_{2\Delta''}} v_1(l_2'') &\geq \frac{\alpha_{1\Delta''} + \Delta''}{\alpha_{2\Delta''}} u_1(c_1', l_1') + u_2(c_2', l_2') - \frac{\Delta''}{\alpha_{2\Delta''}} v_1(l_2')
\end{aligned}$$

Definition of $\boldsymbol{\alpha}_\Delta$ implies that for all Δ

$$\frac{\alpha_{1\Delta} + \Delta}{\alpha_{2\Delta}} = \frac{\alpha_1^*}{\alpha_2^*},$$

and therefore,

$$\begin{aligned}
\frac{\alpha_1^*}{\alpha_2^*} u_1(c_1', l_1') + u_2(c_2', l_2') - \frac{\Delta'}{\alpha_{2\Delta'}} v_1(l_2') &\geq \frac{\alpha_1^*}{\alpha_2^*} u_1(c_1'', l_1'') + u_2(c_2'', l_2'') - \frac{\Delta'}{\alpha_{2\Delta'}} v_1(l_2'') \quad (39) \\
\frac{\alpha_1^*}{\alpha_2^*} u_1(c_1'', l_1'') + u_2(c_2'', l_2'') - \frac{\Delta''}{\alpha_{2\Delta''}} v_1(l_2'') &\geq \frac{\alpha_1^*}{\alpha_2^*} u_1(c_1', l_1') + u_2(c_2', l_2') - \frac{\Delta''}{\alpha_{2\Delta''}} v_1(l_2') \quad (40)
\end{aligned}$$

These equations imply

$$\left(\frac{\Delta''}{\alpha_{2\Delta''}} - \frac{\Delta'}{\alpha_{2\Delta'}}\right)\left(v_1(l_2') - v_1(l_2'')\right) \geq 0 \quad (41)$$

Finally, from the definition of $\boldsymbol{\alpha}_\Delta$, we have

$$\begin{aligned}
\frac{1 - \alpha_{2\Delta} + \Delta}{\alpha_{2\Delta}} &= \frac{\alpha_1^*}{\alpha_2^*} \\
\frac{1 + \Delta}{\alpha_{2\Delta}} &= 1 + \frac{\alpha_1^*}{\alpha_2^*}
\end{aligned}$$

which implies that $\alpha_{2\Delta}$ is increasing in Δ. Since $1/\alpha_{2\Delta}$ is decreasing in Δ and

$$\frac{1}{\alpha_{2\Delta}} + \frac{\Delta}{\alpha_{2\Delta}} = 1 + \frac{\alpha_1^*}{\alpha_2^*},$$

$\frac{\Delta}{\alpha_{2\Delta}}$ is increasing in Δ. Therefore from (41) $\Delta'' > \Delta'$ implies $l_2'' \leq l_2'$, completing the proof of the claim.

It remains to show that if $\Delta' = 0$, then for any $\Delta'' > 0$, we have $F(\Delta'') > F(\Delta')$. Suppose to obtain a contradiction that $F(\Delta'') = F(0)$. Previous analysis indicated that this is possible only if $l_2'' = l'_2$. But then (39) and (40) imply that

$$\alpha_1^* u_1(c_1', l_1') + \alpha_2^* u_2(c_2', l_2') = \alpha_1^* u_1(c_1'', l_1'') + \alpha_2^* u_2(c_2'', l_2'').$$

We know that $\{c_i', l_i'\}_{i=1,2}$ is a solution to maximizing $\alpha_1^* u_1(c_1, l_1) + \alpha_2^* u(c_2, l_2)$ subject to

$$c_1 + c_2 \leq l_1 + l_2. \tag{42}$$

Since the u_i's are strictly concave, this solution is unique. Therefore, any $\{c_i'', l_i''\}_{i=1,2}$ that satisfies (42) must have

$$\alpha_1^* u_1(c_1', l_1') + \alpha_2^* u_2(c_2', l_2') > \alpha_1^* u_1(c_1'', l_1'') + \alpha_2^* u_2(c_2'', l_2'')$$

leading to a contradiction. ■

The next claim completes the proof of lemma. We state this claim for party 1; the result is identical for party 2.

Claim 3 *Suppose Assumptions 1 and 3 hold. Let* $\{c_i(\boldsymbol{\alpha}, \Delta), l_i(\boldsymbol{\alpha}, \Delta)\}_{i=1,2}$ *be a solution to the problem*

$$\max_{\{c_i, l_i\}} (\alpha_1 + \Delta) u_1(c_1, l_1) + \alpha_2 u_2(c_2, l_2) - \Delta v_1(l_2)$$

subject to

$$c_1 + c_2 \leq l_1 + l_2$$

for some $\Delta \geq 0$. *For any Pareto weight* $\boldsymbol{\alpha}^* \neq \boldsymbol{\alpha}$, *if* $\boldsymbol{\alpha}$ *and* Δ *are such that*

$$\frac{\alpha_1 + \Delta}{1 + \Delta} > \alpha_1^*$$

then

$$u_1(c_1(\boldsymbol{\alpha}, \Delta), l_1(\boldsymbol{\alpha}, \Delta)) - v_1(l_2(\boldsymbol{\alpha}, \Delta)) > u_1(c_1(\boldsymbol{\alpha}^*, 0), l_1(\boldsymbol{\alpha}^*, 0)) - v_1(l_2(\boldsymbol{\alpha}^*, 0)) \tag{43}$$

Proof. Suppose $\alpha_1 < \alpha_1^*$. Let $\tilde{\Delta}$ be such that

$$\frac{\alpha_1 + \tilde{\Delta}}{1 + \tilde{\Delta}} = \alpha_1^*.$$

Since $\frac{\alpha_1+\Delta}{1+\Delta}$ is increasing in Δ, we have $0 < \tilde{\Delta} < \Delta$. From Claim 2,

$$u_1(c_1\left(\boldsymbol{\alpha}, \tilde{\Delta}\right), l_1\left(\boldsymbol{\alpha}, \tilde{\Delta}\right)) - v_1(l_2\left(\boldsymbol{\alpha}, \tilde{\Delta}\right)) > u_1\left(c_1\left(\boldsymbol{\alpha}^*, 0\right), l_1\left(\boldsymbol{\alpha}^*, 0\right)\right) - v_1\left(l_2\left(\boldsymbol{\alpha}^*, 0\right)\right) \quad (44)$$

and from Claim 1,

$$u_1(c_1\left(\boldsymbol{\alpha}, \Delta\right), l_1\left(\boldsymbol{\alpha}, \Delta\right)) - v_1(l_2\left(\boldsymbol{\alpha}, \Delta\right)) \geq u_1(c_1\left(\boldsymbol{\alpha}, \tilde{\Delta}\right), l_1\left(\boldsymbol{\alpha}, \tilde{\Delta}\right)) - v_1(l_2\left(\boldsymbol{\alpha}, \tilde{\Delta}\right))$$

establishing (43).

If $\alpha_1 > \alpha_1^*$, set $\tilde{\Delta} = 0$ and (44) follows from proof of Lemma 4. ■

This result implies that (29) holds and thus establishes the desired contradiction and completes the proof of Theorem 2. ■

Proof of Theorem 2

Suppose the first-best allocation with Pareto weight $\boldsymbol{\alpha}^*$ is sustainable. Without loss of generality, suppose $\alpha_1\left(h^0\right) \geq \alpha_1^*$. Lemmas 4 and 5 imply that $\alpha_1\left(h^t\right)$ is a monotonically decreasing sequence bounded below by α_1^*. Therefore $\alpha_1\left(h^t\right)$ must converge. From the definition of the Pareto weights this implies that $\mu_i(h^t) \to 0$ for $i = \{1, 2\}$. This implies that $\left\{c_j\left(h^t\right), l_j\left(h^t\right)\right\}$ must converge to some allocation $\left\{c_j^*, l_j^*\right\}$ that solves a relaxed problem for which the constraints (10) and (11) are removed. But such $\left\{c_j^*, l_j^*\right\}$ must be a first-best sustainable allocation, which proves the theorem. ■

References

[1] D. Abreu, D. Pearce, E. Stacchetti, Toward a Theory of Discounted Repeated Games with Imperfect Monitoring, Econometrica 58 (1990) 1041–1063.

[2] D. Acemoglu, G. Egorov, K. Sonin, Coalition Formation in Nondemocracies, Rev. Econ. Stud. 75 (4) (2008) 987–1009.

[3] D. Acemoglu, M. Golosov, A. Tsyvinski, Markets Versus Governments: Political Economy of Mechanisms, NBER Working paper #1224, 2006.

[4] D. Acemoglu, M. Golosov, A. Tsyvinski, Political Economy of Mechanisms, Econometrica 76 (3) (2008) 619–641.

[5] D. Acemoglu, M. Golosov, A. Tsyvinski, Markets Versus Governments, J. Monet. Econ. 55 (1) (2008) 159–189.

[6] D. Acemoglu, M. Golosov, A. Tsyvinski, Dynamic Mirrlees Taxation under Political Economy Constraints, Rev. Econ. Stud. 77 (3) (2010) 841–881.

[7] D. Acemoglu, J.A. Robinson, Why did the West Extend the Franchise? Democracy, Inequality and Growth in Historical Perspective, Quart. J. Econ. 115 (2000) 1167–1199.

[8] D. Acemoglu, J.A. Robinson, Economic Origins of Dictatorship and Democracy, Cambridge University Press, 2006.

[9] D. Acemoglu, J.A. Robinson, Economic Backwardness in Political Perspective, Amer. Polit. Sci. Rev. 100 (2006) 115–131.

[10] M. Aguiar, M. Amador, Growth in the Shadow of Expropriation, Mimeo, 2009.

[11] M. Amador, Savings under Political Compromise, Mimeo, 2003.

[12] M. Amador, A Political Economy Model of Sovereign Debt Repayment, Mimeo, 2003.

[13] M. Battaglini, S. Coate, A Dynamic Theory of Public Spending, Taxation and Debt, Amer. Econ. Rev. 98 (1) (2008) 201–236.

[14] D. Baron, Comparative Dynamics of Parliamentary Governments, Amer. Polit. Sci. Rev. 92 (1998) 593–609.

[15] R. Barro, The Control of Politicians: an Economic Model, Pub. Choice 14 (1973) 19–42.

[16] T. Besley, S. Coate, An Economic Model of Representative Democracy, Quart. J. Econ. 112 (1997) 85–114.

[17] T. Besley, S. Coate, Sources of Inefficiency in a Representative Democracy: a Dynamic Analysis, Amer. Econ. Rev. 88 (1998) 139–156.

[18] B.D. Bueno de Mesquita, J.D. Morrow, R.M. Siverson, A. Smith, The Logic of Political Survival, MIT Press, Cambridge, 2003.

[19] A. Dixit, G.M. Grossman, F.Gul, The Dynamics of Political Compromise, J. Polit. Economy 108 (3) (2000) 531–568.

[20] A. Dixit, J. Londregan, Redistributive Politics and Economic Efficiency, Amer. Polit. Sci. Rev. 89 (4) (1995) 856–866.

[21] J. Ferejohn, Incumbent Performance and Electoral Control, Pub. Choice 50 (1986) 5–25.

[22] G.M. Grossman, E. Helpman, Protection for Sale, Amer. Econ. Rev. 84 (1994) 833–850.

[23] G.M. Grossman, E. Helpman, Electoral Competition and Special Interest Politics, Rev. Econ. Stud. 63 (1996) 265–286.

[24] M.H. Hansen, The Athenian Democracy in the Age of Demosthenes, Blackwell, Oxford, 1991.

[25] J. Hassler, S. Mora, K. Storlesseten, F. Zilibotti, Survival of the Welfare State, Amer. Econ. Rev. 93 (2003) 87–112.

[26] N. Kocherlakota, Implications of Efficient Risk Sharing without Commitment, Rev. Econ. Stud. (1996) 595–609.

[27] P. Krusell, J.-V. Ríos-Rull, Vested Interests in a Theory of Growth and Stagnation, Rev. Econ. Stud. 63 (1996) 301–329.

[28] R. Lagunoff, Markov Equilibrium in Models of Dynamic Endogenous Political Institutions, Mimeo, 2005.

[29] R. Lagunoff, Dynamic Stability and Reform of Political Institutions, Games Econ. Behav. 67 (2) (2009) 569–583.

[30] A. Marcet, R. Marimon, Recursive Contracts, Mimeo, 1998.

[31] M. McGuire, M. Olson, The Economics of Autocracy and Majority Rule, J. Econ. Lit. 34 (1996) 72–96.

[32] M.C. Olson, Dictatorship, Democracy and Development, Amer. Polit. Sci. Rev. 87 (1993) 567–575.

[33] T. Persson, G. Tabellini, Political Economics: Explaining Economic Policy, MIT Press, Cambridge, 2000.

[34] T. Persson, G. Roland, G. Tabellini, Separation of Powers and Political Accountability, Quart. J. Econ. 112 (1997) 1163–1202.

[35] T. Persson, G. Roland, G. Tabellini, Comparative Politics and Public Finance, J. Polit. Economy 108 (2000) 1121–1161.

[36] E.C. Prescott, R. Townsend, General Competitive Analysis in an Economy with Private Information, Int. Econ. Rev. 25 (1984) 1–20.

[37] E.C. Prescott, R. Townsend, Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard, Econometrica 52 (1984) 21–45.

[38] A. Przeworski, Democracy and the Limits of Self-Government, Cambridge University Press, Cambridge, New York, 2010.

[39] K. Roberts, Dynamic Voting in Clubs, Mimeo, 1999.

[40] J.A. Robinson, When Is the State Predatory?, Mimeo, 2001.

[41] K. Sonin, Why the Rich May Favor Poor Protection of Property Rights, J. Compar. Econ. 31 (2003) 715–731.

[42] N. Stokey, R.E. Lucas, E.C. Prescott, Recursive Methods in Economic Dynamics, Harvard University Press, Cambridge, 1989.

[43] J. Thomas, T. Worrall, Fluctuations and Asymmetric Information: An Example of Repeated Principle-Agent Problem, J. Econ. Theory 51 (1990) 367–90.

[44] P. Yared, Politicians, Taxes, and Debt, Rev. Econ. Stud. 77 (2) (2010) 806–840.