DISCUSSION PAPER SERIES

No. 8633

A DYNAMIC THEORY OF ELECTORAL COMPETITION

Marco Battaglini

PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP8633.asp

ISSN 0265-8003

A DYNAMIC THEORY OF ELECTORAL COMPETITION

Marco Battaglini, Princeton University and CEPR

Discussion Paper No. 8633
November 2011

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 8633

November 2011

ABSTRACT

A Dynamic theory of electoral competition*

We present a dynamic model of electoral competition to study the determinants of fiscal policy. In each period, two parties choose electoral platforms to maximize the expected number of elected representatives. The electoral platform includes public expenditure, redistributive transfers, the tax rate and the level of public debt. Voters cast their vote after seeing the platforms and elect representatives according to a majoritarian winner take all system. The level of debt, by affecting the budget constraint in future periods, creates a strategic linkage between electoral cycles. We characterize the Markov equilibrium of this game when public debt is the state variable, and study how Pareto efficiency depends on the electoral rule and the underlying fundamentals of the economy.

JEL Classification:
Keywords:

Marco Battaglini
Department of Economics
Princeton University
Fisher Hall
Princeton, NJ 08544
USA

Email: mbattagl@princeton.edu

* I thank Edoardo Grillo Juan Ortner, and Giri Parameswaran for outstanding research assistance. For helpful comments, I also thank David Baron, Steve Coate, Lydia Mechtenberg, Torsten Persson and seminar participants at the Stanford/UWI Political Economy Conference (Kingston, February 2010), the CIRPEE Workshop on Political Economy (Quebec City, April 2010) and at the economics department of the University of Chicago.

Submitted 21 October 2011

1 Introduction

There is a large literature studying the effect of electoral competition on fiscal policy. Among other achievements, this literature has characterized conditions under which the competition for votes induces self-interested politicians to choose Pareto optimal policies, and it has contributed to the understanding of how electoral rules determine public good provision, taxation and shape income distribution. With a few notable exceptions, the models that have guided this research are static.[1] A number of important questions, therefore, remain open. Under what conditions is electoral competition sufficient to generate efficient policies even in dynamic environments? When these conditions are not satisfied, what types of distortions characterize political choices and how do they depend on the electoral rule? To what extent can the insights developed in static environments be applied in dynamic contexts as well?

In this paper we present a dynamic theory of elections to study these questions. The basic framework is a natural extension of the standard probabilistic voting model to an infinite horizon environment. In each period, two parties choose electoral platforms to maximize the expected number of elected representatives. The electoral platform includes public expenditure, redistributive transfers, the tax rate and the level of public debt. Voters cast their vote after seeing the platforms and elect representatives according to a majoritarian winner take all system. The level of debt, by affecting the budget constraint in future periods, creates a strategic linkage between electoral cycles. We characterize the Markov equilibrium of this game when public debt is the state variable, and study how Pareto efficiency depends on the electoral rule and the underlying fundamentals of the economy.

To see why it is important to take a dynamic perspective in studying elections, consider the issue of Pareto efficiency. In static models Pareto efficiency is achieved because the competition for votes induces the candidates to choose policies that maximize the utility of swing voters: the electoral outcome, therefore, can always be rationalized as maximizing a weighted sum of the citizens' utility, where the weights depend on the ex ante distribution of citizens' preferences (see, for example, Lindbeck and Weibull [1987]). A similar phenomenon occurs in a dynamic model. Still, as we show in the analysis below, in a dynamic model electoral competition typically induces a Pareto inefficient allocation. In a Pareto efficient allocation policies are chosen so that the

[1] We will discuss the literature in greater detail in the next section.

marginal cost of public funds, a key measure of the inefficiency of taxation, obeys a martingale: in any period t, therefore, the inefficiency of taxation is equalized to the corresponding expected value at $t+1$. In a political equilibrium, instead, electoral competition induces the marginal cost of public funds to obey a submartingale: in any period t the inefficiency of taxation tends to be smaller than the corresponding expected value at $t+1$. This distortion does not necessarily imply that in the long run the level of debt in a political equilibrium is larger than in a Pareto efficient solution. The defining feature of the Pareto inefficiency generated by the electoral system is that in the steady state policies (debt, taxes and expenditures) are inefficiently volatile.

The intuition for why electoral competition induces inefficiently volatile policies is as follows. As in static models, the candidates identify the electoral districts that are most contendible and choose policies that maximize the expected value functions of their citizens, under-weighting the welfare of the citizens of the other regions. The favored districts, however, are unlikely to remain in this condition in the future: as preferences change over time, other districts will be more likely to be targeted by the candidates. The districts that are targeted, therefore, realize that they can appropriate a larger fraction of resources in the present than they expect in any period in the future. This makes their preferences time inconsistent. Since in every period parties choose policies that pander to time inconsistent districts, policies tend to be time inconsistent. Time inconsistency, however, does not necessarily imply a higher level of debt than in a Pareto efficient solution. Indeed, the degree of time inconsistency is endogenous and depends on the citizens expectations: the higher is debt, the lower is the discretionality of fiscal policy, so the lower is the expected degree of time inconsistency. This is the reason why debt does not become arbitrarily large and can even be lower than in a Pareto efficient solution. What the time inconsistency does is to prevent the government from accumulating a sufficiently large pool of assets that would allow to adequately self insure against the shocks. In this way it induces policies that are too volatile in the steady state. The resulting policies cannot be rationalized as the choice of a planner for any choice of welfare weights.

The strength with which electoral competition pushes toward inefficient policies depends on the electoral rule. The dynamic theory of elections described above, therefore, provides new insights on how electoral rules shape fiscal policy that could not be grasped in static models. To study this issue, we first consider a world in which geographical regions are ex ante symmetric: i.e., they all have the same ex ante distribution of citizens' preferences, and in each of them candidates have the same ex ante likelihood of victory. In this case the analysis suggests that the smaller is the electoral district (in terms of the number of geographical regions it comprises), the smaller is the dynamic distortion. Small districts reduce the parties' discretionality in choosing regions to target with pork transfers, so they reduce time inconsistency. Indeed, we show that even in the presence of preference shocks, policies can be Pareto efficient if electoral districts are sufficiently small.

When electoral districts are ex-ante heterogeneous, the choice of the optimal size of the electoral district is more complicated because an additional factor -that was previously pointed out in the static literature- becomes important. To study this point we consider a simple case in which regions can be divided in three categories: the neutrals, in which both candidates have the same probability of winning; the rightists, in which the right wing candidate has a large advantage; and the leftists, in which the left wing candidate has a symmetric advantage. As first pointed out by Persson and Tabellini [1999], when electoral districts are small the parties are incentivised to always ignore all regions except the neutrals: in the "extremist" regions the probability of winning a majority is either zero or one, the marginal value of an additional vote is therefore zero. With large districts, on the contrary, voters in all regions are valuable at the margin to the candidates, since they likely all contribute to the total vote count. This effect, combined to the dynamic effect described above, creates a trade-off: the optimal size of the electoral district depends on the size of the neutrals, as well as the expected distribution of the preference shocks and other factors determining the dynamic inefficiency. The theory therefore leads to the prediction that countries that have homogenous constituencies should find it optimal to choose a majoritarian system with small districts. Countries with heterogeneous regions have stronger incentive to choose larger districts. The fact that small districts are optimal with homogeneous preferences, and the trade-off between small and large districts with heterogeneous preferences are results that cannot be appreciated in static environments since they depend on the dynamic properties of the electoral systems.

The organization of the remainder of the paper is as follows. Section 1.2 explains how our paper relates to prior research on electoral rules and fiscal policy. Section 2 outlines the model and Section 3 establishes a benchmark by characterizing the Pareto optimal allocations. Sections 4 and 5 characterize the political equilibrium under the assumption that electoral districts are symmetric and no candidate has an electoral advantage in any district. In these sections we discuss how the dynamic inefficiency changes with the size of the electoral college. In Section 6, we consider the case in which districts are heterogeneous with respect to the ex ante preference for the parties. For this environment, we compare the equilibrium in a majoritarian system with one large electoral district to the equilibrium in a majoritarian system with many small districts. Section 7 concludes.

1.1 Related literature

This paper attempts to bring together two literatures in political economy: the literature on probabilistic voting that has mostly focused attention on static environments; and the literature on dynamic political distortions, that has considered electoral systems only in the context of simple economic environments.

With respect to the first literature, Pareto efficiency of the electoral outcome in static probabilistic models has been established by Ledyard [1984], Coughlin and Nitzan [1981], Coughlin [1982], and Lindbeck and Weibull [1987] (see Austen-Smith and Banks [2005] for a discussion of these results). The basic model of electoral competition that we adopt in this paper is taken from a special version of Lindbeck and Weibull [1987] developed by Persson and Tabellini [1999]. Persson and Tabellini [1999] is one of the first to study of the effect of electoral rules on fiscal policy.[2]

With respect to the second literature, Persson and Svensson [1989], Alesina and Tabellini [1990] and Tabellini and Alesina [1990] present the first political economy theories of dynamic fiscal policy. These papers, either do not formally model the election, or they focus on simple unidimensional policy spaces and do not study the effect of alternative electoral rules. They also limit the analysis to two-period environments, and so can not compare the steady state effects of the equilibria. In our paper we adopt a probabilistic voting model to extend the analysis of these papers to a more sophisticated policy space and electoral environments (more suitable to the study of fiscal policy); and we adopt a more general dynamic environment to analyze the long term effects of voting rules.

[2] A contemporaneous study of voting rules on fiscal policy is Lizzeri and Persico [2001]. Rather than adopting a probabilistic voting framework, Lizzeri and Persico consider a simpler policy environment (in which the public good is indivisible) and characterize the mixed strategy equilibrium. In addition to this, an important literature in political science has studied a wide range of electoral rules focusing on static settings with no Pareto inefficiencies: a purely distributive environment (Myerson [1993] and [1999]), or an abstract onedimensional policy space (Cox [1997], Morelli [2004]).

To our knowledge, the effect of alternative electoral rules on fiscal policy in a dynamic model has been previously studied only by Lizzeri [1999]. This paper, however, considers a purely distributive environment and focuses only on how transfers are distributed both within and across periods. A number of recent papers have studied probabilistic voting models in macroeconomic models: Azzimonti [2009], Bachmann and Bai [2009], Sleet and Yeltekin [2007], among others. These papers, however, have assumed that voters' preferences remain constant over time, so they arrive at the conclusion that policies solve a "virtual planning problem," that is they maximize a weighted sum of the agents' lifetime utilities. Our results allow to clarify the special conditions under which these results hold. None of these models studies political equilibria under alternative voting rules.

A related literature has studied dynamic inefficiencies in political economy models in which decisions are not taken in elections: political agency models (Acemoglu, et al. [2008], Yared [2010]), legislative bargaining models (Kalandrakis [2004], Battaglini and Coate [2008]), common pool models (Battaglini et al. [2010]) and others (Acemoglu and Robinson [2001], Lagunoff [2008], Bai and Lagunoff [2010]). Battaglini and Coate [2008] study a political economy theory of fiscal policy in an economic environment similar to the environment of this paper. In Battaglini and Coate [2008], however, policies are chosen through a very different political process (legislative bargaining in which elections are not modelled). This leads to different types of distortions in equilibrium. All the papers in this literature focus on only one political system and do not study how fiscal policy varies across political systems.

2 The Model

2.1 The economy

A continuum of infinitely-lived citizens reside in n identical regions indexed by $i = 1, ..., n$, with n odd. The size of the population in each region is normalized to be one. There are three goods - a public good g; consumption z; and labor l. The consumption good is produced from labor according to the technology $z = wl$ and the public good can be produced from the consumption good according to the technology $g = z$. Each citizen's per period utility function is

$$z + Ag^{\alpha} - \frac{l^{(1+1/\varepsilon)}}{1+\varepsilon},$$

where $\alpha \in (0, 1)$ and $\varepsilon > 0$. Citizens discount future per period utilities at rate δ.

There are competitive markets for labor and the public good. The assumptions on technology imply that the competitive equilibrium price of the public good is 1 and the wage rate is w. Moreover, the quasi-linear utility specification implies that the market interest rate is $\rho = 1/\delta - 1$. At this interest rate, citizens will be indifferent as to their allocation of consumption across time and hence their welfare will equal that which they would obtain if they simply consumed their net earnings each period. At wage rate w, each citizen will work an amount $l^*(w) = (\varepsilon w)^\varepsilon$ in each period, so that ε is the elasticity of labor supply. The associated per period indirect utility function is given by $u(w, g; A) = \frac{\varepsilon^\varepsilon w^{\varepsilon+1}}{1+\varepsilon} + Ag^\alpha$.

The relative value of the public good, as measured by A, may change across periods in a random way, reflecting changes in preferences. We assume A is a realization from an i.i.d. process. The probability that the state is A' is $f(A')$. The support of A is a finite set $\mathcal{A} = \{A_1, ..., A_q\}$, where $A_j \leq A_{j+1}$ $\forall j = 1, ..., q-1$, and $f(A_j) > 0$ for all j. For later reference, the maximal realization A_q is denoted $\overline{A}$.

Public policies are chosen by an elected government. The way the election is conducted is presented in the next two sections; here we focus on the description of the policy space. The government can raise revenues in two ways: via a proportional tax on labor income and via borrowing in the capital market. Borrowing takes the form of issuing one period bonds with interest rate ρ. Thus, if the government borrows an amount b in period t, it must repay $b(1+\rho)$ in period $t+1$. Public revenues can be used to finance the provision of the public good g but can also be diverted to finance targeted region-specific transfers $s = (s_1, ..., s_n)$, which are interpreted as (non-distortionary) pork-barrel spending. The legislature can also hold bonds if it so chooses, so that b can be negative.

A public policy is described by an $n+3$-tuple $\{r, g, x, s_1, ..., s_n\}$, where r is the income tax rate, g is the amount of the public good provided, x is the proposed new level of public debt, and s_i is the proposed transfer to region i's residents. The policy must satisfy the budget constraint that revenues must be sufficient to cover expenditures. The revenues raised under the proposal are $x + R(r)$, where $R(r) = nrwl^*((1-r)w) = nrw(\varepsilon w(1-r))^\varepsilon$ is the *tax revenue function.* Letting $B(r, g, x; b) = x + R(r) - g - (1+\rho)b$ denote the difference between revenues and spending on public goods and debt repayment, this requires that:

$$B(r, g, x; b) \geq \sum_i s_i. \tag{1}$$

The set of constraints is completed by the non-negativity constraints: $s_i \geq 0$ for each region i (which rules out financing public spending via region-specific lump-sum taxes).

There are limits on both the amount the government can borrow and the amount of bonds it can hold. Thus, $x \in [\underline{x}, \overline{x}]$ where $\overline{x}$ is the maximum amount that the government can borrow and $-\underline{x}$ is the maximum amount of bonds that it can hold. The limit on borrowing is determined by the unwillingness of borrowers to hold government bonds that they know will not be repaid. If the government were borrowing an amount x such that the interest payments exceeded the maximum possible tax revenues (i.e., $\rho x > \max_r R(r)$), then it would be unable to repay the debt even if it provided no public goods or pork. Thus, the maximum level of debt cannot exceed this level, implying that $\overline{x} \leq \max_r R(r)/\rho$. The limit on the amount of bonds that the government can hold is determined constitutionally. Without loss of generality we assume that the government is allowed to hold no more than the amount of bonds that would allow it to finance the Samuelson level of the public good from interest earnings. Thus, $\underline{x} = -g_S(\overline{A})/\rho$, where $g_S(A)$ is the level of the public good that satisfies the *Samuelson Rule* when the value of the public good is A.[3]

2.2 A model of electoral politics

In every period the ruling government is chosen by a national election. Electoral competition is modelled as probabilistic voting game in a way similar to Persson and Tabellini [1999].[4] Two parties, L and R, run for office. Before the election both parties simultaneously and noncooperatively commit to policy platforms $p^i = \{r^i, g^i, x^i, s_1^i, ..., s_n^i\}$ $\forall i = L, R$ that satisfy the budget constraint (1). Voters vote for their preferred party and the electoral rule determines the winning party. A key difference with the previous literature is that the election is embedded in a dynamic game. The level of debt, by affecting the budget constraint in future periods, creates a strategic linkage between electoral cycles.

The utility of voter l for a policy $p = \{r, g, x, s_1, ..., s_n\}$ in a state A is:

$$W^l(p; v, A) = u(r, g, A) + s_l + \delta E\left[v\left(x; A'\right)\right] \tag{2}$$

where $v\left(x; A'\right)$ is the continuation value function if the debt level is x and the state is A'. In general, this continuation value may depend on the entire history of the game and on the identity

[3] The Samuelson Rule requires that the sum of marginal benefits of the public good be equal to the marginal cost, which implies that $g_S(A)$ satisfies the first order condition that $n\alpha A g^{\alpha-1} = 1$.

[4] See Austen-Smith and Banks [2005] for a general discussion of these models.

of the region. In what follows we will focus on symmetric Markov equilibria in which strategies depend only on A' and x: so there is no loss of generality involved in restricting v to depend on A' and x only. Voters care about the policies and about the intrinsic quality of the candidates. This intrinsic quality may depend on the candidate's perceived character or on some other policies that are not directly modelled: for example, the candidate's standing on non economic issues such as civil rights, religion, etc. Voter l in region j will vote for candidate L iff

$$W^l(p_L; v; A) \geq W^l(p_R; v, A) + \varepsilon^l \tag{3}$$

where the term ε^l is a preference shock. We assume $\varepsilon^l = \widetilde{\theta}^l + \widetilde{\theta}^G$. The term $\widetilde{\theta}^l$ is idiosyncratic to voter l and it reflects the individual ideology of voter l; the term $\widetilde{\theta}^G$ represents the general popularity of party R in period t and it is common to all citizens in all districts. Both $\widetilde{\theta}^l$ and $\widetilde{\theta}^G$ are independent random variables that are realized at the beginning of the period, but are not observed by the candidates. We assume that $\widetilde{\theta}^G$ is a random variable with uniform distribution on $\left[-\frac{1}{2\theta^G}, \frac{1}{2\theta^G}\right]$; while $\widetilde{\theta}^l$ are random variable with uniform distributions on $\left[-\frac{1}{2\theta^j}, \frac{1}{2\theta^j}\right]$. It should be noted that the distribution of the idiosyncratic shock may depend on the region (j in this example) where the citizen resides.

The variance of the distribution of the ideological components is described by the vector of parameters $\theta = \left(\theta^G, \left(\theta^j\right)_j\right)$. We assume that θ is known by all players at the beginning of every period; the distribution of the shocks may, however, change over time: θ^G, and each θ^l are a independent random variables with densities, respectively, $\varphi^G(\cdot)$ and $\varphi(\cdot)$. We assume that $\varphi(\cdot)$ and $\varphi^G(\cdot)$ have full support on a finite set with minimal and maximal elements, respectively, $\underline{\theta}, \overline{\theta}$ and $\underline{\theta}^G, \overline{\theta}^G$. For future reference we denote by $z = (A, \theta)$, a shock to the economy.

In this model the fraction of votes received by candidate L in region j when political platforms are p_L and p_R is easily computed. Given (3), the voter in region j that is indifferent between L and R is characterized by $\theta_j^*(p_L, p_R; \widetilde{\theta}^G) = W^j(p_L; v, A) - W^j(p_R; v, A) - \widetilde{\theta}^G$: candidate L receives the votes of all voters l with $\widetilde{\theta}^l \leq \theta_j^*(p_L, p_R; \widetilde{\theta}^G)$. Given θ^j and $\widetilde{\theta}^G$, as illustrated in Figure 1, the fraction of votes received by L when policy platforms are p_L, p_R in region j is:

$$\frac{1}{2} + \theta^j \left[W^j(p_L; v, A) - W^j(p_R; v, A) - \widetilde{\theta}^G\right]. \tag{4}$$

The votes received by party R will be one minus (4). For simplicity, we assume that, for all platforms p_L, p_H, there is sufficient uncertainty on the citizens' preferences that both the fraction

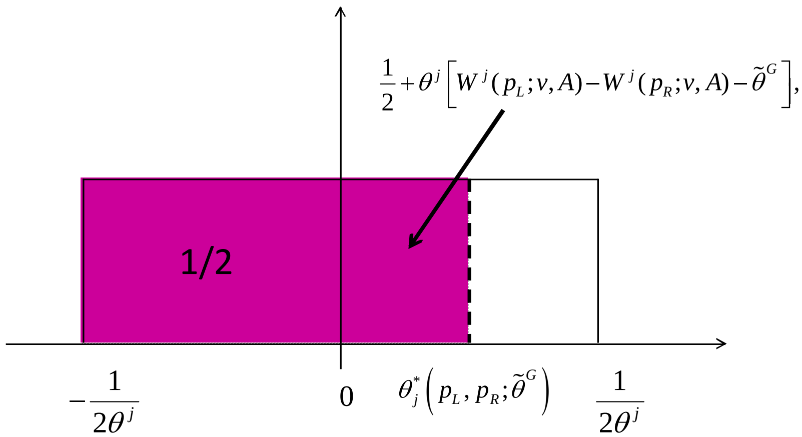

Figure 1: The probability that L wins district j.

of votes received by a candidate and the probability that a candidate receives more than 50% of the votes are interior in $(0, 1)$. It can be shown that the absolute value $\left|W^j(p_L; v, A) - W^j(p_R; v, A)\right|$ is bounded above by a constant ΔW^* (explicitly defined in Appendix 8.1). As we show in Appendix 8.1, a sufficient condition for interior probabilities is that $\overline{\theta}$ and $\overline{\theta}^G$ are sufficiently small:

Assumption 1. $1/2\overline{\theta} > \Delta W^* + 1/2\underline{\theta}^G$ *and* $1/2\overline{\theta}^G > \Delta W^*$

The first condition guarantees that no policy platform can guarantee a party 100% of the votes in a region. The second condition guarantees that there is always a positive probability that a candidate receives 50% or more of votes in a region. Assuming $1/2\overline{\theta}$ and $1/2\overline{\theta}^G$ sufficiently large implies that the distribution of the shocks have sufficiently high variance and so the outcome may sometimes be driven by the shocks regardless of the policies.

2.3 Discussion

To what extent the model described above captures important features of real world elections? Probabilistic voting has been extensively studied both theoretically and empirically in static models. Stromberg [2008] has used a probabilistic voting model to study presidential races in the US, and it has shown that the model does a remarkable job in explaining candidates' behavior. By introducing public debt and extending the time horizon to multiple periods, the model discussed in the previous section is a straightforward extension of this class of models.

A key assumption is that voters' preferences for the candidates change stochastically over time. The idea behind this assumption is that the emergence of new issues both at the national or local levels (a diplomatic crisis, emergence of a new technology with social implications, closure of a local factory, environmental problems, etc.) may shift the voters' party identification and partisanship. Are these changes in partisanship a first order effect and do they matter for elections and candidates' strategies? There is a large literature in political science studying precisely this issue. Changes in voters' preferences have been studied both at the micro level, i.e. following individual voters (Jackson [1975], Franklin and Jackson [1983]); and at the aggregate level (MacKuen et al. [1989] and [1992], Allsop and Weisberg [1988]). As Abramowitz and Saunders [1998] put it: "political scientist have long recognized that party identification has a dynamic component. Major realignments, or shifts in the partisan orientation of the electorate, have occurred periodically throughout American history." Perhaps more pertinently, the more recent literature on "Macropartisanship" has shown that changes in voters' orientation toward parties are not only long term phenomena, but have a systematic and recurrent nature. As MacKuen et al. [1989] say, changes in party affiliation "appear to be a midrange phenomenon, one that appears and disappears in a time frame of a year or two rather than a month or two or alternatively, a decade or two. The movements within this stable alignment period appear substantial, both in magnitude and duration." Does the model described above capture these changes? The key parameters describing the distribution of preferences for parties in the model are θ^j (at the local level) and θ^G (at the national level): they describe the dispersion of partisanship, that is the preferences for the right or left wing party. These variables seem to capture (at least to a first approximation) the phenomena described in the political science literature. Periods with low partisanship corresponds to periods with relatively high θ^j and θ^G, in which shocks are concentrated around zero (in $\left[-1/2\theta^j, 1/2\theta^j\right]$ and $\left[-1/2\theta^G, 1/2\theta^G\right]$) and voters care little about party labels. Periods of high partisanship corresponds to periods with low θ^j and θ^G, in which a larger mass of population have strong party preferences (since the extremes $\pm 1/2\theta^j$ and $\pm 1/2\theta^G$ have a larger absolute value). The political science literature has also identified some of the sources for these fluctuations in party attitudes. They may depend on unforeseen events like wars or economic crises (Ladd and Hadley [1978], Beck and Sorauf [1992], for example), emergence of new political issues (Ladd and Hadley [1978], Carmines and Stimsons [1989]), geographic mobility and cohort replacement (Converse [1976]), mobilization of previously disenfranchised voters (Campbell

[1985], Carmines and Stimsons [1989]). These phenomena may follow long term trends: but, as the literature described above clearly shows, their combined effect is typically best described as a stochastic process that is fundamentally unpredictable. Our assumption that θ^j and θ^G follow an i.i.d process is a convenient simplification, but it capture the essence of the phenomenon.[5]

3 The normative benchmark

Before turning to the analysis of the electoral systems, it is useful to discuss the Pareto efficient solution that would be chosen by a benevolent government. In a Pareto efficient allocation policies are chosen to maximize a weighted sum of the citizens' utilities. The government's problem and the corresponding equilibrium can be formulated recursively. Let $v_i^\circ(b; A)$ denote the value function of a citizen of district i in a state in which the current level of public debt is b and the state of the economy is A. The government chooses a policy $\{r, g, x, s_1, ..., s_n\}$ to solve:

$$\max_{(r,g,x,s)} \left\{ \begin{array}{c} \sum_i \mu_i \left[u\left(r, g; A\right) + s_i + \delta E v_i^\circ\left(x; A'\right) \right] \\ s.t. \;\; s_i \geq 0 \;\; \text{for all } i, \; \sum_i s_i \leq B(r, g, x; b), \; \& \; x \in [\underline{x}, \overline{x}] \end{array} \right\}. \tag{5}$$

where $\mu_i \geq 0$ is the weight associate to region i (we assume without loss of generality that $\sum \mu_i = 1$). Given the government choice $r^\circ(b; A), g^\circ(b; A), x^\circ(b; A), (s_i^\circ(b; A))_{i=1}^n$, the utility of a district is immediately defined by:

$$v_i^\circ(b, A) = u\left(r^\circ(b; A), g^\circ(b; A); A\right) + s_i^\circ(b; A) + \delta E v_i^\circ\left(x^\circ(b; A); A'\right). \tag{6}$$

A Pareto efficient solution consists of a collection of policy functions $r^\circ(b; A)$, $g^\circ(b; A)$, $x^\circ(b; A)$, $(s_i^\circ(b; A))_{i=1}^n$ and value functions $(v_i^\circ(b, A))_{i=1}^n$ such that $r^\circ(b; A)$, $g^\circ(b; A)$, $x^\circ(b; A)$, $(s_i^\circ(b; A))_{i=1}^n$ solve (5) for some choice of weights μ, given $(v_i^\circ(b, A))_{i=1}^n$; and, for any i, $v_i^\circ(b, A)$ satisfies (6) given $r^\circ(b; A)$, $g^\circ(b; A)$, $x^\circ(b; A)$, $(s_i^\circ(b; A))_{i=1}^n$. For the following discussion, it is useful to define the expected value function $v_i^\circ(b) = E v_i^\circ(b, A')$.

By standard methods, we can show that, for any μ, a Pareto efficient solution is associated with a unique well-behaved welfare function $V(b) = \sum_i \mu_i v_i^\circ(b)$. The welfare function is well-behaved in the sense that it is continuous, concave and almost everywhere differentiable in b. The planner's solution is characterized by two set of conditions. First, the planner equalizes the

[5] In a previous version of this paper, the shocks A, θ^j and θ^G were assumed to be positively correlated over time. See Battaglini [2010].

marginal benefit of resources across alternative uses. The first order conditions with respect to g, x and r imply:

$$\alpha A g^{\circ}(b;A)^{\alpha-1} = \frac{1}{n}\frac{1-r^{\circ}(b;A)}{1-r^{\circ}(b;A)(1+\varepsilon)} = -\delta\sum_{i}\mu_i\left[v_i^{\circ}\right]'(x)\,; \tag{7}$$

and the first order condition with respect to s_i, gives:

$$\frac{1}{n}\frac{1-r^{\circ}(b;A)}{1-r^{\circ}(b;A)(1+\varepsilon)} \geq \mu^{*} \text{ with equality if } B(r,g,x;b)>0. \tag{8}$$

where $\mu^{*} = \max_i \mu_i$. To interpret these conditions, note that the middle term in (7) is the *marginal cost of public funds.* The marginal cost of public funds is the compensating variation for a marginal increase in tax revenues, it therefore measures the social cost of taxation. In our economy it is simply equal to $(1-r)/1-r(1+\varepsilon)\geq 1$ (strict if $r>0$), independently on how the tax rate is chosen). In light of this, the first two equations state that the marginal benefit of one extra dollar spent in the public good should be equal to the marginal cost of financing it with either discretionary taxation or debt. Condition (8) requires that if there are transfers, then the marginal social benefit to the districts who receive them should be equal to the marginal cost of financing them. The planner would make only transfer to the districts with the highest social weight, so that the marginal social benefit is μ^{*}.

Second, the planner equalizes the opportunity cost of resources over time:[6]

$$\frac{1}{n}\frac{1-r^{\circ}(b;A)}{1-r^{\circ}(b;A)(1+\varepsilon)} = \frac{1}{n}E\left[\frac{1-r^{\circ}(x^{o}(b,A);A')}{1-r^{\circ}(x^{o}(b,A);A')(1+\varepsilon)}\right]. \tag{9}$$

where the left hand side is the marginal cost of public funds in state b, A; and the right hand side is the expected marginal cost of public funds. Intuitively, if the marginal cost of public funds at t where larger than the corresponding expected level at $t+1$, the planner would find it optimal to reduce taxes and increase debt, thus reducing the average intertemporal cost of taxation. Condition (9) implies, that in a Pareto efficient solution, the marginal cost of public funds obeys a martingale.

Together (7)-(8) and (9), determine the short run and long run behavior of the Pareto allocation. As it can be formally proven studying (5), along the transition to the steady state the optimal tax rate and debt level are increasing functions of b and A; and the optimal public good

[6] This condition follows from the envelope theorem and (7). Indeed, by the envelope theorem, $-\delta\sum_i \mu_i\left[v_i^{\circ}\right]'(x^{o}(b,A))$ is equal to the right hand side of (9).

level is decreasing in b and increasing in A. Since the marginal cost of public funds is concave in r, the planner wants to smooth taxes over time. This implies saving in periods of low demand and issuing debt in periods of high demand.

The long run behavior of equilibrium policies follows from (9). Since the marginal cost of public funds is a convex function of the tax rate, (9) implies that the tax rate is a submartingale and therefore converges to a constant with probability one. To see the implication of this for the steady state, let $\mu^* = \max_i \mu_i$. We have:

Proposition 1. *A benevolent equilibrium converges to a steady state in which the tax rate and the public good level are:*

$$r_{\mu^*} = \frac{n\mu^* - 1}{n\mu^*(1+\varepsilon) - 1}, \text{ and } g_{\mu^*}(A) = \left(\frac{\alpha A}{\mu^*}\right)^{\frac{1}{1-\alpha}}; \tag{10}$$

and debt is lower or equal than $x_{\mu^*} \geq \underline{x}$*, where* x_{μ^*} is defined by $B(r_{\mu^*}, g_{\mu^*}(\overline{A}), x_{\mu^*}; x_{\mu^*}) = 0$. *Moreover, any policy* $r_{\mu^*}, g_{\mu^*}(A), x$ *with* $x \in [\underline{x}, x_{\mu^*}]$ *is a steady state.*

The intuition for Proposition 1 is as follows. To keep the social cost of taxation constant, the planner accumulates resources when A is low to self insure against future shocks. This leads to a gradual accumulation of resources. In the steady state the accumulated resource are sufficiently high that the government can provide transfers after any shock. In this case (8) holds as equality, implying that the tax rate is r_{μ^*}; condition (7), then, implies that the public good is $g_{\mu^*}(A)$. These policies are possible only if $B(r_{\mu^*}, g_{\mu^*}(A), x_{\mu^*}; b) \geq 0$ for any A, so only if $b \leq x_{\mu^*}$.

4 A majoritarian electoral system

In this section we study the equilibrium in a majoritarian electoral system. In the majoritarian system, there are $m \geq 1$ *electoral districts* each comprising $n/m \geq 1$ geographical regions, where m is assumed odd and n/m an integer. A party wins a district if and only if it receives a majority of votes in the district. Parties choose policy in order to maximize the expected number of their representatives elected in the parliament. When $m = 1$, the model can be interpreted as describing a "presidential" system, in which voters elect the head of the executive branch.

4.1 Definition and existence of a political equilibrium

We focus on symmetric Markov equilibria (SME) in which voters use weakly stage undominated strategies. In these equilibria parties find it optimal to use the same strategies when committing to policy platforms, and these strategies depend only on the payoff relevant state variables b, z; and voters vote for L if and only if (3) is satisfied. A SME can be described by a collection of proposal functions $r(b; z)$, $g(b; z)$, $x(b; z)$, $s(b; z)$ and a value function $v(b; z)$. Here $r(b; z)$ is the proposed tax rate, $g(b; z)$ is the public good level, $x(b; z)$ is the new level of public debt, and $s(b; z) = \left\{s^1(b; z), ..., s^n(b; z)\right\}$ is a vector of transfers offered to the regions by each candidate. Associated with an equilibrium, is a value function $v(b; z)$ that specifies the expected future payoff of a citizen in a period in which the state is (b, z). In what follows, it will prove useful to define $v(b) = E\left[v(b; z')\right]$.

In a political equilibrium there is a reciprocal feedback between the policy proposals $r(b; z)$, $g(b; z)$, $x(b; z)$, $s(b; z)$ and the associated expected value function $v(b)$. On the one hand, the value function $v(b)$ is determined by the equilibrium policy proposals. A citizen's expected value function $v(b)$ can be determined recursively given $r(b; z), g(b; z), x(b; z)$ as:

$$v(b) = E\left[u\left(r(b; z'), g(b; z'); A'\right) + \frac{B(r(b; z'), g(b; z'), x(b; z'); b)}{n} + \delta v(x(b; z'))\right] \tag{11}$$

where $E\left[\cdot\right]$ be the expectation with respect to z'. Intuitively, citizens are all affected in the same way by the tax rate and the public good. Pork transfers are not generally distributed in uniform ways (since they depend on the realized preference shocks θ). Since, however, citizens are ex ante symmetric, and candidates treat them anonymously, they all receive the same amount of transfers in expectation: $B\left(r(b; z), g(b; z), x(b; z); b\right)/n$.

On the other hand, for a given $v(b)$, policies are chosen to maximize the expected number of representatives in the parliament. Party L wins district i if:

$$\sum\nolimits_{j \in H_i} \left[1/2 + \theta^j \left[W^j(p_L; v, A) - W^j(p_R; v, A) - \widetilde{\theta}^G\right]\right] \geq 1/2\frac{n}{m} \tag{12}$$

where H_i is the set of regions in district i. Condition (12) requires that the share of votes received in district i (the left hand side) is larger than $\frac{n}{2m}$, half of the population living in district i. If we denote by $H(j)$, the set of regions that belongs to the same district as j, condition (12) can be rewritten in a way that will prove convenient in the following analysis:

$$\widetilde{\theta}^G \leq \sum\nolimits_{j \in H_i} \frac{\theta^j}{\sum\nolimits_{l \in H(j)} \theta^l} \left[W^j(p_L; v, A) - W^j(p_R; v, A)\right]$$

The larger is the utility that L promises to region j (i.e. $W^j(p_L; v, A)$), the larger is the probability that L wins the district where j is located. The sensitivity of voters in region j is measured by

θ^j: that is, an additional dollar of transfers to region j by party L, ceteris paribus, will cause a fraction θ^j more voters to vote for party L. The pertinent issue for political parties however is the sensitivity of voters in region j relative to the sensitivity of voters in other regions from the same district, $\theta^j / \sum_{l \in H(j)} \theta^l$.[7]

Given a realized z, the expected number of representatives for party L is:

$$\sum_{i=1}^{m} E\left[sign\left(\sum_{j \in H_i} \frac{\theta^j}{\sum_{l \in H(j)} \theta^l}\left(W^j(p_L; v, A) - W^j(p_R; v, A) - \widetilde{\theta}^G\right)\right)\right] \tag{13}$$

where the function $sign(x)$ is equal to one if $x \geq 0$ and zero otherwise.[8] Using the distribution of $\widetilde{\theta}^G$, this expression can be written as:

$$\sum_{i=1}^{m}\left[1/2 + \theta^G \sum_{j \in H_i} \frac{\theta^j}{\sum_{l \in H(j)} \theta^l}\left(W^j(p_L; v, A) - W^j(p_R; v, A)\right)\right] \tag{14}$$

Simplifying and eliminating irrelevant constraints, candidate $L's$ problems becomes:

$$\max_{(r,g,x,s)} \left\{ \begin{array}{c} \sum_{i=1}^{m} \sum_{j \in H_i} \frac{\theta^j}{\sum_{l \in H(j)} \theta^l} W^j(p_L; v, A) \\ \text{s.t.} \quad B(r, g, x; b) \geq \sum s_j, \quad x \in [\underline{x}, \overline{x}]. \end{array} \right\}. \tag{15}$$

Naturally, party R wins an expected number of candidates equal to n minus (13): so he chooses $p_R = (r_R, g_R, x_R, s_R)$ to minimize the objective function of (15) given p_L and the same constraints as in (15). By symmetry, R's problem is therefore equivalent to L's problem: R finds it optimal to choose the same tax rate, public good level, and vector of transfers as as candidate L.

Definition 4. *A political equilibrium consists of a collection of policy functions $r(b; z)$, $g(b; z)$, $x(b; z)$, $s(b; z)$ and a value function $v(b)$ such that $r(b; z)$, $g(b; z)$, $x(b; z)$, $s(b; z)$ solve (15) given $v(b)$; and $v(b)$ satisfies (11) given $r(b; z)$, $g(b; z)$, $x(b; z)$.*

Problem (15) can be conveniently simplified. Without loss of generality, we can assume $B(r, g, x; b) = \sum s_j$. At the margin a dollar spent in pork transfers in region j has a value $\theta^j / \sum_{l \in H(j)} \theta^l$. Candidate L, therefore choose the region i with the highest relative weight. Ignoring constant terms, and using the fact that $\sum_{j \in H_i} \left[\theta^j / \sum_{l \in H(j)} \theta^l\right] = 1$, candidate L's problem

[7] Note that Assumption 1 guarantees that the probability with which a candidate wins district i is in $(0, 1)$ for all $i = 1, ..., m$.

[8] The expectation is taken with respect to the random variable $\widetilde{\theta}^G$, since the parameters of the distribution θ^j, θ^G and A are known by the candidates at this stage.

can be written as:

$$\max_{(r,g,x,s)} \left\{ \begin{array}{c} u(r,g;A) + \max_j \left[\theta^j / \sum_{l \in H(j)} \theta^l\right] B(r,g,x;b)/m + \delta v(x) \\ \text{s.t.} \quad B(r,g,x;b) \geq 0, \quad x \in [\underline{x}, \overline{x}]. \end{array} \right\}. \tag{16a}$$

We can therefore characterize the equilibrium focusing only on (16a). We say that an equilibrium is *well-behaved* if $v(b)$ is continuous and weakly concave on $[\underline{x}, \overline{x}]$. In what follows, we will always study well-behaved equilibria. Henceforth, when we refer to an "equilibrium," it is to be understood that it is a well-behaved, symmetric Markov equilibrium. We have the following result:

Proposition 2. *A well-behaved political equilibrium exists in a majoritarian system for any m.*

The existence problems associated with models of probabilistic voting have been extensively discussed by Lindbeck and Weibull [1987], who derive a sufficient condition for the existence of a pure strategy equilibrium in a static model. Since the electoral game here is embedded in a stochastic game in which debt is the state variable, existence is in no way implied by these results: the usual complications associated with stochastic games (for which no general existence theorem is available) are valid. The strategy to prove existence in Proposition 2 is as follows. Problem (16a) defines a correspondence $T(v)$ that maps a bounded, continuous and weakly concave functions to a subset of bounded, continuous and weakly concave functions. In general $T(v)$ is not convex valued. Indeed, the candidate parties may be indifferent among a variety of policies: these policies however may generate different expected utility levels for the citizens; and the convex combination of these utilities is not necessarily in $T(v)$. The key steps in the proof of Proposition 2 consists in proving that we can find a selection $T^*(v)$ of $T(v)$ that is convex valued (and non empty, compact and with closed graph). Given this, we can prove that a fixed-point $v^* \in T^*(v^*)$ exists by the Glicksberg-Fan theorem.

4.2 Characterization

The equilibrium can be understood by observing the objective function in (16a). For a given policy platform of candidate R, candidate L chooses the policy that maximizes the expected number of elected representatives. The budget constraint imposes a trade off. Candidate L can offer a more attractive policy by choosing a platform with generously high g, low tax r, low debt level x, or a combination of all of these. These policies increase L's popularity uniformly in all regions. Alternatively L can be more conservative in the choice of g, r, x, and reserve a higher level of resources for pork transfers. Transfers are less efficient because they do not affect all regions. Candidate L, however, can target them to the regions in which he expects there are more swing voters. This trade off is present in the static versions of the voting games that have been studied by Lindbeck and Weibull [1987] and Persson and Tabellini [1999]. A key difference of our model with respect to these models is that, in a dynamic environment, the trade-off depends on the state of the economy, and this changes endogenously over time.

Two cases are possible. First, candidate L finds it optimal to provide pork transfers. In this case taxes are chosen so that the marginal benefit of making a transfer equals the marginal cost of a higher tax rate. From the first order conditions with respect to r, the optimal tax is:

$$\frac{1}{n}\left[\frac{1-r_m(\theta)}{1-r_m(\theta)\,(1+\varepsilon)}\right]=\frac{1}{m}\max_j\left[\theta^j/\sum\nolimits_{l\in H(j)}\theta^l\right]. \tag{17}$$

To interpret this condition, note that the left hand side is the marginal cost of public funds. The right hand side, measures the political benefit of making a pork transfer.

Similarly, the level of the public good and the tax rate are chosen so that the marginal benefit of the public good, the marginal cost of debt and the marginal cost of public funds are equal:

$$\alpha A g_m(A,\theta)^{\alpha-1}=\frac{1}{n}\left[\frac{1-r_m(\theta)}{1-r_m(\theta)\,(1+\varepsilon)}\right]=-\delta v'(x_m(\theta)). \tag{18}$$

In this case, as it can be seen from (17) and (18), the policies $r_m(\theta), g_m(A,\theta), x_m(\theta)$ are independent on the level of debt. This follows from the fact that at the margin, resources are used for pork transfers and the marginal benefit of pork is independent of b. Obviously, the policies depend on θ and A- the preference parameters.

The second case is when candidate L finds it optimal to have no transfers. In this case $B(r,g,x;b)=\sum s_j=0$. It follows that $\max_j\left[\theta^j/\sum_{l\in H(j)}\theta^l\right]\cdot B(r,g,x;b)$ is equal to $\frac{1}{n}\cdot B(r,g,x;b)$ (both are zero), and it is *as if* the policies in state b, z are chosen to maximize:

$$\max_{(r,g,x)}\left\{\begin{array}{c} u\,(r,g;A)+\frac{B(r,g,x;b)}{n}+\delta v\,(x) \\ \\ s.t.\;\; B(r,g,x;b)\geq 0\;\&\; x\in[\underline{x},\overline{x}]. \end{array}\right\}. \tag{19}$$

Note that the solution of (19), $r(b;A)$, $x(b;A)$ and $g(b;A)$, depends only on b and on A: since pork transfers are zero, all agents are treated equally, and their preferences for the candidates are irrelevant.

The first case is clearly possible whenever the policies defined by (17) and (18) are feasible, so

$$B(r_m(\theta), g_m(A,\theta), x_m(\theta); b) \geq 0. \tag{20}$$

Since $g_m(A,\theta)$ is increasing in A and since $B(r_m(\theta), g_m(A,\theta), x_m(\theta); b)$ is decreasing in g, there must be an $A_m(\theta,b)$ such that (20) is satisfied if and only if $A \leq A_m(\theta,b)$. We can conclude:

Proposition 3. *There is a threshold $A_m(\theta,b)$, such that for $A > A_m(\theta,b)$ equilibrium policies $r(b;z)$, $x(b;z)$ and $g(b;z)$ are equal to the solution of (19), $r(b;A)$, $x(b;A)$ and $g(b;A)$; and expected pork transfers are zero. For $A \leq A_m(\theta,b)$ policies $r(b;z)$, $x(b;z)$ and $g(b;z)$ are equal to $r_m(\theta)$, $x_m(\theta)$, $g_m(A,\theta)$ uniquely defined by (17)-(18), and expected pork transfers are $B(r_m(\theta), x_m(\theta), g_m(A,\theta); b)/n > 0$.*

Proposition 3 provides a clear description of how candidates choose policies and of the equilibrium dynamics. When $A > A_m(\theta,b)$ the citizens' preferences for the individual candidates are irrelevant, only A and b affect the policies. In this region, because candidates don't have enough money to chase the districts with cash transfers, they find it optimal to treat citizens equally: given the equilibrium v, candidates choose policies as if they were maximizing aggregate welfare. Using (19), it is easy to show that $r(b;A)$, and $x(b;A)$ are increasing in A and in b, while $g(b;A)$ is decreasing in b and increasing in A. [9]

When $A \leq A_m(\theta,b)$, the level of public good $g_m(\theta)$ depends on A, θ and the specific electoral rule (as measured by m), but not on b; similarly taxation $r_m(\theta)$ and debt $x_m(\theta)$ depend only on θ and m. In this region, the candidates allocate pork transfers according to the regions's preferences for candidates, so it is not surprising that the variation of the citizen's preferences for the candidates (as measured by θ) and the voting rule are important. The effect of the citizens's preferences for the candidate depends on m: the smaller is m the larger is the effect of θ on these policies. The reason is intuitive. In multi-region districts, parties can trade off voter share gains in some regions against others, since all they care about is the total votes in the district. Hence the relative density amongst regions matters as to where pork is directed. In single region districts, no trade-off is possible within districts.[10] Furthermore, it is impossible to trade-off votes between districts, and parties want to maximize the number of districts won.

[9] This observation follows from the fact that, given $v(b;A)$, problem (19) is a standard concave programming problem.

[10] Indeed, when $m = n$, $H(j) = \{j\}$ and $\max_j \left[\theta^j / \sum_{l \in H(j)} \theta^l\right] = 1 \ \forall \theta$.

Given these policies, the effect of a shock to the economy propagates over time through the impact on debt. An increase in A at t, induces an increase in b at t. In the following periods, even if A remains constant, the higher level of debt induces lower public good provision, higher taxes and higher debt. This occurs for two reasons. First, because whenever $A > A_m(\theta, b)$ the higher is b the higher is $r(b; A)$, $x(b; A)$ (and the lower is $g(b; A)$). Second, because (as it can be seen from 20) an increase in b causes $A_m(\theta, b)$ to decrease, it is more likely that $A > A_m(\theta, b)$. When A falls, the state follows a similar dynamics but in reverse. The decrease at t induces an decrease in b; even if A remains constant in the following periods, the decrease in b will induces further decreases in the following periods until a positive shock on A arrives.

The dynamics described above can be very complicated, since shocks hit the economy in every period and their effects accumulate over time. In the long run the economy converges to a steady state distribution. The properties of the steady state distribution of the policies provide a relatively simple way to describe the effect of political distortion on fiscal policy. To study the properties of this distribution, we need to study how resources are allocated over time in more detail.

5 Long term distribution and welfare

As in the planner's solution, in a political equilibrium the first order condition with respect to b can be written as:

$$\frac{1}{n}\left[\frac{1-r(b;z)}{1-r(b;z)\,(1+\varepsilon)}\right] = -\delta v'(x(b;z)) \tag{21}$$

To see the implications of this condition, it is important to study how the expected value function changes following a marginal change in b. From (11) and Proposition 3, we know that when $A < A_m(\theta, b)$, we have:

$$E\left[v(x;z')\right] = E\left[u\left(r_m(\theta'), g_m(A',\theta'); A'\right) + 1/n \cdot B(r_m(\theta'), g_m(A',\theta'), x_m(\theta'); x)) + \delta v(x_m(\theta'))\right]$$

where

$$B(r_m(\theta'), g_m(A,\theta'), x_m(\theta'); x)) = x_m(\theta') + R(r_m(\theta')) - g_m(A',\theta') - (1+\rho)x.$$

It follows that $-\delta E v'(x(b;z);z') = 1/n$, since $\delta\,(1+\rho) = 1$. Intuitively, when $A < A_m(\theta, b)$ a reduction in debt frees 1 unit of resources to spend in pork transfers, with an expected benefit of

$1/n$ in a symmetric equilibrium. When $A \geq A_m(\theta, b)$, instead, $E\left[v(x(b;z);z')\right]$ is given by (19). Using the envelope theorem, we find that in a state z':

$$-\delta v'(x(b;z);z') = \frac{1}{n}\frac{1-r(x(b;z);z')}{1-r(x(b;z);z')\,(1+\varepsilon)} \tag{22}$$

The derivative in this case is higher than $1/n$ because the increase in debt does not only imply a transfer of resources from $t+1$ to t, but also a social loss. Given that pork transfer are zero at $t+1$, the agents will be forced to increase taxes or debt, or reduce g or t: all measures that generate a loss in social welfare. We conclude that

$$-\delta v'(x) = 1/n \cdot G(x) + E\left[\frac{1}{n}\frac{1-r(x;z')}{1-r(x;z')\,(1+\varepsilon)}\,|A' > A_m(\theta', x)\right]\cdot[1-G(x)] \tag{23}$$

where $G(x)$ is the probability that $A' \leq A_m(\theta', x)$.

Using (17) and (23), we can rewrite (21) as:[11]

$$\begin{aligned}\frac{1}{n}\frac{1-r(b;z)}{1-r(b;z)\,(1+\varepsilon)} &= \frac{1}{n}E\left[\frac{1-r(x(b;z);z')}{1-r(x(b;z);A')\,(1+\varepsilon)}\right] \\ &\quad -G(x(b;z))\cdot\frac{1}{n}E\left[\left(\frac{n}{m}\max_j\frac{\theta^j}{\sum_{l\in H(j)}\theta^l}-1\right)\right]\end{aligned} \tag{24}$$

From this expression we draw two conclusions. First, since $\frac{n}{m}\max_j\left[\theta^j/\sum_{l\in H(j)}\theta^l\right] > 1$, we have:

$$\frac{1}{n}\frac{1-r(b;z)}{1-r(b;z)\,(1+\varepsilon)} \leq \frac{1}{n}E\left[\frac{1-r(x(b;z);A')}{1-r(x(b;z);A')\,(1+\varepsilon)}\right] \tag{25}$$

when $m < n$, so the marginal cost of public funds is a submartingale. Indeed we can be more specific. We say that the marginal cost of public funds is a *strict submartingale* if for any initial state, there is a positive probability of reaching a state at which (25) is strict. We have:

Proposition 4. *In a political equilibrium, the marginal cost of public funds is a strict submartingale when $m < n$.*

Comparing (25) to (9), we can see that resources are inefficiently allocated over time: politicians tend to shift too much of the burden of taxation to the future.

The second conclusion that we can draw from (24) is that the wedge between the marginal cost of public funds at time t and at time $t+1$ depends on the electoral rule and it is state contingent.

[11] To write (24) we add and subtract $G(x)E\left[\frac{1}{n}\left[\frac{1-r(x;z')}{1-r(x;z')(1+\varepsilon)}\right]|A' \leq A_m(\theta', x)\right]$ in (23). The added term contributes to form the first term in (24); the subtracted term is equal to the second term in (24) by (17).

It is decreasing in m and it is zero when $m = n$, since $\frac{n}{m} \max_j \left[\theta^j / \sum_{l \in H(j)} \theta^l\right] = 1$. When $m < n$ there is no choice of Pareto weights that could rationalize the allocation, since the candidates tend to systematically transfer the burden of taxation to the future. This bias, however, does not lead to an arbitrarily large accumulation of debt. As b increases, the probability that $A' > A_m(\theta', x(b; z))$ (i.e. $G(x(b; z))$ decreases, and so the wedge decreases as well. For high levels of b, the marginal cost of public funds starts to behave as in a Pareto efficient solution, so debt tends to go down. This observation has important implications for the steady state distribution of policies in the political equilibrium. Let $r_m = \min_\theta r_m(\theta)$, $g_m = \max_{\theta, A} g_m(A, \theta)$, $x_m = \min_\theta x_m(\theta)$. We say that a stationary distribution of policies is nondegenerate if the policies are not constant. We have:

Proposition 5. *The political equilibrium is Pareto efficient if and only if $m = n$. For any initial state A_0, b_0:*

- *If $m = n$, the political equilibrium coincides with the utilitarian optimum in which $\mu_i = 1/n$ for all $i = 1, ..., n$: $r(b; z)$ and $g(b; z)$ converge, respectively, to 0 and $(\alpha n A)^{1/1-\alpha}$,* and *public debt converges to $\underline{x}$.*
- *If $m < n$, $r(b; z)$ and $g(b; z)$ converge to a non degenerate stationary distributions with support, respectively, $[r_m, \overline{r}]$ and $[0, g_m]$,* and *public debt $x(b; z)$ converges to a nondegenerate distribution in $[x_m, \overline{x}]$.*

When $m < n$, each district is comprised of more than one region and the candidates can be opportunistic in making transfers to the region that is most susceptible to being influenced. The marginal benefit of this policy is equal to $\frac{1}{m} \max_j \left[\theta^j / \sum_{l \in H(j)} \theta^l\right]$, so it changes over time with θ. Although $\frac{1}{m} \max_j \left[\theta^j / \sum_{l \in H(j)} \theta^l\right]$ may go up or down over time, and so changes the incentive to make transfers, (24) shows that this mechanism creates a *systematic* bias toward shifting the burden of taxation in the future. The reason is that whatever the value of θ at t, the marginal benefit of making a transfer at t is always larger than the expected benefit of making it at $t + 1$. This happens because, for almost any realization of θ, $\frac{1}{m} \max_j \left[\theta^j / \sum_{l \in H(j)} \theta^l\right] > \frac{1}{n}$ (the exception being the unlikely event in which θ^j is identical for all regions in the electoral district); the expected benefit of a dollar in future transfers to a region, however, is $1/n$ (since in expectation all regions have the same probability to receive it). As m increases, the number of regions per district decreases, and the arbitrage opportunities decrease as well. When $m = n$, we have only one region per district, so $\frac{1}{m}\max_j\left[\theta^j/\sum_{l\in H(j)}\theta^l\right]$ is constant at 1: in this case the marginal benefit of making a transfer is constant over time, and the benefit of a transfer at t equals the expected benefit at $t+1$.

Comparing the equilibrium to the steady state reached in a Pareto efficient allocation, we can note two differences. First, in the long term the level of debt may either be higher or lower in a political equilibrium. From Proposition 1 and Proposition 5 we can see that debt will be uniformly higher in the political equilibrium only if $x_{\mu^*} < x_m$. As it can be easily verified, this is the case if and only if:

$$\mu^* < \max_\theta \frac{1}{m}\max_j\left[\theta^j/\sum\nolimits_{l\in H(j)}\theta^l\right]$$

If this condition is not satisfied, the debt will fluctuate above and below the steady state in the Paretian allocation. Although this condition is always satisfied for a utilitarian social welfare function (in which $\mu^* = 1/n$) it does not need to hold for general social weights, leaving open the possibility that even high level of debt can be rationalized as optimal in a Paretian sense.

The second and most important difference is that when $m < n$ the steady state in the political equilibrium is more volatile than the steady state of Pareto efficient equilibrium. For any set of welfare weights, the Paretian planner always finds it optimal to accumulate sufficient resources to perfectly self insure. In the political equilibrium, on the contrary, self insurance is imperfect (except when $m = n$).

There are three lessons to draw form Proposition 5. First, electoral competition is not sufficient per se to guarantee Pareto efficiency. This is in contrast to the standard results in static models where electoral outcomes are Pareto efficient for some choice of welfare weights (see Lindbeck and Weibull [1987]).

Second, what distinguishes a political equilibrium form a Pareto efficient solution is not necessarily a higher level of debt, but more volatile policies in the steady state.

Finally, the political distortion generated by political competition is higher in electoral systems in which the electoral districts are large and comprised of regions with poorly correlated shocks. When the shocks are uncorrelated, the candidates are tempted to pick and choose where to make the transfers: this generates inefficiently large transfers and volatility. When the shocks are correlated, the arbitrage opportunities are smaller because the θ^j move together. The case in which $m = n$ should be seen as a limit in which the districts are designed to comprise only regions with perfectly correlated shocks (indeed there is a common shock with $m = n$): in this case the arbitrage opportunity is zero, and resources are allocated efficiently.

6 Heterogeneous regions

In the analysis presented above, regions are ex-ante symmetric: they all have the same distribution of preferences for candidates; and none of them has an ex ante bias for a candidate (i.e. $E\varepsilon^l = 0$ $\forall l$). As we have seen, in this case it is unambiguously optimal to design electoral districts with the smallest number of regions. Small districts are optimal because they limit the ability of politicians to target funds opportunistically.

Regions are often asymmetric with respect to the candidates: some regions are ex-ante neutral and so easily contendible; others have strong and persistent biases for one candidate or the other. In this case the choice of the size of the district is less straightforward. Small electoral districts may allow the candidates to focus resources only on regions that can be easily swung with targeted transfers, ignoring the regions where they have an ex ante small (or high) probability of winning. When districts are large, on the contrary, even biased regions have a marginal value for the candidates. In districts comprised only of right-wing regions, for example, the L candidate never wins. In a district with mixed regions, both candidates may win and the right wing regions may contribute to the total vote count *at the margin*: so they should not be ignored. This idea has been exploited by Persson and Tabellini [1999] to argue that a system with one large district may generate a higher level of public good expenditure and lower taxation than a system with many small districts.

To study this environment with heterogeneous regions, we now assume that the regions can be partitioned in three groups: *neutral regions*, that is regions with $E\varepsilon^l = 0$, as described in the previous sections; *right-wing regions*, with a bias for R; and *left-wing regions* with a bias for L. In particular we assume that for the right-wing regions and the left wing-regions the idiosyncratic shock $\widetilde{\theta}^l$ has a uniform distribution with support in, respectively, $\left[-\frac{1}{2\theta^l} + \sigma, \frac{1}{2\theta^l} + \sigma\right]$ and $\left[-\frac{1}{2\theta^l} - \sigma, \frac{1}{2\theta^l} - \sigma\right]$. As in the previous sections, we assume that all candidates receive an interior fraction of votes in (0,1) in each region; now we also assume that σ is sufficiently large that candidate R (respectively, L) never receives more than 50% of the votes in a left-wing (respectively, right-wing) region. As we show in the appendix, a sufficient condition for this is:

Assumption 2. $\sigma > \Delta W^* + \frac{1}{2\underline{\theta}^G}$, $1/2\overline{\theta} > \Delta W^* + \sigma + \frac{1}{2\underline{\theta}^G}$.

where ΔW^* is defined as in Assumption 1. The first inequality makes sure that candidate L (respectively, R) never wins a majority in a right-wing (respectively, left-wing) region; the second inequality corresponds to Assumption 1 and guarantees that all candidates receive an interior fraction of votes in all regions. The fraction of regions that are neutral is β (with βn assumed to be an integer for simplicity); and the fraction of right- and left-wing regions is $(1-\beta)/2$.

In this section we compare two alternative voting rules: a rule with $m = 1$ and a rule with $m = n$. We interpret the first rule as a *presidential system* in which only one executive position is elected; the second is a *majoritarian system* in which n representatives are elected. Persson and Tabellini [1999,2004] have done a similar exercise in a static model with 3 regions (a neutral-, a right-wing region ad a left-wing region). They concluded that a system with $m = 1$ generates higher g, lower r and higher utilitarian welfare. Here we show that, although their intuition is correct, another force is at play in a dynamic model**;** and the opposite result may occur in the steady state of a dynamic model. This highlights the importance of taking a dynamic perspective in the study of institutions.

In the *presidential system* with $m = 1$ the analysis is very similar to the analysis presented above. Although candidate L cannot win a majority in the right-wing regions, the votes gained in right-wing regions contribute towards the overall vote count. Candidate L, therefore, finds it optimal to direct transfers to the regions that are more reactive to the transfers at the margin, regardless of the average ideology. Policies maximize the same objective function as in (16a):

$$u(r,g;A) + \max_j \left[\theta^j / \sum_{i=1}^{n} \theta^i\right] B(r,g,x;b)/m + \delta E\left[v\left(x;A'\right)|A\right] \tag{26}$$

under the same constraints as in (16a).[12] Note that this objective function depends on θ because the candidate will choose the region with the highest θ^j.

In a *majoritarian system* with $m = n$ districts, the bias σ is relevant. By Assumption 2, candidate L knows that he never wins right-wing regions, so all votes received in these regions are lost. It is also irrelevant to target resource to left-wing regions, since these are won for sure. This implies that, as we show in the appendix, candidate L maximizes:

$$u(r,g;A) + \frac{1}{\beta n} B(r,g,x;b) + \delta E\left[v\left(x;A'\right)|A\right] \tag{27}$$

[12] The expression in (26) is a special case of (16a) for $m = n$.

The candidates care only about the βn regions that are contendible. They moreover treat those regions symmetrically because as we have discussed above, the relative sensitivities $\theta^j / \sum_{l \in H(j)} \theta^l$ in the regions are identical (and equal to 1) when $m = n$. (This follows since in a majoritarian system there is no possibility of trading off excess votes between districts). Comparing (26) to (27), we can make two observations. When $1/\beta n > \max_j \left[\theta^j / \sum_{i=1}^n \theta^i\right]$, the candidates tend to overestimate the benefit of pork in a majoritarian system (and underestimate the benefit of g and the cost of r and b). This is the point that was made by Persson and Tabellini [1999] in their 3 region example. In a dynamic model, however, there is another difference between a presidential system and a majoritarian system. The objective function (26) depends on the realization of θ, and so it fluctuates over time. As shown in the previous section, the equilibrium is dynamically inefficient. When $m = n$, instead, the objective function is independent from θ, and the policy is time consistent. There is therefore a trade off. In a majoritarian system policies are time consistent, but tend to ignore the fraction $1 - \beta$ of "extremist" regions; in a presidential system, regions are treated in an ex ante symmetric way, but ex post the allocation depends on θ, because the candidates choose the more sensitive regions as targets of pork transfers.[13]

Let r_n, g_n and x_n be defined as in Proposition 4 of Section 5 when $m = n$. We have:

Proposition 6. *Under Assumption 2:*

- *In simple majoritarian system ($m = n$), $r(b; z)$ and $g(b; z)$ converge, respectively, to $r_\beta = \frac{1-\beta}{1+\varepsilon-\beta}$ and $g_\beta(A) = (\alpha \beta n A)^{\frac{1}{1-\alpha}}$ and debt to a level lower than or equal* to $\mathbf{x}_\beta = \left[R(r_\beta) - g_\beta(\overline{A})\right] / \rho$. *In the steady state, neutral regions receive $B(r_\beta, g_\beta(A), x_\beta; b)/\beta n$ in pork transfers and the other regions receive zero. Any policy $r_\beta, g_\beta(A), x$ with $x \in [\underline{x}, x_\beta]$ is a steady state. Policies are Pareto efficient.*

- *In a presidential system ($m = 1$), $r(b; z)$ and $g(b; z)$ converge to a non degenerate stationary distributions with support, respectively, $[r_n, \overline{r}]$ and $[0, g_n]$, and public debt $x(b; z)$ converges to a nondegenerate distribution in $[x_n, \overline{x}]$. Regions receive the same expected transfer $B(r_\beta, g_\beta(A), x_\beta; b)/n$. Policies are Pareto inefficient.*

This Proposition has implications for the comparison of alternative electoral systems. In

[13] Clearly in a majoritarian system we may have $m < n$, in which case the policies would be Pareto inefficient. To the extent that $m > 1$, however, the Pareto inefficiency would be smaller than in the presidential system with $m = 1$.

choosing the political system there is a trade off. In the majoritarian system the allocation of resources is targeted to only neutral regions. This tends to keep tax rates higher and public good provision lower because neutral regions like to tax the entire community and appropriate all the tax revenues; they also underestimate the value of the public good because they do not internalize its value for non-neutral regions. The policies, however, are Pareto efficient. Because of this, the economy tends to have a level of debt sufficiently low to eliminate any volatility in policy choices. In the proportional system, on the contrary, regions are treated symmetrically from an ex-ante point of view. Policies are however Pareto inefficient: this reduces the "size of the pie" for everybody. Naturally the larger is β, the smaller is the inequality in a proportional system and the larger are the benefits of the majoritarian system. When β is sufficiently large even a utilitarian planner would prefer a majoritarian system. From this discussion, it follows that:

Proposition 7. *There is a $\beta^* < 1$ such that for all $\beta > \beta^*$ the expected level of the public good and of the tax rate in the invariant distribution are, respectively, higher and lower in a majoritarian system (m=n) than in a presidential system (m=1). When $\beta > \beta^*$, moreover, the steady state of the majoritarian system is strictly preferred by a utilitarian planner to the steady state of a presidential system.*

The analysis in this section makes clear the importance of considering dynamic models when studying fiscal policy in political systems, even if the analysis is limited to the comparison of steady states and so the dynamics of the equilibria are ignored. The phenomena described by Propositions 5 and 6 follows from the fact that a system with $m = 1$ is dynamically inefficient, so it leads to too much debt accumulation. In a system with $m = n$, under Assumption 2, candidates ignore a fraction $1-\beta$ of the electorate, the allocation, however is dynamically efficient. In a static environment both systems would be Pareto efficient since they would differ only with respect to the distribution of resources in the economy.

7 Conclusion

In this paper we have presented a dynamic theory of electoral competition to study the determinants of fiscal policy. The basic framework we adopted is a generalization of a standard probabilistic voting model first used by Persson and Tabellini [1999] to study fiscal policy in a static environment in which public debt is a state variable that creates a strategic link between policy-making periods. We have shown that while in static environments fiscal policy is generally Pareto efficient, in dynamic environments Pareto efficiency depends on the electoral rule. In the benchmark case in which regions are symmetric, the smaller are the electoral districts, the more Pareto efficient is the policy outcome. Large electoral districts tend to be more dynamically inefficient because they give candidates more discretion to target fiscal policies opportunistically towards the regions that are easier to swing. When regions are heterogeneous and there are regions with ex ante bias towards a candidate, there is a trade off. Large electoral districts are more Pareto inefficient, but they tend to be more inclusive because they do not provide incentives to ignore regions with biased preferences (that would not be contendible in the case of small districts). The theory therefore predicts that countries that have homogenous constituencies should find it optimal to choose a majoritarian system with small districts. Countries with heterogeneous regions have stronger incentive to choose larger districts. This result extends the analysis of static models that associate large districts with superior policies. The fact that small districts are optimal with homogeneous preferences, and the trade off between small and large districts with heterogeneous preferences are results that cannot be appreciated in static environments since they depend on the dynamic properties of the electoral systems.

The analysis can be extended in several directions. First, in this paper we have restricted the analysis to majoritarian electoral system, and we have considered only changes in the size of the electoral district. Clearly, it would be interesting to develop dynamic theories of electoral competition for a wider range of electoral rules. Second, we have restricted the analysis to environments with symmetric regions and to environments with a very simple type of heterogeneity in preferences; we have moreover focused on symmetric equilibria. It would be interesting to develop a more general theory that applies to asymmetric environments in which, for example, a party has a dominant position; or in which, a bias can persist persists for a while (e.g. shocks that are mean zero autoregressive). Finally, it would be interesting to extend the analysis to study dynamic electoral competition with more than two candidates.

References

Abramowitz, Alan and Kyle Saunders (1998), "Ideological Realignment in the U.S. Electorate," *The Journal of Politics*, 634-652, 60:3.

Acemoglu, Daron and James Robinson (2001) "A Theory of Political Transitions," *American Economic Review*, 91(4), pp. 938-963.

Acemoglu, Daron, Michael Golosov and Aleh Tsyvinsky (2008), "Political Economy of Mechanisms," *Econometrica*, 76(3), pp. 619-641.

Alsop, Dee and Herbert Weisberg (1988), "Measuring change in party identification in an election campaign," *American Journal of Political Science*," 32: 996-1017.

Austen-Smith, David (2000), "Redistributing Income under Proportional Representation," *Journal of Political Economy*, 108(6), pp. 1235-69.

Austen-Smith, David and Jeffrey Banks (2005), *Positive Political Theory II*, University of Michigan Press.

Azzimonti, Marina (2009), "Barriers to investment in polarized societies" University of Texas, mimeo.

Bachman Rudiger and Jinhui Bai (2010), "Government Purchases Over the Business Cycle: the Role of Heterogeneity andWealth Bias in Political DecisionMaking," Georgetown University, mimeo.

Bai Jinhui and Roger Lagunoff (2010), "On the Faustian Dynamics of Policy and political Power," *Review of Economic Studies*, forthcoming.

Battaglini Marco (2009), "Dynamic Electoral Competition and Constitutional," Princeton University, mimeo.

Battaglini, Marco. and Stephen Coate, (2008), "A Dynamic Theory of Public Spending, Taxation and Debt," *American Economic Review*, 98(1), 201-236.

Battaglini Marco, Salvatore Nunnari and Thomas Palfrey (2010), "Political Institutions and the Dynamics of Public Investment," Princeton University, mimeo.

Beck paul Allen and Frank Souauf (1992), *Party Politics in America.* 7th edition, New York: Harper Collins.

Besley T. and S. Coate (1998), "Sources of Inefficiency in a Representative Democracy: A Dynamic Analysis, " *American Economic Review*, 88(1), pp. 139-56.

Besley T. and T. Persson (2010), "State capacity, conflict and development," *Econometrica*, forthcoming.

Converse Phillip (1976), *The Dynamics of Party Support: Cohort-Analysing Party Identification.* Sage Publications.

Carmines, Edward and James Stimsons (1989), *Issues Evolution: Race and the Transformation of American Politics*, Princeton: Princeton University Press.

Cox, Gary (1994), "Strategic Voting Equilibria under the Single Nontransferable Vote," *American Political Science Review*, 88(3), pp. 608-21.

Franklyn, Charles and John Jackson (1982), “The Dynamics of Party Identification,” *American Political Science Review*, 957-973, 77:4.

Jackson, Kark (1975), “Issues, Party Choices and Presidential Votes,” *American Journal of Political Science*, 19, pp161-185.

Kalandrakis, Tasos (2004), “A Three-player Dynamic Majoritarian Bargaining Game,” *Journal of Economic Theory*, 116(2), pp. 294-322.

Ladd, Everet and Charles Adley (1978), *Transformation of the American Party System*, 2nd edition, New York: Norton.

Lagunoff, Roger (2008), “ Markov Equilibrium in Models of Dynamic Endogenous Political Institutions,” Georgetown University, mimeo.

Lindbeck, Assar and Weibull, Jurgen. “Balanced-Budget Redistribution as the Outcome of PoliticalCompetition.” *Public Choice*, 1987, 52, pp.273-97.

Lizzeri, Alessandro (1999), “Budget Deficits and Redistributive Politics,” *Review of Economic Studies*, 66(4), pp. 909-28.

Lizzeri, Alessandro and Nicola Persico (2001), “The Provision of Public Goods under Alternative Electoral Incentives,” *American Economic Review*, March 2001, v. 91(1), pp. 225-39.

MacKuen, Michael, Robert Erikson and James Stimson (1989), “Macropartisanship,” *American Political Science Review*, 1125-1142, 83:4.

Morelli, Massimo (2004), “Party Formation and Policy Outcomes under Different Electoral Systems,” *Review of Economic Studies*, 71(3), pp. 829-53.

Myerson, Roger (1993), “Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems.” *American Political Science Review*, 87(4), pp. 856-69.

Myerson, Roger (1999), “Theoretical Comparisons of Electoral Systems,” *European Economic Review*, 43(4-6), pp. 671-97.

Persson, Torsten and Tabellini, Guido (1999), “The Size and Scope of Government: Comparative Politics with Rational Politicians,” *European Economic Review*, 43(4-6), pp.699-735.

Persson, Torsten and Tabellini, Guido (2003), *The Economic Effects of Constitutions*, Munich Lectures in Economics. Cambridge and London: MIT Press.

Persson, Torsten and Tabellini, Guido (2004), “Constitutional Rules and Fiscal Policy Outcomes,” *American Economic Review*, v. 94, iss. 1, pp. 25-45.

Sleet Christopher and Sevin Yeltekin (2008), “Politically Credible Social Insurance, ” *Journal of Monetary Economics* 55, 129-151.

Shiryaev, A.N. (1991), *Probability*, New York: Springer-Verlag.

Smart, D. R. (1974), *Fixed Point Theorems*, Cambridge University Press.

Stokey, Nancy, Robert Lucas and Edward Prescott. 1989. *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

Stromberg, David (2008), "How the Electoral College Influences Campaigns and Policy: The Probability of Being Florida," *American Economic Review*, 98(3), pp. 769-807.

Yared, Pierre (2010), "Politicians, Taxes and Debt," *Review of Economic Studies*, 77, pp. 806-840.

8 Appendix

8.1 Assumptions 1 and 2

In this section we formally derive and motivate Assumptions 1 and 2. Consider Assumption 1 first. Let $v^*(b,A)$ be the value function when the policies are chosen by a benevolent utilitarian planner, as discussed in Section 3 (assuming $\mu_i = 1/n$ for all i). Define

$$W^*(b) = \max_{(r,g,x,s)} \left\{ \begin{array}{c} [u(r,g;A) + s_i + \delta v^*(x)] \\ s.t. \;\; s_i \geq 0 \;\; \text{for all } i, \; \sum_i s_i \leq B(r,g,x;b), \; \& \; x \in [\underline{x}, \overline{x}] \end{array} \right\}$$

where $v^*(x) = Ev^*(x, A')$. This is the maximal expected utility a citizen may receive in a symmetric equilibrium. Note that both $v^*(b)$ and $W^*(b)$ are bounded, and $W^j(p_L; v, A) - W^j(p_R; v, A) \leq \Delta W^* = W^*(\underline{x})$. It follows that:

$$\frac{1}{2} - \overline{\theta}\left[\Delta W^* + 1/2\underline{\theta}^G\right] \leq \frac{1}{2} + \theta^j \left[\Delta W^j(p; v, A) - \widetilde{\theta}^G\right] \leq \frac{1}{2} + \overline{\theta}\left[\Delta W^* + 1/2\underline{\theta}^G\right]$$

where $\Delta W^j(p; v, A) = W^j(p_L; v, A) - W^j(p_R; v, A)$. We conclude that $\frac{1}{2} + \theta^j\left[\Delta W^j(p; v, A) - \widetilde{\theta}^G\right]$ is in $(0,1)$ if the first inequality in Assumption 1 is satisfied. A candidate wins 50% of the votes if and only if $\widetilde{\theta}^G \leq \Delta W^j(p; v, A)$. The probability of this event is in $(0,1)$ if $1/2\overline{\theta}^G > \Delta W^*$, the second inequality in Assumption 1. Consider Assumption 2 presented in Section 6. Following the same steps as above, we can show that candidate L (respectively, R) wins less that $1/2$ of the votes in a right-wing (respectively, left-wing) region if $\frac{1}{2} + \overline{\theta}\left[\Delta W^* - \sigma + 1/2\underline{\theta}^G\right] < 1/2$, which implies the first inequality in Assumption 2. By a similar logic, we can show that the share of votes received by any candidate is always nonnegative in any region if the second inequality of Assumption 2 is satisfied.

8.2 Proof of Proposition 1

Since it is relatively standard, the proof that a well-behaved Pareto efficient solution exists is omitted. Let $V^o(b) = \sum_i \mu_i v_i^\circ(x)$ be the social welfare function. To derive conditions (7)-(9), note that from the first order conditions of (5) with respect to g and x we have:

$$\begin{aligned} \alpha A g^\circ(b;A)^{\alpha-1} &= \frac{1}{n}\frac{1 - r^\circ(b;A)}{1 - r^\circ(b;A)(1+\varepsilon)}; \\ \frac{1}{n}\frac{1 - r^\circ(b;A)}{1 - r^\circ(b;A)(1+\varepsilon)} &\geq -\delta\left[V^o\right]'(x^o(b,A)) \text{ with equality if } x^o(b,A) < \overline{x}; \end{aligned} \tag{28}$$

and from the first order conditions with respect to r, we have:

$$\frac{1}{n}\frac{1-r^{\circ}(b;A)}{1-r^{\circ}(b;A)\,(1+\varepsilon)} \geq \mu^{*} \text{ with equality if } B(r,g,x;b)>0. \tag{29}$$

To prove (8), we now show that (28) is satisfied as equality. We wish to prove that for any state (b,A) with $b<\overline{x}$ there is an $\epsilon\,(b,A)>0$ such that $x^{o}(b;A)<\overline{x}-\epsilon\,(b,A)$. Assume that there is a state (b,A) such that $x^{o}(b;A)$ is arbitrarily close to $\overline{x}$; that is, $x^{o}(b,A)=\overline{x}-\zeta$, where ζ is arbitrarily small. We can write $g^{o}(b;A)=\phi\,(r(b;A))$ where $\phi(r)$ is a continuous function implicitly defined by the solution of the equation $\alpha nAg^{\alpha-1}=[\frac{1-r}{1-r(1+\varepsilon)}]$. Since $x^{o}(b;A)$ is arbitrarily close to $\overline{x}$, we must have:

$$B(r^{o}(b;A),\phi\,(r^{o}(b;A))\,,x^{o}(b;A);b)=0. \tag{30}$$

Indeed, assume by contradiction that there is a sequence $\zeta^{\tau}\longrightarrow 0$ such that (30) is not true. Then along this sequence r and g are given by $\frac{1}{n}\frac{1-r^{\circ}(\mu)}{1-r^{\circ}(\mu)(1+\varepsilon)}=\mu^{*}$ and $\alpha Ag^{\circ}(\mu)^{\alpha-1}=\mu^{*}$. But then there must be a $\overline{\tau}$ such that for $\tau>\overline{\tau}$ we have $B\left(r^{\circ}(\mu),g^{\circ}(\mu),\overline{x}-\zeta^{\tau}\right)<0$, a contradiction. Thus, by (30) we can express all the policy choices as a function of ζ, where $x^{o}(b,A)=\overline{x}-\zeta=x(\zeta)$, $r^{o}(b;A)=r\,(\zeta)$ solves (30) and $g^{o}(b;A)=\phi\,(r\,(\zeta))=g\,(\zeta)$. Note that as $\zeta\to 0$, we have $r\,(\zeta)\to\widetilde{r}<1/(1+\varepsilon)$. For if $r\,(\zeta)\to 1/(1+\varepsilon)$, then $g\,(\zeta)\to 0$ and (30) would not be satisfied since $b<\overline{x}$. Moreover, $r\,(\zeta)\to\widetilde{r}$ implies $g\,(\zeta)\to\widetilde{g}>0$. From the first order condition on debt, we have that:

$$\begin{aligned}
\left(\frac{1-r\,(\zeta)}{1-r\,(\zeta)\,(1+\varepsilon)}\right) &\geq -\beta\,[V^{o}]'\,(x(\zeta)) \\
&\geq -\beta f(\overline{A})\,[V^{o}]'\,(x(\zeta);\overline{A})=-\beta f(\overline{A})\left(\frac{1-r(x(\zeta),\overline{A})}{1-r(x(\zeta),\overline{A})(1+\varepsilon)}\right).
\end{aligned}$$

It is easy to see that, for any A, $r(x(\zeta),A)\to 1/(1+\varepsilon)$ as $\zeta\to 0$. This implies that the right hand side of the previous inequality diverges to infinity, while the left hand side converges to a finite value: a contradiction. Given this, to prove (9) note that using the envelope theorem we can show that:

$$-\delta\,[V^{o}]'\,(x^{o}(b,A))=E\left[\frac{1-r^{\circ}(x^{o}(b,A);A')}{1-r^{\circ}(x^{o}(b,A);A')\,(1+\varepsilon)}\right].$$

The result then follows from (7).

We now prove Proposition 1. From condition (9) it can be seen that the tax rate $r^{o}(b,A)$ defines a submartingale $(r^{o}_{\tau})_{\tau>0}$ with $\sup_{\tau}|r_{\tau}|<\infty$. By Theorem 1 in Shiryaev (1991, ch.

VII.4), the limit $\lim r^o_\tau = r^o_\infty$ exists with probability one. From *(7)*, it follows that the limit $\lim g^o_\tau = g^o_\infty$ exists with probability one as well, where the random sequence $(g^o_\tau)_{\tau>0}$ is defined by $g^o(b, A)$. In these limits, it must be that the sequence of Lagrangian multipliers $(\lambda_\tau)_{\tau>0}$ defined by the Lagrangian multiplier $\lambda^o(b, A)$ of the constraint $B(r, g, x; b) \geq 0$ in (5) converges to zero: otherwise the tax rate would depend on the realization of the shock A for an arbitrarily large τ. Conditions *(7)-(8)* imply that the tax rate converges to $r_\mu = \frac{n\mu^*-1}{n\mu^*(1+\varepsilon)-1}$ and that g converges to a function $g_\mu(A)$ such that: $\alpha A\left[g_\mu(A)\right]^{\alpha-1} = \mu^*$. Such a solution is possible if and only if, in the steady state, debt is lower or equal than x_μ, where x_μ is defined by $B(r_\mu, g_\mu(\overline{A}), x_\mu; x_\mu) = 0$, and therefore is such that $x_\mu > \underline{x}$. Let n^* be the number of agent that have maximal μ, and let $s^\circ_i(A, x) = \frac{B(r_\mu, g_\mu(A), x; x)}{n^*}$ if $\mu_i = \mu^*$, $s^\circ_i(A, x) = 0$ otherwise. It is easy to verify that any r_μ, $g_\mu(A)$ and $x \in [\underline{x}, x_\mu]$ with corresponding pork transfers $(s^\circ_i(A, x))^n_{i=1}$ is a steady state of the planner's solution. ■

8.3 Proof of Proposition 2

Let F be the metric space of real valued continuous and bounded, weakly concave functions of b in $[\underline{x}, \overline{x}]$ endowed with the sup norm, $\|f\| = \sup_{b\in[\underline{x},\overline{x}]} |f|$. It is useful to represent a state in the compact form as $z = \left(A, \left(\theta^j\right)_j, \theta^G\right)$. Following the discussion in Section 4.1, the problem of a candidate or a given value function v can be expressed as (16a). Let $p(b; v, z) = \{r(b; v, z), g(b; v, z), x(b; v, z)\}$ be a solution to (16a), and let $P(b; v, z) \subseteq P$ be the set of solutions, where $P = [0, 1] \times [0, \overline{g}] \times [\underline{x}, \overline{x}]$ is the feasible policy space (note that we can assume $g \in [0, \overline{g}]$ without loss of generality since in every period resources are bounded by $\max_r R(r) + \overline{x}$). Define $P^*(b; v, z)$ as the set of policies $p^*(b; v, z) = \{r^*(b; v, z),\ g^*(b; v, z),\ x^*(b; v, z)\} \in P(b; v, z)$ such that: 1. $p^*(b; v, z)$ is continuous and monotonic in b, where monotonic means that $r^*(b; v, z)$, $x^*(b; v, z)$ are non-decreasing in b and $g^*(b; v, z)$ is non increasing in b; and 2. such that $B(r^*(b; v, z), g^*(b; v, z), x^*(b; v, z); b)$ is weakly convex (and continuous) in b.

Lemma A.2. *In all states b* and *z, $P^*(b; v, z)$ is a non empty, compact and convex valued correspondence from F* into *P.*

Proof. We proceed in two steps.

Step 1. We first show $P^*(b; v, z)$ is non empty. Fix a state z and let $\lambda(b; v, z)$ be the Lagrangian multiplier associated to the constraint $B(r, g, x; b) \geq 0$ in (16a). Let $\Omega_m(\theta) =$

$\frac{1}{m}\max_j\left[\theta^j/\sum_{l\in H(j)}\theta^l\right]$ and, for any state $z=(A,\theta)$, define r_z^* and g_z^* as

$$[\frac{1-r_z^*}{1-r_z^*(1+\varepsilon)}]=n\Omega_m(\theta)\,,\text{ and }\alpha A\left(g_z^*\right)^{\alpha-1}=\Omega_m(\theta)\,.$$

and let

$$\mathcal{X}_z(v)=\arg\max_{x\in[\underline{x},\overline{x}]}\left\{\Omega_m(\theta)\,x+\delta v(x)\right\}. \tag{31}$$

If $\lambda(b;v,z)=0$, then the solution of (16a) is $r(b;v,z)=r_z^*$, $g(b;v,z)=g_z^*$ and $x(b;v,z)\in\mathcal{X}_z(v)\cap\{x\,|B(r,g,x;b)\geq 0\}$. Let $\overline{x}(v,z)=\max_x\{x\,|x\in\mathcal{X}_z(v)\}$. It is immediate to verify that $\lambda(b;v,z)=0$ for all $b\leq\overline{b}(\overline{x}(v,z);z)$, where $\overline{b}(x;z)$ is defined by $B(r_z^*,g_z^*,x;\overline{b}(x;z))=0$ $\forall x\in[\underline{x},\overline{x}]$. Since v is weakly concave, $\mathcal{X}_z(v)$ is a non empty, compact and convex set. Choose a $x_z^*\in[\underline{x},\overline{x}(v,z)]$ and define $\widetilde{x}(b;z)$ as the solution of $B(r_z^*,g_z^*,\widetilde{x}(b;z);b)=0$. Note that $\widetilde{x}(b;z)$ is linear in b; moreover we have: $\overline{b}(x_z^*;z)\leq\overline{b}(\overline{x}(v,z);z)$. Set

$$x^*(b;v,z)=\begin{cases} x_z^* & b\leq\overline{b}(x_z^*;z) \\ \widetilde{x}(b;z) & b\in\left[\overline{b}(x_z^*;z),\overline{b}(\overline{x}(v,z);z)\right]\end{cases}.$$

It is easy to see that $x^*(b;v,z)\in\mathcal{X}_z(v)\cap\{x\,|B(r,g,x;b)\geq 0\}$ for any $b\in\left[\underline{x},\overline{b}(\overline{x}(v,z);z)\right]$: so r_z^*,g_z^*, $x^*(b;v,z)$ is a solution of (16a) in $b\in\left[\underline{x},\overline{b}(\overline{x}(v,z);z)\right]$. Moreover we have

$$B(r_z^*,g_z^*,x^*(b;v,z);b)=\max\left\{0,B(r_z^*,g_z^*,x_z^*;b)\right\}$$

for any $b\in\left[\underline{x},\overline{b}(\overline{x}(v,z);z)\right]$. Note, finally, that $\widetilde{x}(\overline{b}(\overline{x}(v,z);z);z)=\overline{x}(v,z)$.

Consider now $b>\overline{b}(\overline{x}(v,z);z)$. In this case $\lambda(b;v,z)>0$, so (16a) can be written as:

$$\max_{(r,g,x)}\left\{\begin{array}{c} u(r,g;A)+\delta v(x) \\ \text{s.t. } B(r,g,x;b)=0\ \&\ x\in[\underline{x},\overline{x}]. \end{array}\right\}. \tag{32}$$

This problem admits a unique solution, so $P(b;v,z)$ is non empty, compact and convex valued and continuous in for all $b\in(\overline{b}(\overline{x}(v,z)),\overline{x}]$. Since the solution of (16a) is upper hemicontinuous in $\left[\overline{b}(\overline{x}(v,z)),\overline{x}\right]$ it must be that for any sequence $(b_n)_{n=1}^{\infty}$ with $b_n>b(v,z)$ converging to $b(v,z)$, we have $\lim_{n\to\infty}p(b_n;v,z)\in P(\overline{b}(\overline{x}(v,z));v,z)$. Since $P(\overline{b}(\overline{x}(v,z));v,z)$ is a singleton, this implies that $P(b;v,z)$ coincides with $r^*(b;v,z),g^*(b;v,z),x^*(b;v,z)$ at $\overline{b}(\overline{x}(v,z))$. We can therefore extend $r^*(b;v,z),g^*(b;v,z),x^*(b;v,z)$ to $b\in(\overline{b}(\overline{x}(v,z)),\overline{x}]$, by setting it equal to the solution of (32). The resulting policy $r^*(b;v,z),g^*(b;v,z),x^*(b;v,z)$ is continuous in $[\underline{x},\overline{x}]$. It can also be verified that $r^*(b;v,z), x^*(b;v,z)$ are non decreasing in b, and $g^*(b;v,z)$ is non increasing in b. Finally, $B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z);b) = \max\{0, B(r_z^*, g_z^*, x_z^*;b)\}$, which is weakly convex in b for any $b \in [\underline{x}, \overline{x}]$.

Step 2. Consider a sequence of policies $(p^n)_{n=1}^{\infty}$ such that $p^n \in P^*(b;v,z)\ \forall n$, and define $p^\infty = (r^\infty, g^\infty, x^\infty) = \lim_{n\to\infty} p^n$. To see that $P^*(b;v,z)$ is compact, note that by the theorem of the maximum, $p^\infty \in P(b;v,z)$; moreover p^∞ is continuous, monotonic and $B(r^\infty(b), g^\infty(b), x^\infty(b);b)$ is weakly convex. It follows that $p^\infty \in P^*(b;v,z)$. To see convexity, let $p^1(b) \in P^*(b;v,z)$, $p^2(b) \in P^*(b;v,z)$, and $p^\phi(b) = \phi p^1(b) + (1-\phi)p^2(b)$ for some $\phi \in (0,1)$. Since $v \in F$ it must be that $p^\phi(b) \in P(b;v,z)$; moreover $p^\phi(b)$ is continuous, monotonic and $B(r^\phi(b), g^\phi(b), x^\phi(b);b)$ is weakly convex. So $p^\phi \in P^*(b;v,z)$. ■

For a given $v \in F$ and state z, let $\Upsilon(v,z)$ be defined by:

$$\Upsilon(v,z) = \left\{ \widetilde{v} \;\middle|\; \begin{array}{c} \widetilde{v}(b) = u(r(b), g(b); A) + \frac{B(r(b),g(b),x(b);b)}{n} + \delta v\left(x(b)\right) \\ \text{for some } \left(r(b), g(b), x(b)\right) \in P^*(b;v,z) \end{array} \right\}.$$

This expression defines a correspondence with the following properties:

Lemma A.3. For any given z, $\Upsilon(v,z)$ *is a non empty, convex and compact valued correspondence from* F *into* F.

Proof. For any given $(r(b), g(b), x(b)) \in P^*(b;v,z)$, we can write:

$$\begin{aligned}
\widetilde{v}(b) &= u(r^*(b;v,z), g^*(b;v,z);A) + \frac{B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z);b)}{n} \\
&\quad + \delta v\left(x^*(b;v,z)\right) && (33) \\
&= u(r^*(b;v,z), g^*(b;v,z);A) + \Omega_m(\theta)\, B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z);b) && (34) \\
&\quad + \delta v\left(x^*(b;v,z)\right) + \left(\frac{1}{n} - \Omega_m(\theta)\right) B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z);b) && (35) \\
&= \max_{(r,g,x)} \left\{ \begin{array}{c} u(r,g;A) + \Omega_m(\theta)\, B(r,g,x;b) + \delta v(x) \\ \text{s.t. } B(r,g,x;b) \geq 0 \ \&\ x \in [\underline{x}, \overline{x}]. \end{array} \right\} && (36) \\
&\quad + \left(\frac{1}{n} - \Omega_m(\theta)\right) B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z);b). && (37)
\end{aligned}$$

It can be verified that the first term of the expression above is concave. By Lemma A.2, $B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z);b)$ is weakly convex. Since $\left(\frac{1}{n} - \Omega_m(\theta)\right) \leq 0$, the second term of (37) is weakly concave. It follows that $\widetilde{v}(b)$ is bounded, continuous and weakly concave in b. Since $P^*(b;v,z)$ is non empty, then $\Upsilon(v,z)$ is a non empty correspondence of F^q into F.

To show that $\Upsilon(v,z)$ is convex valued, consider two functions $v_1 \in \Upsilon(v,z)$, $v_2 \in \Upsilon(v,z)$, and let $v_\alpha = \alpha v_1 + (1-\alpha)\, v_2$. Let $\lambda(b;v,z)$ be the Lagrangian multiplier associated to the constraint $B(r,g,x;b) \geq 0$ in (16a). If $\lambda(b;v,z) > 0$, then $v_1 = v_2 = v_\alpha$. If $\lambda(b;v,z) = 0$, then

$$\begin{aligned} v_\alpha(b) &= u(r_z^*, g_z^*; A) + \left(\frac{1}{n} - \Omega_m(\theta)\right) B(r,g,x_z^\alpha;b) + \alpha\left[\Omega_m(\theta)\, B(r_z^*, g_z^*, x_z^1; b)) + \delta v\left(x_z^1\right)\right] \\ &\quad + (1-\alpha)\left[\Omega_m(\theta)\, B(r_z^*, g_z^*, x_z^2; b)) + \delta v\left(x_z^2\right)\right] \end{aligned}$$

where, for any $i = 1, 2$, x_z^i is the optimal value of x associated with v_i, and $x_z^\alpha = \alpha x_z^1 + (1-\alpha)\, x_z^2$. Since v is weakly concave, it must be that $\Omega_m(\theta)\, B(r_z^*, g_z^*, x; b) + \delta v(x)$ is constant in $\left[x_z^1, x_z^2\right]$, so we can write: $v_\alpha(b) = u(r_z^*, g_z^*; A) + \frac{B(r_z^*, g_z^*, x_z^\alpha; b)}{n} + \delta v\left(x_z^\alpha\right)$. Since $(r_z^*, g_z^*, x_z^\alpha) \in P^*(b;v,z)$ by Lemma A.2, we have $v_\alpha(b) \in \Upsilon(v,z)$. The fact that $\Upsilon(v,z)$ is compact valued follows from the compactness of $\{B' \,|\exists (r,g,x) \in P^*(b;v,z) \text{ s.t. } B' = B(r,g,x;b)\}$.

Using standard methods we can also show that:

Lemma A.4. For any given z, $\Upsilon(v,z)$ *has a closed graph.*

Proof. The proof of this result is available form the author upon request. ■

Let $\eta(z)$ be the probability of z (that is well defined given $f, \left(\varphi^j\right)_j, \varphi^G$). Define the correspondence from F into F:

$$T(v) = \left\{ (\widetilde{v}(b) \left| \exists l(b;z) \in \Upsilon(v,z) \text{ s.t. } \widetilde{v}(b) = \sum\nolimits_z l(b;z)\eta(z) \right. \right\}.$$

We have:

Lemma A.5. *The correspondence $T(v)$ admits a fix point $v^* = T(v^*)$.*

Proof. It can be verified that Lemma A.3 implies that $T(v)$ is a non empty, compact and convex valued correspondence form F to F with a closed graph. The result therefore follows from the Glicksberg-Fan Theorem (see Theorem 9.2.2 in Smart [1974]). ■

For any $v^* = T(v^*)$, let $l^*(b;z) \in \Upsilon(v^*, z)$ be the associated functions such that for any A: $v^*(b) = \sum\nolimits_z l^*(b;z)\eta(z)$; and $(r^*(b;v^*,z), g^*(b;v^*,z), x^*(b;v^*,z)) \in P^*(b;v^*,z)$, the associated policy such that:

$$l^*(b;z) = u(r^*(b;v,z), g^*(b;v,z); A) + \frac{B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z); b)}{n} + \delta v\left(x^*(b;v,z)\right).$$

Moreover let $\mathcal{H}(\theta) = \left\{j \left|\theta^j \geq \theta^i \; \forall i = 1, ..., n\right.\right\}$, with cardinality $|\mathcal{H}(\theta)|$; and define $s(b; v^*, z) = \left\{s^1(b; v^*, z), ..., s^n(b; v^*, z)\right\}$, with $s^j(b; v^*, z) = 0 \; \forall j \notin \mathcal{H}(\theta)$ and

$$s^j(b; v^*, z) = \frac{B(r^*(b; v, z), g^*(b; v, z), x^*(b; v, z); b)}{|\mathcal{H}(\theta)|}$$

otherwise. Then $r^*(b; v^*, z), g^*(b; v^*, z), x^*(b; v^*, z), s^*(b; v^*, z)$ is an optimal strategy for a candidate when v^* is the citizen's value function. Moreover, when the optimal strategy is $r^*(b; v^*, z)$, $g^*(b; v^*, z)$, $x^*(b; v^*, z)$, $s^*(b; v^*, z)$, then the expected value function is v^*. We conclude that $v^*(b)$, $r^*(b; v^*, z)$, $g^*(b; v^*, z)$, $x^*(b; v^*, z)$, $s^*(b; v^*, z)$ is a political equilibrium.

8.4 Proof of Proposition 3

For a given equilibrium value function v the properties of the policy functions depend on the solution of (16a). Let $\lambda(b; z)$ be the Lagrangian multiplier of the first constraint in this problem in state b, z. If $\lambda(b; z) > 0$, then (16a) admits a solution $(r(b, A), g(b, A), x(b, A))$ independent of $\left(\theta^j\right)_j, \theta^G$, with $r(b, A), x(b, A)$ increasing in b and A, and $g(b, A)$ increasing in A and decreasing in b. Consider now the case $\lambda(b; z) = 0$. We start from a preliminary result. For a state z, let $\mathcal{X}_z(v)$ be defined as in (31). We have:

Lemma A.5. *The set* $\mathcal{X}_z(v)$ is a singleton for any z.

Proof. Assume not. Since v is weakly concave, there must be constants $\underline{\beta}, \overline{\beta}$ such that $\mathcal{X}_z(v) = \left[\underline{\beta}, \overline{\beta}\right]$. Let $\widehat{x} = \min_z \left[R(r^*_z) - g^*_z\right]/\rho$. Assume first $\widehat{x} < \overline{\beta}$. Then there must be a z' such that $R(r^*_{z'}) - g^*_{z'} - \rho\overline{\beta} < 0$, and so $R(r^*_{z'}) - g^*_{z'} + x(b; z') - (1 + \rho)\,\overline{\beta} < 0$ for any $x(b; z') \leq \overline{\beta}$. Then there is an $\varepsilon > 0$ such that for all $b \in \left[\overline{\beta} - \varepsilon, \overline{\beta}\right]$: $R(r^*_{z'}) - g^*_{z'} + x(b; z') - (1 + \rho)\, b < 0$. It follows that in state z', for any $b \in \left[\overline{\beta} - \varepsilon, \overline{\beta}\right]$ we have $B(r(b; z'), g(b; z'), x(b; z'); b) = 0$, and we can write:

$$v(b; z') = \max_{(r,g,x)} \left\{ \begin{array}{c} u(r, g; A) + B(r, g, x; b) + \delta v\,(x) \\ \\ \text{s.t. } B(r, g, x; b) \geq 0 \;\&\; x \in [\underline{x}, \overline{x}]. \end{array} \right\}$$

for all $b \in \left[\overline{\beta} - \varepsilon, \overline{\beta}\right]$. It follows that $v(b; z')$ is strictly concave in $\left[\overline{\beta} - \varepsilon, \overline{\beta}\right]$, and therefore $\delta v\,(x)$ is strictly concave in $\left[\overline{\beta} - \varepsilon, \overline{\beta}\right]$ as well. This implies that it is not possible that $\left[\overline{\beta} - \varepsilon, \overline{\beta}\right] \subseteq \mathcal{X}_z(v)$, a contradiction.

We conclude that $\widehat{x} \geq \overline{\beta}$. In this case for all states z and $b \in \left[\underline{\beta}, \overline{\beta}\right]$ we have $x(b; z) \in \left[\underline{\beta}, \overline{\beta}\right]$ is optimal and $B(r_z^*, g_z^*, x(b;z); b) \geq 0$, with strict inequality for some state b, z. We can write:

$$v(x(b;z);z') = \max_{(r,g,x)} \left\{ \begin{array}{c} u(r,g;A') + \Omega_m(\theta') B(r,g,x;x(b;z)) + \delta v(x) \\ \text{s.t. } B(r,g,x;x(b;z)) \geq 0 \ \& \ x \in [\underline{x}, \overline{x}]. \end{array} \right\} \tag{38}$$
$$- \left[\Omega_m(\theta') - \frac{1}{n}\right] \cdot B(r_z^*, g_z^*, x(x(b;z);z'); x(b;z)).$$

Let's choose a θ' such that $\Omega_m(\theta') - \frac{1}{n} > 0$. This is always possible since $\underline{\theta} < \overline{\theta}$. Since v is concave, the previous expression has a right and left derivative. Since for all $b \in (\underline{\beta}, \overline{\beta})$, $b \in \mathcal{X}(V)$, the right and left derivatives with respect to b of $\Omega_m(\theta') b + \delta v(b)$ must be zero, implying that it must be differentiable. By the Envelope Theorem the derivative of the first term in (38) is $-(1+\rho)\Omega_m(\theta')$ (since the constraint $B(r,g,x;x(b;z)) \geq 0$ is not binding). The derivative of the second is $\left[\Omega_m(\theta') - \frac{1}{n}\right][(1+\rho) - Edx(b;z)/db]$, where $Edx(b;z)/db$ is the derivative of $x(b;z)$ at b (it must be differentiable otherwise $v(b)$ would not be differentiable at $x(b;z)$). We must therefore have the first order necessary condition:

$$\Omega_m(\theta') + \delta\left[-(1+\rho)\Omega_m(\theta') + \left[\Omega_m(\theta') - \frac{1}{n}\right][(1+\rho) - Edx(b;z)/db]\right] = 0 \tag{39}$$

Rewriting (39) and using the fact that $\delta(1+\rho) = 1$, we have: $Edx(b;z)/db = 1/\delta > 1$ for any $b \in [\underline{\beta}, \overline{\beta}]$. This implies a contradiction. To see this, note first that $Ex(\underline{\beta};z') \geq \underline{\beta}$, since $x(\underline{\beta};z') \geq \underline{\beta}$ for all z'. This observation, plus $Edx(b;z)/db > 1$ implies that there is a z such that $x(\overline{\beta};z) > \overline{\beta}$: but then, again, $x(\overline{\beta};z) \notin \mathcal{X}(V)$, a contradiction. We conclude that $\mathcal{X}(V)$ is a singleton. ■

By Lemma A.5 this solution $x_m(\theta)$ of (18) is uniquely defined. If $\lambda(b;z) = 0$, the solution of (16a) is then $r_m(\theta), g_m(A,\theta), x_m(\theta)$, given by (17)-(18). Note that $r(b,A) \geq r_m(\theta)$, $g(b,A) \leq g_m(A,\theta)$ & $x(b,A) \geq x_m(\theta)$ for any b, z. The case $\lambda(b;z) = 0$ is possible if and only if $B(r_m(\theta), g_m(A,\theta), x_m(A,\theta); b) \geq 0$. If we define $A_m(\theta, b)$ from $B(r_m(\theta), g_m(A_m(\theta,b),\theta), x_m(A,\theta); b) = 0$. We conclude that the solution will be $r_m(\theta)$, $g_m(A,\theta)$, $x_m(A,\theta)$ for $A \leq A_m(\theta,b)$ and $r(b,A)$, $g(b,A)$, $x(b,A)$ otherwise, as stated in the proposition. ■

8.5 Proof of Proposition 4

The fact that the marginal cost of public funds is a submartingale follows from the discussion in Section 5. We now show that for any initial state, there is a positive probability of reaching a state z, b in which the martingale is strict. From the discussion in Section 5 it is clear that this is proven if we show that *for any initial state* A_0, h_0, b_0*, the set* $S = \{z \,|B(b;z) > 0\}$ *is reached with probability one* (where $B(b;z) = B(r(b;z), g(b;z), x(b;z); b)$). *To this goal we show that* there is an $\varepsilon > 0$ and a $T \geq 1$ such that for any initial state z_0,b_0: $\Pr\left\{(z_t)_{t=0}^T \,|B(b_T; z_T) > 0\right\} \geq \varepsilon$. Assume not. Then, since $B(b,z) = 0$ with probability one, we can recursively write the value function as:

$$v(b; A, h) = \max_{(r,g,x)} \left\{ \begin{array}{c} u(r, g; A) + \frac{B(r,g,x;b)}{n} + \delta E\left[v\left(x; A', h'\right) |A\right] \\ \text{s.t. } B(r, g, x; b) \geq 0 \;\&\; x \in [\underline{x}, \overline{x}]. \end{array} \right\}. \tag{40}$$

Note that (40) is a contraction. It admits a unique fixpoint $v(b; A)$ that corresponds to the planner's solution in the case in which $\mu_i = 1/n$ for all i. By Proposition 1, then, we have that the tax rate converges to 0, g to $g_{1/n}(A)$ and debt to $\underline{x}$. In correspondence to this solution, moreover $B(0, g_{1/n}(A), \underline{x}; b) > 0$, a contradiction. ■

8.6 Proof of Proposition 5

Define the state space $\mathcal{S} = [\underline{x}, \overline{x}]$ with associated σ-algebra $\mathcal{B}$, where $\mathcal{B}$ is the family of Borel sets that are subsets of $[\underline{x}, \overline{x}]$. For any subset $S \in \mathcal{B}$, let $\mu_t(S)$ denote the probability that b lies in S in period t. The probability distribution μ_1 depends on the initial level of debt b_0. Let $Q(S\,|b)$ be the probability that a set S is reached in one step if the initial state is b. The probability distribution μ_t $\forall t \geq 2$ is defined inductively as: $\mu_t(S) = \int_b Q(S\,|b)\mu_{t-1}(db)$. The probability distribution μ^* is an invariant distribution if: $\mu^*(S) = \int_b Q(S\,|b)\mu^*(db)$. We now show that the sequence of distributions $\langle \mu_t \rangle_{t=1}^{\infty}$ converges strongly to a unique invariant distribution. Let $Q^1(S\,|b) = Q(S\,|b)$ and define recursively: $Q^n(S\,|b) = \int_{b'} Q(S\,|b')Q^{n-1}(db'\,|b)$. Thus, $Q^n(S\,|b)$ is the probability that a set S is reached in n steps if the initial state is b. With standard methods we can prove:

Lemma A.5 *There exists a state* (x^*)*, an integer* $N \geq 1$ *and a number* $\varepsilon > 0$*, such that for any initial state* b*,* $Q^N(x^*\,|b) > \varepsilon$*.*

Proof. The proof of this result is available form the author upon request. ■

Lemma A.5 implies the transition function Q satisfies the Doeblin's condition (see Condition D in Section 11.4, Stokey, Lucas and Prescott [1989]). The statement of Proposition 5 therefore follows by Theorem 11.12 in Stokey, Lucas and Prescott [1989]. There are two cases to consider.

Case 1: $m < n$**.** The fact that the support of the distribution is as described in the proposition follows from the discussion in Section 4.2. To see that the distribution is non degenerate when $m <$

n, note that taxes and public expenditures can be constant only if the constraint $B(r, g, x; b) \geq 0$ in (16a) is never binding in the stationary distribtuion. This is true only if $A_m(b, \theta) = \overline{A}$, and so $-\delta v'(x_m(A, \theta); A) = 1/n$. But then we would have a contradiction with (17)-(18), since $\frac{1}{m} \max_j \left[\theta^j / \sum_{l \in H(j)} \theta^l\right] > 1/n$. The fact that the equilibrium can not be Pareto efficient is then implied by Proposition 1.

Case 2: $m = n$**.** Note that when $m = n$ the policies maximize a welfare function with weights $\mu_i = 1/n \; \forall i$. The properties of the equilibrium, therefore, follow from Proposition 1. ■

8.7 Proposition 6

As it can be easily verified, under assumption 2, the optimal reaction function of the candidates corresponds to the maximization of a welfare function in which $\mu_i = 1/\beta n$ if i is a neutral agent, and $\mu_i = 0$ otherwise. The result in the first bullet, therefore, follows from Proposition 1. For the result in the second bullet, note that under assumption 2, σ is irrelevant for the marginal effect of a policy on the fraction of votes received in a district. The candidates, therefore solve (16a) with $m = n$. The result follows from Proposition 4. ■