This revision: March 2013

A dynamic theory of electoral competition*

Abstract

We present a dynamic theory 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 platforms include 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.

Marco Battaglini
Department of Economics
Princeton University
Princeton NJ 08544
mbattagl@princeton.edu

*I thank Nemanja Antic, Edoardo Grillo Juan Ortner, and Giri Parameswaran for outstanding research assistance. For helpful comments, I also thank David Baron, Steve Coate, Lydia Mechtenberg, Martin Osborne, Torsten Persson, seminar participants at the Stanford/UWI Political Economy Conference (Kingston, February 2010), the CIRPEE Workshop on Political Economy (Quebec City, April 2010), the economics department of the University of Chicago and three anonymous referees. I am grateful for financial support from a NSF CAREER Award (SES-0547748), and the hospitality of the Einaudi Institute for Economics and Finance (EIEF) while working on this paper.

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 office-motivated politicians to choose Pareto optimal policies, and it has contributed to the understanding of how electoral rules determine public good provision, taxation and the shape of the 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 votes 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 parties 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' utilities, where the weights depend on the ex ante distribution of citizens' preferences (see, for example, Coughlin and Nitzan [1981] and 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

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

allocation policies are chosen so that the marginal cost of public funds at time t, a key measure of 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 be systematically smaller than the corresponding expected value at $t+1$. This distortion limits the ability of society to store resources and self insure against shocks to the economy: this induces steady state policies that are inefficiently volatile.

The intuition for why electoral competition induces inefficiently volatile policies is as follows. As in static models, to win an electoral district the parties identify the municipalities that are most contestable and choose policies that maximize the expected value functions of their citizens, under-weighting the welfare of the citizens of the other municipalities.[2] The favored municipalities, however, are unlikely to remain in this condition in the future: as preferences change over time, other municipalities will be more likely to be targeted by the parties. The municipalities 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 the preferences of their citizens time inconsistent. Since in every period parties choose policies that pander to time inconsistent districts, policies tend to be time inconsistent.

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 municipalities are ex ante symmetric: i.e., they all have the same ex ante distribution of citizens' preferences, and in each of them parties have the same ex ante likelihood of victory. In this case the analysis suggests that the smaller the electoral district is (in terms of the number of municipalities it comprises), the smaller the dynamic distortion is. Small districts reduce the parties' freedom in choosing municipalities to target with pork barrel 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 municipalities are ex ante heterogeneous, the choice of the optimal size of the electoral district is more complicated because an additional factor –which was previously pointed out in

[2] Here and in the reminder of the paper, we assume that each electoral district comprises one or more municipalities.

the static literature– becomes important. To study this point we consider a simple case in which municipalities can be divided in three categories: neutrals, in which both parties have the same probability of winning; rightists, in which the right wing party has a large advantage; and leftists, in which the left wing party 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 municipalities except the neutrals: in the "extremist" municipalities the probability of winning a majority is either zero or one, so the marginal value of an additional vote is zero. With large districts, in contrast, voters in all municipalities are valuable at the margin to the parties, since they likely all contribute to the total vote count. This effect, combined with 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. Our theory therefore leads to the novel prediction that countries that have homogenous constituencies should find it optimal to choose a majoritarian system with small districts. Countries with heterogeneous constituencies have stronger incentive to choose larger districts. The fact that small districts are optimal with homogeneous preferences and the existence of trade-offs between small and large districts with heterogeneous preferences are new results that cannot be anticipated from 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 1.1 explains how our paper relates to prior research on electoral rules and fiscal policy. Section 2 outlines the model, 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 party 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. The proofs that are omitted in the text are presented in the appendix.

1.1 Related literature

This paper attempts to bring together two literatures in political economy: the literature on probabilistic voting, which has mostly focused attention on static environments; and the literature on dynamic political distortions, which 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].[3] The basic model of electoral competition that we adopt in this paper is taken from a special version of Lindbeck and Weibull [1987] proposed by Persson and Tabellini [1999]. Persson and Tabellini [1999] is one of the first to study the effects of electoral rules on fiscal policy.[4] An empirical test of these theories is presented by Stromberg [2008].

With respect to the second literature, Persson and Svensson [1989], Alesina and Tabellini [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 cannot 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.

The effect of alternative electoral rules has been studied by Lizzeri [1999] in a two period model. This paper studies a purely redistributive environment with lump sum taxation: policies, therefore, do not affect welfare (that is the main focus of the comparison of electoral rules in our work). A number of recent papers have studied probabilistic voting models in macroeconomic models.[5] These papers, however, have assumed that voters' preferences remain constant over

[3] See Section 7.5 in Austen-Smith and Banks [2005] and Coughlin [2011] for discussions of these results.

[4] 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 one-dimensional policy space (Cox [1997], Morelli [2004]).

[5] See Azzimonti [2009], Bachmann and Bai [2009], Sleet and Yeltekin [2007], among others.

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 us to clarify the special conditions under which these results hold. None of these models studies political equilibria under alternative voting rules.[6]

A recent literature has studied Markov equilibria in sophisticated political economy models in which decisions are not taken in elections.[7] All the papers in this literature focus on only one political system and do not study how fiscal policy varies across alternative electoral rules.

2 The Model

2.1 The economy

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

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

where $\alpha \in (0, 1)$, $\varepsilon > 0$ is a measure of the elasticity of labor supply, and A is a measure of the relative importance of the public good g with respect to consumption and leisure. Citizens discount future per period utilities at rate δ. The parameter A may change across periods in a random way, capturing changes in preferences. We assume A is a realization from an i.i.d. process. Specifically, in each period, A is the realization of a random variable with range $[\underline{A}, \overline{A}]$ (where $0 < \underline{A} < \overline{A}$) and cumulative distribution function $G(A)$. The function G is continuously differentiable and its associated density is bounded uniformly below by some positive constant $\xi > 0$, so that for any pair of realizations such that $A < A'$, the difference $G(A') - G(A)$ is at least as big as $\xi(A' - A)$.

[6] Different types of models of elections in infinite horizons with moral hazard, adverse selection or both are presented by Ferejohn [1986], Banks and Sundaram [1991], Duggan [2000], Bernhardt et al [2004], Schwabe [2011] and Camara [2012], among others. In these model policy making periods are not strategically linked by a state variable that affects the set of feasible policies as public debt. None of these papers, moreover, studies alternative voting rules.

[7] In this literature we find models of political agency (Acemoglu, et al. [2008], Yared [2010]), legislative bargaining (Kalandrakis [2004], Battaglini and Coate [2007] and [2008], Battaglini et al. [2012a], Duggan and Kalandrakis [2012], Battaglini and Palfrey [2012]), common pools (Harstad [2011], Battaglini et al. [2012b] and [2012c]) and other political processes (Lagunoff [2008], Bai and Lagunoff [2010]).

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 they can also be diverted to finance targeted municipality-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 therefore described by an $n+3$-tuple $(r, g, x, s_1, ..., s_n)$, where r is the proposed income tax rate, g is the proposed amount of the public good provided, x is the proposed level of public debt, and s_i is the proposed transfer to municipality i's residents.

There are competitive markets for the consumption good, labor and the public good. The consumption good is the numeraire and its price is set to 1. 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 are indifferent to their allocations of consumption across time: thus, their welfare is equal to what they would obtain if they simply consumed their net earnings each period. The consumer chooses c and l to maximize her utility given the private budget constraint $c = (1-r)wl$, taking $(r, g, x, s_1, ..., s_n)$ and ρ as given. At net wage rate $(1-r)w$, each citizen works an amount $l^*((1-r)w) = (\varepsilon(1-r)w)^\varepsilon$ each period, so that ε is the elasticity of labor supply. The associated per period indirect utility function is given by

$$u(r, g; A) = (1-r)^{\varepsilon+1}\frac{\varepsilon^\varepsilon w^{\varepsilon+1}}{1+\varepsilon} + Ag^\alpha.$$

The public policy must satisfy the budget constraint requiring that revenues are 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, the budget constraint requires that

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

The set of constraints is completed by the non-negativity constraints: $s_j \geq 0$ for each municipality j (which rules out financing public spending via municipality-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 $-\underline{x} > 0$ 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 $\max_r R(r)/\rho$. For technical reasons it is convenient to assume that $\overline{x}$ is close to $\max_r R(r)/\rho$, but strictly smaller. 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.[8]

2.2 A model of electoral politics

In every period the ruling government is chosen by a national election. Electoral competition is modelled as a probabilistic voting game in a way similar to Persson and Tabellini [1999]. Two parties, L and R, run for office. Before the election, both parties simultaneously and noncooperatively commit to policy platforms $p^i = (r^i, g^i, x^i, s_1^i, ..., s_n^i)$, $i = L, R$ that satisfy the budget constraint (1). Voters are subdivided in $m \leq n$ electoral districts, each comprising n/m municipalities. One representative is elected in each district by majority rule, the policy platform of the party with the largest number of representatives is chosen. A key difference with the previous literature is that the election is embedded in a dynamic game. Because the level of debt affects the budget constraint in future periods, it creates a strategic linkage between electoral cycles.

The per period utility of voter ν in municipality j when party i is elected depends on the tax rate r^i, the level of the public good g^i, the transfer s_j^i received in municipality j and the identity

[8] 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 $n\alpha A g^{\alpha-1} = 1$.

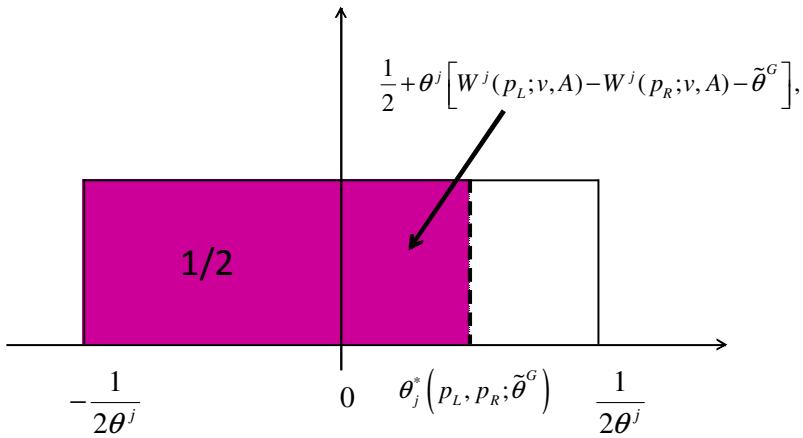

Figure 1: The probability that L wins district j.

of party i:

$$u(r^i, g^i; A) + s_j^i + \varepsilon^{\nu i},$$

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

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; θ may, however, change over time: θ^G and each θ^j are independent random variables with respective densities $\widehat{\varphi}(\cdot)$ and $\widehat{\varphi}^G(\cdot)$. We assume that $\widehat{\varphi}(\cdot)$ and $\widehat{\varphi}^G(\cdot)$ have full support on a finite set with respective minimal and maximal elements $\underline{\theta}^G, \overline{\theta}^G$ and $\underline{\theta}, \overline{\theta}$. For future reference we denote a shock to the economy by $z = (A, \theta)$, and the distribution of θ by $\varphi(\theta) = \widehat{\varphi}^G(\theta^G) \prod_j \widehat{\varphi}(\theta^j)$.

In what follows we focus on symmetric Markov equilibria in which strategies depend only on

z and b. In this case the expected value of choosing a level b of debt can be represented as a function v with $v(b) = E[v(b;z)]$, where $v(b;z)$ is the value function given the state z and the expectation is taken with respect to z.

The expected continuation value of a voter in municipality j for a policy $p^i = (r^i, g^i, x^i, s_1^i, ..., s_n^i)$ in a state A net of the preference shock is $W^j(p^i; v, A) = u(r^i, g^i; A) + s_j^i + \delta v\left(x^i\right)$. Because there are only 2 candidates, voters find it optimal to vote for the candidate that promises the highest expected continuation value. Voter ν in municipality j votes for party L if and only if

$$W^j(p_L; v, A) \geq W^j(p_R; v, A) + \varepsilon^\nu. \tag{2}$$

The fraction of votes received by party L in municipality j when political platforms are p_L and p_R is easily computed. Given (2), the voter in municipality j who is indifferent between L and R is characterized by $\theta_j^*(p_L, p_R; \widetilde{\theta}^G) = W^j(p_L; v, A) - W^j(p_R; v, A) - \widetilde{\theta}^G$: party L receives the votes of all voters ν with $\widetilde{\theta}^\nu \leq \theta_j^*(p_L, p_R; \widetilde{\theta}^G)$. 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 municipality j is:

$$1/2 + \theta^j \left[W^j(p_L; v, A) - W^j(p_R; v, A) - \widetilde{\theta}^G\right]. \tag{3}$$

The votes received by party R are one minus (3). For simplicity, we assume that, for all platforms (p_L, p_R), there is sufficient uncertainty about the citizens' preferences that both the fraction of votes received by a party and the probability that a party 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 the appendix). As we show in the appendix, a sufficient condition for interior probabilities is that $\overline{\theta}$ and $\overline{\theta}^G$ are sufficiently small:

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

The first condition guarantees that no policy platform can guarantee a party 100% of the votes in a municipality. The second condition guarantees that there is always a positive probability that each party receives a vote share of at least 50% in a municipality. Assuming $1/\left(2\overline{\theta}\right)$ and $1/\left(2\overline{\theta}^G\right)$ sufficiently large implies that the distribution of the shocks has sufficiently high variance that the outcome is sometimes driven by the shocks regardless of the policies.

2.3 Discussion

To what extent does the model capture important features of real world elections? Probabilistic voting has been extensively studied both theoretically and empirically in static models.[9] Our model is an extension of this class of models that introduces public debt and extends the time horizon to multiple periods. To make the dynamic extension of the model tractable, we have made a number of assumptions. In the following we discuss some of these assumptions.

Changing preferences. A key assumption is that voters' preferences for parties 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 parties' strategies? There is a large literature in political science studying this issue.[10] As Abramowitz and Saunders [1998] put it: "political scientists 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." 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.[11] Does the model 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 correspond 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 correspond to periods with low θ^j and θ^G, in which a larger mass of the population has strong party preferences (since the

[9] Stromberg [2008] uses a probabilistic voting model to study presidential races in the US, and shows that the model does a remarkable job in explaining candidates' behavior.

[10] 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], Allsop and Weisberg [1988]).

[11] 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."

extremes $\pm 1/(2\theta^j)$ and $\pm 1/(2\theta^G)$ have larger absolute values).[12]

Symmetric and serially uncorrelated preference shocks. In the model, parties are ex ante symmetric because the shocks $\left(\theta^i\right)_i$ and $\widetilde{\theta}^G$ have mean zero; the shocks, moreover, are serially uncorrelated: the fact that a party has a favorable realization at time t does not imply that it will continue to be favored. These assumptions keep the analysis tractable. If we had persistence, then the players' equilibrium strategies would depend on the past realizations of θ^j and $\widetilde{\theta}^G$ and we also would not be able to focus on symmetric equilibria, since parties would have asymmetric expectations about the future.[13]

Quasi-linear preferences. We are assuming quasi-linear preferences that are linear in pork transfers. This helps in two ways. With quasi-linear preferences, the equilibrium interest rate is only a function of δ because the marginal utility of consumption is constant. Without quasi-linearity, the interest rate would depend on consumption and hence on current and expected fiscal policy.[14] Quasi-linearity also helps us to characterize how the politicians choose to distribute pork; we discuss this aspect in Section 4.1.

Uniform distribution of the preference shocks and Assumption 1. As in static models, the assumption of a uniform distribution of the shocks guarantees that the parties' objective functions are concave with respect to the policy variables, and therefore it is instrumental in obtaining a pure strategy equilibrium.[15] Assumption 1 is used to avoid dealing with corner solutions in which a party wins all the votes in a municipality. A limitation of the assumption is that it implies that some voters would not change their vote for any policy platforms, and that the preferences of a significant fraction of voters have a strong bias for one party or the other. This strong assumption is made only to make the model tractable.[16] Assumption 1 would be unnecessary

[12] 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 Stimson [1989]), geographic mobility and cohort replacement (Converse [1976]), and mobilization of previously disenfranchised voters (Campbell [1985], Carmines and Stimsons [1989]).

[13] In a previous version of this paper (Battaglini [2009]), the shocks A, θ^j and θ^G were assumed to be positively correlated over time.

[14] An analysis of the case with more general non-linear utility functions (for a simpler political decision process) is presented in Battaglini and Barsheghyan [2012].

[15] See Lindbeck and Weibull [1987] and Persson and Tabellini [1999] for a discussion of this aspect in static models.

[16] It is, however, unclear the extent to which this assumption is unrealistic. First, as noted above, it is natural to rule out situations in which a candidate obtains 100 per cent of the votes: it seems realistic to assume that

with an unbounded distribution of shocks, for example, with a normal distribution. Assuming an unbounded distribution of the shocks that satisfies the condition for existence by Lindbeck and Weibull [1987] , however, would likely make the analysis intractable without adding very much to the basic results of the paper.

Unobservable $\left(\theta^j\right)_j$ *and* $\widetilde{\theta}^G$. In line with the static literature on probabilistic voting, we assume that parties do not know $\left(\theta^j\right)_j$ and $\widetilde{\theta}^G$. As in static version of the model, lack of knowledge of these variables is important to obtain an equilibrium in pure strategies.

3 The normative benchmark

Before turning to the analysis of the political equilibrium, it is useful to discuss the Pareto efficient solution that would be chosen by a benevolent government.[17] 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 expected value function of a citizen of municipality i in a state in which the current level of public debt is b and the preference parameter is A. The government chooses a policy $(r, g, x, s_1, ..., s_n)$ to solve

$$\begin{gathered} \max_{(r,g,x,s)} \sum\nolimits_i \mu_i \left[u\left(r,g;A\right) + s_i + \delta v_i^\circ\left(x;A'\right)\right] \\ s.t.\ \ s_i \geq 0 \ \text{ for all } i,\ \sum\nolimits_i s_i \leq B(r,g,x;b),\ \&\ x \in [\underline{x}, \overline{x}] \end{gathered} \tag{4}$$

where $\mu_i \geq 0$ is the weight associated with municipality i (we assume without loss of generality that $\sum_i \mu_i = 1$), and the expectation is taken with respect to A'. Given the government choice $\left(r^\circ(b;A), g^\circ(b;A), x^\circ(b;A), (s_i^\circ(b;A))_{i=1}^n\right)$, the value function of a municipality is immediately defined by

$$v_i^\circ(b;A) = u\left(r^\circ(b;A), g^\circ(b;A);A\right) + s_i^\circ(b;A) + \delta E v_i^\circ\left(x^\circ(b;A);A'\right). \tag{5}$$

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 $(r^\circ(b;A),\ g^\circ(b;A),\ x^\circ(b;A),\ (s_i^\circ(b;A))_{i=1}^n)$

shocks have sufficiently high variance to rule this out. Second, it does not seem implausible to assume that a large fraction of voters have strong party biases and vote on that basis.

[17] A similar discussion of the benevolent planner's case is presented in Aiyagari et al [2002] and Battaglini and Coate [2008]. These papers, however, assume a utilitarian planner. In what follows we will characterize the entire Pareto efficient solution. This will prove useful in Section 6 when we discuss the equilibrium with heterogeneous municipalities.

solve (4) 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 (5) given $(r^\circ(b;A),\, g^\circ(b;A),\, x^\circ(b;A),\, (s_i^\circ(b;A))_{i=1}^n)$.

Using standard methods, one can show that, for any μ, a Pareto efficient solution is associated with a unique well-behaved welfare function $V(b) = E\sum_i \mu_i v_i^\circ(b;A')$. The welfare function is well-behaved in the sense that it is continuous, concave and differentiable in b. The planner's solution is characterized by two sets of conditions. First, the planner equalizes the marginal benefit of resources across alternative uses. The first-order conditions with respect to g, r and s_i imply

$$\alpha A g^\circ(b;A)^{\alpha-1} = \frac{1}{n}\frac{1-r^\circ(b;A)}{1-r^\circ(b;A)\,(1+\varepsilon)} \geq \mu^* \text{ with equality if } B(r,g,x;b) > 0 \tag{6}$$

where $\mu^* = \max_i \mu_i$. The first-order condition with respect to x implies:

$$\frac{1}{n}\frac{1-r^\circ(b;A)}{1-r^\circ(b;A)\,(1+\varepsilon)} \geq -\delta\sum_i \mu_i\left[v_i^\circ\right]'(x) \quad (= \text{ if } x^o(b;A) < \overline{x}). \tag{7}$$

To interpret these conditions, note that the middle term in (6) is the *marginal cost of public funds*. The marginal cost of public funds is the compensating variation for a marginal increase in tax revenues, and 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 equation states that the marginal benefit of one extra dollar spent on the public good should be equal to the marginal cost of financing it with discretionary taxation. The second condition states that either transfers are zero (that is, if $B(r,g,x;b) = 0$) or the marginal social benefit to the municipalities that receive them should equal the marginal cost of financing them with taxes: the planner would make transfers only to the municipalities with the highest social weight, so that the marginal social benefit is μ^*. Condition (7) requires that the marginal cost of raising a dollar with taxes is equal to the expected marginal cost of issuing debt (if feasible, that is, if $x < \overline{x}$).

Second, the planner equalizes the opportunity cost of resources over time,[18]

$$\frac{1-r^\circ(b;A)}{1-r^\circ(b;A)\,(1+\varepsilon)} \geq E\left[\frac{1-r^\circ(x^o(b;A);A')}{1-r^\circ(x^o(b;A);A')\,(1+\varepsilon)}\right] \quad (= \text{ if } x^o(b;A) < \overline{x}). \tag{8}$$

The left-hand side of (8) is the marginal cost of public funds in state (b,A), and the-right hand side is the expected marginal cost of public funds in the following period when the debt is $x^o(b;A)$.

[18] 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 (8).

Intuitively, if the marginal cost of public funds at t exceeded the corresponding expected level at $t+1$ and $x^o(b;A) < \overline{x}$, then the planner would find it optimal to reduce taxes and increase debt, thus reducing the average intertemporal cost of taxation.

Together (6), (7) and (8) determine the short- and long-run behavior of the Pareto efficient allocation. 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. Moreover, straightforward analysis of (4) reveals the following result:

Result 1. *Optimal tax and debt increase in b, while the optimal amount of the public good falls in b and increases in A.*

The long-run behavior of equilibrium policies follows from (8). Since the marginal cost of public funds is a convex function of the tax rate, (8) implies that the tax rate is a supermartingale and therefore converges to a constant with probability one. We have

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

$$r_{\mu^*} = \frac{n\mu^* - 1}{n\mu^*(1+\varepsilon) - 1}, \text{ and } g_{\mu^*}(A) = \left(\frac{\alpha A}{\mu^*}\right)^{\frac{1}{1-\alpha}} \tag{9}$$

and debt is at most $x_{\mu^*} \geq \underline{x}$, 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 resources are sufficiently high that the government can provide transfers after any shock. In this case (6) holds as equality, implying that the tax rate is r_{μ^*}; condition (6), moreover, implies that the public good provision 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 $b \leq x_{\mu^*}$.[19]

[19] When $b \leq x_{\mu^*}$ the government assets are higher or equal than the amount needed to perfectly self insure against shocks, i.e. x_{μ^*}, given the first-best level of expenditure, g_{μ^*}. Indeed, for any $b \leq x_{\mu^*}$, taxation and public expenditures are unaffected by the level of debt: a marginal change in debt, only affect monetary transfers. Because the citizens' utilities are linear in transfers, and the equilibrium interest rate is such that $\delta(1+\rho) = 1$, the government is indifferent as to when resources that are not needed for self insurance $(x_{\mu^*} - b)$ are consumed: this is the reason why multiple steady states of debt are possible.

4 A majoritarian electoral system

In the majoritarian system, there are $m \geq 1$ *electoral districts* where m is odd. Each electoral district elects one representative to parliament. We assume $m \leq n$ where n/m is assumed to be an integer. Electoral districts coincide with the municipalities described in the previous sections if only if $m = n$. When $m < n$, each electoral district is divided in n/m municipalities. A party wins a district if and only if it receives a majority of votes in the district. As standard in the literature on probabilistic voting, we assume parties are office seeking: they choose policy in order to maximize the expected number of their representatives elected in the parliament.[20] When $m = 1$, the model can be interpreted as a "presidential" system, in which voters elect the head of the executive branch. The set of municipalities in district i is denoted H_i.

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; voters vote for L if and only if (2) 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 $v(b)$. 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 $s(b;z) = (s_1(b;z), ..., s_n(b;z))$ is a vector of transfers offered to the municipalities by each candidate. Associated with an equilibrium is a value function v that specifies the expected future payoff of a citizen in a period in which the debt is b.

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 $v(b)$. The value function $v(b)$ is determined by the equilibrium policy proposals. A citizen's expected value function $v(b)$ can be

[20] In most parliamentary systems parties desire to maximize the expected number of representatives since their power in the legislature depends on it even when they are in the minority (in terms of their ability to sustain a filibuster, or in terms of the reimbursements they receive, and in terms of providing jobs to local party leaders). With the exception of the case in which $m = 1$, assuming that parties choose policy in order to maximize the expected number of their representatives is not the same as assuming that they maximize the probability of winning a majority of representatives. An analysis of this case is presented in Battaglini [2009]. Assuming the existence of a pure strategy equilibrium, the analysis in this case is similar to the analysis presented here. Even in static versions of the model, however, the conditions for the existence of a pure strategy equilibrium are more demanding.

determined recursively given $(r(b;z), g(b;z), x(b;z))$ as

$$v(b) = E\left[u\left(r(b;z'), g(b;z'); A'\right) + \frac{B(r(b;z'), g(b;z'), x(b;z'); b)}{n} + \delta v(x(b;z'))\right], \tag{10}$$

where the expectation is taken 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). However, since citizens are ex ante symmetric, and parties treat them anonymously, they all receive the same amount of transfers in expectation, $B\left(r(b;z), g(b;z), x(b;z); b\right)/n$.

For a given $v(b)$, policies are chosen to maximize the expected number of representatives in the parliament. Recall that in electoral district i there are n/m municipalities. Municipalities are heterogeneous because they are associated with different realizations of θ^j for $j \in H_i$, so they react differently to L's policies: L's share of votes received in i is obtained by integrating over the different municipalities. Party L wins district i if

$$\sum\nolimits_{j\in H_i}\left[1/2 + \theta^j\left[W^j(p_L; v, A) - W^j(p_R; v, A) - \widetilde{\theta}^G\right]\right] \geq \frac{1}{2}\frac{n}{m}. \tag{11}$$

Condition (11) requires that the share of votes received in district i (the left-hand side) is at least half of the population living in district i, $n/(2m)$. If we denote the set of municipalities that belong to the same district as j by $H(j)$, condition (11) can be rewritten in a way that proves convenient in the analysis:

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

The larger is the utility that L promises to municipality 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 municipality j is measured by θ^j: that is, an additional dollar of transfers to municipality j by party L, ceteris paribus, induces a fraction θ^j more voters to vote for party L. The pertinent issue for political parties is the sensitivity of voters in municipality j relative to the sensitivity of voters in other municipalities from the same district, $\theta^j/(\sum_{\iota\in H(j)}\theta^\iota)$.[21]

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

$$\sum\nolimits_{i=1}^{m} E\left[sign\left(\sum\nolimits_{j\in H_i}\frac{\theta^j}{\sum_{\iota\in H(j)}\theta^\iota}\left(W^j(p_L; v, A) - W^j(p_R; v, A) - \widetilde{\theta}^G\right)\right)\right], \tag{12}$$

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

where the function $sign(x)$ is equal to one if $x \geq 0$ and zero otherwise.[22] Using the distribution of $\widetilde{\theta}^G$, this expression can be written as:

$$\sum_{i=1}^{m} \left[1/2 + \theta^G \sum_{j \in H_i} \frac{\theta^j}{\sum_{\iota \in H(j)} \theta^\iota} \left(W^j(p_L; v, A) - W^j(p_R; v, A) \right) \right]. \tag{13}$$

Simplifying and eliminating irrelevant constraints, party L's best response problem becomes

$$\begin{array}{c} \max\limits_{(r,g,x,s)} \sum_{i=1}^{m} \sum_{j \in H_i} \frac{\theta^j}{\sum_{\iota \in H(j)} \theta^\iota} W^j(p_L; v, A) \\ \text{s.t.} \quad B(r, g, x; b) \geq \sum s_j, \quad x \in [\underline{x}, \overline{x}]. \end{array} . \tag{14}$$

Naturally, party R wins an expected number of districts equal to m minus (12): so it chooses $p_R = (r_R, g_R, x_R, s_R)$ to minimize the objective function of (14) given p_L and the same constraints as in (14). 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 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* $(r(b; z), g(b; z), x(b; z), s(b; z))$ *solves (14) given* $v(b)$ *and* $v(b)$ *satisfies (10) given* $(r(b; z)$, $g(b; z)$, $x(b; z))$.

Problem (14) can be simplified. Without loss of generality, we can assume $B(r, g, x; b) = \sum_i s_j$. At the margin a dollar spent in pork transfers in municipality j has a value $\theta^j / (\sum_{l \in H(j)} \theta^l)$. Party L therefore chooses the municipality i with the highest relative weight. Ignoring constant terms, and using the fact that $\sum_{j \in H_i} [\theta^j / (\sum_{\iota \in H(j)} \theta^\iota)] = 1$, party L's problem can be written as

$$\begin{array}{c} \max_{(r,g,x)} \left[u(r, g; A) + \max_j \left(\theta^j / \left(\sum_{\iota \in H(j)} \theta^\iota \right) \right) \cdot B(r, g, x; b)/m + \delta v\,(x) \right] \\ \text{s.t.} \quad B(r, g, x; b) \geq 0, \quad x \in [\underline{x}, \overline{x}]. \end{array} \tag{15}$$

The fact that parties find it optimal to make monetary transfers only to the municipalities that have the highest relative sensitivity follows from the assumption of quasi-linear preferences that are linear in monetary transfers. If this were not the case, then all municipalities would receive some transfers, although clearly the municipalities with the highest relative sensitivities would receive more.

We can characterize the equilibrium focusing only on (15). We say that an equilibrium is *well-behaved* if $v(b)$ is continuous and weakly concave on $[\underline{x}, \overline{x}]$. Without continuity and concavity

[22] 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.

of the value function, it would be hard to obtain general qualitative results on equilibrium policies. Proving the existence of well-behaved equilibria is therefore an essential step for the study of fiscal policy. 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 in static environments are extensively discussed by Lindbeck and Weibull [1987], who derive a sufficient condition for the existence of a pure strategy equilibrium. 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: besides the usual complications associated with stochastic games (for which no general existence theorem is available), the existence proof is complicated by the fact that we are interested in well-behaved equilibria.[23] The strategy to prove Proposition 2 is as follows. Problem (15) defines a correspondence $T(v)$ that maps a bounded, continuous and weakly concave function to a subset of bounded, continuous and weakly concave functions. In general $T(v)$ is not convex-valued. Indeed, the parties may be indifferent among a variety of policies: what complicates matters is that these policies may generate different expected utility levels for the citizens, and the convex combination of these utilities is not necessarily in $T(v)$. The key step in the proof of Proposition 2 consists in proving that we can find a convex-valued sub-correspondence $T^*(v)$ of $T(v)$ that maps a compact subset of bounded, continuous and weakly concave functions to itself. Given this, we can prove that a fixed-point $v^* \in T^*(v^*)$ exists by the Glicksberg-Fan theorem.

In what follows, we 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.

4.2 Characterization

The equilibrium can be understood by observing the objective function in (15). For a given policy platform of party R, party L chooses the policy that maximizes the expected number of elected representatives. The budget constraint imposes a trade-off. Party L can offer a more attractive policy by choosing a platform with a generously high g, a low tax r, a low debt level x, or a combination of all of these. These policies increase L's popularity uniformly in all municipalities. Alternatively, L can be more conservative in the choice of g, r, x, and reserve a

[23] Note, moreover, that the electoral game described above is not a zero sum game. The voters' choices depend on their expectations of future policies, as incorporated in their expected value functions.

higher level of resources for pork transfers. Transfers are less efficient because they do not affect all municipalities. Party L, however, can target them to municipalities with more swing voters. This trade-off is also 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 between these models and ours 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, party 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 and g, the optimal tax $r_m(\theta)$ and level of public expenditure $g_m(A,\theta)$ are

$$\frac{1}{n}\left[\frac{1-r_m(\theta)}{1-r_m(\theta)\,(1+\varepsilon)}\right]=\frac{1}{m}\max_j\frac{\theta^j}{\sum_{\iota\in H(j)}\theta^\iota}=\alpha A g_m(A,\theta)^{\alpha-1} \tag{16}$$

To interpret this condition, note that the first expression is the marginal cost of public funds, and the last one is the marginal benefit of the public good: both must equal the political benefit of making a pork transfer, as measured by $[\theta^j/(\sum_{\iota\in H(j)}\theta^\iota)]/m$.

Similarly, debt $x_m(\theta)$ is chosen so that:

$$x_m(\theta)\in\arg\max_{x\in[\underline{x},\overline{x}]}\left\{\frac{1}{m}\cdot\max_j\left(\frac{\theta^j}{\sum_{\iota\in H(j)}\theta^\iota}\right)\cdot x+\delta v\left(x\right)\right\}. \tag{17}$$

The marginal benefit of debt is the political benefit of making a pork transfer; the marginal cost is given by the change in the expected value function $v\left(x\right)$.[24]

In this case, as can be seen from (16) and (17), the policies $(r_m(\theta), g_m(A,\theta), x_m(\theta))$ are independent of 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 the preference parameters, θ and A.

The second case is when party L finds it optimal to have no transfers. In this case, $B(r,g,x;b)=\sum_i s_j=0$. It follows that

$$\max_j\left[\theta^j/\left(\sum\nolimits_{\iota\in H(j)}\theta^\iota\right)\right]B(r,g,x;b)=\frac{1}{n}B(r,g,x;b)$$

[24] As we will see in the next section $v(x)$ is differentiable in b. The argument in this section, however, does not need this property.

and both are zero. It is *as if* the policies in state b, z are chosen to maximize:

$$\begin{gathered}\max_{(r,g,x)} \left[u\left(r,g;A\right)+B(r,g,x;b)/n+\delta v\left(x\right)\right] \\ s.t.\ \ B(r,g,x;b)\geq 0 \text{ and } x\in[\underline{x},\overline{x}].\end{gathered} \tag{18}$$

The solution of (18), $r(b;A)$, $x(b;A)$ 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 parties are irrelevant.

The first case is clearly possible whenever the policies defined by (16) and (17) are feasible, so

$$B(r_m(\theta),g_m(A,\theta),x_m(\theta);b)\geq 0. \tag{19}$$

Since $g_m\left(A,\theta\right)$ 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 (19) 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 $A\geq A_m(\theta,b)$ equilibrium policies $r(b;z)$, $x(b;z)$ and $g(b;z)$ are equal to the solution of (18), $r(b;A)$, $x(b;A)$ and $g(b;A)$, and expected pork transfers are zero. For $A<A_m(\theta,b)$ policies $r(b;z)$, $x(b;z)$ and $g(b;z)$ are equal to $r_m(\theta)$, $x_m(\theta)$ and $g_m(A,\theta)$, uniquely defined by (16)-(17); 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 parties choose policies, and of the resulting equilibrium dynamics. 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 case, the parties allocate pork transfers according to the municipalities' preferences for parties, so it is not surprising that the variation of the citizen's preferences for the parties (as measured by θ) and the voting rule are important. The effect of the citizens's preferences for the party depends on m: the smaller is m the larger is the effect of θ on these policies. The reason is intuitive. In districts with multiple municipalities, parties can trade-off voter share gains in some municipalities against others, since all they care about is the total votes in the district. Hence, the relative density amongst municipalities affects where pork is directed. In single municipality districts, no trade-off is possible within districts,[25] and the margin of victory within a district is irrelevant since parties are only interested in maximizing the number of district won.

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

When $A > A_m(\theta, b)$, the citizens' preferences for the individual parties are irrelevant; only A and b affect the policies. In this case, because parties don't have enough money to chase the municipalities with cash transfers, they find it optimal to treat citizens equally: given the equilibrium v, parties choose policies as if they were maximizing aggregate welfare.

From this discussion and a straightforward analysis of (18) we have the following result, which parallels Result 1 obtained for the planner's solution.

Result 2. *In a political equilibrium with m districts, the equilibrium levels of tax and debt increase in b, while the equilibrium level of public goods falls in b and increases in A.*

Given the policies described above, the effect of a shock to the economy propagates over time through its impact on debt. An increase in A at t induces an increase in b at t. In subsequent 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 are $r(b; A)$ and $x(b; A)$ (and the lower is $g(b; A)$). Second, (as can be seen from (19)) because 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 similar dynamics, but in reverse. The decrease at t induces a decrease in b: even if A remains constant in the following periods, the decrease in b 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 steady-state distribution, we need to study how resources are allocated over time in more detail.

5 Long-term distribution and welfare

To study the long-term properties of equilibrium policies, it is useful to characterize how the expected value function changes following a marginal change in b. From (10) and Proposition 3, we know that when $A < A_m(\theta, b)$, we have

$$v(b; z) = u\left(r_m(\theta), g_m(A, \theta); A\right) + \frac{1}{n} \cdot B(r_m(\theta), g_m(A, \theta), x_m(\theta'); x)) + \delta v(x_m(\theta)),$$

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 $-\delta v'(b;z) = 1/n$ when $A < A_m(\theta, b)$, 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, $v(b;z)$ is given by (18). Using the envelope theorem, we find that in a state z

$$-\delta v'(b;z) = \frac{1}{n}\frac{1-r(b;z)}{1-r(b;z)(1+\varepsilon)}. \tag{20}$$

The derivative in this case exceeds $1/n$ because the increase in debt not only implies a transfer of resources from $t+1$ to t, but also a social loss. Given that pork transfers are zero at $t+1$, the agents are forced to increase taxes or debt, or to reduce g at t, all measures that generate a loss in social welfare. We conclude that $v(x)$ is differentiable with derivative

$$-\delta v'(x) = \sum\nolimits_{\theta'} \left[\frac{1}{n}\cdot G(x,\theta') + E\left[\frac{1}{n}\frac{1-r(x;z')}{1-r(x;z')(1+\varepsilon)}\,|A' > A_m(\theta', x)\right]\cdot[1-G(x,\theta')]\right]\varphi(\theta'), \tag{21}$$

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

We can now characterize the first-order necessary and sufficient conditions of (17) and (18) with respect to debt. In both cases, we must have

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

for any b such that $x(b;z) < \overline{x}$. Using (16) and (21), we can rewrite (22) as[26]

$$\begin{aligned}\frac{1}{n}\frac{1-r(b;z)}{1-r(b;z)(1+\varepsilon)} &= \frac{1}{n}E\left[\frac{1-r(x(b;z);z')}{1-r(x(b;z);A')(1+\varepsilon)}\right] \\ &\quad -\frac{1}{n}E\left[\left(\frac{n}{m}\max_j \frac{\theta^j}{\sum_{\iota\in H(j)}\theta^\iota} - 1\right)\right]\sum\nolimits_{\theta'} G(x(b;z),\theta')\varphi(\theta').\end{aligned} \tag{23}$$

As long as $m < n$, then

$$\frac{n}{m}\max_j \sum\nolimits_{l\in H(j)} \theta^l > 1,$$

so that we have the following result.

[26] To write (23) we add and subtract

$$\sum\nolimits_{\theta'} E\left[\frac{1}{n}\frac{1-r(x;z')}{1-r(x;z')(1+\varepsilon)}\,|A' \leq A_m(\theta', x)\right]\cdot G(x,\theta')\varphi(\theta')$$

in (21). The added term contributes to form the first term in (23); the subtracted term contributes to the second term in (23) (as it can be seen using (16)).

Proposition 4. *In a political equilibrium, when $m < n$:*

$$\frac{1}{n}\frac{1-r(b;z)}{1-r(b;z)\,(1+\varepsilon)} \leq \frac{1}{n}E\left[\frac{1-r(x(b;z);z')}{1-r(x(b;z);z')\,(1+\varepsilon)}\right], \tag{24}$$

for any b such that $x(b;z) < \overline{x}$. Moreover, there is a positive probability of reaching a state in which the inequality is strict.

Comparing (24) to (8), we can see that resources are inefficiently allocated over time: politicians tend to shift too much of the burden of taxation to the future.

It is interesting to note that the wedge between the marginal cost of public funds at time t and at time $t+1$ depends on the electoral rule and is state contingent. It is decreasing in m and is zero when $m = n$, since $(n/m)\max_j[\theta^j/(\sum_{\iota\in H(j)}\theta^\iota)] = 1$. When $m < n$ there is no choice of Pareto weights that could rationalize the allocation, since the parties tend to systematically transfer the burden of taxation to the future. As b increases, however, the probability that $A' > A_m(\theta', x(b;z))$ (i.e. $G\left(x(b;z),\theta'\right)$) 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 the following result:

Proposition 5. *The political equilibrium is Pareto efficient if and only if $m = n$. For any initial state z_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 $(\alpha n A)^{\frac{1}{1-\alpha}}$,* and *public debt converges to $\underline{x}$.*

- *If $m < n$, $r(b;z)$ and $g(b;z)$ converge to non-degenerate stationary distributions with associated supports $[r_m, \overline{r})$ 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 comprises more than one municipality and the parties can be opportunistic in making transfers to the municipality that is most susceptible to being influenced. The marginal benefit of this policy is equal to $\max_j[\theta^j/(\sum_{\iota\in H(j)}\theta^\iota)]/m$, so it changes over time

with θ. Although $\max_j[\theta^j/(\sum_{\iota\in H(j)}\theta^\iota)]/m$ may go up or down over time, and this changes the incentive to make transfers, (24) shows that this mechanism creates a *systematic* bias toward shifting the burden of taxation to 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, $\max_j[\theta^j/(\sum_{\iota\in H(j)}\theta^\iota)]/m > 1/n$ (the exception being the event in which θ^j is identical for all municipalities in the electoral district); the expected benefit of a dollar in future transfers to a municipality, however, is $1/n$ (since in expectation all municipalities are equally likely to receive it). As m increases, the number of municipalities per district decreases, and the arbitrage opportunities decrease as well. When $m = n$, we have only one municipality per district, so $\max_j[\theta^j/(\sum_{\iota\in H(j)}\theta^\iota)]/m$ 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 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 is uniformly higher in the political equilibrium only if $x_{\mu^*} < x_m$. As can easily be verified, this is the case if and only if

$$\mu^* < \frac{1}{m}\max_\theta\left(\max_j \frac{\theta^j}{\sum_{\iota\in H(j)}\theta^\iota}\right)$$

If this condition is not satisfied, the debt fluctuates 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 high levels 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 a 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 from Proposition 5. First, electoral competition is not sufficient per se to guarantee Pareto efficiency. This contrasts with 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 from a Pareto efficient solution is not necessarily a higher level of debt, but rather 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 comprises municipalities with different distributions of preference shocks that may change over time. The case in which $m = n$ should be seen as a limit case in which the districts are designed to comprise only perfectly homogeneous municipalities with the same distribution of preference shocks: in this case the arbitrage opportunity is zero, and resources are allocated efficiently.

5.1 Discussion

Before turning to the analysis of heterogeneous municipalities, it is useful to discuss the inefficiency results presented in this section in the context of the previous literature on the political economy of public debts and the literature on probabilistic voting.

The economy described in Section 2.1 is similar to the economy adopted in Battaglini and Coate [2008], who present a model of public debt in which policies are chosen by a legislature.[27] It is therefore not surprising that there are similarities between the two papers. What distinguishes these papers is the political system, which induces important differences in the equilibrium policies and the types of political inefficiencies that characterize the equilibrium. In Battaglini and Coate [2008],[28] the equilibrium is inefficient for any choice of non-unanimous voting rule. This is not the case with probabilistic voting: as Proposition 5 makes clear, the policy may be efficient if electoral districts are sufficiently small.[29] Second, political distortions, when present, are different from the distortions in Battaglini and Coate [2008]. In Battaglini and Coate [2008] the equilibrium policy either coincides with the policy that a planner would choose (given the equilibrium value function), or it depends only on A. As Proposition 3 makes clear, this is no

[27] In Battaglini and Coate the representatives in the legislature have the same preferences as the citizens in the district they represent, so their election is not modelled.

[28] In Battaglini and Coate [2008] the legislature deliberates by q-rule, so a policy must be approved by q out of n legislators.

[29] As we will discuss more extensively in the next section, this insight is important for the study of constitutional design with heterogeneous municipalities because it suggests a trade off between electoral systems with small districts, which induce more efficient policies but tend to disenfranchise municipalities that are not pivotal; and electoral systems with large districts, which are less efficient, but force candidates to propose policies that benefit all municipalities.

longer the case when we have elections, as modelled as in this paper. As in Battaglini and Coate [2008], equilibrium policies coincide with the policy that a utilitarian planner would choose when $A \geq A_m(\theta, b)$ (given the equilibrium value function). When $A < A_m(\theta, b)$, however, policies depend on θ. This dependency on the preferences for the parties is the source of an inefficiency that is not present in Battaglini and Coate [2008]: changes in θ induce an inefficient volatility in r and g that all citizens would like to avoid.

The presence of dynamic distortions distinguishes this paper from the previous literature on probabilistic voting, which typically stresses the result that policies are Pareto efficient. This result is always true in static models, but it is true in dynamic models only under the assumption that the distribution of preference shocks is perfectly constant over time, an assumption that is questionable from an empirical point of view. Proposition 5 shows that Pareto efficiency can occur even with changing preference distributions, and characterizes precise conditions for Pareto efficiency and for the size of distortions when Pareto efficiency does not hold.

6 Heterogeneous municipalities

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

Municipalities are often asymmetric with respect to parties: some municipalities are ex ante neutral and so easily contestable; while others have strong and persistent biases for one party or the other. In this case the size of the optimal district is less straightforward. Small electoral districts may allow the parties to focus resources only on municipalities that can be easily swung with targeted transfers, ignoring municipalities where they have an ex ante small (or high) probability of winning. When districts are large, in contrast, even biased municipalities have a marginal value for parties. In districts comprising only right-wing municipalities, for example, party L never wins. In a district with mixed municipalities, both parties may win and the right wing municipalities 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 an environment with heterogeneous municipalities, we consider a setting where the municipalities can be partitioned in three groups: *neutral municipalities*, that is municipalities with $E\varepsilon^{\nu} = 0$, as described in the previous sections; *right-wing municipalities*, with a bias for R; and *left-wing municipalities*, with a bias for L. In particular, we assume that for the right-wing municipalities and the left wing-municipalities the idiosyncratic shocks $\widetilde{\theta}^{l}$ have uniform distributions with respective supports $\left[-1/(2\theta^{l}) + \sigma, 1/(2\theta^{l}) + \sigma\right]$ and $\left[-1/(2\theta^{l}) - \sigma, 1/(2\theta^{l}) - \sigma\right]$. As in the previous sections, we assume that all parties receive an interior fraction of votes in (0,1) in each municipality; but now we assume that σ is sufficiently large that party R (respectively, L) never receives more than 50% of the votes in a left-wing (respectively, right-wing) municipality. As we show in the appendix, the following condition is sufficient for this.

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

where ΔW^{*} is defined as in Assumption 1. The first inequality ensures that party L (respectively, R) never wins a majority in a right-wing (respectively, left-wing) municipality; the second inequality corresponds to Assumption 1 and guarantees that all parties receive an interior fraction of votes in all municipalities. The fraction of municipalities that are neutral is β (whose βn is assumed to be an integer for simplicity), and the fraction of right- and left-wing municipalities 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; and the second is a *majoritarian system* in which n representatives are elected. Persson and Tabellini [1999, 2004] did a similar exercise in a static model with 3 municipalities (a neutral, a right-wing municipality and a left-wing municipality). They conclude that a system with $m = 1$ generates higher g, lower r and higher utilitarian welfare. Here we show that another force is at play in a dynamic model, and the opposite result may occur in the steady state. This highlights the importance of a dynamic perspective in the study of institutions.

In the *presidential system* with $m = 1$ the analysis is very similar to the earlier analysis. Although party L cannot win a majority in the right-wing municipalities, the votes gained in right-wing municipalities contribute towards the overall vote count. Party L, therefore, finds it optimal to direct transfers to the municipalities that are more reactive to the transfers at the

margin, regardless of the average ideology. Policies maximize the same objective function as in (15):

$$u(r,g;A)+\max_j\left(\frac{\theta^j}{\sum_{i=1}^n\theta^i}\right)\cdot B(r,g,x;b)+\delta v\left(x\right) \tag{25}$$

under the same constraints as in (15). Note that this objective function depends on θ because the party chooses the municipality with the highest θ^j.

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

$$u(r,g;A)+\frac{1}{\beta n}B(r,g,x;b)+\delta v\left(x\right). \tag{26}$$

The parties care only about the βn neutral municipalities. They treat those municipalities symmetrically because, as discussed above, the relative sensitivities $\theta^j/(\sum_{\iota\in H(j)}\theta^\iota)$ in the municipalities are identical (and equal to 1) when $m=n$.[30] Comparing (25) with (26), we can make two observations. When $1/(\beta n)>\max_j\left[\theta^j/\left(\sum_{i=1}^n\theta^i\right)\right]$, the parties 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 point was made by Persson and Tabellini [1999] in their example with 3 municipalities. In a dynamic model, however, there is another difference between presidential system and majoritarian systems. The objective function (25) depends on the realization of θ, and so fluctuates over time. As shown in the previous section, the equilibrium is dynamically inefficient. However, when $m=n$ the objective function is independent of 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" municipalities; in a presidential system, municipalities are treated in an ex ante symmetric way, but ex post the allocation depends on θ, because the parties target the more sensitive municipalities for pork transfers.[31]

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:*

[30] This follows since in a majoritarian system there is no possibility of trading off excess votes between districts

[31] 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 a presidential system where $m=1$.

- *In simple majoritarian system ($m = n$), $r(b; z)$ and $g(b; z)$ converge, respectively, to $r_\beta = (1-\beta)/[1+\varepsilon-\beta]$ and $g_\beta(A) = (\alpha\beta n A)^{\frac{1}{1-\alpha}}$ and debt to a level lower than or equal* to $x_\beta = \left[R(r_\beta) - g_\beta(\overline{A})\right]/\rho$. *In the steady state, neutral municipalities receive*

$$\frac{1}{\beta n} B(r_\beta, g_\beta(A), x_\beta; b)$$

in pork transfers and the other municipalities 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 non-degenerate distribution in $[x_n, \overline{x}]$. Municipalities 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 choosing a political system there is a trade-off. In the majoritarian system resources are targeted only to neutral municipalities. This tends to keep tax rates higher and public good provision lower because neutral municipalities 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 municipalities. The policies, however, are Pareto efficient. Because of this, the economy tends to have a low enough level of debt to eliminate any volatility in policy choices. In contrast in the proportional system municipalities are treated ex ante symmetrically. However, policies are 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 a majoritarian system. When β is sufficiently large even a utilitarian planner would prefer a majoritarian system. From this discussion, we have the following result:

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

The analysis in this section emphasizes 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 follow 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, parties ignore a fraction $1 - \beta$ of the electorate, but the allocation 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

This paper presents a dynamic theory of electoral competition to study the determinants of fiscal policy. We have shown that while fiscal policy is generally Pareto efficient in static environments, Pareto efficiency depends on the electoral rule in dynamic environments. In the benchmark case in which municipalities 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 parties more discretion to target fiscal policies opportunistically toward the municipalities that are easier to swing. When municipalities are heterogeneous and there are municipalities with ex ante biases toward some candidate, there is a trade-off. While large electoral districts are more Pareto inefficient, they tend to be more inclusive because they do not provide incentives to ignore municipalities with biased preferences (that would not be contestable were the districts smaller). The theory therefore suggests that countries that have homogenous constituencies should find it optimal to choose a majoritarian system with smaller districts. Countries with heterogeneous municipalities have a 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 anticipated from static analyses since they depend on the dynamic properties of the electoral systems.

The analysis can be extended in several directions. First, we restrict the analysis to a majoritarian electoral system, and we consider only changes in the size of electoral districts. Clearly, it would be interesting to develop dynamic theories of electoral competition for a wider range of electoral rules. Second, we restrict the analysis to environments with symmetric municipalities and to environments with a very simple type of heterogeneity in preferences; we also focus 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 persists for a while (e.g., autoregressive shocks that have mean zero). Finally, it would be interesting to extend the analysis to study dynamic electoral competition with more than two parties.

References

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

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

Aiyagari, Rao, Albert Marcet, Thomas Sargent and Juha Seppala, (2002), "Optimal Taxation without State-Contingent Debt," *Journal of Political Economy*, 110, 1220-1254.

Alesina, Alberto and Guido Tabellini (1990), "A Positive Theory of Fiscal Deficits and Government Debt." *Review of Economic Studies*, 57, 403-414.

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), 1235-69.

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

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.

Banks, Jeffrey and Rangarajan Sundaram (1993), "Adverse Selection and Moral Hazard in a Repeated Elections Model," in William A. Barnett , Melvin J. Hinich and Norman J. Schofield (eds.) *Political Economy: Institutions, Information, Competition, and Representation*, pp. 295-311. Cambridge: Cambridge University Press.

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

Battaglini, Marco and Levon Barseghyan (2012), "Growth and Fiscal Policy: a Positive Theory," Princeton Economic Theory Center Working Paper n. 41.

Battaglini, Marco and Stephen Coate (2007) "Inefficiency in Legislative Policy-Making: A Dynamic Analysis," *American Economic Review*, 97(1), 118-149.

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 (2012a), "Legislative Bargaining and the Dynamics of Public Investment," *American Political Science Review*, 106(2), 407-429.

Battaglini, Marco, Salvatore Nunnari, and Thomas Palfrey (2012b), "The Free Rider Problem: a Dynamic Analysis," NBER Working Paper No. 17926.

Battaglini, Marco, Salvatore Nunnari, and Thomas Palfrey (2012c), "The Dynamic Free Rider Problem: A Laboratory Study," Princeton Economic Theory Center Working Paper n.48.

Battaglini Marco and Thomas Palfrey, "The Dynamics of Redistributive Politics," *Economic Theory*, 49 (3), 739-777.

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

Bernhardt Dan, Sangita Dubey and Eric Hughson (2004), "Term Limits and Pork Barrel Politics," *Journal of Public Economics*, 88:2383-2422.

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

Besley Timothy and Torsten Persson (2010), "State capacity, conflict and development," *Econometrica*, 78, 1-34.

Camara, Odilon (2012), "Economic Policies of Heterogeneous Politicians," University of Southern California, mimeo.

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

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

Coughlin, Peter (1982), "Pareto optimality of policy proposals with probabilistic voting," Public Choice, *Public Choice* 39:427-433

Coughlin Peter (2011), "Probabilistic and Spatial Models of Voting," In: Kenneth J. Arrow, Amartya Sen and Kotaro Suzumura, Editor(s), Handbook of Social Choice and Welfare, Elsevier, Amsterdam, Volume 2, Pages 833-896.

Coughlin, Peter and Shmuel Nitzan (1981), "Electoral outcomes with probabilistic voting and Nash social welfare maxima," *Journal of Public Economics*, 15, 113-121.

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

Duggan, John (2000), "Repeated Elections with Asymmetric Information," *Economics and Politics*, 12(2), 109-135.

Duggan, John and Tasos Kalandrakis (2012) "Dynamic Legislative Policy Making," *Journal of Economic Theory*, 147(5), 1653-1688.

Ferejohn, John (1986), "Incumbent Performance and Electoral Control," *Public Choice*, 50: 5-25.

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

Harstad, Bard (2012), "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations." *Review of Economic Studies*, 79 (4), 1527-1557.

Jackson, Kark (1975), "Issues, Party Choices and Presidential Votes," *American Journal of Political Science*, 19, 161-185.

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

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

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 Political Competition." *Public Choice*, 1987, 52, 273-97.

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

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

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

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

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

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

Persson, Torsten and Lars Svensson (1989), "Why a Stubborn Conservative Would Run a deficit: policy with Time-inconsistent Preferences." *Quarterly Journal of Economics*, 104, 325-346

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

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

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

Schwabe, Reiner (2009), "Reputation and Accountability in Repeated Elections" Princeton University, mimeo.

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

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

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

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

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

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

8 Appendix

9 Assumptions 1 and 2

In this section we formally derive the upper bound ΔW^* introduced in Section 2.2 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 $W^*(b)$ as the value of the following problem:

$$\max_{(r,g,x,s)} \left[u\left(r, g; \overline{A}\right) + s_i + \delta v^*(x)\right]$$

$$s.t. \;\; s_i \geq 0 \;\text{ for all } i, \; \sum\nolimits_i s_i \leq B(r, g, x; b), \;\&\; x \in [\underline{x}, \overline{x}]$$

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 $W^j(p_L; v, A) - W^j(p_R; v, A) \leq \Delta W^* = W^*(\underline{x})$, since $W^*(b)$ is decreasing in b. It follows that:

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

where $\Delta W^j(p; v, A) = W^j(p_L; v, A) - W^j(p_R; v, A)$. We conclude that $1/2 + \theta^j\left[\Delta W^j(p; v, A) - \widetilde{\theta}^G\right]$ is in $(0, 1)$ if the first inequality in Assumption 1 is satisfied. Party L 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 party L (respectively, R) wins less that $1/2$ of the votes in a right-wing (respectively, left-wing) municipality if $1/2 + \overline{\theta}[\Delta W^* - \sigma + 1/(2\underline{\theta}^G)] < 1/2$, which implies the first inequality in Assumption 2. By a similar logic, we can show that the share of votes received by any party is always nonnegative in any municipality if the second inequality of Assumption 2 is satisfied.

10 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 (6)-(7), note that from the first order conditions of (4) with respect to g and x we have:

$$\alpha A g^\circ(b; A)^{\alpha - 1} \;\; = \;\; \frac{1}{n}\frac{1 - r^\circ(b; A)}{1 - r^\circ(b; A)\,(1 + \varepsilon)}; \tag{27}$$

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

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

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

From the envelope theorem applied to (4) we have

$$-\delta\left[V^{o}\right]'(b)=\frac{1}{n}\frac{1-r^{\circ}(b;A)}{1-r^{\circ}(b;A)\,(1+\varepsilon)} \tag{30}$$

Condition (8) follows from (28), (30) and the fact that in equilibrium $\delta\,(1+\rho)=1$.

Because $(1-r)\,/\,[1-r\,(1+\varepsilon)]$ is a convex function in r, condition (8) implies that the tax rate $r^{o}(b;A)$ defines a supermartingale $(r^{o}_{\tau})_{\tau>0}$ 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 (27), 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 (4) converges to zero: otherwise the tax rate would depend on the realization of the shock A for an arbitrarily large τ. Conditions (27)-(29) imply that the tax rate converges to $r_{\mu}=(n\mu^{*}-1)\,/\,[n\mu^{*}(1+\varepsilon)-1]$ 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 $s^{\circ}_{i}(A,x)=B(r_{\mu},g_{\mu}(A),x;x)/n^{*}$ 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. ■

11 Proof of Proposition 2

We say that a function is κ-Lipschitz if it is Lipschitz continuous with constant of continuity κ. In this section we prove that there is an equilibrium such that the value function $v(b)$ is uniformly bounded, weakly concave and K-Lipschitz with a constant of continuity K specified below. To define this constant, let $g(\kappa)$ and $r(\kappa)$ be such that: $\alpha A\left[g(\kappa)\right]^{\alpha-1}=\kappa$ and $(1/n)\left[1-r(\kappa)\right]/\left[1-r\left(\kappa\right)(1+\varepsilon)\right]=\kappa$ and let $\Delta x=\max_{r}R(r)/\rho-\overline{x}>0$. Define, moreover,

$$B(\kappa)=R\left(r(\kappa)\right)-g\left(\kappa\right)-\max_{r}R(r)+\rho\Delta x.$$

Note that $B(\kappa)$ is continuous, increasing in κ and such that $\lim_{\kappa\to\infty} B(\kappa) = \rho\Delta x > 0$ and $\lim_{\kappa\to 1/(\delta n)} B(\kappa) < 0$ when Δx is small (i.e. $\Delta x < [\max_r R(r) + g(1/(\delta n)) - R(r(1/(\delta n)))]/\rho$). Define K such that $B(K) = 0$. Note that $K \in (1/(n\delta), \infty)$ for $\Delta x > 0$.

Let F be the metric space of real valued, uniformly bounded, weakly concave and K-Lipschitz 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 party or a given value function v can be expressed as (15). Note that for any b, the policy that solves (15) must be in P, where $P = [0,1] \times [0,\overline{g}] \times [\underline{x}, \overline{x}]$.[32] Let $P^*(v,z)$ be the correspondence that maps $v \in F$ to the set of bounded and continuous functions $p^*(b;v,z) = (r^*(b;v,z),\ g^*(b;v,z),\ x^*(b;v,z))$ such that: (1) $p^*(b;v,z)$ is a solution to (15), for all b; (2) $p^*(b;v,z)$ 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; (3) $p^*(b;v,z)$ is such that $B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z); b)$ is weakly convex in b; and (3) $p^*(b;v,z)$ is such that $B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z); b)$ is $1/\delta$-Lipschitz in b.

Lemma A.1. *The correspondence $P^*(v,z)$ is non-empty and convex-valued.*

Proof. We proceed in two steps.

Step 1. We first show $P^*(v,z)$ is non-empty by constructing a function $p^*(b;v,z) \in P^*(v,z)$. 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 (15). Let $\Omega_m(\theta) = \max_j \left[\theta^j / \left(\sum_{i\in H(j)} \theta^i\right)\right]/m$ and, for any state $z = (A,\theta)$, define r_z^* and g_z^* as

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

and let

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

If $\lambda(b;v,z) = 0$, then the solution of (15) is $r(b;v,z) = r_z^*$, $g(b;v,z) = g_z^*$ and $x(b;v,z) \in \mathcal{X}_z(v) \cap \{x \,|\, B(r,g,x;b) \geq 0\}$. Since v is weakly concave, $\mathcal{X}_z(v)$ is a non-empty, compact and convex set. 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 $B(r_z^*, g_z^*, x; \overline{b}(x;z)) = 0\ \forall x \in [\underline{x}, \overline{x}]$. Set $x^*(b;v,z) = \overline{x}(v,z)$ for $b \leq \overline{b}(\overline{x}(v,z);z)$. It follows that $\{r_z^*, g_z^*, x^*(b;v,z)\}$ is a solution of (15) in $b \in \left[\underline{x}, \overline{b}(\overline{x}(v,z);z)\right]$.

[32] We can assume $g \in [0,\overline{g}]$, with $\overline{g} < \infty$ without loss of generality since in every period resources are bounded by $\max_r R(r) + \overline{x}$.

Consider now $b > \overline{b}(\overline{x}(v,z);z)$. In this case $\lambda(b;v,z) > 0$, so (15) can be written as:

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

This problem admits a unique solution, so the set of solutions $P(b;v,z)$ is non-empty, convex-valued and continuous in b for all $b \in (\overline{b}(\overline{x}(v,z),z),\overline{x}]$.[33] Since the solution of (15) is upper hemicontinuous in $\left[\overline{b}(\overline{x}(v,z),z),\overline{x}\right]$ it must be that for any sequence $(b_n)_{n=1}^{\infty}$ with $b_n > \overline{b}(\overline{x}(v,z),z)$ converging to $\overline{b}(\overline{x}(v,z),z)$, we have $\lim_{n\to\infty} p(b_n;v,z) \in P(\overline{b}(\overline{x}(v,z),z);v,z)$. Since $P(\overline{b}(\overline{x}(v,z),z);v,z)$ 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 $\overline{b}(\overline{x}(v,z),z)$. We can therefore extend $(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z))$ to $b \in (\overline{b}(\overline{x}(v,z),z),\overline{x}]$, 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,

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

which is weakly convex in b for any $b \in [\underline{x},\overline{x}]$. To see that $B(r^*(b;v,z), g^*(b;v,z), x^*(b;v,z);b)$ is also $1/\delta$-Lipschitz in b, note that it is differentiable in b for $b > \overline{b}(\overline{x}(v,z);z)$, with the derivative equal to zero; and it is differentiable for $b < \overline{b}(\overline{x}(v,z);z)$ with the derivative equal to $-(1+\rho) = -1/\delta$.

Step 2. Let $p^1(b) \in P^*(v,z)$, $p^2(b) \in P^*(v,z)$, 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 $p^\phi(b)$ is an optimal reaction function. It can also be easily verified that $p^\phi(b)$ is continuous and monotonic. Moreover, $B(r^\phi(b), g^\phi(b), x^\phi(b);b)$ is weakly convex and $1/\delta$-Lipschitz in b. Therefore $p^\phi \in P^*(v,z)$. ■

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

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

[33] The fact that (32) admits a unique solution follows from the observation that if we had 2 different solutions, say (x_1,r_1,g_1), and (x_2,r_2,g_2), we must have $(r_1,g_1) \neq (r_2,g_2)$: otherwise we would have $(r_1,g_1) = (r_2,g_2)$ and only $x_2 \neq x_1$, implying that, in one of the two solutions, $B(r_i,g_i,x_i;b) \neq 0$. Given this, the result follows from the fact that $u(r,g;A) + \frac{1}{n}B(r,g,x;b)$ is strictly concave in r and g and weakly concave in x.

This expression defines a correspondence with the following properties:

Lemma A.2. *For any given z and K-Lipschitz v, any $\widetilde{v} \in \Upsilon(v, z)$ is K-Lipschitz.*

Proof. We proceed in three steps.

Step 1. Consider first the subset of the support in which $b \leq \overline{b}(\overline{x}(v, z); z)$. From the proof of Lemma A.1 we know that for any solution of (15) we must have a $\overline{b}(\overline{x}(v, z); z)$ such that for $b \leq \overline{b}(\overline{x}(v, z); z)$ the value function is:

$$\widetilde{v}(b; z) = u\left(r_z^*, g_z^*; A\right) + \frac{1}{n} \cdot B(r_z^*, g_z^*, x(b; v, z); x)) + \delta v(x(b; v, z)), \tag{33}$$

where $B(r_z^*, g_z^*, x^*(b; v, z); b))$ is $1/\delta$-Lipschitz and $x(b; v, z) \in \mathcal{X}_z(v)$. We can write (33) as:

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

The derivative of the first term is (by the envelope theorem) $-\Omega_m(\theta)(1 + \rho)$. Since $B(r_z^*, g_z^*, x(b; v, z); b)$ is $1/\delta$-Lipschitz, we have:

$$\left|\widetilde{v}(b_1; z) - \widetilde{v}(b_2; z)\right| / \left|b_1 - b_2\right| \leq \left| -(1 + \rho)\Omega_m(\theta) + \frac{1}{\delta}\left(\Omega_m(\theta) - \frac{1}{n}\right)\right| = \frac{1}{\delta n} \leq K,$$

for any b_1, $b_2 \leq \overline{b}(\overline{x}(v, z); z)$.

Step 2. Consider the subset of the support in which $b > \overline{b}(\overline{x}(v, z); z)$. When $b > \overline{b}(\overline{x}(v, z); z)$, the value function can be written as (32). By the envelope theorem applied to this maximization problem, we have $\widetilde{v}'(b; z) = -(1 + \rho)(1 + \lambda(b; z))/n$, where $\lambda(b; z)$ is the Lagrangian multiplier of the first constraint. We now prove that $|\widetilde{v}'(b; z)| \leq K$ for $b > \overline{b}(\overline{x}(v, z); z)$. Assume by contradiction that $(1 + \rho)(1 + \lambda(b; z))/n > K$ for some $b > \overline{b}(\overline{x}(v, z); z)$. Then, from the first-order necessary conditions of (32) with respect to r and g, we have $r(b; z) > r(K)$, $g(b; z) < g(K)$. Moreover, from the necessary conditions of (32) with respect to x, $x(b; z)$ must be such that if $x(b; z) < \overline{x}$, then

$$\frac{1}{n}(1 + \lambda(b; z)) \leq -\delta v^+(x(b; z)),$$

where $v^+(x(b; z))$ is the right derivative of $v(b)$ at $x(b; z)$ (well defined since $v(b)$ is concave). But by the initial assumption that we intend to contradict and the fact that v is K-Lipschitz, we have:

$$\frac{1}{n}(1 + \rho)(1 + \lambda(b; z)) > K \geq -v^+(x(b; z)),$$

Dividing both sides by $(1+\rho)$ and remembering that the equilibrium interest rate must be such that $\delta(1+\rho)=1$, we have $(1+\lambda(b;z))/n > -\delta v^{+}(x(b;z))$, implying $x(b;z)=\overline{x}$. It follows that the budget is

$$\begin{aligned} B(b,z) &= x(b;z) + R(r(b;z)) - g(b;z) - (1+\rho)\overline{x} \\ &> R(r(K)) - g(K) - \max_{r} R(r) + \rho\Delta x = B(K) \geq 0, \end{aligned}$$

where $\Delta x = \max_r R(r)/\rho - \overline{x}$ and the last inequality follows from the definition of K. We conclude that $\lambda(b;z)=0$, so $(1+\rho)(1+\lambda(b;z))/n = 1/(\delta n) \leq K$, a contradiction.

Step 3. We now prove that for any b_1, b_2 in $[\underline{x},\overline{x}]$, we have $|\widetilde{v}(b_1;z)-\widetilde{v}(b_2;z)| \leq K\cdot|b_1-b_2|$. The result was proven above when both b_1 and b_2 are less than $\overline{b}(\overline{x}(v,z);z)$. If both $b_1$and b_2 exceed $\overline{b}(\overline{x}(v,z);z)$, the result follows immediately from the fact that $|\widetilde{v}'(b;z)| \leq K$ in $[b_1,b_2]$, and the mean value theorem. Assume therefore that $b_1 \leq \overline{b}(\overline{x}(v,z);z)$, and $b_2 \geq \overline{b}(\overline{x}(v,z);z)$, with one of the two inequalities being strict. We can write

$$\begin{aligned} |\widetilde{v}(b_1;z)-\widetilde{v}(b_2;z)| &\leq \left|\widetilde{v}(b_1;z)-\widetilde{v}(\overline{b}(\overline{x}(v,z);z);z)\right| + \left|\widetilde{v}(\overline{b}(\overline{x}(v,z);z);z)-\widetilde{v}(b_2;z)\right| \\ &\leq K\cdot\left|b_1-\overline{b}(\overline{x}(v,z)\right| + K\cdot\left|\overline{b}(\overline{x}(v,z)-b_2\right| = K\cdot|b_1-b_2|, \end{aligned}$$

where the last equality from the premise that $b_1 \leq \overline{b}(\overline{x}(v,z);z) \leq b_2$. We conclude that for any given z and any K-Lipschitz v, any $\widetilde{v}\in\Upsilon(v,z)$ is K-Lipschitz. ■

Using Lemma A.1, we obtain:

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

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

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

$$+\left(\frac{1}{n}-\Omega_m(\theta)\right)B(r^*(b;v,z),g^*(b;v,z),x^*(b;v,z);b). \tag{36}$$

It can be verified that the first term of the expression above is concave. By Lemma A.1, $B(r^*(b;v,z),g^*(b;v,z),x^*(b;v,z);b)$ is weakly convex. Since $(1/n-\Omega_m(\theta)) \leq 0$, the second term of (36) is weakly concave. It follows that $\widetilde{v}(b)$ is uniformly bounded, weakly concave and, by Lemma A.2, K-Lipschitz in b. Since $P^*(v,z)$ is non empty, then $\Upsilon(v,z)$ is a non empty correspondence of F 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 (15). 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: $v_\alpha(b) = u(r_z^*,g_z^*;A) + B(r_z^*,g_z^*,x_z^\alpha;b)/n + \delta v(x_z^\alpha)$. Since $(r_z^*,g_z^*,x_z^\alpha) \in P^*(b;v,z)$ by Lemma A.1, we have $v_\alpha(b) \in \Upsilon(v,z)$. ■

We can also show that:

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

Proof. Consider a sequence of functions v^n that uniformly converges to v^∞. Let $\widetilde{v}^n \in \Upsilon(v^n,z)$ be a corresponding sequence and define $\widetilde{v}^\infty = \lim_{n\to\infty}\widetilde{v}^n$. By definition, to each $\widetilde{v}^n$ we have an associated policy: $p^n = (r^n,g^n,x^n) \in P^*(v^n,z)$. Define $p^\infty = \lim_{n\to\infty} p^n$. We have:

$$\begin{aligned} \widetilde{v}^\infty(b;z) &= \lim_{n\to\infty}\widetilde{v}^n(b;z) = \lim_{n\to\infty}\left\{\begin{array}{c} u(r^n(b;z),g^n(b;z)) + \frac{B(r^n(b;z),g^n(b;z),x^n(b;z);b)}{n} \\ +\delta v^n\left(x^n(b;z)\right) \end{array}\right\} \\ &= u(r^\infty(b;z),g^\infty(b;z);A) + \frac{B(r^\infty(b;z),g^\infty(b;z),x^\infty(b;z);b)}{n} \\ &\quad +\delta v^\infty\left(x^\infty(b;z)\right), \end{aligned}$$

for any $b \in [\underline{x},\overline{x}]$, so we have $\widetilde{v}^\infty \in \Upsilon(v^\infty,z)$, if $p^\infty \in P^*(v^\infty,z)$. Assume by contradiction $p^\infty \notin P^*(v^\infty,z)$. It is easy to verify that $v^\infty \in F$, so $P^*(v^\infty,z)$ is non empty by Lemma A.1. Therefore

there exists a $\eta > 0$ such that for any solution $p(b; v^\infty; z) = \{r(b; v^\infty; z), g(b; v^\infty; z), x(b; v^\infty; z)\}$ such that $p(b; v^\infty; z) \in P^*(v^\infty, z)$, we have:

$$\begin{aligned} & u(r(b; v^\infty; z), g(b; v^\infty; z); A) + \Omega_m(\theta) B(b; v^\infty, z)) + \delta v^\infty(x(b; v^\infty; z)) \\ \geq \quad & u(r^\infty(b; z), g^\infty(b; z); A) + \Omega_m(\theta) B^\infty(b; z) + \delta v^\infty(x^\infty(b; z)) + \eta, \end{aligned} \tag{37}$$

for some b, where $B(b; v^\infty, z)) = B(r(b; v^\infty; z), g(b; v^\infty; z), x(b; v^\infty; z); b)$, and

$$B^\infty(b; z) = B(r^\infty(b; z), g^\infty(b; z), x^\infty(b; z); b).$$

Since $v^\infty \in F$, there must be a n_1 such that for $n > n_1$:

$$\begin{aligned} & u(r^\infty(b; z), g^\infty(b; z); A) + \Omega_m(\theta) B^\infty(b; z) + \delta v^\infty(x^\infty(b; z)) \\ \geq \quad & u(r^n(b; z), g^n(b; z); A) + \Omega_m(\theta) B^n(b; z) + \delta v^\infty(x^n(b; z)) - \frac{\eta}{4}, \end{aligned}$$

where $B^n(b; z) = B(r^n(b; z), g^n(b; z), x^n(b; z); b)$. Since $v^n \to v^\infty$ in sup norm, for any η we must have a $n_2 > n_1$ such that for $n > n_2$ we have

$$\delta v^\infty(b) - \frac{\eta}{8} \leq \delta v^n(b) \leq \delta v^\infty(b) + \frac{\eta}{8} \quad \forall b \in [\underline{x}, \overline{x}].$$

This implies that for $n > n_2$:

$$\begin{aligned} & u(r(b; v^\infty; z), g(b; v^\infty; z); b); A) + \Omega_m(\theta) B(b; v^\infty, z) + \delta v^n(x(b; v^\infty; z)) \\ \geq \quad & u(r^n(b; z), g^n(b; z); b); A) + \Omega_m(\theta) B^n(b; z) + \delta v^n(x^n(b; z)) + \frac{\eta}{2}. \end{aligned}$$

But then $\{r^n(b; z), g^n(b; z), x^n(b; z)\} \notin P^*(v^n, z)$, a contradiction. ■

Let $\Gamma(z)$ be the distribution function associated to z (that is well-defined given $G(A)$ and $\varphi(\theta)$). Define the correspondence from F into F:

$$T(v) = \left\{ (\widetilde{v}(b) \middle| \exists l(b; z) \in \Upsilon(v, z) \text{ s.t. } \widetilde{v}(b) = \int l(b; z) \Gamma(dz) \right\}.$$

We have:

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

Proof. The fact that for $v \in F$, $\widetilde{v} \in T(v)$ is uniformly bounded, weakly concave, K-Lipschitz function is immediate. Moreover, since $\Upsilon(v, z)$ is non-empty, convex-valued, then $T(v)$ has these properties as well. To see that $T(v)$ has a closed graph, consider a sequence v_n with $\lim_{n\to\infty} v_n = v$

and the associated sequence $\widetilde{v}_n$ such that $\widetilde{v}_n = \int l_n(b;z)\Gamma(dz)$ and $l_n(b;z) \in \Upsilon(v_n, z)$ for all n. Then since v_n are all uniformly bounded, by the Lebesgue's bounded convergence theorem we have $\lim_{n\to\infty} \int l_n(b;z)\Gamma(dz) = \int \lim_{n\to\infty} l_n(b;z)\Gamma(dz)$. Define $l(b;z) = \lim_{n\to\infty} l_n(b;z)$. Since $\Upsilon(v,z)$ is a correspondence with closed graph, we have $l(b;z) \in \Upsilon(v,z)$. It follows that $\widetilde{v} = \lim_{n\to\infty} \widetilde{v}_n = \int l(b;z)\Gamma(dz) \in T(v)$. We conclude that $T(v)$ is a non empty and convex valued correspondence from F to F with a closed graph (and therefore upper hemicontinuous, since F is compact). The result 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: $v^*(b) = \int l^*(b;z)\Gamma(dz)$; and let $(r^*(b;v^*,z), g^*(b;v^*,z), x^*(b;v^*,z)) \in P^*(b;v^*,z)$, be 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 $\mathcal{H}(\theta) = \left\{ j \,\middle|\, \theta^j \geq \theta^i \ \forall i = 1, ..., n \right\}$, 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 party when v^* is the citizen's value function. Moreover, when the optimal strategy is

$$\left(r^*(b;v^*,z), g^*(b;v^*,z),\ x^*(b;v^*,z), s^*(b;v^*,z)\right),$$

then the expected value function is v^*. We conclude that $v^*(b)$ and $(r^*(b;v^*,z),\ g^*(b;v^*,z),\ x^*(b;v^*,z),\ s^*(b;v^*,z))$ constitute a political equilibrium.

12 Proof of Proposition 3

For a given equilibrium value function v the properties of the policy functions depend on the solution of (15). 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 (15) admits a solution $(r(b;A), g(b;A), x(b;A))$ independent of $\left(\theta^j\right)_j, \theta^G$, with $r(b;A)$ and $x(b;A)$ 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 with a preliminary result. For a state z, let $\mathcal{X}_z(v)$ be defined as in (31). We have:

Lemma A.6. *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 $b \in N_\varepsilon(\overline{\beta})$, $z \in N_\varepsilon(z')$: $R(r_z^*) - g_z^* + x(b;z) - (1+\rho)\,b < 0$ (where $N_\varepsilon(x)$ is a neighborhood of diameter ε centered at x). It follows that in states $b \in N_\varepsilon(\overline{\beta}) \cap \left[\underline{\beta}, \overline{\beta}\right]$, $z \in N_\varepsilon(z')$ we have $B(r(b;z), g(b;z), x(b;z); b) = 0$, and we can write $v(b;z)$ as the value of:

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

It follows that for any $b \in N_\varepsilon(\overline{\beta}) \cap \left[\underline{\beta}, \overline{\beta}\right]$, $z \in N_\varepsilon(z')$, we have $v(x;z)$ is strictly concave and so that $\delta v\,(x)$ is strictly concave in $b \in N_\varepsilon(\overline{\beta}) \cap \left[\underline{\beta}, \overline{\beta}\right]$ as well. This implies that it is not possible that $N_\varepsilon(\overline{\beta}) \cap \left[\underline{\beta}, \overline{\beta}\right] \subseteq \mathcal{X}_z(v)$, 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 that equilibrium debt must be $x(b;z) \in \left[\underline{\beta}, \overline{\beta}\right]$ and $B(r_z^*, g_z^*, x(b;z); b) \geq 0$ is never binding. Note that since $v(b)$ is concave, it has a right derivative, $v^+(x(b;z))$. Moreover, since v is a fixpoint of $T(v)$, as defined in the proof of Proposition 2, from Step 1 of Lemma A.2 we have $-v^+(x) \leq (1+\rho)/n$, or equivalently $-\delta v^+(x) \leq 1/n$. Since the previous inequality must be true for all z (and so all θ), we can then choose a θ such that $1/n < \Omega_m(\theta)$ (this is always possible since $\underline{\theta} < \overline{\theta}$). It follows that: $-\delta v^+(x) < \Omega_m(\theta)$ for all $x \in \left[\underline{\beta}, \overline{\beta}\right]$. This implies that $x(b;z) > \overline{\beta}$, a contradiction. ■

By Lemma A.6, the solution $x_m(\theta)$ of (17) is uniquely defined. If $\lambda(b;z) = 0$, the solution of (15) is then $r_m(\theta), g_m(A,\theta), x_m(\theta)$, given by (16)-(17). 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 $B(r_m(\theta), g_m(A,\theta), x_m(\theta); b) \geq 0$. If we define $A_m(\theta, b)$ from $B(r_m(\theta), g_m(A_m(\theta,b),\theta), x_m(\theta); b) = 0$. We conclude that the solution will be $r_m(\theta)$, $g_m(A,\theta)$ and $x_m(\theta)$, for $A \leq A_m(\theta,b)$, and $(r(b;A), g(b;A), x(b;A))$, otherwise, as stated in the proposition. ■

13 Proof of Proposition 4

We show that for any initial state, there is a positive probability of reaching a state z, b in which (24) 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 (32). Note that (32) is a contraction. It admits a unique fixed-point $v(b;A)$ 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 steady state we have $B(0, g_{1/n}(A), \underline{x}; \underline{x}) > 0$ for all realizations of A, a contradiction. ■

14 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. It is easy to prove that the transition function defined by $Q(S\,|b)$ has the Feller Property and that it is monotonic in b (see Ch. 12.4 in Stokey, Lucas and Prescott (1989) for definitions). Following the same steps as in Proposition 3 in Battaglini and Coate, we can moreover prove that $Q^n(S\,|b)$ satisfies the following condition:

Mixing Condition: *There exists an $\epsilon > 0$ and $m \geq 1$, such that $Q^m([\underline{x}, x^*]\,|\overline{x}) \geq \epsilon$ and $Q^m([x^*, \overline{x}]\,|\underline{x}) \geq \epsilon$.*

This condition requires that, for any $\epsilon > 0$, if we start out with the highest level of debt $\overline{x}$, then we will end up at x^* with probability greater than ϵ after m periods; while if we start out with the lowest level of debt $\underline{x}$, we will end up above x^* with probability greater than ϵ in m periods. Given these properties, the fact that there is a unique invariant distribution follows by Theorem 11.12 in Stokey, Lucas and Prescott [1989].

To characterize the remaining properties of the distribution, there are two cases to consider: $m < n$ and $m = n$. In the first case, the fact that the support of the distribution is as described in the proposition follows from Proposition 3. We only need to show that for all $b \in [\underline{x}, \overline{x}]$ and $A \in \left[\underline{A}, \overline{A}\right]$, we have $g(b; A) > 0$ and $r(b, A) < \overline{r}$. To see the first inequality, note first that if $A \leq A_m(\theta, b)$, then $r(b; z) = r_m(\theta) < \overline{r}$ and $g(b; A) \geq g_m(\underline{A}, \theta) > 0$. Assume, therefore, $A > A_m(\theta, b)$. In this case, by the first order necessary condition, we must have $\alpha A \left(g(b; z)\right)^{\alpha - 1} = 1 + \lambda(b; z)$, where $\lambda(b; z)$ is the Lagrangian multiplier of the budget constraint in (32). As shown in Step 2 of Lemma A.2, $1 + \lambda(b; z) \leq \delta n K$, where K is a constant defined in the proof of Proposition 2. It follows that $g(b; A) \geq [(\underline{A}\alpha)/K]^{\frac{1}{1-\alpha}} > 0$. The proof that $r(b, A) < \overline{r}$ is analogous and therefore omitted. To see that the distribution is non degenerate when $m < n$ assume by contradiction that it converges to a constant b^*. If $A_m(\theta, b^*) < \overline{A}$, then debt is a function of A with probability $1 - G(A_m(\theta, b^*)) > 0$; if $A_m(\theta, b^*) = \overline{A}$, then $-\delta v'(x_m(\theta); A) = 1/n$. Note that can always find a θ such that $\max_j [\theta^j / (\sum_{i \in H(j)} \theta^i)]/m > 1/n$. It follows that we must have a θ such that the steady state is $x_m(\theta) = \overline{x}$, and since the steady state is constant, this must be true for all θ. When $b^* = \overline{x}$, however, $A_m(\theta, b^*) < \overline{A}$, a contradiction. For the second case, 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. ■

15 Proof of Proposition 6

For the result in the first bullet, note that under Assumption 2, the optimal reaction function of the parties 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, therefore, follows from Proposition 1. For the result in the second bullet, note that under Assumption 2, σ is irrelevant for the fraction of votes received by a party in a district. The parties, therefore solve (15) with $m = 1$. The result follows from Proposition 3. ■