

A Political Economy Theory of Partial Decentralization*

John William Hatfield Gerard Padró i Miquel
Graduate School of Business London School of Economics
Stanford University NBER and BREAD

December 2008

Abstract

We revisit the classic problem of tax competition in the context of federal nations, and derive a positive theory of partial decentralization. A capital poor median voter wants to use capital taxes to provide public goods. This results in redistributive public good provision. As a consequence, when all public goods are provided by the central government, capital taxes and public good provision are high. The expectation of high capital taxes, however, results in a small capital stock which lowers returns to redistribution. The median voter would therefore like to commit to a lower level of capital taxes. Decentralization provides such a commitment: local governments avoid using capital taxes due to the pressure of tax competition. We therefore obtain that the median voter favors a *partial degree* of decentralization. The equilibrium degree of decentralization is non-monotonic in inequality, increasing in the redistributive efficiency of public good provision, and decreasing in capital productivity. When public goods are heterogeneous in their capacity to transfer funds, all voters agree that goods with high redistributive capacity should be decentralized.

*We are grateful to Tim Besley, Ernesto Dal Bó, Dennis Epple, Maitreesh Ghatak, Bard Harstad, Henrik Kleven, Torsten Persson, Andrea Prat, Ken Shotts, Jaume Ventura, Romain Wacziarg, John Wallis, Barry Weingast, David Wildasin and seminar participants at Universitat Pompeu Fabra and Stanford for helpful comments. The usual disclaimer applies. Any comments or suggestions are welcome and may be emailed to hatfield@stanford.edu or g.padro@lse.ac.uk.

1 Introduction

In any given political unit, citizens both participate in and are subject to public decisions taken at many different levels of government. In most countries the responsibility to provide goods to citizens is partially decentralized, with some goods and services provided and funded at the local level, while other goods are provided at the central level or even at an international level such as the European Union. In the United States, for instance, education is mostly funded at the local level, using real estate taxes, while spending on parks or highways is mostly decided at the federal level. An extensive theoretical literature provides normative analyses based on trading off different combinations of benefits and costs such as local preference diversity, economies of scale, public good or budgetary spillovers across districts, political agency considerations, and risk-sharing between districts.¹ The usual approach in this literature is to derive normative statements from comparing welfare in the polar cases of full centralization and full decentralization.

Tax competition has also received attention in this literature. Whether it is considered a positive or a negative consequence of decentralization depends on whether governments are treated as benevolent social planners or as Leviathan institutions populated by rent-seeking agents. In the former case, Oates (1972) first articulated the idea that tax competition between subnational units for mobile factors of production can force benevolent governments to engage in a “race to the bottom”, and either switch to more distortive sources of public funds or reduce public provision levels below optimality.² In the latter case, decentralization is positive insofar as tax competition helps restrain self-serving governments.³ Both of these literature streams

¹There is large early body of work assuming benevolent governments. This literature is normally referred to as fiscal federalism. An overview is provided by Oates (1999). More relevant to our analysis are studies that explicitly consider political agents whose incentives depend on the constitutional structure they face. Persson and Tabellini (1994, 1996), Seabright (1996), Alesina and Spolaore (1997), Bolton and Roland (1996), Lockwood (2002), Besley and Coate (2003), Tommasi and Weinschelbaum (2007) among many others balance several combinations of the trade-offs listed above from a political economy perspective. See Inman and Rubinfeld (1997), Oates (2005) and Weingast (2006) for reviews of this second-generation approach.

²Wilson (1986), Zodrow and Mieszkowski (1986) and Wildasin (1988) provide the first formal accounts of this mechanism and Wilson (1999) provides a review of the literature on tax competition in this tradition.

³Leviathan views of government envisage bureaucrats and politicians primarily preoccupied with increasing the size of government and therefore their own rents. This view, therefore, considers tax competition a welcome restraint on the size of government. Edwards and Keen (1996) and Wilson (1999) provide models in which competition improves the performance of Leviathan governments. Besley and Smart (2007) bridge these two views with an analysis of fiscal restraints in the presence of both benevolent and rent-seeking politicians.

follow tradition in that they normatively compare the extreme cases of complete decentralization and complete centralization. However, this question has seldom been approached from a positive perspective: what is the structure of the state that results from a constitutional political game in the presence of tax competition?

In this paper we propose a positive theory of partial decentralization. In this theory, the constitutional game results in a degree of decentralization that balances the desire for redistribution with the need to avoid highly distortive taxes. The framework we examine departs from the previous literature in two important ways. First, we assume that policies are determined by citizens in a political contest. Therefore governments are neither benevolent nor rent-seeking: they simply implement the policies that result from political competition. Second, our model allows us to focus on the *degree* of decentralization as opposed to the polar cases examined previously. This turns out to be crucial: in most circumstances, the result of the political contest is an intermediate degree of decentralization.

More specifically, we consider a setting with a central government and a large number of identical sub-units. Each sub-unit is populated by a continuum of citizens that differ in their capacity to deploy capital. There is a continuum of local public goods. Some of these public goods are to be funded and provided at the local level and the remainder of these goods are funded and provided at the central level. Each level of government has access to two sources of revenue: it can either tax capital invested within its jurisdiction or it can raise money using non-distortive head taxes. Taxes and public good provision levels are decided by majority voting among all citizens affected. Capital is raised before policies are voted on, but it is mobile and can therefore be invested in the local sub-unit that offers the best after-tax returns.

We show that despite the availability of non-distortive head taxes, centrally provided public goods are funded with capital taxes. The reason is that a relatively poor median voter prefers to shift the burden of taxation to large capital owners. In essence, by using capital taxes public good provision becomes redistributive in favor of capital-poor citizens.⁴ Because the median voter does not face the full marginal cost of taxation, she also votes for an excessive supply of centrally provided public goods. The level of capital taxes at the central level is therefore increasing in two variables: inequality in capital holdings—in particular the ratio of average to median

⁴Besley and Coate (1991) show that public provision of goods can serve as a redistribution device. In Alesina and Rodrik (1994) such redistribution can also occur through productive public goods. Hatfield (2006) considers how decentralization affects the provision of productive public goods. For an early take on the redistributive properties of public good provision see Buchanan (1964).

capital—and redistributive efficiency of public good provision.⁵ Intuitively, inequality naturally captures the demand for redistribution and redistributive efficiency captures the ease with which utility can be transferred using public good provision.

In contrast, competitive pressures ensure that public goods provided at the local level are funded via head taxes and consequently supplied at the efficient level. In essence, capital mobility across localities forces the median voter to renounce using decentralized public goods as redistributive channels because taxing capital implies losing it to neighboring districts.

Now consider a constitutional stage of the game where voters decide on the federal architecture of the country. In the simple framework under consideration, this reduces to a vote over the degree of decentralization, i.e. the fraction of public goods to be provided by the central government. Imagine that this vote takes place before capital, taxes, and public good decisions are made. The discussion above suggests that the median voter would like full centralization of public good provision, as this would enable the highest degree of redistribution funded with capital taxes unencumbered by tax competition among sub-units. However, this misses the fact that the capital stock is generated endogenously: in a fully centralized country voters set high capital taxes and this expectation distorts aggregate capital supply downwards. The median voter therefore faces a trade-off; increasing centralization allows for better redistribution of capital rents, but it also depresses capital supply thereby reducing the pool from which to redistribute. The solution to the constitutional vote balances these two forces and yields as the equilibrium a partially decentralized government structure.

Crucially, the constitutional stage allows voters to commit to a limited degree of capital taxation. In other words, voters use the federal structure of the constitution to partially tie their own hands *ex ante* and rein in their *ex post* desire for capital taxes.⁶ It follows that the stronger is the temptation to tax capital *ex post*, the higher the degree of decentralization that results from the constitutional vote. In our analysis, we find two interesting determinants of decentralization. First, decentralization increases monotonically with redistributive efficiency. This follows because high re-

⁵If the marginal utility of public good consumption diminishes very fast, public good provision is a poor redistributive channel. Conversely, if this marginal utility diminishes very slowly, public good provision behaves like a lump-sum transfer and therefore it is a channel with high redistributive efficiency.

⁶Obviously, this argument hinges on the fact that constitutional features are more resilient to change than policies such as taxes. This difference between policies and institutions is crucial in modern theories of institutional change such as Acemoglu and Robinson (2006). For another argument why federalism can be self-sustaining in equilibrium, see de Figueiredo and Weingast (2005).

distributive efficiency induces a strong temptation to set high capital taxes. Second, the equilibrium degree of decentralization is non-monotonically related to inequality; when inequality is small, centralization is not as dangerous as incentives to redistribute are moderate. At this low level, a marginal increase in inequality results in less decentralization as the median voter wants to increase redistribution at the margin. At high inequality, however, the temptation to heavily tax capital is much stronger. Hence, as inequality increases, the median voter is forced to commit to a moderate level of capital taxes. The resulting degree of decentralization is therefore decreasing in inequality when inequality is low, but increasing in inequality when inequality is high.

When we allow public goods to be heterogeneous in redistributive efficiency, we obtain a striking result. All voters agree on the ordering in which public goods should be decentralized. As a consequence, the set of Pareto optimal decentralization schemes is easy to characterize. In particular, we show that public goods with high redistributive efficiency are the best to decentralize as they are the most oversupplied when provided by the central government, and therefore constitute the worst temptation to raise capital taxes. The only point of disagreement between voters is over the amount of power that is to be devolved to the districts; the richer the agent is, the more decentralization she desires. It thus follows from our analysis that political parties that represent capital owners should favor increased levels of decentralization, as the Republican Party does in the United States.

We contribute to the literature in several ways. We analyze the determinants of federal structures and focus on the *degree* of decentralization as opposed to the comparison of institutional extremes that dominates previous work. Most importantly, we link the federalism literature with the commitment literature that is at the base of many explanations of endogenous political institutions.⁷ In our parsimonious model, we assume homogenous districts, no externalities, and direct democracy as a political mechanism. As a consequence, we abstract from the usual trade-offs examined in the federalism literature. These factors are, of course, important in determining the optimal structure of a federation, but are already well understood. Instead, our goal is to emphasize the overlooked dimension of commitment. Just as it is important in understanding other institutional arrangements, it can have a key role in determining the federal structure of a polity.

⁷The use of institutional arrangements to provide commitment to dynamically inconsistent policies is long known in economics. It is at the base of classic arguments for central bank independence (see, e.g., Barro and Gordon, 1983). It also features prominently in recent models of institutional change such as Acemoglu and Robinson (2001, 2006) and Besley and Persson (2007).

The underlying economic mechanism is closely related to the classic dynamic inconsistency problem in capital taxation in the macroeconomics literature.⁸ Indeed, Rogoff (1985), Kehoe (1989) and Tabellini (1990) show that policy coordination between countries might not be desirable because tax competition imposes potentially beneficial low capital taxes. We show that partial centralization provides a natural way to voters of trading off the benefits and the costs of this competition.⁹

Our analysis is also related to a small literature that is interested in the positive determinants of the structure of federations. Crémer and Palfrey (2000) examine whether voters support stringent control of district policy by federal authorities. Closer to our analysis, Crémer and Palfrey (1999) use a one-dimensional median voter framework to determine the relative weight that centrally voted policies versus district-level policies have on citizens' utility. This weight can be interpreted as the degree of decentralization. Their mechanism differs from ours in that uncertainty as opposed to commitment plays a prominent role.¹⁰ Bodenstein and Ursprung (2005) provide a numerical analysis of the degree of decentralization that results from a similar constitutional choice. In their paper, however, the emphasis is on the benefits that political yardstick competition conveys. Finally, Dixit and Londregan (1998) take the structure of government as given but also consider the redistributive consequences of federal arrangements.

The remainder of the paper is organized as follows. The next section describes the general model of partial decentralization. Section 3 provides the analysis of the model. Section 4 discusses the main intuitions of the paper. The following section takes a few functional form assumptions in order to examine the degree of decentralization favored by each voter and provide rich comparative statics. Section 6 provides an extension of the model to heterogeneous public goods. The last section concludes. All proofs are in the Appendix.

2 The General Model

The economy is divided into J identical districts, each with its own local government, and each with a total mass 1 of individuals. There are two levels of government,

⁸For an early account, see Kydland and Prescott (1977) and Fischer (1980). Klein et al (2004) provide a useful comprehensive analysis and literature review.

⁹Persson and Tabellini (2000, Ch. 12) discuss that federalism can be a solution to commitment problems. We extend this insight to explicitly analyze the *degree* of decentralization that results in a constitutional game.

¹⁰See also Crémer and Palfrey (1996) for a discrete version of the same mechanism.

the district level (which captures municipality or state level) and the central government (which captures the federal level or an international level such as the European Union). These administrations use their revenues to provide local public goods to the citizens within their jurisdiction. There is a continuum of size 1 of such homogeneous public goods. District governments are responsible for providing goods $[0, \lambda]$, and the national government is responsible for provision on goods $(\lambda, 1]$. λ therefore constitutes a measure of the *degree of decentralization* in this economy and, for the moment, we take it as given.

Administrations must raise revenues to meet their expenses in public goods. We consider a simple economy with a single factor of production, k . Governments have access to both a tax on capital within their jurisdiction and a head tax.¹¹ Denote by τ and T the tax level on capital and the head tax levied by the central government and by τ_j and T_j the level of taxation levied by district j .¹² We shall assume that taxes are constrained to be nonnegative. Denote by $s(p)$ the amount of spending on public good $p \in [0, 1]$ by the administration responsible for its provision (we denote by $s_j(p)$ the level of spending on good p by district j). With this notation, the budget constraint for the district government is given by

$$\int_0^\lambda s_j(p) dp = \tau_j k_j + T_j \quad (1)$$

where k_j denotes the amount of capital invested in district j . The budget constraint for the central government is given by

$$\int_\lambda^1 s(p) dp = \tau k + T \quad (2)$$

where k is the economy-wide average amount of capital holdings.¹³

Consumption goods are produced by a continuum of firms at the district level using capital. Denote by $F(k_j)$ the production function. $F(k_j)$ is increasing, weakly concave, and smooth. Since it does not necessarily display constant returns to scale, we assume that the district accrues the returns that are not captured by capital owners. Such returns are shared equally by residents of district j .¹⁴ Denote by ρ_j the

¹¹To ensure the existence of a Condorcet winner in the policy space, we constrain the tax on capital to be linear. See Romer (1975), Roberts (1977), and Meltzer and Richard (1981).

¹²Our results do not change if τ_j and τ are taxes on capital returns and not capital investment.

¹³The uniform provision of local public goods allows us to write the budget constraint at the district level. The global budget constraint for the central government would be $J \left(\int_\lambda^1 s(p) dp \right) = (\tau k + T) J$ which is obviously the same.

¹⁴These can be unmodeled returns to land or unskilled and non-mobile labor. The assumption of

pre-tax rate of return to capital in district j . Competition within districts implies that capital captures its marginal contribution to production, or

$$\rho_j = F'(k_j) \quad (3)$$

for each district. Moreover, we assume that capital is perfectly mobile across districts. This implies that after-tax returns must be equalized. For large J , it follows that

$$r = \rho_j - \tau_j - \tau \quad (4)$$

where r is the net return to capital and in equilibrium it is uniform throughout the economy.

We can now proceed to describe the preferences of citizens and their economic opportunities. Citizens are endowed with some wealth that they use as collateral to obtain capital. Afterwards they invest this capital somewhere in the economy. The initial wealth endowment β is the only source of heterogeneity in the model. Wealth is identically distributed across districts according to some cumulative distribution function $H(\cdot)$ on $[\beta^{\min}, \beta^{\max}]$. An agent n in district j is thus endowed with β^n . If she wants to invest k^n units of capital, she needs to raise $k^n - \beta^n$. We assume that there is a credit market friction such that the repayment interest rate for her loan, l , is increasing in the leverage ratio of her investment: $l\left(\frac{k^n}{\beta^n}\right)$ where $l(\cdot)$ is increasing and convex.¹⁵ Therefore, her total cost of investing k^n amounts to $(k^n - \beta^n) l\left(\frac{k^n}{\beta^n}\right)$.¹⁶

Agent n 's preferences are given by

$$u(c^n, s(j), k^n, k_j; \beta^n) = c_j^n + \int_0^1 G(s(p)) dp - (k^n - \beta^n) l\left(\frac{k^n}{\beta^n}\right)$$

where c^n denotes agent n 's consumption and $G(\cdot)$ is a smooth, increasing, and concave

equal sharing in these returns allows us to focus on different capital holdings as the sole source of inequality in this economy. In any case, as the example in section 3 shows, the presence of these returns is by no means essential to our argument.

¹⁵There are many examples of models of credit markets with frictions that yield interest rates decreasing in wealth. See, for instance, Ghatak, Morelli and Sjöström (2007) or Banerjee (2003). See Stiglitz and Weiss (1981) for a seminal paper that analyzes credit markets with imperfect information, and Tirole (2006) for a textbook approach to the issue.

¹⁶Imperfect capital markets is just a simple way of introducing endogenous and unequally held capital in the economy. We want to capture the idea that available productive capital in an economy is a function of past investment and savings decisions. Obviously, heterogeneous capital holdings could be endogenized in many other ways. In the case of human capital, β could capture a differential cost of attending school. In the case of past labor supply decisions, β could simply scale labor supply costs. All is needed is a capital generation cost $\nu(k^n; \beta^n)$ increasing and convex in k^n and satisfying a single crossing condition with respect to β^n .

function of spending in the publicly provided goods that the agent will enjoy.¹⁷ Given the taxation patterns described above, agent n in district j enjoys consumption equal to

$$c_j^n = rk^n + F(k_j) - \rho_j k_j - T_j - T \quad (5)$$

where rk^n are the net returns to her capital investment, which she can invest anywhere in the economy. $F(k_j) - \rho_j k_j$ are her returns to living in district j . Finally, $-T_j - T$ are head taxes she incurs living in district j .

The timing of the model is as follows.

1. Each agent n in each district j decides how much capital to raise, k^n .
2. By simple majority, taking the net rate of return to capital r , and the district budget constraint as given, the citizens in each district choose the Condorcet winner in their policy space $(\tau_j; T_j; s_j(p), p \in [0, \lambda])$.
3. By simple majority, taking the budget constraint as given, all the citizenry chooses the Condorcet winner in the policy space of the central government $(\tau; T; s(p), p \in (\lambda, 1])$.
4. After observing taxation patterns across the economy, agents decide in which district to invest their productive capital, k^n .

This model has a number of noteworthy features. First, note that the local public goods we discuss here could also be publicly provided private goods, given that we do not allow for citizen mobility. In particular, note that a citizen of district j does not obtain any utility from resources spent by district i in these goods. Therefore we abstract from cross-district spillovers.

Second, while citizens are heterogeneous in their endowment β , districts are identical because the distribution of β is the same across districts. Hence, we also abstract from district heterogeneity.

Third, we use a quasilinear utility function for consumption and publicly provided goods. With this assumption, we ensure that there are no income effects in the enjoyment of such public goods and all citizens enjoy them in the same degree. Any tension in deciding the provision level will thus come from the use of taxes that affect citizens differently. Another helpful consequence of quasilinearity is the fact that

¹⁷For an agent living in district j , $s(p) = s_j(p)$ for goods $[0, \lambda]$. Goods $(\lambda, 1]$, being provided by the central government, have a funding level of $s(p)$ that is equal across districts.

the outcomes of the political contest at the district level become separable from the political contest at the central level.

Finally note that for simplicity we focus on the case where J is large and therefore all districts take the net rate of return on capital r as given.

2.1 Definition of Equilibrium

For a given level of decentralization λ , a subgame perfect equilibrium is a capital investment decision function $k(r, \beta)$ for $\beta \in [\beta^{\min}, \beta^{\max}]$, policy decisions $T, \tau, s(p)$ for $p \in [\lambda, 1]$, $\{T_j, \tau_j, \rho_j\}_{j=1, \dots, J}$, $\{s_j(p) \text{ for } p \in [0, \lambda]\}_{j=1, \dots, J}$, an after-tax rate of return on capital r , and investment location decisions such that

1. Capital markets are perfectly competitive both intra- and interdistrict: $\rho_j = F'(k_j)$ in each district and r , the after-tax rate of return in each district, is $\rho_j - \tau - \tau_j$.
2. The district citizens, taking the rate of return on capital r and their budget constraint as given, choose the Condorcet winner in their policy space.
3. The citizenry, taking the central government budget constraint as given, chooses the Condorcet winner in its policy space.
4. Agents choose to invest an amount of capital $k(r, \beta)$, $\beta \in [\beta^{\min}, \beta^{\max}]$ to maximize their utility.

We note that, as in many models where the tax base reacts to expected taxation levels, there may be multiple equilibria. In particular, there may exist equilibria on the wrong side of the Laffer curve, where the total revenue of the government is locally decreasing in the level of taxation. In such situations, we only consider equilibria in the increasing side of the Laffer curve, i.e. that a small increase in the (expected) capital tax rate yields an increase in government revenues. We call such equilibria, *standard equilibria*.

3 Characterization of Equilibrium

3.1 The Problem of the Central Government

We begin the analysis at stage 3. At this stage, the whole of the citizenry decides on capital taxes, head taxes and public good provision by the central government.

This vote takes place after capital raising decisions have been taken and therefore the resulting policy will depend on the predetermined distribution of capital across voters.

We first introduce some notation to describe such capital holdings. Let $k(\beta^n, r)$ be the amount of capital held by an agent with endowment β^n who expects an after-tax rate of return on capital r . It follows that

$$k(r) \equiv \int_{\beta^{\min}}^{\beta^{\max}} k(\beta, r) dH(\beta)$$

is the amount of capital held in a district when the expected net return is r . Since there is a mass of size 1 of citizens in each district, $k(r)$ is also the average amount of capital per voter. Also, we let $k^{\text{med}} \equiv k(\beta^{\text{med}}, r)$ be the amount of capital held by the citizen with the median endowment, β^{med} .¹⁸ We shall assume that

$$1 < \Phi(r) \equiv \frac{k(r)}{k^{\text{med}}} < \infty$$

for all r . Therefore we consider unequal capital holdings such the median voter has less capital than the average taxpayer.¹⁹ $\Phi(r)$ thus captures the level of *ex post* inequality (after capital decisions are taken).

Inequality generates political conflict between voters. Capital rich voters prefer to use head taxes to fund public good provision so that every recipient of the public good pays the same amount for it. In contrast, capital poor voters prefer to use capital taxes. Since capital is unequally held but public good enjoyment is uniform, public good provision funded by capital taxes becomes a redistributive tool.

To characterize the result of the political contest, we shall first consider the preferred policy from the perspective of the median voter, i.e., the agent with endowment β^{med} , and later show that the median voter's ideal point is indeed a Condorcet winner within the policy space of the central government.

The problem for the median voter in some district j is

$$\max_{T, \tau, s(p)} \left\{ c_j^{\text{med}} + \int_0^\lambda G(s_j(p)) dp + \int_\lambda^1 G(s(p)) dp - (k^{\text{med}} - \beta^{\text{med}}) l\left(\frac{k^{\text{med}}}{\beta^{\text{med}}}\right) \right\}$$

¹⁸Therefore, $H(\beta^{\text{med}}) = \frac{1}{2}$.

¹⁹This is a standard assumption in voting settings with inequality. See, for instance, Persson and Tabellini (2000) and references contained therein.

subject to the central government's budget constraint

$$\int_{\lambda}^1 s(p) dp = \tau k + T.$$

However, the relevant maximand for the median voter is much smaller for two reasons. First, k^{med} is determined before this stage, and hence the costs of raising it are irrelevant at the time of voting. Second, the utility function is quasilinear in consumption. As a result, additive separability allows us to drop many terms. Using the expression for citizen's consumption (5) and the capital markets condition (4), the median voter's problem reduces to

$$\max_{T, \tau, s(p)} \left\{ -\tau k^{\text{med}} - T + \int_{\lambda}^1 G(s(p)) dp \right\}$$

subject to the national government's budget constraint.

Consider first the decision of how to spend in public goods $s(p)$ a fixed amount of raised revenues. The concavity of G makes this problem particularly simple: all voters agree that any revenues raised at the central government should be spent equally across the $1 - \lambda$ public goods the central government is responsible for. It follows that $s(p) = \frac{\tau k + T}{1 - \lambda}$.

Since average capital holdings k are fixed at the time of voting, capital taxes are not distortive *ex post*. The median voter thus faces a choice between two nondistortive forms of taxation. This choice is easy: a median voter with less than average capital always prefers capital taxes over head taxes. This is because capital taxes allow her to shift the tax burden to citizens with higher capital holdings. Hence, as long as $\Phi \equiv \frac{k}{k^{\text{med}}} > 1$, the preferred tax policy of the median voter is

$$\begin{aligned} G' \left(\frac{\hat{\tau} k}{1 - \lambda} \right) &= \Phi^{-1} < 1 \\ \hat{T} &= 0. \end{aligned} \tag{6}$$

With quasilinear preferences and available head taxes, the efficient level of public provision is set at $G'(\cdot) = 1$. Hence, according to (6), the median voter favors excess provision of public goods. To build intuition for this result, recall that the median voter is using public provision of goods as a redistributive tool from the average capital owner to herself. Since she faces less than the average tax burden, she increases redistribution by voting for an expanded supply of public goods.

Condition (6) also shows that the total revenue raised in capital taxes, $\hat{\tau} k$, in-

creases in *ex post* inequality, Φ . This is natural: the less capital the median voter holds, the more she benefits from using capital taxes. In the limit in which she held no capital, she would face no tax burden and therefore she would vote for infinite taxation. Therefore, a larger degree of inequality increases the desire for redistribution in a natural result reminiscent of Meltzer and Richard (1981) and Roberts (1977).

Finally, note that the shape of $G(\cdot)$ also affects the degree to which central public goods are overprovided. Since in this model redistribution is channeled through public good provision, $G(\cdot)$ ultimately affects how efficiently utility can be transferred from rich voters to poor voters. If $G(\cdot)$ is very concave, then the marginal utility of public goods diminishes very fast and the median voter does not gain much from an excess supply of public goods. In such a case, we say that *redistributive efficiency* is low.²⁰ Conversely, if $G(\cdot)$ is close to linear, large quantities of public good can be provided before diminishing marginal utility sets in. In such a case of high redistributive efficiency, (6) implies that capital tax revenues are much larger and central public goods are grossly oversupplied.

In sum, given λ , capital taxes and central public good provision are increasing in the demand for redistribution—captured by inequality, Φ —and in the ease with which utility can be transferred—which we call redistributive efficiency.

It remains to be shown, however, that the preferences of the median voter constitute the Condorcet winner under majority voting. Notice that this is not immediate given that the problem is multidimensional. To see that the policy preferred by the median agent is indeed a Condorcet winner, consider figure 1. In this figure we consider taxation units such that iso-budget curves slope at minus 45 degrees.²¹ Consider any policy other than the most favored by the median voter. If it involves a higher capital tax, as well as a positive head tax (a policy in region A of figure 1), then the policy with the same capital tax and no head tax will be favored by all voters; at its turn, this policy is dominated by the policy with no head taxes and a capital tax $\hat{\tau}$ in the eyes of all voters with an amount of capital greater than or equal to the median amount of capital. Hence, by individual transitivity of preferences, $(\hat{\tau}, 0)$ is preferred to any alternative policy in region A by a majority of voters. If the alternative policy has a higher total level of funding, but a lower capital tax (a policy in region B of

²⁰This is probably true of expenditures in public parks or basic infrastructure such as the judiciary. Simply put, a bloated judicial system is clearly a very poor way of transferring utility from rich voters to poor voters. Redistributive efficiency using this channel is therefore very low.

²¹As discussed above, given a set amount of revenue, it is clear that all voters wish to spend it in the most efficacious way on public goods. Therefore the political tension is over the level of T and τ .

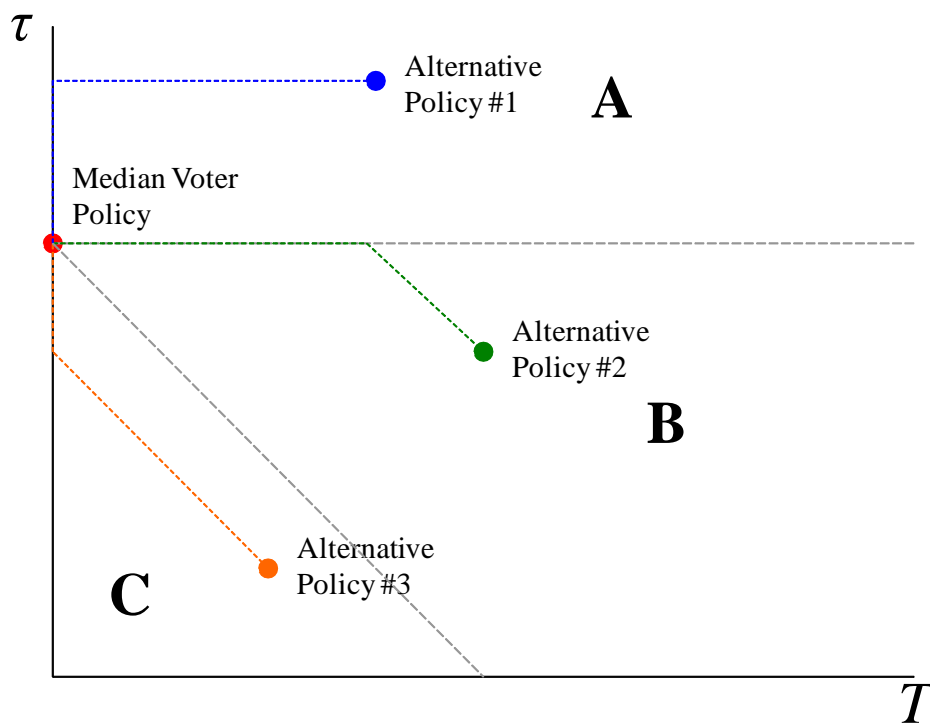


Figure 1: The graph shows possible policy alternatives to the median voter's ideal point, and how the median voter's ideal point must be preferred by a majority of voters: see discussion in text.

figure 1) then the policy that holds funding constant, while decreasing the head tax and increasing the capital tax to $\hat{\tau}$ is favored by all voters with less capital than the mean. At this point, reducing the head tax to 0 is favored by all voters, since with a capital tax of $\hat{\tau}$ each public good has a marginal return of less than one. Hence, again by individual transitivity of preferences, $(\hat{\tau}, 0)$ is preferred to any policy in region B by a majority of voters. Finally, consider a policy with a lower total level of funding than $(\hat{\tau}, 0)$ (a policy in region C). Again, decreasing the head tax to zero and increasing the capital tax while holding funding constant is favored by all voters with less capital than average. After that, increasing the capital tax is favored by all voters with less than or equal to the capital of the median voter. Hence, by individual transitivity of preferences, $(\hat{\tau}, 0)$ is preferred to any policy in region C by a majority of voters.

We summarize these results in the following proposition:

Proposition 1 *The central government will exclusively use capital taxes, and will set $G' \left(\frac{\tau k}{1-\lambda} \right) = \Phi^{-1}$, which provides more than the efficient amount of public good.*

3.2 The Problem of the District Government

The political game at the district level is very different from the conflict with respect to central government policies. The ability of capital to flee from districts with high capital taxation plays a crucial role in this difference. Recall that district policies are decided after capital is raised but before citizens decide in which district to invest.

Consider the district tax and spending policy favored by a given agent n with endowment β^n . The problem of this generic voter n in district j is:

$$\max_{k_j, T_j, \tau_j, s_j(p)} \left\{ c_j^n + \int_0^\lambda G(s_j(p)) dp + \int_\lambda^1 G(s(p)) dp - (k^n - \beta^n) l\left(\frac{k^n}{\beta^n}\right) \right\}.$$

Again, the relevant maximand is much simpler. k^n , the amount of capital that agent n commands, is determined before this stage; this is in contrast to k_j , the amount of capital invested in district j , which *is* affected by tax policy. Quasilinearity of preferences also implies again that many terms drop from the problem. Using the expression for citizen's consumption (5), we can reduce the program to

$$\max_{k_j, T_j, \tau_j, s_j(p)} \left\{ F(k_j) - \rho_j k_j - T_j + \int_0^\lambda G(s(p_j)) dp_j \right\} \quad (7)$$

subject to the district budget constraint

$$\int_0^\lambda s_j(p) dp = \tau_j k_j + T_j$$

and the constraint from capital mobility

$$r = \rho_j - \tau_j - \tau.$$

This program has two important and related features that distinguish it from the median voter's problem of the last subsection. First, there is a direct relationship between τ_j and k_j given by the capital mobility constraint. A district that chooses high capital taxes suffers from lower capital supply as capital flees to neighboring districts. This is costly because district j voters care about k_j . Through $F(k_j) - \rho_j k_j$ citizens can appropriate any returns to production that are not assigned to capital. In addition, any revenues collected using capital taxes, $\tau_j k_j$, can reduce head taxes and help provide district public goods. For these two reasons citizens want to attract capital to their districts. In contrast, since aggregate capital k is fixed at the central level, there is no relationship between τ and k and hence there is no direct cost of

increasing central capital taxes.

Second, note that k^n (and therefore β^n) is absent from (7). The reason is that each district is too small to affect aggregate net returns to capital r . Capital owners in district i are therefore safe from high capital taxes in their home district as they can obtain the rate of return r simply by moving their capital to district j . Hence, the fact that agents command different amounts of productive capital is inconsequential, and agent heterogeneity drops out of the district problem. This implies that there is no political conflict within the district: all citizens agree on tax and spending decisions at the district level. In contrast, at the central level capital cannot escape taxation. Central capital taxes reduce net returns to capital thereby generating a political conflict between large and small capital holders.

Since all agents share the same preferences with respect to district policies, the Condorcet winner is simply the policy most preferred by every single agent.

Note again that it is immediate from the concavity of G that $s_j(p) = \frac{\tau_j k_j + T_j}{\lambda}$. Plugging this condition in the objective function we obtain

$$\max_{k_j, T_j, \tau_j} \left\{ F(k_j) - \rho_j k_j - T_j + \lambda G\left(\frac{\tau_j k_j + T_j}{\lambda}\right) \right\}$$

Now, taking the first order condition with respect to T_j it is immediate that

$$G'\left(\frac{\tau_j k_j + T_j}{\lambda}\right) = 1 \tag{8}$$

which implies the Samuelson condition: each good is provided at the efficient level as the opportunity cost of public funds equals 1. The first order condition with respect to k_j , plugging in (8), yields

$$F'(k_j) - \rho_j + \tau_j = 0$$

which, using (3), immediately implies $\tau_j = 0$. Therefore public goods at the district level are entirely funded by head taxes. Intuitively, as long as any head tax is used, (8) implies that the level of public provision is fixed. Moreover, districts want to attract capital in order to reduce their head tax burden. This leads to competition across districts. Assuming that districts are price takers (i.e. they take r as given), competition for capital is perfect and hence all rents at the district level are dissipated in equilibrium: capital taxes are driven to 0 and local public goods are fully financed

by head taxes.²² We have established the following proposition:

Proposition 2 *For any level of decentralization λ , each district government will efficiently provide each public good (i.e. $G'(s_j(p)) = 1$ for all $j = 1, \dots, J, p \in [0, \lambda]$) using only a head tax.*

3.3 The Initial Problem of the Citizen

At the initial stage of the game, citizens decide how much capital to raise. The problem of agent n in district j is simply to pick the k^n level that maximizes $u(c^n, s(j), k^n, k_j; \beta^n)$ for an equilibrium-consistent expectation over r, τ_j, τ, T_j , and T . Again, because quasilinearity implies additive separability and individual decisions do not affect aggregate k and k_j , this problem reduces to

$$\max_{k^n} \left\{ r k^n - (k^n - \beta^n) l \left(\frac{k^n}{\beta^n} \right) \right\}$$

for an equilibrium-consistent expected return r . Note that due to capital mobility, this decision is independent of the district j where the agent lives. Thanks to the convexity of $l(\cdot)$, this objective function is globally concave. The first order condition is

$$r - l \left(\frac{k^n}{\beta^n} \right) - \frac{k^n}{\beta^n} l' \left(\frac{k^n}{\beta^n} \right) = 0 \tag{9}$$

which implicitly defines the capital holding function $k(\beta^n, r)$. Note that the objective function is supermodular in k^n and β^n . It follows that citizens with larger endowments β will raise more capital. It is also immediate from the maximization problem that the citizens will, regardless of endowment, demand more capital the higher the expected after-tax rate of return. This implies that a high expected level of capital taxes results in a lower level of aggregate capital. This is at the core of the time inconsistency problem that we discuss in the next section.

²²Essentially, the district is trying to maximize returns at the district level buying capital from a competitive market for capital at price $r + \tau$. Efficiency then requires that capital is rented up to the level where $F'(k_j) = r + \tau$. If the district taxes capital at any positive level, capital mobility condition (4) requires that the district obtains less capital than would be efficient. Therefore, it is optimal to refrain from capital taxation and raise revenues using exclusively head taxes.

4 Time Inconsistency, Capital Taxation, and Decentralization

Capital decisions are made before voting on taxes. As a consequence, the median voter takes the amount of capital in the economy as given, and her desired taxes only consider the trade-off between her individual consumption lost to capital taxes and her desire for public provision of goods. To see this, condition (6) can be rewritten as

$$kG'(\cdot) - k^{\text{med}} = 0. \quad (10)$$

The first term captures the marginal gains from increasing capital taxes, namely an increased level of public good spending. The only marginal cost associated to such increase is the capital tax that the median voter herself has to pay, captured by the second term.

In contrast, if the median voter had to decide on capital taxes *ex ante*, before capital decisions are made, she would consider an additional cost of high marginal taxes. This additional cost is given by the fact that an increase in capital taxes depresses capital stocks and therefore reduces the pool from which to redistribute. To establish this, consider the following *ex ante* problem of capital taxation at the central level.

$$\max_{\tau} \left\{ \begin{array}{l} (F'(k) - \tau) k^{\text{med}} + (F(k) - F'(k)k) - T_j \\ + \lambda G\left(\frac{T_j}{\lambda}\right) + (1 - \lambda) G\left(\frac{\tau k}{1 - \lambda}\right) \\ - (k^{\text{med}} - \beta^{\text{med}}) l\left(\frac{k^{\text{med}}}{\beta^{\text{med}}}\right) \end{array} \right\}$$

where we already take into account that districts use head taxes and the central government will only use capital taxes. The first order condition that determines the level of capital taxes that the median voter prefers *ex ante* is

$$kG'(\cdot) - k^{\text{med}} + (\tau G'(\cdot) - F''(k)(k - k^{\text{med}})) \frac{\partial k}{\partial \tau} = 0. \quad (11)$$

The third term in (11), which is the only difference with respect to (10), captures the effect of capital taxes on aggregate capital supply. Obviously $\frac{\partial k}{\partial \tau} < 0$, which implies that this term is negative.²³ Within the parentheses, we see that the median voter cares about aggregate capital k for three reasons. First, $\tau G'(\cdot)$ captures the

²³Note that the envelope theorem ensures that the effect of changing τ on the median voter's utility through the change in k^{med} is zero.

fact that a larger capital stock implies more public goods for the same level of taxes. Second, $-F''(k)k$ captures the marginal effect on the gains to the district when capital supply is higher. Finally, the only cost of an increased capital supply is the fact that the marginal return to capital diminishes. The size of this effect on the median voter is given by $F''(k)k^{\text{med}}$. Since $k^{\text{med}} < k$, it is clear that the first two forces dominate the third, and the median voter prefers an enlarged aggregate stock of capital. As a consequence, the capital tax that she would vote for *ex ante* is lower than the capital tax she ends up voting for *ex post*.

We have established the following result.

Lemma 1 *The median voter's preferred capital tax rate at the beginning of the game is lower than the Condorcet winning capital tax rate ex post.*

Since agents are forward looking, they expect high capital taxes and respond by reducing the capital stock, which hurts the redistribution flows that the median voter perceives *ex post*. As a consequence, in equilibrium, despite the fact that the preferred *ex post* rate of the median voter is implemented, she would gain from the ability to commit to lower capital taxes *ex ante*.

It is, however, notoriously difficult to commit to policies such as taxes. There are many reasons why governments want to tailor taxation policies to time-varying circumstances and as a consequence tax schedules are typically determined on a yearly basis. As long as the horizon of capital investments is longer than the interval between tax changes, voters cannot commit not to tax capital once it is deployed, and hence they suffer from this lack of commitment.

This simple model, however, suggests a second-best solution to this commitment problem. In the analysis of the model in Section 3, we have taken the degree of decentralization λ as a given institutional parameter. However, it is easy to show that λ affects overall capital taxes in equilibrium because the public goods that are provided by the central government are funded via capital taxes, while district public goods are funded using head taxes. It follows that an increase in decentralization should lead to a reduction in overall capital taxation. The following proposition establishes that this is indeed the case when we have a unique standard equilibrium.

Proposition 3 *If Φ does not increase too fast in τ , any standard equilibrium is unique, and the equilibrium capital tax rate $\tau + \tau_j$ is a decreasing function of decentralization:*

$$\frac{d(\tau + \tau_j)}{d\lambda} = \frac{d\tau}{d\lambda} < 0$$

The intuition behind this proposition follows directly from the equilibrium structure of taxation. District governments face tax competition for capital and hence they always set $\tau_j = 0$, irrespective of λ . In contrast, the median voter always votes for capital taxes to be used by the central government. However, as λ increases, the central government is responsible for less public goods. Since there are diminishing returns to each good, the same tax revenues spread across less goods provide less utility, which reduces incentives to redistribute. As a consequence, the median voter prefers to reduce τ when decentralization increases.

In most countries, the allocation of responsibilities between the different layers of government is either enshrined in the Constitution or requires major legislative efforts to change. This stands in contrast with taxation policy, which is typically decided every year during the budgetary process, and hence is subject to the commitment problem highlighted here. Institutional arrangements can not be so finely tailored as budgetary policy, but this lack of flexibility comes with an important benefit: the electorate, by voting on constitutional arrangements, is able to precommit to institutional features in a way it can not with respect to policy. Insofar as the allocation of public good provision between local and central governments is an institutional feature, Proposition 3 suggests that it provides a strong commitment device to a reduced level of capital taxation.

The question then becomes: if we allow for an *ex ante* vote on the Constitution, should we expect the median voter to support a decentralized Constitution?

By the argument above, the median voter should be willing to commit at least to *some* decentralization *ex ante*. However, complete decentralization will not typically be optimal from her perspective: every good that is transferred from the central government to the district governments cannot be used as a redistributive device. An increase in decentralization thus reduces the ability to redistribute and it is therefore costly to the median voter. However, by decentralizing a few goods, the median voter can commit to a limited level of capital taxation and can thereby increase capital stocks. These larger capital holdings might compensate for the lower level of capital taxes and generate more revenue to redistribute. Therefore, the median voter typically prefers an interior solution to the constitutional problem that entails a partial level of decentralization.²⁴

Note that this argument does not rely on the $1 - \lambda$ centralized goods being *directly*

²⁴It is also important to note that the partial decentralization solution is not perfect from the point of view of the median voter. It implies an inefficiency because some public goods will receive higher funding than others. Indeed, the preferred constitution by the median citizen would be one in which she could commit to an upper bound to capital taxes.

provided by the central government. What matters is which level of government has the ability to determine the quantity supplied of a given public good, and the responsibility to raise funds to meet this need. Specifically, suppose that the central government raises funds and then transfers this money to the districts to provide a centrally determined level $s(p)$ of public good p . In this case the actual providers would be the districts, but the incentives for capital taxation would remain high. Hence, in the context of the model, this good p would count as centrally provided. Note also that in this model, transfers of centrally raised money to be spend at the will of the district do *not* count as decentralization either, because tax competition does not impair the ability of raising such funds via capital taxes.²⁵ To help in the commitment problem, a decentralized good must be characterized by having both the quantity supply decision and the fund-raising done at the district level.

In the following section we assume specific functional forms which allow us to explicitly solve for the level of decentralization preferred by the median voter. In such an example we can explore comparative statics on how the preferred level of decentralization varies with factors such as productivity, inequality, or the redistributive nature of publicly provided goods.

5 The Equilibrium Level of Decentralization

With general functional forms, the study of the equilibrium level of decentralization is impaired by two problems. First, since $\Phi(r)$ is endogenous, multiple equilibria could exist, which obviously complicates comparative statics. Second, some concepts such as redistributive efficiency lack a clear parameter of reference. Hence, to examine the full constitutional game, we consider a particular case of the model developed above.

Assume that technology is linear in k , $F(k) = Ak$, where A captures the general level of productivity in this economy. Moreover, let $l\left(\frac{k^n}{\beta^n}\right) = \frac{k^n}{\beta^n}$. Finally, assume that $G(s(p)) = \frac{[s(p)]^\alpha - 1}{\alpha}$, for $\alpha < 1$. As before, we consider a distribution of types $H(\cdot)$ such that the expected value, $\bar{\beta}$ is greater than the median value, β^{med} . It is straightforward to see that these functional assumptions satisfy the conditions of the general model.

Note that α captures *redistributive efficiency*. If α is very close to 1, redistribution

²⁵These types of arrangements are prevalent in a number of countries in Latin America, including Argentina, Brazil, and Columbia. In Argentina, for instance, while more than 80% of revenues were generated at the federal level in the early 1990s, less than half of the expenditures were done at the federal level. Please see Ter-Minassian (1997) for further details of the institutional arrangements for these and other countries with regards to fiscal decentralization.

through public provision performs very similarly to the classic case of proportional taxes on income and lump-sum equal transfers back to all citizens. Conversely, if α is close to 0, public funds are not easy to transfer through public good provision as marginal utility for such goods diminishes very quickly. The justice system might be a good example of a public service with a very small α , while a public health system would have relatively high α .²⁶

We therefore consider a constitutional game with the following timing:

1. By simple majority a Constitution is chosen such that the degree of decentralization, λ^* , is determined.
2. Each agent n in each district j decides how much capital to raise, k^n .
3. By simple majority, taking the net rate of return to capital, r , and the district budget constraint as given, the citizens in each district choose the Condorcet winner in their policy space $(\tau_j; T_j; s_j(p), p \in [0, \lambda])$.
4. By simple majority, taking the budget constraint as given, all the citizenry chooses the Condorcet winner in the policy space of the central government $(\tau; T; s(p), p \in (\lambda, 1])$.
5. After observing taxation patterns across the economy, agents decide in which district to invest their productive capital, k^n .

We begin the analysis at the second stage, using the results in the previous section, before moving to the Constitutional vote.

The problem of agent n in district j simplifies to

$$\max_{k^n} \mathbb{E} \left\{ k(A - \tau - \tau_j) - T - T_j - (k^n - \beta^n) \frac{k^n}{\beta^n} \right\} \quad (12)$$

and yields a solution that is proportional to her endowment, β^n

$$k^n = \frac{\beta^n}{2} [\mathbb{E}(A - \tau - \tau_j) - 1] \quad (13)$$

As before, an agent with a richer endowment, expecting a net rate of return $r = \mathbb{E}(A - \tau - \tau_j)$, raises more capital because her collateral allows her to access loans at lower rates.²⁷

²⁶Note that for the case where waste and bureaucratic expenses increase more than proportionately in the funds to be disbursed, direct lump-sum transfers can also be captured by a high α .

²⁷For simplicity, we shall assume A large enough so that $k^n > \beta^n$

With linear technology there are no returns to the district. As a consequence, the problem of the district government is particularly simple

$$\max_{T_j, \tau_j} \left\{ -T_j + \lambda \frac{\left(\frac{T_j + \tau_j k_j}{\lambda} \right)^\alpha - 1}{\alpha} \right\}.$$

This objective function already assumes that public spending will be equally distributed across the λ goods. Note that with linear technology, the capital mobility constraint is particularly tight. In particular, district d only receives any capital investment at all if $\tau_d = \min \{ \tau_j \}$. In short, districts compete à la Bertrand for capital. Not surprisingly, such competition between districts yields $\tau_j = 0 \forall j \in J$. Hence, the voting equilibrium at the district level is $(\tau_j, T_j) = (0, \lambda)$.

The problem of the median voter for central government policies is slightly more involved. Integrating (13), and taking into account $\mathbb{E}(\tau_j) = 0$ in equilibrium, we obtain the average level of capital as a function of expected taxation. This yields $\bar{k} = \int \frac{\beta^n}{2} [A - \mathbb{E}(\tau) - 1] h(\beta) d\beta = \bar{\beta} \frac{A - \mathbb{E}(\tau) - 1}{2}$, where $\bar{\beta}$ is the expected value of the distribution $H(\cdot)$ of abilities to generate capital. With this, we can write

$$\max_{T, \tau} k^{\text{med}} (A - \tau) - T + (1 - \lambda) \frac{\left(\frac{T + \tau \bar{k}}{1 - \lambda} \right)^\alpha - 1}{\alpha}$$

subject to $T \geq 0$. Again, since \bar{k} and k^{med} are predetermined when this vote takes place (in stage 4), the median voter sets $T = 0$ whenever $\frac{\bar{k}}{k^{\text{med}}} > 1$. An advantage of the linear-quadratic formulation is that *ex post* inequality equals *ex ante* inequality and is independent of r . We thus have

$$\frac{\bar{k}}{k^{\text{med}}} = \frac{\bar{\beta}}{\beta^{\text{med}}} \equiv \Phi > 1$$

thanks to our initial assumptions on $H(\cdot)$. Hence, no head taxes are used. Taking the first order condition with respect to τ yields

$$\tau = \Phi^{\gamma+1} \frac{1 - \lambda}{\bar{k}} \tag{14}$$

where $\gamma = \frac{\alpha}{1 - \alpha}$. It is clear from this condition that the total revenues collected with capital taxes, $\tau \bar{k}$, equal $(1 - \lambda) \Phi^{\gamma+1}$. Therefore, total capital tax revenue is increasing in inequality Φ and in redistributive efficiency α . As in the general model, this is natural: inequality refers to the median voter's desire for redistribution, and

redistributive efficiency speaks to the feasibility of redistribution. Note also that capital tax revenue is decreasing in λ , the degree of decentralization, as was discussed in the previous section. This formulation thus confirms the intuitions built with the general model.

The equilibrium level of capital taxes and capital generation can be found by solving the non-linear system of equations (13) and (14). By doing so, one obtains a well-defined capital generation function $k^n(\beta^n, \alpha, \Phi, \lambda)$ for the standard equilibrium of this model.

Proposition 4 *The linear-quadratic model admits a unique standard equilibrium. This equilibrium features the following comparative statics for all agents n :*

1. $\frac{\partial k^n}{\partial \lambda} > 0$
2. $\frac{\partial k^n}{\partial \alpha} < 0$
3. $\frac{\partial k^n}{\partial \Phi} < 0$
4. $\frac{\partial k^n}{\partial A} > 0$

Since (13) is proportional to β^n and τ is common for all voters, the equilibrium capital level of any agent n is proportional to \bar{k} . This is the reason why the comparative statics in Proposition 4 are common for all voters. These are also true for the aggregate level of capital held in the economy. Note again that an increase in decentralization λ reduces the level of capital taxation expected by agents, thereby generating a bigger stock of productive capital. Holding λ constant, however, we find other natural comparative statics.

As redistributive efficiency α increases, aggregate capital contracts. As discussed at length above, as α increases, the median voter prefers an increased supply of public goods to redistribute in her favor. As a consequence, she is tempted to increase capital taxation. These expectations depress capital generation.

Similar intuition lies behind the result that capital generation is decreasing in inequality, Φ . Again, as inequality increases, the temptation to expand public spending in order to redistribute becomes stronger because the tax burden is more unevenly allocated. In such circumstances, expected capital taxation increases which reduces incentives to hold capital.

Finally, an increase in productivity increases returns to capital. Since in this linear model *ex post* inequality Φ is constant, these excess returns are not taxed away and therefore capital reacts positively.

5.1 The Constitutional Problem

We now examine the initial stage in which voters decide on λ^* , the level of decentralization that they want enshrined in the Constitution. Since the Constitution is decided by a majority vote, we first consider the problem of the median voter before showing that her preferred position is again the Condorcet winner in the constitutional stage.

Taking as given the taxation and investment decisions that will follow, her problem can be written as

$$\max_{\lambda} k^{\text{med}}(A - \tau) - (k^{\text{med}} - \beta^{\text{med}}) \frac{k^{\text{med}}}{\beta^{\text{med}}} - T_j + (1 - \lambda) \frac{\left(\frac{\tau \bar{k}}{1 - \lambda}\right)^{\alpha} - 1}{\alpha} + \lambda \frac{\left(\frac{T_j}{\lambda}\right)^{\alpha} - 1}{\alpha}$$

The first two terms correspond to her private returns to capital. The third term is her expected head tax and the last two terms correspond to her enjoyment of publicly provided goods. By using the results in the previous subsection, we can further simplify this expression to

$$\max_{\lambda} k^{\text{med}}(A - \tau) - (k^{\text{med}} - \beta^{\text{med}}) \frac{k^{\text{med}}}{\beta^{\text{med}}} - \lambda + (1 - \lambda) \frac{\Phi^{\gamma} - 1}{\alpha} \quad (15)$$

This program is well behaved, and it always yields a unique solution. In particular, because k^{med} is a concave function of λ , we obtain:

Lemma 2 *Program (15) is globally concave.*

Hence, we can proceed to examine the first order condition of this problem:

$$-k^{\text{med}} \frac{d\tau(\lambda, \bar{k})}{d\lambda} - 1 - \frac{\Phi^{\gamma} - 1}{\alpha} = 0$$

The last two terms encapsulate the costs that the median voter suffers when decentralization increases: first, her head taxes increase as decentralized goods are funded with such taxes. This has a constant marginal cost of 1. Second, the level of provision of decentralized goods is lower than for centralized goods due to the lost incentive to redistribute. This effect is captured by the last term and is larger if inequality Φ and transferability α are high.

The gains that the median voter obtains from an increase in decentralization are in the first term.²⁸ Clearly, these gains come from the fact that τ decreases as λ

²⁸Due to the envelope condition, the benefits that the median voter perceives from increased decentralization are not related to her own adjustment in k^{med} .

increases. Note that using (14), we can write $\tau(\lambda, \bar{k})$. It follows that this adjustment of τ has two components:

$$\frac{d\tau(\lambda, \bar{k})}{d\lambda} = \frac{\partial\tau}{\partial\lambda} + \frac{\partial\tau}{\partial\bar{k}} \frac{\partial\bar{k}}{\partial\lambda} < 0$$

The first component is the *direct effect*: keeping the capital stock constant, as decentralization increases capital taxes mechanically decrease as they are to be spent on less goods. This is the effect we have emphasized in Section 4. However, there is a second force that corresponds to aggregate capital adjustment: as there is more capital in the economy, lower capital taxes can raise larger amounts of revenue. We call this second channel the *indirect effect* of decentralization on capital taxes.

Using (14) to derive the direct and indirect effects we obtain

$$\frac{d\tau(\lambda, \bar{k})}{d\lambda} = -\frac{\Phi^\gamma}{k^{\text{med}}} - \Phi^{\gamma+1} \frac{1 - \lambda \frac{\partial\bar{k}}{\partial\lambda}}{(\bar{k})^2}.$$

Hence the direct and indirect gains of the change in τ for the median voter are both increasing in Φ and α . This is intuitive as inequality and redistributive efficiency increase τ in equilibrium. It is then natural that the reduction in τ caused by an increase in decentralization is bigger when Φ and α are high. Using this expression, the first order condition of program 15 can be simplified to:

$$\Phi^\gamma + \Phi^\gamma \frac{1 - \lambda \frac{\partial\bar{k}}{\partial\lambda}}{\bar{k}} = 1 + \frac{\Phi^\gamma - 1}{\alpha} \quad (16)$$

It follows that the median voter's incentives to decentralize do not have an obvious relationship with Φ and α as both benefits and costs are increasing in Φ and α .

On the one hand, the marginal cost of decentralizing, in the right hand side of (16), is increasing in both α and Φ for most values. This is again because an increase in either parameter increases the returns to redistribution. And obviously, if the median voter wants to increase redistribution, she should favor *less* decentralization at the constitutional stage.

On the other hand, these exacerbated incentives to redistribute worsen the commitment problem and further contract aggregate capital. As a consequence, the *ex ante* marginal gains to decentralization, in the left hand side of (16), also increase in α and Φ . According to this commitment problem, the median voter would gain more from an *increase* in decentralization when α and Φ are high. The following proposition characterizes the optimal degree of decentralization and explores the comparative

statics that result from these conflicting incentives.

Proposition 5 *Program (15) defines a unique optimal level of decentralization λ^* which can be expressed in closed form as*

$$\lambda^* = 1 - \frac{A^2 \beta^{med} (\Phi^\gamma - 1)}{\Phi^\gamma} \frac{\Phi^\gamma - 1 + \gamma \Phi^\gamma}{(2(\Phi^\gamma - 1) + \gamma \Phi^\gamma)^2} \quad (17)$$

λ^* features the following comparative statics

1. $\frac{\partial \lambda^*}{\partial A} \leq 0$
2. $\frac{\partial \lambda^*}{\partial \alpha} \geq 0$
3. For each $\alpha \in (0, 1)$, exists a $\tilde{\Phi}(\alpha)$ such that $\frac{\partial \lambda^*}{\partial \Phi} < 0$ for $\Phi < \tilde{\Phi}(\alpha)$ and $\frac{\partial \lambda^*}{\partial \Phi} > 0$ for $\Phi > \tilde{\Phi}(\alpha)$.

Given the complex forces that the median voter faces, it is quite striking that we obtain some unambiguous comparative statics. The intuition for the first result is, however, clear. An increase in productivity A increases the incentives to generate capital while keeping the incentives to tax it constant. In that case, the median agent can afford to reduce the level of decentralization: this allows the median voter to redistribute some of the returns to this additional capital accumulated.

Despite the opposite incentives that the median voter faces with respect to redistributive efficiency α , the commitment problem dominates. An increase in α implies high *ex post* incentives to raise τ and the median voter prefers to increase decentralization to avoid the capital contraction that these expectations generate.

This is not always true, however, for an increase in inequality. Note, in particular, that when the median voter expects to have the same amount of capital as the median voter, i.e. $\Phi = 1$, she prefers full decentralization, i.e. $\lambda^* = 1$, to ensure that only head taxes are used. From this point, if inequality marginally increases, the median voter wants to centralize a few goods: since inequality is still small, expected capital taxes are small and hence capital accumulation distortions are not large enough to make the median voter relinquish this opportunity for redistribution. For higher levels of inequality, however, these distortions increase and eventually the *ex post* temptation to redistribute is too costly. At that point, the median voter prefers to gradually decentralize to avoid such costs. Figure 2 shows the evolution of λ^* as inequality increases, for a given level of transferability $\gamma = 1$.

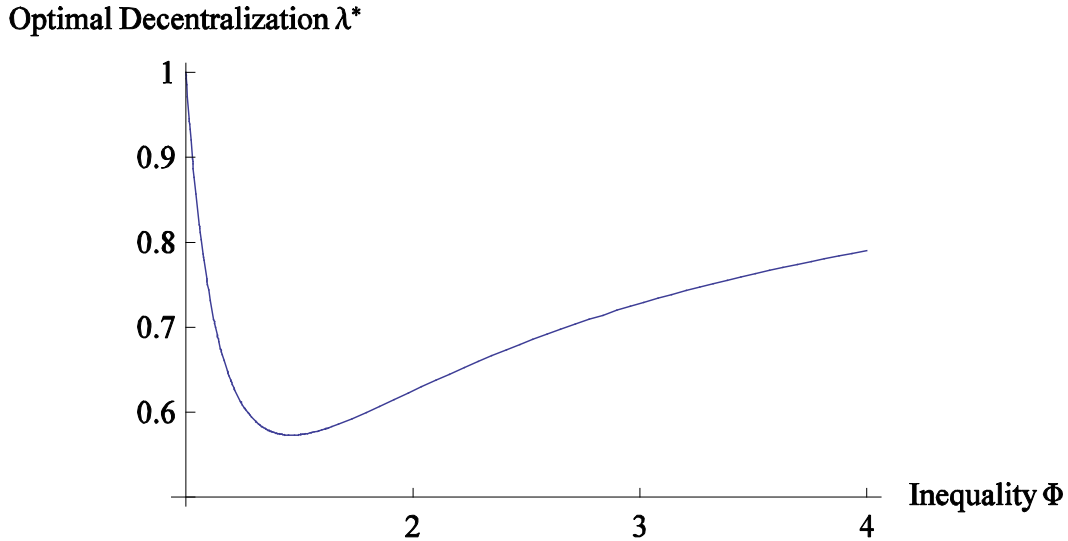


Figure 2: Graph of the optimal level of decentralization as a function of inequality. Note that for no inequality, full decentralization is optimal, while with some inequality, the median voter prefers less than full decentralization. However, as inequality grows, the median voter prefers more decentralization as a means of precommitment.

To see how the median voter’s optimal level of decentralization evolves with inequality and redistributive efficiency, Figure 3 shows the three-dimensional plot. As can be readily seen, for any given level of inequality, the optimal degree of decentralization is (weakly) increasing in the redistributive efficiency of the public good, which is parameterized by γ . Furthermore, when inequality is zero, the median voter always wishes to fully decentralize, so as to ensure the use of nondistortive head taxes. However, as redistributive efficiency increases, the need to commit becomes more important to the median voter; hence, for higher levels of γ , optimal decentralization rises sooner with respect to the level of inequality.

5.2 Preferences for Decentralization

In the previous subsection we have shown that, in general, the very median agent that decides on taxation patterns *ex post* prefers to tie her hands *ex ante* by voting for an interior level of decentralization. We still need to determine, however, that the preferences of the median voter are the Condorcet winner at the Constitutional stage. For this, we need to determine the preferences for decentralization for the rest of the citizens. This turns out to be a simple problem. In particular, we can write the constitutional program that the median agent solves (15), for a generic agent with

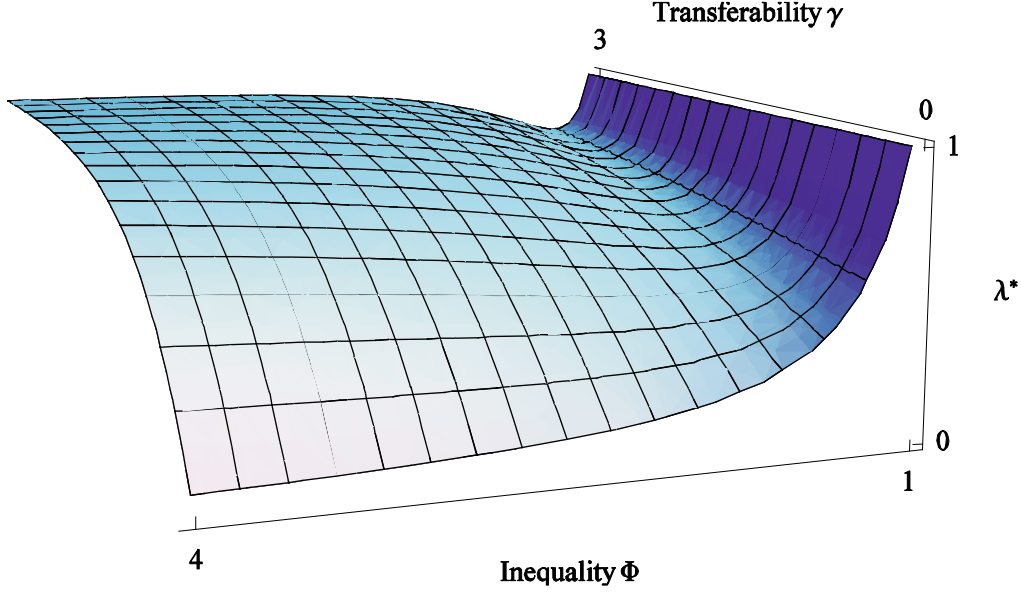


Figure 3: Graph of the optimal level of decentralization as a function of inequality and transferability. Note that for public goods with a quickly diminishing marginal utility (i.e. low γ), the optimal level of decentralization is decreasing in inequality over a larger range.

endowment β . Denote by $\lambda(\beta)$ the solution to the following program:

$$\max_{\lambda} k(\beta) [A - \tau(\lambda)] - (k(\beta) - \beta) \frac{k(\beta)}{\beta} - \lambda + (1 - \lambda) \frac{\Phi^{\gamma} - 1}{\alpha} \quad (18)$$

The last two terms do not depend on β due to the fact that taxation and redistribution decisions will *ex post* be decided by the median agent, conditional on constitutional arrangements. Hence the tension over λ only depends on $k(\beta)$. It is intuitive, then, that the higher the ability to raise capital, the lower the level of capital taxation preferred and therefore the higher the degree of decentralization favored.

Proposition 6 *Program (18) admits a unique solution, $\lambda(\beta)$. This solution is such that*

$$\frac{\partial \lambda(\beta)}{\partial \beta} > 0$$

The unique solution follows from the fact that (18) is concave in λ . Therefore voter's preferences are single-peaked. The proposition states that their ideal point is increasing in β and hence it follows that at the Constitutional stage, the degree of decentralization favored by the median agent is the Condorcet winner. Moreover, this proposition also implies that there should be a relationship between capital ownership,

preferences for capital taxation and public spending, and preferences for decentralization. In particular, political parties that represent capital owners should put forward platforms that favor low taxes on capital, a low level of public expenditure, and a higher degree of decentralization. In our model, decentralization becomes the way of obtaining the first two items on this agenda. The position in these dimensions of parties such as the Republican Party in the United States are therefore consistent with a political economy view of the degree of decentralization.

6 Heterogenous Public Goods

In this section we explore whether the commitment problem prescribes a distribution of public goods between central and local administrations that is orthogonal to spillovers or differences in taste. We use the same utility function over public goods in the previous section, $G(s(p)) = \frac{[s(p)]^{\alpha}-1}{\alpha}$. However, we now assume that goods are heterogeneous. In particular, there are many types of public goods, and each is characterized by $\alpha_h \in (0, 1]$, $h \in \mathbb{H} = \{1, 2, \dots, H\}$ with $\alpha_h < \alpha_{h'}$ if $h < h'$. Furthermore, for each type of good, there is a continuum of size 1 of these goods.

A *decentralization scheme* is now a set of $\{\lambda_h\}_{h \in \mathbb{H}}$, such that $0 \leq \lambda_h \leq 1$. λ_h is the degree of decentralization of goods of type h : local governments will provide λ_h of this type of public good, while the central government will provide $1 - \lambda_h$. If g_j^h is the amount spent per good on goods of type h in district j , and g^h is the amount on goods of type h by the central government, then the utility of agent n in district j is given by

$$u^n(c_j^n, k_j^n, \vec{g}_j, \vec{g}) = c_j^n + \sum_{h=1}^H \lambda_h \frac{\left(\frac{g_j^h}{\lambda_h}\right)^{\alpha_h} - 1}{\alpha_h} + \sum_{h=1}^H (1 - \lambda_h) \frac{\left(\frac{g^h}{\lambda_h}\right)^{\alpha_h} - 1}{\alpha_h} - (k^n - \beta^n) l\left(\frac{k^n}{\beta^n}\right)$$

Since governments must balance their budget, total tax revenues must equal total expenditures. Hence

$$\sum_{h=1}^H g_j^h = \tau_j k_j + T_j$$

with an equivalent definition for the central government.

Given this inherently multidimensional set up, we obtain a striking result: all voters agree on the optimal structure of the Constitution.

Proposition 7 *All Pareto optimal decentralization schemes are characterized by an*

α^* , such that all public goods with $\alpha_h < \alpha^*$ are completely centralized and all public goods with $\alpha_h > \alpha^*$ are completely decentralized.

In other words, all voters agree that the best structure of decentralization is one in which goods with high redistributive efficiency are decentralized and goods with low redistributive efficiency are centralized. The intuition for the result is as follows. Consider a decentralization scheme such that a good with a high α' is centralized but one with a low α is decentralized. Now consider decentralizing ε of the first good, and centralizing δ of the second, such that the equilibrium capital tax remains the same. Since $\alpha' > \alpha$, each centralized α' good pushes τ up by more than each centralized α good. This implies that to keep τ fixed, it must be true that $\varepsilon < \delta$. Since this change implies that more goods are centralized, citizens save in head taxes and only lose at the margin in public good provision. An invariant τ also means that the amount of *ex post* redistribution and *ex ante* capital generation remains the same and hence this is a Pareto improving change.

As shown above in the case of a single α , high redistributive efficiency is dangerous because it distorts taxation decisions *ex post*. As a consequence, it is the goods with the most ‘redistributive power’ which are the most important to decentralize as they exacerbate the temptation to expand public provision *ex post* as redistribution. Furthermore, all voters agree on this because all of them want to minimize distortions at the capital generation margin. This is counter the usual arguments that the goods that need to be centralized are those with a high redistributive content. Here, the greater the redistributive power, the bigger the problem of lack of commitment.

Hence, during the constitutional stage of the game, we can see that the conflict between rich and poor should not be on which particular goods to be decentralized but rather over the extent of decentralization. In particular, the richer the agent, the more she benefits from increased decentralization and hence she prefers more goods to be decentralized (lower α^*). Letting $\alpha^*(\beta)$ be the minimal $\alpha \geq 0$ that corresponds to an optimal decentralization scheme for an agent of type β , we have that:

Proposition 8 $\alpha^*(\beta)$ is a weakly decreasing function, and $\alpha^*(\beta^{\min}) < 1$.

Note that even the poorest citizen, one with no capital, does not want full centralization. In particular, all agents agree that goods for which α is close to 1 should be fully decentralized. Otherwise, this good will be used *ex post* to fully redistribute wealth across society. While this may be good for the poorest agent *ex post*, it means that *ex ante* no capital investment takes place and hence such redistributive capacity is useless.

7 Conclusion

In this paper, we have proposed a theory of federalism that does not rely on assumptions about spillovers and taste heterogeneity. While these issues are important, many public goods, such as fire protection, sewers, etc. lack both significant spillover effects and substantial taste heterogeneity. To build this theory we have focused on tax competition within a state. In this view, the main distinction between a centralized and decentralized state is the existence of constraints on policymaking due to the effects of competitive pressures. The competitive effects of federalism are many and varied, and have spawned a literature both decrying the effects and touting the virtues of such competition.²⁹

We contribute to that literature by considering the classic dynamic inconsistency in capital taxation problem. We have shown that federalism provides a tool for a nation to precommit to certain taxation policies that it would not choose to implement *ex post*. We also show that this commitment comes at a cost for the median voter and hence this theory predicts a partial degree of decentralization even when all public goods to be provided are homogeneous.

To illustrate our argument, we have focused on the redistributive properties of public good provision. However, it has long been noticed in the literature that benevolent governments who simply maximize static welfare have very similar precommitment problems.³⁰ It is easy to see that in this case our argument would have similar predictions: by decentralizing the provision of some public goods, the nation can effectively precommit to fund those goods using instruments other than capital taxes. While *ex post* this may require the use of inefficient tax instruments, *ex ante* it provides assurances to those who would choose to invest in capital generation. Hence, in a world without commitment, federalism is a second-best solution to the problem of choosing tax policy and public investment.

We choose to focus on political economy issues because this allows us to link the expected degree of decentralization to the level of inequality, the redistributive properties of public goods and the level of productivity in the economy. Moreover, the virtues of partial federalism in solving these kind of redistributive commitment problems have not, to our knowledge, been studied in detail before.

In other respects, our model is obviously restrictive. We do not, for instance,

²⁹For a summary of the competitive effects of federalism, see McKinnon and Nechyba (1997).

³⁰If the head tax in our model was slightly distortionary, a welfare-maximizing government would choose to use only capital taxes, even if agents were completely homogenous. For arguments built on similar structures (capital versus labor taxes), see Fischer (1980).

allow for citizen mobility which would only add to the competitive pressures that the districts face. Also, further research should explore the interaction of capital taxes and distortive labor taxes in a model with redistribution but without access to lump-sum taxes. In that case our argument would still go through, but public provision at the local level would be distorted downwards and hence welfare analysis would be more nuanced. Finally, note that in an age of open capital markets, the central government might also face capital flight fears. The argument would then hinge on easier capital mobility within a country than across borders. This is probably true, but an explicit consideration would allow for informative comparative statics with respect to global capital market integration.

References

- [1] Acemoglu, Daron, and Robinson, James. 2001. "A Theory of Political Transitions." *American Economic Review*, 91: 938-963.
- [2] Acemoglu, Daron, and Robinson, James. 2006. *Economic Origins of Dictatorship and Democracy*. Cambridge and New York: Cambridge University Press.
- [3] Alesina, Alberto, and Rodrik, Dani. 1994. "Distributive Politics and Economic Growth." *Quarterly Journal of Economics*, 109: 465-490.
- [4] Alesina, Alberto, and Spolaore, Enrico. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics*, 112:1027-1056.
- [5] Banerjee, Abhijit V. 2003. "Contracting Constraints, Credit Markets, and Economic Development." In L. Hansen, M. Dewatripont and S. Turnovsky, eds. *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*. Cambridge and New York: Cambridge University Press.
- [6] Barro, Robert J. and Gordon, David B. 1983. "A Positive Theory of Monetary Policy in a Natural Rate Model." *Journal of Political Economy*, 91:589-610.
- [7] Besley, Timothy, and Coate, Stephen. 1991. "Public Provision of Private Goods and the Redistribution of Income." *The American Economic Review*, 81: 979-984.
- [8] Besley, Timothy, and Coate, Stephen. 2003. "Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Approach." *Journal of Public Economics*, 81: 2611-2637.

- [9] Besley, Timothy J., and Persson, Torsten. 2007. "The Origins of State Capacity: Property Rights, Taxation, and Politics ." Forthcoming in *American Economic Review*.
- [10] Besley, Timothy, and Smart, Michael. 2007. "Fiscal Restraints and Voter Welfare." *Journal of Public Economics*, 91:755-773.
- [11] Bodenstein, Martin, and Ursprung, Heinrich W. 2005. "Political Yardstick Competition, Economic Integration, and Constitutional Choice in a Federation: A Numerical Analysis of a Contest Success Function Model." *Public Choice*, 124: 329-352.
- [12] Bolton, Patrick and Gérard Roland. 1996. "Distributional Conflicts, Factor Mobility, and Political Integration." *American Economic Review*, 86: 99-104.
- [13] Buchanan, James M. 1964. "Fiscal Institutions and Efficiency in Collective Outlay." *The American Economic Review*, 54: 227-235.
- [14] Crémer, Jacques, and Palfrey, Thomas R. 1996. "In or Out? Centralization by Majority Vote." *European Economic Review*, 40: 43-60.
- [15] Crémer, Jacques, and Palfrey, Thomas R. 1999. "Political Confederation." *American Political Science Review*, 93: 69-83.
- [16] Crémer, Jacques, and Palfrey, Thomas R. 2000. "Federal Mandates by Popular Demand." *Journal of Political Economy*, 108: 905-927.
- [17] de Figueiredo, Rui, and Weingast, Barry. 2005. "Self-Enforcing Federalism." *Journal of Law, Economics, and Organization*, 21: 103-35.
- [18] Dixit, Avinash, and Londregan, John. 1998. "Fiscal Federalism and Redistributive Politics." *Journal of Public Economics*, 68: 153-180.
- [19] Edwards, Jeremy, and Keen, Michael. 1996. "Tax competition and Leviathan." *European Economic Review*, 40: 113-134.
- [20] Fischer, S. 1980. "Dynamic Inconsistency, Cooperation and the Benevolent Disassembling Government." *Journal of Economic Dynamics and Control*, 2:93-107.
- [21] Hatfield, John William. 2006. "Federalism, Taxation, and Economic Growth." Stanford GSB Research Paper #1929.

- [22] Inman, Robert P., and Rubinfeld, Daniel L. 1997. "Rethinking Federalism" *Journal of Economic Perspectives*, 11: 43-64
- [23] Ghatak, Maitreesh, Morelli, Massimo and Sjoström, Thomas. 2007. "Entrepreneurial Talent, Occupational Choice and Trickle up Policies." *Journal of Economic Theory*, 137: 27-48.
- [24] Kehoe, Patrick J. 1989. "Policy Cooperation Among Benevolent Governments May Be Undesirable." *The Review of Economic Studies*, 56: 289-296.
- [25] Klein, Paul, Krusell, Per, and Ríos-Rull, José Víctor. 2004. "Time-Consistent Public Expenditures." CAERP working paper, 15.
- [26] Kydland, Finn E., and Prescott, Edward C. 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *The Journal of Political Economy*, 85: 473-492.
- [27] Lockwood, Ben. 2002. "Distributive Politics and the Benefits of Decentralization." *Review of Economic Studies*, 69: 313-37.
- [28] McKinnon, Ronald, and Nechyba, Thomas. 1997. "Competition in Federal Systems." in *The New Federalism: Can the States be Trusted?*, ed. John Ferejohn and Barry R. Weingast. Stanford: Hoover Institution Press.
- [29] Meltzer, Allan H., and Richard, Scott F. 1981. "A Rational Theory of the Size of Government." *The Journal of Political Economy*, 89: 914-927.
- [30] Oates, Wallace E. 1972. *Fiscal Federalism*. Harcourt Brace Jovanovich: New York.
- [31] Oates, Wallace E. 1999. "An Essay on Fiscal Federalism." *Journal of Economic Literature*, 37:1120-1149.
- [32] Oates, Wallace E. 2005. "Toward A Second-Generation Theory of Fiscal Federalism." *International Tax and Public Finance*, 12: 349-373.
- [33] Persson, Torsten, and Tabellini, Guido. 1994. "Does Centralization Increase the Size of Government?" *European Economic Review*, 38: 765-773.
- [34] Persson, Torsten, and Tabellini, Guido. 1996. "Federal Fiscal Constitutions: Risk Sharing and Redistribution." *Journal of Political Economy*, 104: 979-1009.
- [35] Persson, Torsten, and Tabellini, Guido. 2000. *Political Economics: Explaining Economic Policy*. The MIT Press: Cambridge.

- [36] Roberts, Kevin W. S. 1977. "Voting over an Income Tax Schedule." *Journal of Public Economics*, 8: 329-340.
- [37] Rogoff, Kenneth. 1985. "Can International Monetary Policy Cooperation Be Counterproductive?" *Journal of International Economics*, 18: 199-217.
- [38] Romer, Thomas. 1975. "Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax." *Journal of Public Economics*, 4:163-185.
- [39] Seabright, Paul. 1996. "Accountability and Decentralisation in Government: An Incomplete Contracts Model." *European Economic Review*, 40:61-89.
- [40] Stiglitz, Joseph and Weiss, Andrew. 1981. "Credit Rationing in Markets with Imperfect Information." *The American Economic Review*, 71: 393-410.
- [41] Tabellini, Guido. 1990. "Domestic Politics and the International Coordination of Fiscal Policies." *Journal of International Economics*, 28: 245-65.
- [42] Ter-Minassian, Theresa. 1997. *Fiscal Federalism in Theory and Practice*. International Monetary Fund: Washington.
- [43] Tirole, Jean. 2006. *The Theory of Corporate Finance*. Princeton and Oxford: Princeton University Press.
- [44] Tommasi, Mariano, and Weinschelbaum, Federico. 2007. "Centralization vs. Decentralization: A Principal-Agent Analysis." *Journal of Public Economic Theory*, 9:369-389.
- [45] Weingast, Barry 2006. "Second Generation Fiscal Federalism: Implications for Decentralized Democratic Governance and Economic Development." Working Paper, Hoover Institution, Stanford University.
- [46] Wildasin, David E. 1988. "Nash Equilibria in Models of Fiscal Competition." *Journal of Public Economics*, 19: 296-315.
- [47] Wilson, John D. 1986. "A Theory of Interregional Tax Competition." *Journal of Urban Economics*, 19: 356-370.
- [48] Wilson, John D. 1999. "Theories of Tax Competition." *National Tax Journal*, 52: 269-304.

- [49] Zodrow, George, and Peter Mieszkowski. 1986. “Pigou, Property Taxation, and the Underprovision of Local Public Goods.” *Journal of Urban Economics*, 19: 356-370.

8 Appendix

Proof of Proposition 1. Consider the problem of the voter with the median amount of capital who sets central government policy.

$$\max_{T, \tau, s(p)} \left\{ -\tau k^{\text{med}} - T + \int_{\lambda}^1 G(s(p)) dp \right\}$$

subject to the budget constraint

$$\int_{\lambda}^1 s(p) dp = \tau k + T$$

and the constraint on land taxes (with Lagrange multiplier ν)

$$T \geq 0$$

Since it is immediate that spending will be equal across all national public goods, we have that the maximization problem is

$$\max_{T, \tau} \left\{ -\tau k^{\text{med}} - T + (1 - \lambda) G\left(\frac{\tau k + T}{1 - \lambda}\right) \right\}$$

and taking the first-order conditions we have

$$\begin{aligned} -1 + G'(\cdot) &= \nu \\ G'(\cdot) &= \frac{k^{\text{med}}}{k} = \Phi^{-1} \end{aligned}$$

Hence, since $\Phi^{-1} < 1$, $G'(\cdot) < 1$ and the constraint on land taxes binds. So we have that

$$\begin{aligned} T &= 0 \\ G'\left(\frac{\tau k}{1 - \lambda}\right) &= \Phi^{-1} < 1 \end{aligned}$$

That this is a Condorcet winner in the policy space is shown in the text. ■

Proof of Proposition 2. The problem for each agent within the district, if she were allowed to choose policy is

$$\max_{k_j, T_j, \tau_j} \left\{ F(k_j) - \rho_j k_j - T_j + \lambda G\left(\frac{\tau_j k_j + T_j}{\lambda}\right) \right\}$$

by substituting in the budget constraint. Subject to the constraint from capital mobility (with Lagrange multiplier μ)

$$r = F'(k_j) - \tau_j - \tau$$

Taking the first order conditions we find (noting that $\rho_j = F'(k_j)$)

$$\begin{aligned} -F''(k_j) k_j + \tau_j G'(\cdot) + \mu F''(k_j) &= 0 \\ G'(\cdot) &= 1 \\ G'(\cdot) k_j - \mu &= 0 \end{aligned}$$

(Note that the second order conditions are satisfied, so we are at a maximum.) Thus we can calculate the taxes using the constraints as

$$\begin{aligned} \tau_j &= 0 \\ G'\left(\frac{T_j}{\lambda}\right) &= 1 \end{aligned}$$

■

Proof of Proposition 3. Any standard equilibrium is characterized by two equations. The first is (6), given by the maximization problem of the median voter. The second is the definition of *ex post* inequality

$$\Phi(\tau) = \frac{k(\tau)}{k^{\text{med}}(\tau)}$$

That is, given the expected tax rate τ , the agents' optimal investment decisions must produce a level of inequality that supports that tax rate. The left hand side in (6) is strictly decreasing in τ because $G(\cdot)$ is concave and $\frac{\partial(\tau k)}{\partial \tau} > 0$ in the range where standard equilibria exist. $\Phi(\tau)$, however, might be increasing or decreasing. Differentiating both sides of (6) with respect to τ , it is easy to see that as long as

$$\frac{\partial \Phi}{\partial \tau} < -\frac{\frac{\partial(\tau k)}{\partial \tau} \Phi^2 G''\left(\frac{\tau k}{1-\lambda}\right)}{1-\lambda} \quad \forall \tau \tag{19}$$

the two functions of τ can only cross once in the relevant range. This proves uniqueness. Note that since the right hand side is positive, any $\frac{\partial \Phi}{\partial \tau} < 0$ satisfies this condition.

Implicitly differentiating the expression for capital taxes (6) with respect to λ , we have

$$\begin{aligned} \frac{d}{d\lambda} \left[G' \left(\frac{\tau k}{1-\lambda} \right) = \Phi^{-1} \right] \\ G'' \left(\frac{\tau k}{1-\lambda} \right) \frac{\frac{\partial(\tau k)}{\partial \tau} \frac{\partial \tau}{\partial \lambda} (1-\lambda) + \tau k}{(1-\lambda)^2} &= \frac{-1}{\Phi^2} \frac{\partial \Phi}{\partial \tau} \frac{\partial \tau}{\partial \lambda} \\ \frac{\frac{\partial(\tau k)}{\partial \tau} \frac{\partial \tau}{\partial \lambda} (1-\lambda) + \tau k}{(1-\lambda)^2} &= \frac{-1}{\Phi^2 G'' \left(\frac{\tau k}{1-\lambda} \right)} \frac{\partial \Phi}{\partial \tau} \frac{\partial \tau}{\partial \lambda} \\ \frac{\partial \tau}{\partial \lambda} &= - \frac{\tau k}{\frac{\partial(\tau k)}{\partial \tau} (1-\lambda) + \frac{(1-\lambda)^2}{\Phi^2 G'' \left(\frac{\tau k}{1-\lambda} \right)} \frac{\partial \Phi}{\partial \tau}} < 0 \end{aligned}$$

so long as

$$\frac{\partial \Phi}{\partial \tau} < - \frac{\frac{\partial(\tau k)}{\partial \tau} \Phi^2 G'' \left(\frac{\tau k}{1-\lambda} \right)}{1-\lambda} \quad (20)$$

and $\frac{\partial(\tau k)}{\partial \tau} > 0$, i.e. we are in a standard equilibrium. Since (20) is obviously implied by (19), we obtain that (19) implies capital taxes decreasing in λ . ■

Proof of Proposition 4. (13) holds for any β^n , so take the condition for β^{med} . Also, rewrite (14) using $\bar{k} = \Phi k^{\text{med}}$. From these two conditions, obtain a second degree equation on k^{med} . The largest solution to this equation is

$$k^{\text{med}} = \frac{1}{2} \left[A \beta^{\text{med}} + \sqrt{A^2 (\beta^{\text{med}})^2 - 4 \beta^{\text{med}} \Phi^\gamma (1-\lambda)} \right] \quad (21)$$

This is the only solution consistent with a standard equilibrium. The comparative statics for k^{med} are immediate from this expression. Since $\frac{k^n}{k^{\text{med}}} = \frac{\beta^n}{\beta^{\text{med}}}$, k^n is proportional to k^{med} and hence the comparative statics for k^{med} are common to all k^n . ■

Proof of Lemma 2. Substituting (14) in (15) obtain:

$$\max_{\lambda} k^{\text{med}} A - \frac{(k^{\text{med}})^2}{2\beta^{\text{med}}} - (1-\lambda) \Phi^\gamma - \lambda + (1-\lambda) \frac{\Phi^\gamma - 1}{\alpha}$$

The first two terms are a concave function of k^{med} . From (21) it is clear that $\frac{\partial^2 k^{\text{med}}}{\partial \lambda^2} < 0$. Hence we have a concave function of a concave function of λ followed by a linear function of λ . The sum of two weakly concave functions is weakly concave and hence

it follows that the second order condition of (15) holds. ■

Proof of Proposition 5. First substitute $\bar{k} = \Phi k^{\text{med}}$ in both (14) and (13). From these, obtain a second degree equation on τ . The smallest solution to this equation is

$$\tau = \frac{1}{2\beta^{\text{med}}} \left[\beta^{\text{med}} A - \sqrt{(\beta^{\text{med}} A)^2 - 4\beta^{\text{med}} \Phi^\gamma (1 - \lambda)} \right] \quad (22)$$

which is the solution consistent with a standard equilibrium. The first order condition (16) can be written as:

$$\tau \frac{\partial k^{\text{med}}}{\partial \lambda} = \frac{\Phi^\gamma - 1}{\gamma}$$

Substituting in (22) and $\frac{\partial k^{\text{med}}}{\partial \lambda}$ obtained from (21) we obtain a linear equation in λ that after much algebra can be reduced to

$$\lambda^* = 1 - \frac{A^2 \beta^{\text{med}} (\Phi^\gamma - 1)}{\Phi^\gamma} \frac{\Phi^\gamma - 1 + \gamma \Phi^\gamma}{(2(\Phi^\gamma - 1) + \gamma \Phi^\gamma)^2}$$

Since $\Phi^\gamma \geq 1$ in the relevant range, λ^* is weakly decreasing in A . Furthermore, we obtain

$$\frac{d\lambda^*}{d\gamma} = \frac{A^2}{\Phi^\gamma} \frac{\beta^{\text{med}}}{\gamma (\Phi^\gamma \gamma + 2\Phi^\gamma - 2)^3} F(\Phi^\gamma, \gamma)$$

where $F(\Phi^\gamma, \gamma)$ is a function of Φ^γ and γ . In particular, $F(\Phi^\gamma, \gamma)$ can be expressed as a quadratic in γ

$$\begin{aligned} & \ln \Phi^\gamma (-2 + 6\Phi^\gamma - 6\Phi^{2\gamma} + 2\Phi^{3\gamma}) + \\ & \gamma \ln \Phi^\gamma (3\Phi^\gamma - 6\Phi^{2\gamma} + 3\Phi^{3\gamma}) + \\ & \gamma^2 \Phi^{2\gamma} [(\Phi^\gamma - 1 - \ln \Phi^\gamma) + \ln \Phi^\gamma (\Phi^\gamma - 1)] \end{aligned}$$

hence if the constant term (first line) and the quadratic coefficient (third line) are positive $F(\Phi^\gamma, \gamma)$ cannot take on negative values. Since $\Phi^\gamma \geq 1$, $\ln \Phi^\gamma \geq 0$. Moreover the cubic form $-2 + 6\Phi^\gamma - 6\Phi^{2\gamma} + 2\Phi^{3\gamma}$ is strictly positive for $\Phi^\gamma \geq 1$. Therefore the constant term is weakly positive. For the same reason, since $\ln \Phi^\gamma \leq \Phi^\gamma - 1$, the quadratic coefficient must also be weakly positive. Hence $\frac{d\lambda^*}{d\gamma} \geq 0$ which implies $\frac{d\lambda^*}{d\alpha} \geq 0$.

Finally, we obtain

$$\frac{d\lambda^*}{d\Phi} = \frac{A^2}{\Phi^{\gamma+1}} \frac{\gamma \beta^{\text{med}}}{(\Phi^\gamma \gamma + 2\Phi^\gamma - 2)^3} [(\Phi^{3\gamma} (3\gamma + 2 + \gamma^2) - \Phi^{2\gamma} (6\gamma + 6 + 2\gamma^2) + \Phi^\gamma (3\gamma + 6) - 2)]$$

and the sign of this derivative is given by the square parenthesis. At $\Phi^\gamma = 1$ the sign is negative. Since the coefficient of the cubic term is positive, we know there is a Φ^γ large enough that the sign is positive for any larger Φ^γ . Finally note that the derivative of the expression in square brackets at $\Phi^\gamma = 1$ equals $-\gamma^2$. It follows that $\frac{d\lambda^*}{d\Phi}$ starts negative and can only switch signs once. ■

Proof of Proposition 6. First note that the objective function is globally concave in λ because $\tau(\lambda)$ is a convex function given by (22). Hence preferences are single-peaked.

Differentiate the objective function of program 18 with respect to β .

$$\frac{\partial k}{\partial \beta} \left[A - \tau(\lambda) - \frac{2k(\beta)}{\beta} + 1 \right] + \frac{(k(\beta))^2}{\beta^2}$$

The first term is always 0 due to capital being endogenously chosen. Hence the cross-derivative is simply

$$\frac{2k(\beta)}{\beta^2} \frac{\partial k}{\partial \lambda} > 0$$

as established in Proposition 4. By Topkis' Theorem, the ideal point of voters is increasing in β . ■

Proof of Proposition 7. Consider any decentralization scheme $\{\lambda_h\}_{h=1}^H$ such that there exists a h', h such that $h < h', \lambda_h > 0, \lambda_{h'} < 1$. The utility of the agent n is given by

$$rk(\beta^n, r) + F(\bar{k}(r)) - r\bar{k}(r) - \sum_{h=1}^H \lambda_h + \sum_{h=1}^H (1 - \lambda_h) \frac{\Phi^{\gamma_h} - 1}{\alpha_h} - (k(\beta^n, r) - \beta^n) l \left(\frac{k(\beta^n, r)}{\beta^n} \right)$$

Now consider another decentralization scheme $\{\hat{\lambda}_h\}_{h=1}^H$ such that

$$\begin{aligned} \hat{\lambda}_h &= \lambda_h - \delta \\ \hat{\lambda}_{h'} &= \lambda_{h'} + \varepsilon \\ \hat{\lambda}_{\tilde{h}} &= \lambda_{\tilde{h}} \text{ for all } \tilde{h} \neq h, h' \end{aligned}$$

where

$$\frac{\delta}{\varepsilon} = \frac{\Phi^{\gamma_{h'}+1}}{\Phi^{\gamma_h+1}} = \frac{\Phi^{\gamma_{h'}}}{\Phi^{\gamma_h}}$$

and δ is small enough that $\hat{\lambda}_{h'} < 1$ and $\hat{\lambda}_h > 0$. Note that this change in the decentralization scheme holds the equilibrium capital tax constant, and hence holds equilibrium investment capital decisions constant. Hence, we need only calculate

changes in utility due to differences in public goods and head taxes. Furthermore, the change in outcomes affects each agent in the same way, so all agents will agree on whether it is good or bad.

For public goods of type h , there will be more provision, since these have now been centralized. In particular, agents will gain $\delta \frac{\Phi^{\gamma_h} - 1}{\alpha_h}$ in utility from these newly centralized goods. Similarly, agents will lose $\varepsilon \frac{\Phi^{\gamma_{h'}} - 1}{\alpha_{h'}}$ from the fact that less of goods of type h' will now be provided. Finally, head taxes will decrease by $\delta - \varepsilon = (\Phi^{\gamma_{h'} - \gamma_h} - 1) \varepsilon$. Adding together these three effects and dividing by ε gives us

$$\begin{aligned} (\Phi^{\gamma_{h'} - \gamma_h} - 1) + \Phi^{\gamma_{h'} - \gamma_h} \frac{\Phi^{\gamma_h} - 1}{\alpha_h} - \frac{\Phi^{\gamma_{h'}} - 1}{\alpha_{h'}} &> \Phi^{\gamma_{h'} - \gamma_h} \left(\frac{\alpha_h - 1}{\alpha_h} \right) + \Phi^{\gamma_{h'}} \left(\frac{1}{\alpha_h} - \frac{1}{\alpha_{h'}} \right) \\ &> \Phi^{\gamma_{h'}} \left(\frac{1}{\gamma_h} - \frac{1}{\gamma_{h'}} \right) (1 - \Phi^{-\gamma_h}) \\ &> 0 \end{aligned}$$

as $\Phi > 1$ and $\gamma_h < \gamma_{h'}$. ■

Proof of Proposition 8. For the first part, we wish to solve for each agent

$$\max_{\{\lambda_h\}_{h=1}^H} \left\{ \begin{array}{l} rk(\beta^n, r) + F(\bar{k}(r)) - r\bar{k}(r) - \sum_{h=1}^H \lambda_h + \sum_{h=1}^H (1 - \lambda_h) \frac{\Phi^{\gamma_h} - 1}{\alpha_h} \\ - (k(\beta^n, r) - \beta^n) l\left(\frac{k(\beta^n, r)}{\beta^n}\right) \end{array} \right\}$$

For simplicity, let

$$v(k, \beta) = (k(\beta^n, r) - \beta^n) l\left(\frac{k(\beta^n, r)}{\beta^n}\right)$$

and note that $\frac{\partial^2 v}{\partial k \partial \beta} < 0$. It is enough to show, given the previous proposition, that for each λ_h , the cross-partial of the objective function with respect to λ_h and β is positive, as then the most-preferred λ_h must be increasing in the parameter β^i by Topkis' theorem.

Taking the derivative of the objective function with respect to β^i , we have

$$(r - \nu_1(k(\beta^i, r), \beta^i)) k_1(\beta^i, r) - \nu_2(k(\beta^i, r), \beta^i) = -\nu_2(k(\beta^i, r), \beta^i)$$

where the equality comes from the the condition that the agent is optimizing his capital choice given the rate of return, so that $r = \nu_1(k(\beta^i, r), \beta^i)$. Taking the derivative now with respect to λ_h , we have

$$-\nu_{12}(k(\beta^n, r), \beta^n) k_2(\beta^n, r) \frac{dr}{d\lambda_h}$$

However, $-\nu_{12}(k(\beta^n, r), \beta^n)$ is positive by assumption, and the proof of proposition 3 tells us that the derivative of k with respect to r is positive. Finally, Proposition 1 and 4 show that $\frac{dr}{d\lambda_h}$ is positive, and hence by Topkis' theorem we are done.

To see the second part, note that if $\alpha_H < 1$, then $\alpha^*(\beta^{\min}) < 1$ is less than 1 by definition. Otherwise, the median voter will use this perfectly redistributive instrument and set the capital tax equal to the pre-tax return on capital *ex post*. Hence, there will be no capital investment *ex ante*. Then any agent could be better off if every public good was decentralized, ensuring a capital tax of zero so he not be worse off. ■