# Fiscal Federalism and Electoral Accountability*

Toke S. Aidt† Jayasri Dutta‡

August 15, 2015

**Abstract**

We evaluate how governance uncertainty – exemplified by turnout uncertainty – affects the trade off between internalization of externalities and political accountability in the design of the fiscal state. We show that centralization only weakens political accountability in the presence of negative externalities. Unlike positive externalities, negative externalities allow federal politicians to extract higher rents. This yields two new insights. First, decentralization can only Pareto dominate centralization in economies with negative externalities. Second, centralization may not be Pareto efficient in economies with positive externalities despite the fact that policy can be tailored to regional taste differences and centralization internalizes the positive externality.

*Keywords*: Fiscal federalism, local public goods, externalities, performance voting, turnout uncertainty, electoral accountability.

*JEL Classifications*: D72; D78; H41.

## 1 Introduction

Should a society apt for a centralized fiscal system where spending decisions are made by a central authority and financed from general tax revenues or should it apt for a decentralized system where fiscal choices are made by local authorities and financed by local taxes? Oates (1972) answered this question in his Decentralization Theorem which states that decentralization is desirable if externalities are weak and regional differences in taste are large.[1] The design of the fiscal state also entails many political economy trade

*We thank Robert Evans, Paola Profeta, Dalibor Eterovic and Lars Feld for useful comments and suggestions. This research was supported by a grant from the Barcelona Institute of Economics (IEB) under its research programme in Fiscal Federalism.

†Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, Tel.: +44 1223 335231. E-mail: Toke.Aidt@econ.cam.ac.uk.

‡Department of Economics, University of Birmingham. E-mail: J.Dutta@bham.ac.uk.

[1] Harstad (2007) provides a rationale for why it might be politically optimal to select uniform federal policies.

offs.[2] This paper contributes to the political economy of fiscal federalism by exploring how governance uncertainty affects the trade off between internalization of externalities and the perceived benefits of electoral accountability under regionalism. Governance uncertainty arises when there is uncertainty about which region in a federation will determine the outcome of federal elections.

The general framework of our analysis is the common agency model with governance uncertainty studied by Aidt and Dutta (2004). This model portrays a society populated by heterogenous groups of voters (e.g., living in different regions) with conflicting policy preferences. The groups of voters (the principals) use elections to hold an opportunistic politician (the agent) accountable for his policy choices while in office. They do so by voting retrospectively in an infinite sequence of elections, as in Ferejohn (1986), Persson *et al.* (1997), Coate and Morris (1999), Aidt and Magris (2006), or Aidt and Dutta (2007). The critical new feature of the analysis is that before each election the politician is uncertain about which group will be pivotal in deciding the outcome of the election. We call this governance uncertainty. Governance uncertainty has many different sources. To be concrete, however, we relate it to randomness in the electoral turnout of voters in different groups.[3] These random turnout shocks introduce uncertainty from the point of view of the politician as to which group holds the majority amongst those voters who actually turn out to vote in any given election. An example is random fluctuations in weather conditions. Such fluctuations can affect the turnout rate in some geographical locations and not in others or keep certain types of voters, such as the poor, at home (Roemer, 1998, and Gomez *et al.*, 2007). Artés (2014), for example, reports that election day rainfall decreases turnout in Spanish general elections and that high turnout harms the conservatives and benefits the smaller parties. Lind (2014) also reports evidence that rain on election day affects voter turnout and that this, in turn, affects the party composition of the elected councils in Norwegian local elections. Along similar lines, Arnold and Freier (2015) show that rain shocks affect turnout in the German state of North-Rhine Westphalia and that lower turnout benefits the conservatives and harms the social democrats. These examples illustrate that random fluctuations in turnout, induced by rain shocks, can affect election outcomes.

Based on this general framework, we develop a specific political economy model of fiscal federalism in a country with two regions. Provision of local public goods in one region has spillover effects on the other region. Centralization encourages internalization, but introduces governance uncertainty because the federal politician cannot know for sure *ex ante* which of the two regions will decide his reelection. We use this model to study when provision of local public goods should be centralized? The model also provides insights into the forces that stabilize and destabilize federal fiscal structures.

[2] We discuss this literature in Section 2.

[3] Dhillon and Peralta (2002) survey the literature on voter turnout. Aldashev (2015) studies the link between political rents and voter turnout.

The main result of the analysis is an asymmetry between positive and negative externalities with regard to how much rent politicians in a federation can extract. With negative externalities, a federal politician can extract more rents from his citizens than the collective of regional politicians. With positive externalities, this is not the case. Intuitively, the asymmetry arises because it is more expensive for the federal politician to keep all regions satisfied when public goods provided in one region exhibit a negative externality than when the externality is positive. To avoid being "ignored" when fiscal externalities are negative, voters in the two regions must accept that the federal politician collects higher rents than with positive externalities. In the presence of negative externalities, the classical result that federalism is better for citizens than regionalism can even be turned on its head. Since federal politicians can extract more rents, federalism can be worse than regionalism for all citizens, despite the fact that regionalism does not internalize the negative externality. With positive externalities, federal politicians are not able to extract additional rents. This implies, in sharp contrast to the case with negative externalities, that regionalism cannot Pareto dominate federalism. The reason is that externalities get internalized with federalism, which is good for all voters. The only downside of federalism is that the federal politician may be induced to redistribute, but that, by definition, benefits one region at the expense of the other.

Section 2 discusses the related literature. Section 3 presents the general political common agency model with governance uncertainty. Section 4 presents the equilibrium characterization results from Aidt and Dutta (2004). Section 5 tailors the general model to the case of local public goods and fiscal federalism and presents the main results of the paper. Section 6 discusses the implications for fiscal integration and disintegration. Section 7 summarizes and discusses a number of extensions. The supplementary online appendix contains complete proofs of all propositions and details on robustness checks related to distortionary taxation.

## 2 Related Literature

The paper is most directly related to the work by Seabright (1996), Tommasi and Weinschelbaum (2007), Bordignon *et al.* (2008), and Hindriks and Lockwood (2009).[4] Seabright (1996) argues that political accountability is weakened when fiscal decisions are centralized. The reason is that politicians only need to win a simple majority of regions in a federal election and this allows them to capture more rents at the federal level than at the regional level. Our analysis also suggests that accountability is lower when fiscal decisions are centralized, but it differs in two regards. Firstly, Seabright (1996) measures the accountability effect as the reduction in the probability that a given region can determine the reelection of the government. In contrast, we conceptualize the political clout of a region

[4] Inman and Rubinfeld (1997) and Lockwood (2006) survey the literature on the political economy of fiscal federalism.

by the probability that voters of that region holds the majority among those who turn out to vote in the federation. Importantly, we show that the accountability effect depends on the nature of the externalities associated with provision of local public goods. Although an asymmetry between the net benefit of centralization with positive and negative externalities can also be found in Lockwood (2002), the observation that electoral accountability operates differently under the two types externalities is a new insight which is not captured in the previous literature. Second, Seabright (1996) does not explicitly model the benefits of centralization (he simply assumes that ego-rents are higher in a federation). Our analysis explicitly models the benefit of centralization related to internalization of externalities. We emphasize the trade off between loss of accountability and internalization of external costs.

Tommasi and Weinschelbaum (2007) and Bordignon *et al.* (2008) study fiscal federalism in a common agency model. They allow the citizens of the regions within the country to offer monetary rewards to either the federal politician (under federalism) or to the regional politicians (under regionalism).[5] They identify a trade off between internalization of externalities and the coordination failure that arises when fiscal decisions are centralized. One can interpret the trade off that we highlight in a similar way. One key difference, however, is that we focus on the implicit incentives provided by elections rather than the explicit incentives provided by monetary payments. Like us, Bordignon *et al.* (2008) find that the distinction between positive and negative externalities matters. The reason is, however, very different: lobbying under regionalism may partly compensate for the fact that local public goods are under-provided, but only if the externality is positive.

Hindriks and Lockwood (2009) stress that voters are often poorly informed about policy outcomes. In addition to their disciplining role, elections can, therefore, serve as a selection devise. As in our context, fiscal centralization may reduce electoral accountability, but this effect is counteracted by a selection effect that encourages "bad" incumbents to pretend to be "good". Our analysis abstracts from selection in order to stress the effect of governance uncertainty on electoral accountability, but future work could consider adding an adverse selection component to the framework.

Besley and Coate (2003) identify two important political effects of centralization. Both are related to the legislative procedures operating at the federal level. First, centralization makes it uncertain whether or not the representative from a particular region will be included in the minimum winning coalition that determines federal policy. Second, regional voters may have an incentive to delegate strategically and elect a politician who cares more about public spending than they do to represent them at the federal level. In both cases, centralization involves a trade off between the political distortion (uncertainty or strategic delegation) and the benefits of internalizing (positive) externalities. Besley and Coate (2003) find that centralization is, typically, beneficial if the externality is strong

[5] Tommasi and Weinschelbaum (2007) assume that the set of lobby groups is fixed. Redoano (2010) shows the number of lobby groups varies with the level of centralization.

enough.[6] Lülfesmann *et al.* (2015) argue that Besley and Coate (2003) underplay the scope for bargaining amongst regions and show that decentralization tends to dominate centralization when this is taken into account.[7] Like Besley and Coate (2003), we focus on the uncertainty that arises when fiscal decisions are centralized, but we stress governance uncertainty rather than uncertainty about being included in the minimum winning coalition. It is interesting to notice that decentralization, in our model, can only Pareto dominate centralization in the presence of a negative externality – a case that Besley and Coate (2003) do not consider.

Lockwood (2002) argues that centralization leads to inefficient outcomes when regional representatives vote over agendas that contain sets of region-specific projects. The reason is that the political choice is not tailored sufficiently to within-region benefits. Thus, centralization entails a classical trade off between catering for regional differences and internalizing externalities. Importantly, however, the political distortions imply that weaker externalities and greater heterogeneity between regions need not increase the efficiency gain from decentralizations. In our model, there is no regional differences with regard to the benefits of public goods. Nonetheless, we find an interesting asymmetry between positive and negative externalities which provides a complementary example of how politics can change the classical trade offs in surprising ways. Lockwood (2008) further explores ways in which the Decentralization Theorem may break down under majority voting even when federal policy, by assumption, cannot reflect regional preferences.

# 3 A General Model of Governance Uncertainty

Society consists of two groups of voters, $i = 1, 2$; politicians are indexed by 0. A group is defined as a subset of voters who are affected in the same way by public policy. In the context of fiscal federalism, the groups represent two regions within a federation. Per-period utility, $u_{it}$, is discounted with the common discount factor $\beta \in (0, 1)$. Lifetime utility is

$$V_{0i} = \sum_{t=0}^{\infty} \beta^t u_{it}; \quad i \in \{0, 1, 2\}. \tag{1}$$

There are $n_1$ voters in group 1 and $n_2$ voters in group 2. We assume that $n_1 \geq n_2$ and define $n = n_1 + n_2$.

Each period, the politician collects taxes up to a maximum of $T$, spends some of this on providing amenities to his electorate, and keeps the rest for himself as rent. Denoting the cost of providing utilities to the two groups of voters by $c_t$, we can write the politician's per-period payoff as

$$u_{0t} = T - c_t \tag{2}$$

[6] Dur and Roelfsema (2005) show that centralization may fail to internalize externalities if the cost of public goods cannot be shared among the regions.

[7] Similar conclusions are reached by Cheikbossian (2000) and Lülfesmann (2002).

if in office, and $u_{0t} = 0$ otherwise.

The cost of providing utility to voters is determined by the *political cost function.* We define $C(x_{1t}, x_{2t})$ as the minimum cost to the politician of providing utility levels $u_{1t} \geq x_{1t}$ and $u_{2t} \geq x_{2t}$ simultaneously to voters in the two groups. Likewise, we define $C_i(x_{it})$ as the minimum cost of providing the utility level $u_{it} \geq x_{it}$ to group $i$, $i = 1, 2$, in isolation. We begin by specifying the political cost function directly, but shall derive it from more fundamental considerations in the application to fiscal federalism that follows. We make the following assumptions.

**Assumption 1** *The political cost functions are monotonically increasing in each argument, i.e.,*

$$(\mathbf{M}) \quad \begin{array}{c} x_t > x_t' \Rightarrow C(x_t) \geq C(x_t') \\ x_{it} > x_{it}' \Rightarrow C_i(x_{it}) \geq C_i(x_{it}') \end{array}$$

*where* $x_t = (x_{1t,}x_{2t})$. *Further,* $\lim_{x_i \to \infty} C(x_1, x_2) = \lim_{x_i \to \infty} C_i(x_{it}) = \infty$.

**Assumption 2** *The political cost functions are continuous, i.e.,*

$$(\mathbf{K}) \quad \begin{array}{c} C(x_{1t,}x_{2t}) \in \mathcal{C}^1 \\ C_i(x_{it}) \in \mathcal{C}^1 \end{array} .$$

The first assumption says that it is costly for the politician to generate utility for each group of voters. This is clearly the case whenever tax resources that could otherwise have been extracted as rents have to be devoted to the task. However, when the politician can provide public goods, the cost functions may not be strictly increasing. The second assumption rules out discontinuities in the cost of generating utilities. Both of these assumptions can be relaxed.

The property of the political cost function that really matters is whether it is sub- or super-additive. The political cost function is sub-additive if

$$(\mathbf{C}^+) \quad C(x_{1t}, x_{2t}) \leq C_1(x_{1t}) + C_2(x_{2t}) \tag{3}$$

and super-additive if

$$(\mathbf{C}^-) \quad C(x_{1t}, x_{2t}) > C_1(x_{1t}) + C_2(x_{2t}). \tag{4}$$

A sub-additive political cost function makes it cheaper to provide utility to all voters jointly than to provide the same utility levels to the two groups separately. In public finance, sub-additivity is, typically, associated with pure public goods or positive externalities. A super-additive cost function makes it more expensive to please all groups of voters jointly than to please them separately. Super-additivity is caused by negative externalities associated with, for example, provision of local public goods, pollution, or envy effects.

The politician, elected at $t$, cannot make binding promises on the level and pattern of public spending before he enters office. Since his own payoff decreases with $c_t$, he would, in the absence of electoral incentives, choose $c_t = 0$ and provide no amenities to the electorate. Voters know this, and threaten to vote retrospectively against a politician who does not provide them with a minimum level of utility. At the beginning of each period, voters in each group announce simultaneously a performance standard, denoted $x_{1t}$ and $x_{2t}$. They then vote in favor of reelection of the incumbent politician if, and only if the policy implementation observed at the end of the period generates at least that level of utility, i.e., if, and only if $u_{it} \geq x_{it}$.

The key feature of the model is that politicians are exposed to governance uncertainty. At the most general level this means that the incumbent cannot be sure ex ante which of the two groups is decisive in determining his reelection. Governance uncertainty can arise for many reasons. A leading example is electoral turnout uncertainty, and this is the interpretation we shall follow here for concreteness. In particular, we assume that none of the groups can guarantee to turn out in full force at elections. Consequently, a politician may deliver on the performance standard set by group 1, who, say, holds the majority ex ante, by incurring the cost $C_1(x_{1t})$, but fail to deliver on the standard set by group 2 ($u_{2t} < x_{2t}$). On the day of the election, $\tilde{n}_{it}$ voters from group $i$ actually show up to vote, and the politician can lose his bid for reelection if $\tilde{n}_{2t} > \tilde{n}_{1t}$. The central assumption of our analysis is that electoral turnout is uncertain, and that individual voters vote according to the announced performance standards *if they show up to vote*, but that they cannot, as a group, guarantee a particular turnout rate. This is captured by the next assumption.

**Assumption 3** *Electoral turnout, $\tilde{n}_t = (\tilde{n}_{1t}, \tilde{n}_{2t})$, is random. The ex ante probability that the turnout of group 1 is greater than that of group 2, $P(\tilde{n}_{1t} \geq \tilde{n}_{2t})$, is equal to $p_1$ and constant over time. The complementary probability is $p_2 = 1 - p_1$. We assume that $p_1 \in (0, 1)$.*

It is important that neither group can guarantee reelection. This is more likely to be the case when turnout shocks are correlated within groups and when differences in group sizes are not too large.

We stress that governance uncertainty can arise for many other reasons than turnout uncertainty. It may, for example, reflect fluctuations in inter-group power relations with one group becoming more powerful and, therefore, more pivotal than another due to unpredictable events. The lobbying power of social groups may fluctuate in this way. Under this interpretation, the probability of being pivotal, $p_i$, is a manifestation of randomness in the cost of political mobilization. Combined with the insights from Olson (1965), a minority could be as likely as a majority group to be pivotal, not because it may in fact hold the majority among those who turn out to vote, but because it is better at organizing an effective lobby group. Another example is random preferences. Suppose that some people like education spending while others want spending on care for the elderly and that the proportions of individuals of these two types fluctuate in unpredictable ways. In this case, $p_i$ represents the probability that one of the "preference types" is pivotal.

The game between the incumbent politician and the two groups of voters unfolds over time as follows. At the beginning of each period, voters in each group announce the (utility) standard that the politician needs to satisfy to get their votes in the next election. The standards are chosen by the two groups non-cooperatively and at the same time. The politician observes the standards and determines whether to comply, and if so, how many standards to meet. We denote the set of actions available to the politician by $A = \{(00), (10), (01), (11)\}$ with elements $a_t = (00)$ (meet neither standard); $a_t = (10)$ (meet group 1's standard only); $a_t = (01)$ (meet group 2's standard only); and $a_t = (11)$ (meet both standards). At the end of the period, a new election is held and voters randomly turn out to vote. Those who turn out vote according to the announced performance standard. The politician either wins or loses. In the latter case, he is replaced by an identical challenger; in the former case, he gets (at least) another term in office. After the election, the game continues to the next period where a similar sequence of events takes place. We restrict attention to history-independent subgame perfect Nash equilibria of this game. In addition, we assume that the politician, if indifferent between two or more actions (which are then preferred to the remaining ones), chooses the action that maximizes his reelection chance. Below, when we refer to equilibrium, this is what we mean.

# 4 Equilibrium Paths

The political agency model presented above is a special case of the general model studied in Aidt and Dutta (2004). We can, therefore, apply Theorem 1 from that paper to characterize the set of equilibria.[8] The Theorem says that all equilibrium paths of the political game described above have a property called *strategic consensus:* the politician prefers to meet all performance standards at all times, all those voters who turn out to vote in the election vote for the incumbent, and the incumbent is reelected with certainty, irrespective of turnout shocks. While this outcome, perhaps, is to be expected when the political cost function is sub-additive and it is cheaper for the politician to satisfy the standards jointly than separately, it is surprising that the same result obtains with super-additive costs. In this case, the fact that it is *more* expensive to satisfy the standards jointly than separately suggests that "partisan" outcomes would be more likely. This intuition is, however, wrong.

To grasp logic behind the consensus result, consider the special case where the only policy instrument is a group-specific transfer. This makes the political cost function additive. To please voters, the politician must either be partisan and give transfers to one group only or seek consensus and give to both. The two groups of voters announce their standards simultaneously. Suppose that group 1 announces a standard that is so high that the politician prefers to take his chances and offer transfers only to group 2. This cannot be an equilibrium. This is because group 1 gets nothing and it would do better by

[8] Supplementary online appendix I restates Theorem 1 in a slightly generalized form along with a proof tailored to the political agency model.

reducing its standard to a level such that it is in the best interest of the politician to offer it a transfer. In other words, whenever the politician is willing to implement a "partisan" policy, the disfavored group has an incentive to lower its standard to induce the politician to implement a "partisan" policy in its favor. This logic continues until the standards are such that the politician is just willing to implement a policy that satisfies both groups. The result is strategic consensus. Importantly, it does not follow from this logic that the two groups will "under-bid" each other until the politician captures the entire rent. This would only happen if the two groups were "perfect substitutes" in the sense that either of them can guarantee reelection for sure (see Ferejohn, 1986). In our model, however, the two groups of voters avoid Bertrand-style competition precisely because they are not "perfect substitutes" from the point of view of the politician: both are needed to secure reelection with certainty. As a consequence, voters retain some control power. A similar logic applies when the political cost function is either sub- or super-additive.

Although all equilibrium paths display strategic consensus, the distribution of payoffs depends critically on the properties of the political cost function. Let $X = \{x_{1t}, x_{2t}\}_{t=0}^{\infty}$ be a sequence of *equilibrium* performance standards. In an economy with sub-additive political costs, the following characterization result holds.

**Proposition 1 (Sub-additive Costs)** *If the political cost functions satisfy assumptions [M] and [K] and are sub-additive, then X must satisfy*

$$(\mathbf{SC_1^+}) \quad C(x_{1t}, x_{2t}) = \beta T;$$

$$(\mathbf{SC_2^+}) \quad C_1(x_{1t}) \geq \beta p_1 T;$$

$$(\mathbf{SC_3^+}) \quad C_2(x_{2t}) \geq \beta p_2 T.$$

*Moreover,* $(\mathbf{SC_2^+})$ *and* $(\mathbf{SC_3^+})$ *hold with equality for additive political cost functions. Along all equilibrium paths, the politician receives payoffs* $(1-\beta)T$ *per period.*

The politician *always* gets per period payoff $(1-\beta)T$, while the remaining share of tax revenues, $\beta T$, is devoted to generate utilities to voters. Importantly, this distribution of resources is unaffected by turnout uncertainty. Thus, strategic consensus provides the politician with "full insurance" against random voter turnout and voters with insurance against "partisan" choices by the politician. When the political cost function is additive, the payoffs are uniquely determined by $p_1$ and $p_2$. In contrast, economies with strictly sub-additive costs exhibit multiple equilibria in performance standards at each $t$. Moreover, any equilibrium allocation what arises with sub-additive costs (weakly) Pareto dominates the utility allocation with additive costs.

In an economy with super-additive political costs, the utility allocation is very different, as shown by the second characterization result.

**Proposition 2 (Super-additive Costs)** *If the political cost functions satisfy assumptions [M] and [K] and are super-additive, then X must satisfy*

$$(\mathbf{SC}_1^-) \quad C(x_{1t}, x_{2t})(1+\eta_1) - C_1(x_1) = \eta_1 T$$

$$(\mathbf{SC}_2^-) \quad C(x_{1t}, x_{2t})(1+\eta_2) - C_2(x_2) = \eta_2 T$$

*where* $\eta_i = \frac{(1-p_i)\beta}{1-\beta}$ *for* $i = 1, 2$. *The politician receives payoffs* $T - C(x_{1t}, x_{2t}) > (1-\beta)T$ *every period. Moreover, if the cost functions are differentiable and* $\frac{\partial C}{\partial x_1 \partial x_2} > 0$, *then the solution to* $(\mathbf{SC}_1^-)$ *and* $(\mathbf{SC}_2^-)$ *is unique.*

With super-additive political costs, the politician receives more than he receives along *any* equilibrium path with sub-additive costs, but his payoff is no longer independent of turnout shocks. Intuitively, super-additive costs make it costly for the politician to implement consensus outcomes. This enables him to extract more rent: the two groups of voters have to lower their standards to prevent "partisan" outcomes.

# 5 A Specific Model of Fiscal Federalism

We now tailor the general model to fiscal federalism. We consider a country with two regions, $i = 1, 2$, inhabited by $n_i$ voters and with $n_1 \geq n_2$. Voters in each region derive utility from local public goods $g_{it}$ and private goods $y_{it}$. Consumption of local public goods in one region generates externalities for voters in the other region. To capture this, we write the utility function of a typical voter living in region $i$ as

$$u_{it} = y_{it} + g_{it} - \gamma g_{-it}. \tag{5}$$

The parameter $\gamma \in (-1, \frac{n_2}{n_1})$ captures the strength of the externality: $\gamma > 0$ corresponds to a negative and $\gamma < 0$ to a positive externality. Public goods are produced by the following technology

$$g_{it} = 2\sqrt{k_{it}}, \tag{6}$$

where $k_{it}$ is the input required to produce the public good, bought at a constant price of one.[9] The maximum revenue that can be collected each period in region $i$ is $T_i$. The maximum revenue that can be collected in the country is $T = T_1 + T_2$. As in Acemoglu and Robinson (2000), we can interpret these limits as the result of convex deadweight costs: up to $T_i$ revenues can be collected at zero cost, while revenues beyond this are prohibitively costly to collect. We return to the question of distortionary taxation in Section 7. We use the convention that politicians raise the maximum revenue each period, spent some of it

[9] We can relax the assumption of a linear utility function and the Cobb-Douglas specification of the production function. What we need is that the policy functions are concave.

on providing local public goods, some on transfers $s_{it} > 0$ to voters, and keep the rest as rent.[10]

We compare two institutional arrangements: Regionalism $[R]$ and federalism $[F]$. Regionalism means that each region elects a regional politician who can finance local public goods, transfers, and rents out of local tax revenues. Federalism means that a single elected politician is in charge of the whole country. The federal politician uses general tax revenues to provide public goods and transfers to the two regions, and to fund rents for himself.

The key assumption is that there is uncertainty at the federal level about which of the two regions will hold the majority in each election. Such uncertainty is, by definition, not present under regionalism. The ex ante probability that voters in region $i$ holds the majority among those who turn out to vote in a federal election is $p_i$ with $p_1 = 1 - p_2$. Again, we stress that turnout uncertainty is not the only valid interpretation. For example, $p_i$ can also be interpreted as a power index that captures the influence of region $i$ in federal decisions. All regions may be pivotal occasionally because of random shifts in power relations, but some regions are more likely to be pivotal than others. In practise, many federal constitutions include mechanisms to protect minorities and smaller regions. This often involves having an upper chamber where regions have equal representation rather than representation in relation to population size. This suggests that governance uncertainty can be induced by constitutional features that aim at protecting minorities.

## 5.1 Utilitarian Planners

As a benchmark, suppose that all public finance decisions were made by utilitarian planners. Under regionalism, two regional planners decide simultaneously how much local public good to provide to their region. They do so by maximizing regional aggregate public goods surplus:

$$s_{it}^{R}(k_{it}; k_{-it}) = 2n_i(k_{it}^{\frac{1}{2}} - \gamma k_{-it}^{\frac{1}{2}}) - k_{it},\ i = 1, 2 \tag{7}$$

subject to the regional revenue constraint and taking the spending decision in the other region as given. In a federation, on the other hand, decisions are made by one utilitarian planner. He maximizes aggregate public goods surplus, i.e., $s_t^F(k_{1t}, k_{2t}) = \sum s_{it}^R(.)$ subject to the total tax revenues raised from the two regions. The public goods surplus cannot be lower under federalism than under regionalism and for $\gamma \neq 0$, the surplus is strictly larger. Since utility is linear in income (and in transfers), the federal planner does not care about redistribution of income across the two regions. These two points together imply that the federal planner, for all $\gamma \neq 0$, is able and willing to devise a tax sharing rule such

[10] Another interpretation is that the politicians set the gross tax rate that collects the maximum revenue and then give voters a tax rebate which determines what the net tax rate is. So, in this restricted sense, taxation is endogenous.

that federalism Pareto dominates regionalism.[11] The intuition is straightforward. With benevolent planners, centralization does not create or eliminate political distortions and federal and regional planners are equality good at catering to local tastes. Moreover, the federal planner is indifferent to redistribution of income but is, unlike the regional planners, willing to internalize externalities. Hence, federalism Pareto dominates regionalism. This provides a clear-cut benchmark against which we measure political distortions.

## 5.2 The Political Cost Functions

To characterize equilibrium allocations, we derive the political cost functions. We focus on the situation in which both federal and regional politicians provide local public goods *and* transfers at equilibrium. This requires that tax revenues are sufficiently large.[12]

Under regionalism, the two regional politicians face a separate performance standard. They make decisions about public spending without (direct) regard for the welfare of voters in the other region. Consider the politician in region $i$ who in period $t$ faces the performance standard $x_{it}$. The minimum cost of satisfying this standard, for a given input to the production of local public goods in the other region, is

$$C(x_{it}; k_{-i}) = \min_{k_{it} \geq 0, s_{it} \geq 0} k_{it} + n_i s_{it} \tag{8}$$

subject to $x_{it} \leq s_{it} + 2k_{it}^{\frac{1}{2}} - 2\gamma k_{-it}^{\frac{1}{2}}$ and the regional budget constraint. It follows that $k_{it} = (n_i)^2$ and $s_{it} = x_{it} - 2(n_i - \gamma n_{-i})$. The political cost functions are

$$C_i^R(x_{it}) = (n_i)^2 + n_i(x_{it} - 2(n_i - \gamma n_{-i})) \quad \text{for } i = 1, 2. \tag{9}$$

We notice that the externality is not internalized: both regions spend on local public goods up to the point where the marginal benefit is equal to the marginal cost for that region. The transfer must, therefore, "compensate" regional voters for the impact of spending on local public goods in the other region. In each region, voters set the performance standard in period $t$ taking the standard of the other region as given. At equilibrium, the standards make each regional politician indifferent between satisfying the standard and getting reelected (for sure) and not satisfying it. In the latter case, he is replaced but keeps local tax revenues $T_i$ for himself as rent. This yields the following stationary equilibrium allocation:

$$x_{it}^R = \frac{\beta T_i}{n_i} - n_i + 2(n_i - \gamma n_{-i}) \text{ for } i = 1, 2. \tag{10}$$

The politician of region $i$ keeps a share, $(1 - \beta)T_i$, of regional tax revenues each period, and uses the rest to provide local public goods and transfers to voters of his region. A

[11] For $\gamma = 0$, the institutional arrangement makes no difference. Supplementary online appendix II derives the sharing rule.

[12] See Supplementary online appendix II for details.

negative externality reduces voters' welfare ($\gamma > 0$), while a positive externality ($\gamma < 0$) enhances their well-being, as one would expect.

Under federalism, the single federal politician faces the performance standards $\{x_{1t}, x_{2t}\}$ set by voters in the two regions. He minimizes the cost of satisfying the two standards jointly by spending $k_{it} = (n_i - n_{-i}\gamma)^2$ on local public goods and by providing transfers $s_{it} = x_{it} - 2n_i(1+\gamma^2) + 4\gamma n_{-i}$ to voters in each of the two regions. The political cost function is, therefore, given by

$$\begin{aligned} C^F(x_{1t}, x_{2t}) &= (n_1 - n_2\gamma)^2 + (n_2 - n_1\gamma)^2 \\ &\quad + n_1\left(x_{1t} - 2n_1\left(1+\gamma^2\right) + 4\gamma n_2\right) \\ &\quad + n_2\left(x_{2t} - 2n_2\left(1+\gamma^2\right) + 4\gamma n_1\right). \end{aligned} \tag{11}$$

If the politician decides to satisfy the standard of one of the regions, say, region $i$, only, then it is clear that $s_{-it} = 0$. However, if local public goods generate a positive externality ($\gamma < 0$), it is cost effective to provide some local public goods to region $-i$. This is not because the politician cares about the welfare of voters in that region as such, but because it is, up to a point, cheaper to provide utility to voters in region $i$ this way than to give them transfers. Hence, for $\gamma < 0$, the cost minimizing choice of spending on local public goods is $k_{it} = (n_i)^2$ and $k_{-it} = (n_i\gamma)^2$ and the transfer to each voter of group $i$ is $x_{it} - 2n_i(1+\gamma^2)$. If, on the other hand, local public goods generate negative externalities, then $k_{-it} = 0$ and the politician spends $k_{it} = (n_i)^2$ on local public goods to region $i$ and provides the voters of that region with the transfer $s_{it} = x_{it} - 2n_i$. With this in mind, the political cost functions are:

$$C_i^F(x_{it}) = (n_i)^2 + n_i(x_{it} - 2n_i) \text{ for } \gamma \geq 0 \tag{12}$$

$$C_i^F(x_{it}) = \left(1+\gamma^2\right)(n_i)^2 + n_i(x_{it} - 2n_i(1+\gamma^2)) \text{ for } \gamma < 0. \tag{13}$$

We notice that for $\gamma < 0$

$$C^F(x_{1t}, x_{2t}) - \sum_i C_i^F(x_{it}) = 4n_2n_1\gamma < 0, \tag{14}$$

and for $\gamma \geq 0$

$$C^F(x_{1t}, x_{2t}) - \sum_i C_i^F(x_{it}) = \gamma\left(4n_1n_2 - \gamma(n_1^2 + n_2^2)\right) > 0. \tag{15}$$

The political cost function is sub-additive for $\gamma < 0$ and additive for $\gamma = 0$. For $\gamma > 0$, the political cost function is super-additive for all $\gamma \in \left(0, \frac{n_2}{n_1}\right)$ as $(3n_1^2 - n_2^2)\frac{n_2}{n_1} > 0$ for $n_1 \geq n_2$.

Below we apply Propositions 1 and 2 to characterize stationary equilibrium allocations. Our main goal is to compare regime [F] and [R] under different assumptions about the magnitude of the externality. We use Pareto efficiency as our welfare criterion. In doing so, we adopt a voter-centric approach and exclude the rents captured by the politicians.

That is, we say that regime [F] Pareto dominates [R] if all voters prefer [F] to [R]. This approach has several advantages. Firstly, in contrast to a criterion based on aggregate public goods surplus, the Pareto criterion has a clear-cut positive implication. If one institutional arrangement Pareto dominates another, *all* voters would support a change in a referendum.[13] Secondly, the comparisons are not distorted by whether politicians can extract more or less rent. Since rent is pure social waste in our model, this seems a reasonable choice from a normative point of view. However, from a positive point of view, it is interesting also to study how regional politicians, who may have a disproportionate say in whether centralization takes place or not, rank the different regimes. We do so in Section 6.

## 5.3 No Externalities

We begin by considering the case without externalities. In this case, political costs are additive and the total rent $((1-\beta)T)$ captured by the federal politician is equal to the sum of those captured by the two regional politicians $((1-\beta)T_1 + (1-\beta)T_2)$. An implication, then, is that the only effect of centralization is to allow redistribution between the regions. With additive political costs, centralization is a zero-sum game. If one region gains, it must be at the expense of the other. Consequently, the two regimes cannot be Pareto ranked.

**Proposition 3 (No externalities $\gamma = 0$)** *Regime [F] and [R] cannot be Pareto ranked. Region $i$ prefers regime [F] to [R] if, and only if*

$$p_i > \frac{T_i}{T} \quad \text{for } i = 1, 2.$$

**Proof.** Use Proposition 1 to derive the equilibrium utility allocation in regime [F]:

$$x_{it}^F(\gamma = 0) = \frac{\beta p_i T}{n_i} + n_i \ \text{ for } i = 1, 2.$$

The utility differences between regime [F] and [R] are

$$x_{it}^F(\gamma = 0) - x_{it}^R = n_i^{-1}\beta\left(Tp_i - T_i\right) = n_i^{-1}\beta(p_i T_{-i} - p_{-i}T_i) \equiv \widehat{\Delta}_i \ \text{ for } i = 1, 2. \qquad (16)$$

where $x_{it}^R$ is defined by equation (10). The proposition follows because $\widehat{\Delta}_1 > 0 \Leftrightarrow \widehat{\Delta}_2 < 0$ ■

Voters in region $i$ receives $\frac{p_i \beta T}{n_i} + n_i$ from the federal politician and $\frac{\beta T_i}{n_i} + n_i$ from the regional politician. Intuitively, therefore, whether a region gains or not from centralization depends on $p_i$ – the probability that each region holds the majority among those who turn out to vote in the federal election – relative to the region's contribution to federal tax revenues. In the absence of externalities, centralization tends to be favored by poor regions or by regions that are likely to be pivotal in federal decision making.

[13] See Crémer and Palfrey (1996), Lockwood (2004), Feld *et al.* (2008), Schnellenbach *et al.* (2010) or Galletta and Jametti (2015) for studies of federalism and direct democracy.

## 5.4 Negative Externalities

The situation is more interesting when local public goods generate a negative externality ($\gamma > 0$) and political costs are super-additive. In this case, centralization is associated with three effects. The first effect is the *redistribution effect* described above: centralization pools revenues from the two regions. This allows redistribution to take place. The second effect is the *internalization effect*: the federal politician internalizes the externality in order to minimize the cost of getting reelected. This benefits all voters. The third effect is the *rent effect.* This effect is new to the literature and arises because political costs are super-additive. Recall from Proposition 2 that the federal politician's share of total revenues, at equilibrium, is *larger* than $(1-\beta)T$. This implies that less is available in total to generate amenities to voters under federalism than under regionalism. This harms all voters. The next proposition isolates the externality and rent effect from the redistribution effect by assuming that $p_1 = \frac{1}{2}$ and that $T_1 = T_2$.

**Proposition 4 (Negative Externalities** $\gamma > 0$**)** *Let* $\theta = \frac{n_1}{n_2} \geq 1$. *Assume that* $p_1 = \frac{1}{2}$ *and* $T_1 = T_2$. *Then for* $\beta > \frac{(1+\theta)(\theta-1)}{2\left(3\theta^2-1\right)}$

1. *[R] is Pareto superior to [F] for* $\gamma \in (0, \frac{4(1-\beta)\theta}{(2-\beta)\left(1+\theta^2\right)})$.

2. *[F] is Pareto superior to [R] for* $\gamma \in (\frac{4(1-\beta)\theta}{(2-\beta)\left(1+\theta^2\right)}, \theta^{-1})$.

**Proof.** Use Proposition 2 and equations (11) and (12) to derive the (unique) stationary utility allocation under $[F]$:

$$x_{it}^{F}(\gamma > 0) = \frac{\beta p_i T + n_i^2 + \gamma\left(1-\beta p_i\right)\left(\gamma n_{-i}^2 - 4 n_i n_{-i} + \gamma n_i^2\right)}{n_i} \quad \text{for } i = 1, 2.$$

The utility differences between regime [F] and [R] are

$$\Delta_{it} = x_{it}^{F}(\gamma > 0) - x_{it}^{R} = \widehat{\Delta}_i + \frac{\gamma\left(\gamma\left(1-\beta p_i\right)\left(n_i^2 + n_{-i}^2\right) - 2 n_i n_{-i}(1-2\beta p_i)\right)}{n_i} \quad \text{for } i = 1, 2,$$

where $\widehat{\Delta}_i$ is defined in equation (16). For $p_1 = \frac{1}{2}$ and $T_1 = T_2$, we get

$$\Delta_{it} = \frac{\gamma\left(\gamma\left(1 - \frac{1}{2}\beta\right)\left(n_i^2 + n_{-i}^2\right) - 2 n_i n_{-i}(1-\beta)\right)}{n_i}.$$

We note that $\Delta_{it} \geq 0 \Leftrightarrow \Delta_{-it} \geq 0$. In particular, $\Delta_{it} < 0$ for $i = 1, 2$ for $\gamma \in (0, \frac{4(1-\beta)\theta}{(2-\beta)\left(1+\theta^2\right)})$ and (weakly) positive for $\gamma \in [\frac{4(1-\beta)\theta}{(2-\beta)\left(1+\theta^2\right)}, \theta^{-1})$ where $\theta = \frac{n_1}{n_2}$. We note that $\frac{4(1-\beta)\theta}{(2-\beta)\left(1+\theta^2\right)} < \theta^{-1} \Leftrightarrow \beta > \frac{(1+\theta)(\theta-1)}{2\left(3\theta^2-1\right)}$. Thus, the condition in the statement of the proposition ensures that the interval $[\frac{4(1-\beta)\theta}{(2-\beta)\left(1+\theta^2\right)}, \theta^{-1})$ exists ■

With negative externalities, centralization might be worse for *all* voters despite the fact that regionalism fails to internalize externalities and does not have any advantage in catering for heterogenous tastes. This stands in sharp contrast to the benchmark case with utilitarian social planners where federalism Pareto dominates regionalism.

The source of this new and surprising result is the rent effect. It implies that the federal politician can, in total, extract more rent than the two regional politicians together. The underlying logic is as follows. With negative externalities, the federal politician must, in part, internalize the externality that provision of local public goods to one region imposes on the other and vice versa if he wants to be reelected by voters in both regions. It follows that fewer public goods are provided under federalism than under regionalism and that voters cannot ask for as much utility from the federal politician as they could from the regional politicians. This, in turn, allows the federal politician to retain more rent.

The rent effect dominates the internalization effect for weak externalities. In this case, both regions are worse off in a federation. However, for $\gamma > \frac{4(1-\beta)\theta}{(2-\beta)\left(1+\theta^2\right)}$, the externality effect is sufficiently strong to dominate the rent effect, and federalism Pareto dominates regionalism. The threshold that determines which effect dominates is decreasing in the relative size of region 1 ($\theta = \frac{n_1}{n_2}$). This means that two unequal-sized regions are *more* likely to benefit from joining a federation than two equal-sized regions. This is because it is relatively expensive for a regional politician to compensate his voters through transfers for any un-internalized externalities when the two regions are of unequal size.

The result that centralization is efficient only with strong negative externalities echoes the classical argument by Oates (1972). The logic, however, is entirely different. Oates focused on the trade off between internalizing externalities and catering for differences in the taste for public goods in different regions. The trade off behind Proposition 4 has nothing to do with heterogenous taste. It is driven by the rent effect. Centralization implies a transfer of resources from voters in the two regions to the federal politician. This is why centralization is not in the interest of any voter if externalities are weak.

Proposition 4 ignores the redistribution effect. This effect, as we noted above, is driven by turnout uncertainty as captured by $p_i$ and differences in tax resources in the two regions. With the redistribution effect, we can define the values of $p_1$ for which the two regions are indifferent between the two regimes as:

$$p_1^1(\gamma, \lambda) = \frac{\beta\lambda T_2 + \gamma\left(2n_1 n_2 - \gamma(n_1^2 + n_2^2)\right)}{\beta\left(T_2\left(1+\lambda\right) + \left(4n_1 n_2 - \gamma(n_1^2 + n_2^2)\right)\gamma\right)}; \tag{17}$$

$$p_1^2(\gamma, \lambda) = \frac{\beta\lambda T_2 + \gamma^2(1-\beta)(n_1^2 + n_2^2) - 2\gamma(1-2\beta)n_1 n_2}{\beta\left(T_2\left(1+\lambda\right) + \left(4n_1 n_2 - \gamma(n_1^2 + n_2^2)\right)\gamma\right)}; \tag{18}$$

where $\lambda = \frac{T_1}{T_2}$. Region 1 prefers regime [F] to [R] if, and only if $p_1 > p_1^1(\gamma, \lambda)$ and region 2 prefers regime [F] to [R] if, and only if $p_1 < p_1^2(\gamma, \lambda)$. The two functions, $p_1^1(\gamma, \lambda)$ and $p_1^2(\gamma, \lambda)$, are drawn in Figure 1 in $(\gamma, p_1)$ space for a given value of $\lambda$. We can identify two main areas. In area 1, regime [R] Pareto dominates [F], while in area 2, the opposite

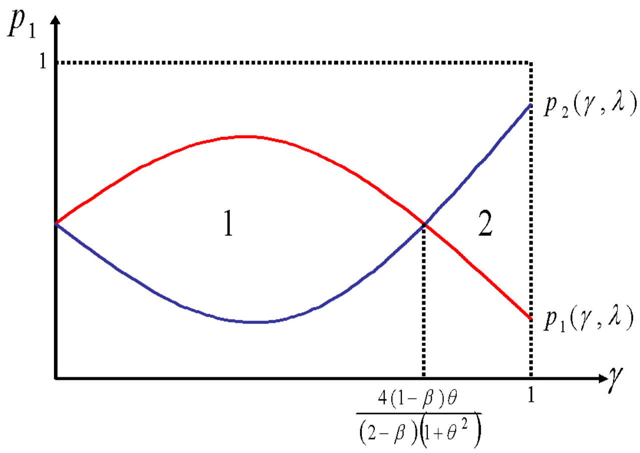


Figure 1: Welfare analysis with super-additive political costs and redistribution. Note: Area 1: regionalism is Pareto superior; Area 2: federalism is Pareto superior.

is true. Outside these areas, the redistribution effect is sufficiently strong to make one of the regions better off at the expense of the other. An increase in $\lambda$ (which makes region 1 relatively richer) shifts $p_1^1(\gamma,\lambda)$ and $p_1^2(\gamma,\lambda)$ up. This makes it less likely that region 1 and more likely that region 2 benefits from federalism.

### 5.5 Positive Externalities

The situation with positive externalities is very different. In this case, political costs are sub-additive and Proposition 1 shows that there exists multiple equilibria under federalism. Along all equilibrium paths, the aggregate utility of the two regions is, however, uniquely determined by

$$n_1x_{1t} + n_2x_{2t} = \beta T + \left(\left(n_1^2+n_2^2\right)\left(1+\gamma^2\right) - 4\gamma n_1 n_2\right). \tag{19}$$

Moreover, the lower bounds on the utility of region $i$ is $x_i \geq \frac{1}{n_i}\left(\beta p_i T + n_i^2\left(1+\gamma^2\right)\right)$. The federal politician collects the rent $(1-\beta)T$ each period. This is the same as the total rent collected by the two regional politicians: there is no rent effect with sub-additive costs. In the absence, then, of redistribution effects (i.e., for $p_1 = \frac{1}{2}$, $T_1 = T_2$), one might expect that centralization is always a Pareto improvement because externalities are internalized. The next proposition, which assumes away the redistribution effect, shows that this is not the case. To state this surprising result, we denote region 1's share of total utility by $\varphi$

and index equilibrium allocations by $\varphi$. We also, for simplicity and without loss of any important insights, assume that $n_1 = n_2 = 1$. In this case $\varphi = \frac{x_{1t}}{\beta T + 2(1-\gamma)^2}$.

**Proposition 5 (Positive externalities $\gamma < 0$)** *Assume that $p_1 = \frac{1}{2}$, $T_1 = T_2$, $n_1 = n_2 = 1$ and $\gamma < 0$. Then there exists a $\overline{\varphi} \in (0, \frac{1}{2})$ such that for $\varphi \in [\overline{\varphi}, 1 - \overline{\varphi}]$, regime [F] Pareto dominates regime [R].*

**Corollary 1** *There exist equilibria in which regime [F] does not Pareto dominate regime [R].*

**Corollary 2** *Regime [R] can never Pareto dominate regime [F].*

**Proof.** Use Proposition 1 to calculate the "best" and the "worst" equilibrium allocation for each region in regime [F]:

$$\begin{aligned} x_{it}^{\max} &= p_i \beta T + 1 + \gamma^2 - 4\gamma \\ x_{it}^{\min} &= \beta p_i T + 1 + \gamma^2 \end{aligned}$$

for $i = 1, 2$. Region $i$ is better off under [R] than under [F] in the "worst" equilibrium if

$$x_{it}^{\min} - x_{it}^{R} = \widehat{\Delta}_i + \gamma^2 + 2\gamma < 0,$$

and is better off under [F] than under [R] in the "best" equilibrium if

$$x_{it}^{\max} - x_{it}^{R} = \widehat{\Delta}_i + \gamma^2 - 2\gamma > 0,$$

where $\widehat{\Delta}_i$ is defined in equation (16). For $p_1 = \frac{1}{2}$ and $T_1 = T_2$, we see that $x_{it}^{\min} - x_{it}^{R} < 0$ and $x_{it}^{\max} - x_{it}^{R} > 0$ for $i = 1, 2$. Thus, at least one region prefers [F] to [R]. Along any equilibrium path

$$x_{1t} + x_{2t} = \beta T + 2(1-\gamma)^2. \tag{20}$$

Define the share of total utility obtained by region $i$ by $\varphi_i$. Region $i$ is then indifferent between the two regimes for

$$\varphi_i = \frac{\beta T_i + 1 - 2\gamma}{\beta T + 2(1-\gamma)^2} \equiv \overline{\varphi}_i.$$

Note that for $T_1 = T_2$, $0 < \overline{\varphi}_1 < 1 - \overline{\varphi}_2 < 1$ and that $\overline{\varphi}_1 = \overline{\varphi}_2 < \frac{1}{2}$. Since $\sum_i \varphi_i = 1$, we conclude that for $\varphi_1 \in (\overline{\varphi}_1, 1 - \overline{\varphi}_2)$ both regions prefer [F] to [R]. Substitution of $\varphi_1 = \varphi$ and $\overline{\varphi}_1 = \overline{\varphi}$ yields the proposition ■

In the absence of the rent and redistribution effect, it is surprising that centralization is not always efficient. The reason is that the selection of equilibria re-opens the door to redistribution. For example, in the "worst" equilibrium under regime [F], the external benefit captured by region 1 is $\gamma^2$. This is less that what it "receives" under regime [R],

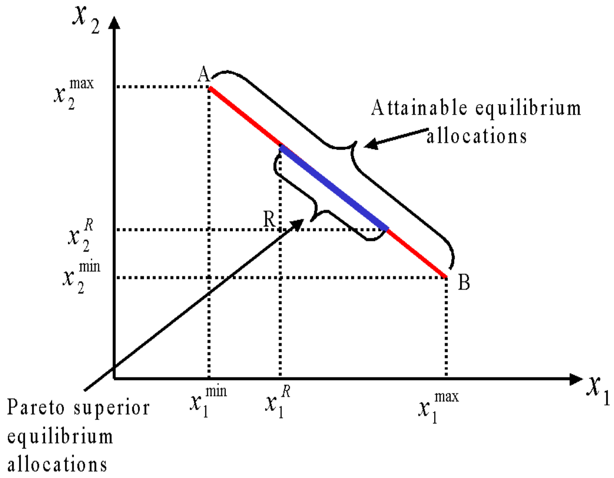


Figure 2: Utility allocations with sub-additive political costs.

namely $-2\gamma$. The point is that in this equilibrium most of the internalization benefits are captured by region 2. As a consequence, region 1 is better off with the "external" benefits unintentionally bestowed upon it by region 2 under regionalism. This – and the proposition more generally – is illustrated in Figure 2. It shows the utility allocations attainable in region [F]. The segment $A-B$, indicated with boldface on the utility frontier, contains the equilibrium allocations that Pareto dominate regionalism (represented by point $R$). The remaining allocations on the frontier cannot be Pareto ranked. In these cases, contrary to the Decentralization Theorem, it is not efficient to centralize despite the absence of regional differences in tastes and the presence of positive externalities.

Another important implication is that regionalism cannot ever Pareto dominate federalism. This stands in sharp contrast to the case with negative externalities discussed above. The reason for this is that the federal politician is unable to extract more rent than the collective of regional politicians: the rent effect is only present with negative externalities. The intuition is that positive externalities makes it "cheaper" for the federal politician to satisfy the demands of all voters. This, in turn, enables them to demand more without risking being ignored.

# 6 Integration and Disintegration of Federations

Our analysis can speak directly to the forces that create and destroy federations. Logically, fiscal integration among otherwise independent regions must either be fully voluntary or forced upon reluctant regions by more powerful neighbors. Voluntary integration leads to a stable fiscal structure, while forced integration is unstable with a tendency to break down over time. Leading examples of the former include Switzerland, where the independent

Cantons in 1848 agreed to form a federation, and the United States in the formative years. As an example of the latter one may point to the United Kingdom. England has traditionally played the leading role within the Union, but over the years her power has gradually been curtailed, first, by Ireland seceding in 1921, and more recently by devolving powers to Wales and Scotland.

## 6.1 A Voter-centric Perspective

In the absence of strong externalities, federations are vehicles for redistribution and must be forced in one way or the other and are likely to be unstable. Voluntary formation of a federation, then, as in Oates (1972), requires strong externalities. Our analysis suggests that the logic leading to a stable federation differs significantly depending on whether externalities are predominately negative or positive. In practice, two of the most fundamental reasons for the formation of a federation are defense and trade. The former is an example of a positive externality insofar as joining forces can deliver more protection at the same or lower cost. The latter is an example of a negative externality insofar as each region in the absence of a trade agreement opts for beggar-thy-neighbor type trade policy. Our model, then, suggests that the forces that pull or push regions apart act very differently depending on the relative strength of the underlying defense and trade motive.

With negative externalities arising from, say, a trade motive, the strength of the externality *is* the key driver of integration: a strong negative externality makes all regions favor a federation and accept the loss of accountability that comes with it. Interestingly, federations are more likely to form among regions of different sizes than among equal-sized regions. This appears counter-intuitive at first, as one might expect that a small region fears being dominated by a large counter-part if it joins a federation. But the reason simply is that it is harder for the federal politician to extract rent in an asymmetric than in a symmetric federation. A recent example of this is the wave of voluntary municipality mergers in Finland where many small municipalities were willing to merge with large neighbors (Saarimaa and Tukiainen, 2015).

In contrast, a strong positive externality, arising from, say, a defense motive, is not sufficient to create a stable federation. The reason is that turnout uncertainty opens up the door for redistribution through equilibrium selection, even among otherwise symmetric regions. Depending on the distributional outcome, some regions may lose out and veto integration even when externalities are strong or, if they are already in the federation, attempt to secede.

## 6.2 A Politician-centric Perspective

We have so far ignored the interests of the regional politicians when making regime comparisons. In practice, however, regional politicians may have disproportionate influence on integration decisions and be able to supersede the interests of the voters they represent.

To consider this possibility, suppose that the fiscal architecture is decided by consent of the two regional politicians, irrespective of what voters want, and that each perceives that there is a probability $q_i$, with $\sum_i q_i = 1$, that he will become the "federal politician".[14] Given that, centralization cannot be voluntary if the externality is (weakly) positive. The reason is that the total rent that can be extracted by the federal politician is equal to the sum of the rents extracted by the two regional politicians. As a consequence, one of them will lose by agreeing to a federation. With negative externalities, the situation is very different. In this case, the aggregate rent that is extracted by the federal politician is greater than the sum of the rents extracted by the two regional politicians. This implies that federalism may be preferred to regionalism by all regional politicians, in particularly so when the negative externality is strong. This is not because they have any interest in internalizing these externalities, but because they can extract extra rent from voters in this case. Combined with Proposition 4, this provides a very strong positive prediction: in the presence of strong negative externalities, e.g., generated by uncoordinated trade policy, all voters and all politicians support a federation. The same is not true for positive externalities, e.g., generated by increasing returns to defense.

# 7 Conclusion and Discussion

We show how governance uncertainty – exemplified by turnout uncertainty – affects the trade off between internalization of externalities and political accountability in the design of the fiscal state. We highlight a novel asymmetry between positive and negative externalities. We show that centralization only weakens political accountability in the presence of negative externalities. This yields two new insights. First, regionalism may Pareto dominate federalism with negative externalities. This is not possible with positive externalities. Second, with positive externalities, federalism may not be Pareto efficient despite the fact that policy can be tailored to regional taste differences and federalism internalizes the positive externality.

These results, however, ignore three potentially important issues. Firstly, we do not consider the benefits of yardstick competition enabled by regionalism. As shown by Besley and Case (1995), voters can use information about what is happening in other jurisdictions to overcome political agency problems. This forces politicians into (yardstick) competition in which they care about what other politicians do. This benefit is, of course, lost if fiscal decisions are centralized. It would be interesting in future research to study the possibility of yardstick competition. Another limitation of the analysis is the focus on the case with two regions. It would be of interest in future work to study more regions. It would also be interesting to consider the effects of inter-regional migration. As show by Kessler *et al.*

[14] If a regional politician has no chance of becoming the federal politician, he will veto any attempt at creating a federation since he will lose his rent. Supplementary online appendix III contains the formal analysis.

(2011) this can have important implications for how citizens would vote in a referendum on the design of the fiscal state.

Second, we ignore tax distortions and collection costs. In reality, taxation creates deadweight cost and tax collection is costly. Our model can be extended to accommodate endogenous distortionary taxation. All our results depend on whether the political cost functions are sub- or super-additive. In the model presented above, this is determined solely by the direction of the externality. This is also true in an economy with distortionary taxation.[15] The assumption of lump sum taxes and transfers greatly simplify our analysis, but it does not do so at the expense of important insights.

Third, a more critical issue is that politicians in our model are perfect substitutes for each other. Voters are, therefore, indifferent between any two candidates at each election. Accordingly, they do not have a strict incentive to vote as they promised. In reality, politicians differ in many ways and voters would like to select politicians with desirable characteristics. This gives them a positive reason to keep reelection promises. This, however, introduces an adverse selection problem and voters need, in general, to resort to more complicated voting strategies along the lines of Banks and Sundaram (1993, 1998). Hindriks and Lockwood (2009) consider an insightful two-period political agency with both moral hazard and adverse selection. They show that centralization reduces the extent of electoral accountability, as the rent seeking federal politician can target good behavior at a "minimum winning coalition" of regions. At the same time, this makes it more profitable for bad incumbents to pool with good ones, thus increasing the probability of electoral discipline occurring at all. In our model, the strategic consensus result shots down the first of these effects. The second effect will, we conjecture, operate if the total rent that can be extracted in the federation is bigger than the rent that can be extracted in the regions. This is the case with negative externalities, but not with positive ones.

We may conclude by noting that the general framework and the characterization results in Aidt and Dutta (2004) can be adopted to many other applications than the one studied here. This includes other public finance problems, e.g., the choice between targeted transfers and universal public goods (Aidt and Dutta, 2009). Applications in many other fields, including corporate governance and labor economics, also come to mind.

## References

Acemoglu, D., and J.A. Robinson (2000) "Why did the west extend the franchise? Democracy, inequality, and growth in historical perspective" *Quarterly Journal of Economics* **115**, 1167-1199.

Aidt, T.S., and J. Dutta (2004) "Strategic consensus" *Journal of Mathematical Economics* **40**, 227-245.

[15] In Supplementary online appendix IV, we introduce convex deadweight costs and show that the core properties of the political cost functions are preserved.

Aidt, T.S., and F. Magris (2006) "Capital taxation and electoral accountability" *European Journal of Political Economy* **22**, 277-291.
Aidt, T.S., and J. Dutta (2007) "Policy myopia and economic growth" *European Journal of Political Economy* **23**, 734-753.
Aidt, T.S., and J. Dutta (2009) "Strategic consensus and heterogenous voters: provision of public goods with random voter turnout" *Public Finance and Management* **9(2)**, 1-38.
Aldashev, G. (2014) "Voter turnout and political rents" *Journal of Public Economic Theory* **17**, 528-552.
Arnold, F., and R. Freier (2015) "The partisan effects of voter turnout. How conservatives profit from rainy election days" Discussion Paper 1463, DIW Berlin, German Institute for Economic Research. Berlin, Germany.
Artés, J. (2014) "The rain in Spain: Turnout and partisan voting in Spanish elections" *European Journal of Political Economy* **34**, 126-141.
Banks, J., and R.K. Sundaram (1993) "Adverse Selection and Moral Hazard in a Repeated Election Model" in *Political Economy: Institutions, Information, Competition and Representation* by W. Barnett, N. Schofield, and M. Hinch, Eds., Cambridge University Press: Cambridge, U.K., 295-311.
Banks, J., and R.K. Sundaram (1998) "Optimal retention in agency problems" *Journal of Economic Theory* **82**, 293-323.
Besley, T., and A. Case (1995) "Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition" *American Economic Review* **85**, 25-45.
Besley. T., and S. Coate (2003) "Centralized versus decentralized provision of local public goods: A political economy approach" *Journal of Public Economics* **87**, 2611-2637.
Bordignon, M., L. Colombo, and U. Galmarini (2008) "Fiscal federalism and lobbying" *Journal of Public Economics* **92**, 2288-2301.
Cheikbossian, G. (2000) "Federalism, distributive politics and representative democracy" *Economics of Governance* **1**, 105-122.
Coate, S., and S. Morris (1999) "Policy persistence" *American Economic Review* **89**, 1327-1336.
Crémer, J., and T.R. Palfrey (1996) "In or out?: Centralization by majority vote" *European Economic Review* **40**, 43-60.
Crémer, J., and T.R. Palfrey (2000) "Federal mandates by popular demand" *Journal of Political Economy* **100**, 905-927.
Dhillon, A., and S. Peralta (2002) "Economic theories of voter turnout" *Economic Journal* **112**, F332-F352.
Dixit, A., and J. Londregan (1998) "Fiscal federalism and redistributive politics" *Journal of Public Economics* **68**, 153-180.
Dur, R., and H. Roelfsema (2005) "Why does centralisation fail to internalise policy externalities?" *Public Choice* **122**, 395-416.
Edwards, J., and M. Keen (1996) "Tax competition and Leviathan" *European Economic*

*Review* **40**, 113-134.
Ferejohn, J., (1986) "Incumbent performance and electoral control" *Public Choice* **50**, 5-25.
Feld, L.P., C.A. Schaltegger, and J. Schnellenbach (2008) "On government centralization and fiscal referendums" *European Economic Review* **52**, 611–645.
Galletta, S., and M. Jametti (2015) "How to tame two Leviathans? Revisiting the effect of direct democracy on local public expenditure in a federation" *European Journal of Political Economy* **39**, 82-93.
Gomez, B.T., T.G. Hansford, and G.A. Krause (2007) "The Republicans should pray for rain: Weather, turnout, and voting in U.S. Presidential Elections" *Journal of Politics* **69**, 647–661.
Harstad, B. (2007) "Harmonization and side payments in political cooperation" *American Economic Review* **97**, 871-889.
Hindriks, J., and B. Lockwood (2009) "Centralization and electoral accountability: Incentives, separation and voter welfare" *European Journal of Political Economy* **25**, 385-397.
Inman, R. P., and D. L. Rubinfeld (1997) "The Political Economy of Federalism" in *Perspectives on Public Choice: a Handbook* by D. Mueller, Ed., Cambridge University Press: Cambridge, U.K.
Kessler, A.S., N.A. Hansen, and C. Lessmann (2011) "Interregional redistribution and mobility in federations: A positive approach" *Review of Economic Studies* **78**, 1345-1378.
Lind, J.T. (2014) "Rainy day politics - An instrumental variables approach to the effect of parties on political outcomes" CESifo Working Paper Series 4911, CESifo Group, Munich.
Lockwood, B. (2002) "Distributive politics and the costs of centralization" *Review of Economic Studies* **69**, 313-337.
Lockwood, B. (2004) "Decentralization via federal and unitary referenda" *Journal of Public Economic Theory* **6**, 79–108.
Lockwood, B. (2006) "Fiscal Decentralization: A Political Economy Perspective" in *The Handbook of Fiscal Federalism* by E. Ahmad and G. Brosio, Eds., Edward Elgar.
Lockwood, B. (2008) "Voting, lobbying, and the decentralization theorem" *Economics and Politics* **20**, 416-431.
Lülfesmann, C. (2002) "Central governance or subsidiarity: A property rights approach to federalism" *European Economic Review* **46**, 1379-1397.
Lülfesmann, C., Kessler A.S., and G. Myers (2015) "The architecture of federations: Constitutions, bargaining, and moral hazard" *Journal of Public Economics* **124**, 18-29.
Oates, W.E. (1972) *Fiscal Federalism.* Harcourt Brace: New York.
Olson, M. (1965) *The Logic of Collective Action.* Harvard University Press: Cambridge, Mass.
Persson, T., G. Roland, and G. Tabellini (1997) "Separation of powers and political accountability" *Quarterly Journal of Economics* **112**, 1163-1202.

Redoano, M. (2010) “Does centralization affect the number and size of lobbies?” *Journal of Public Economic Theory* **12**, 407–435.
Roemer, J.E. (1998) “Why the poor do not expropriate the rich: An old argument in new garb” *Journal of Public Economics* **70**, 399-424.
Saarimaa, T., and Tukiainen, J. (2015) “Common pool problems in voluntary municipal mergers” *European Journal of Political Economy* **38**, 140-152.
Schnellenbach, J., L.P. Feld, and C.A. Schaltegger (2010). “The impact of referendums on the centralisation of public goods provision: A political economy approach” *Economics of Governance* **11**, 3-26.
Seabright, P. (1996) “Accountability and decentralization: An incomplete contacts model” *European Economic Review* **40**, 61-90.
Tommasi, M. and F. Weinschelbaum (2007) “Centralization vs. decentralization: A principal-agent analysis” *Journal of Public Economic Theory* **9**, 369-389.

# Supplementary Online Appendices
# to
# Fiscal Federalism and Electoral Accountability

Toke S. Aidt[16]
Jayasri Dutta[17]

## 8 Supplementary Appendix I

This appendix reproduces and generalizes slightly Theorem 1 and Proposition 1 and 2 from Aidt and Dutta (2004). To prove the main characterization results, we must first prove that all equilibria exhibits strategic consensus. We begin by introducing some extra notation. Let $\underline{x}_{-i}(x_i)$ be the level of utility group $-i$ obtains when the politician provides utility level $x_i$ to group $i$ at minimum cost without regard to the welfare of group $-i$. Then, the following is true:

$$\begin{array}{lrcl} [B_1] & C(x_{1t}, \underline{x}_{2t}(x_{1t})) & = & C_1(x_{1t}) \\ [B_2] & C(\underline{x}_{1t}(x_{2t}), x_{2t}) & = & C_2(x_{2t}). \end{array}$$

A special case of this is when $C(x_{1t}, 0) = C_1(x_{1t})$ and $C(0, x_{2t}) = C_2(x_{2t})$ as assumed in Aidt and Dutta (2004). We also assume that $C(0,0) = C_i(0) = 0$, $i = 1, 2$. We can now state the main Theorem.

**Theorem 1 (Strategic Consensus)** *Assume that* $\beta \in (0,1)$. *Let* $x_t = (x_{1t}, x_{2t})$ *be a pair of performance standards set by the two groups of voters for period* $t$ *and define* $X = \{x_t\}_{t=0}^{\infty}$ *as a sequence of such standards. Let* $a_t^*$ *be the action implemented by the politician in period* $t$*; define* $V_{0t}(a_t)$ *as the politician's payoff.*

1. *A stationary subgame perfect Nash equilibrium exists.*
2. *Suppose* (**M**), (**K**) *and* (**C**$^+$) *hold. Along any stationary equilibrium path,* $X$ *satisfies*

$$(\mathbf{SC}^+) \quad V_{0t}(11) = V_{0t}(00) \geq \max\{V_{0t}(10), V_{0t}(01)\}.$$

*Any sequence* $X$ *satisfying* (**SC**$^+$) *is a stationary subgame perfect Nash equilibrium in performance standards. Along any stationary equilibrium path, the politician chooses* $a_t^* = (11)$ *at every* $t$ *and he is reelected for sure.*

[16] Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, Tel.: +44 1223 335231. E-mail: Toke.Aidt@econ.cam.ac.uk.
[17] Department of Economics, University of Birmingham. E-mail: J.Dutta@bham.ac.uk.

*3. Suppose* $(\mathbf{M})$, $(\mathbf{K})$ *and* $(\mathbf{C}^{-})$ *hold. Along any stationary equilibrium path, X satisfies*

$$(\mathbf{SC}^{-}) \quad V_{0t}(11) = V_{0t}(10) = V_{0t}(01) > V_{0t}(00).$$

*Any sequence X satisfying* $(\mathbf{SC}^{-})$ *is a stationary subgame perfect Nash equilibrium in performance standards. Along any stationary equilibrium path, the politician chooses* $a_t^* = (11)$ *at every t and he is reelected for sure.*

**Corollary 3** *Every stationary subgame perfect Nash equilibrium path displays strategic consensus at each t.*

We prove the Theorem with a series of Lemmas. We begin by introducing some notation. Denote for each action $a_t \in A$, the politician's payoff by $V_{0t}(a_t)$ and write

$$V_{0t}(00) = T; \tag{21}$$

$$V_{0t}(10) = T - C_1(x_{1t}) + p_1 \beta V_{t+1}; \tag{22}$$

$$V_{0t}(01) = T - C_2(x_{2t}) + p_2 \beta V_{t+1}; \tag{23}$$

$$V_{0t}(11) = T - C(x_{1t}, x_{2t}) + \beta V_{t+1}. \tag{24}$$

where $V_{t+1} > 0$ is the value of being reelected at time $t+1$. Note that the politician is only reelected with some probability ($p_1$ or $p_2$) if he chooses to be "partisan" and to satisfy one of the standards only.

Now, suppose, in some period $t$, that the two groups of voters announce the standards $x_t = \{x_{1t}, x_{2t}\}$. Given these standards, the politician chooses an action from the set $\{a_t \in A : \arg\max_{a_t \in A} V_{0t}(a_t)\}$. If the politician is indifferent between two or more actions in this set, he chooses the action that maximizes reelection chances. This is anticipated by the two groups of voters when they, simultaneously, set their standards at the beginning of the period. With these preliminary remarks we can state the first Lemma.

**Lemma 1** *Suppose that* $[\mathbf{M}]$ *and* $[\mathbf{K}]$ *hold. If the performance standards* $x_t = \{x_{1t}, x_{2t}\}$ *constitute a subgame perfect Nash equilibrium at time t, then* $x_t$ *must satisfy*

$$(\mathbf{E0}) \quad V_{0t}(11) \geq \max\{V_{0t}(10), V_{0t}(01), V_{0t}(00)\}.$$

**Proof:** We argue by contradiction. Suppose that $\widetilde{x}_t = \{\widetilde{x}_{1t}, \widetilde{x}_{2t}\}$ constitutes a stationary subgame perfect Nash equilibrium in performance standards *and* that

$$V_{0t}(11) < \max\{V_{0t}(10), V_{0t}(01), V_{0t}(00)\}$$

at time $t$. There are four separate cases to consider. We show in each case that at least one of the two groups of voters has an incentive to deviate from $\widetilde{x}_t$, leading to the required contradiction.

1. Suppose that

$$V_{0t}(10) = \max\{V_{0t}(10), V_{0t}(01), V_{0t}(00)\} > V_{0t}(11)$$

or that

$$V_{0t}(10) = V_{0t}(00) = \max\{V_{0t}(10), V_{0t}(01), V_{0t}(00)\} > V_{0t}(11).$$

Rewrite equations (22) and (24) to get

$$V_{0t}(10) - V_{0t}(11) = C(x_{1t}, x_{2t}) - C_1(x_{1t}) - p_2\beta V_{t+1}.$$

By [**M**] and [**K**], property $[B_1]$ implies that there must exist a $x'_{2t} > \underline{x}_{2t}$ such that

$$C(\widetilde{x}_{1t}, x'_{2t}) - C_1(\widetilde{x}_{1t}) - p_2\beta V_{t+1} < 0.$$

This implies that group 2 can gain by announcing the standard $x'_{2t}$ instead of $\widetilde{x}_{2t}$.

2. Suppose, instead, that

$$V_{0t}(01) = \max\{V_{0t}(10), V_{0t}(01), V_{0t}(00)\} > V_{0t}(11)$$

or that

$$V_{0t}(01) = V_{0t}(00) = \max\{V_{0t}(10), V_{0t}(01), V_{0t}(00)\} > V_{0t}(11).$$

By an argument similar to the previous case, there must exist a $x'_{1t} > \underline{x}_{1t}$ such that

$$C(x'_{1t}, \widetilde{x}_{2t}) - C_2(\widetilde{x}_{2t}) - p_1\beta V_{t+1} < 0.$$

This implies that group 1 can gain by announcing the standard $x'_{1t}$ instead of $\widetilde{x}_{1t}$.

3. Suppose that

$$V_{0t}(00) = \max\{V_{0t}(10), V_{0t}(01), V_{0t}(00)\} > V_{0t}(11).$$

Rewrite equations (21) and (22) to get

$$V_{0t}(00) - V_{0t}(10) = C(x_{1t}, \underline{x}_{2t}) - p_1\beta V_{t+1}.$$

By [**M**] and [**K**] there must exist a $x''_{1t} > 0$ such that

$$C(x''_{1t}, \underline{x}_{2t}) - p_1\beta V_{t+1} < 0.$$

This implies that group 1 can at least gain $x''_{1t} > 0$ by announcing the standard $x''_{1t}$ instead of $\widetilde{x}_{1t}$. A similar argument can be made for group 2.

4. Suppose that

$$V_{0t}(10) = V_{0t}(01) = \max\{V_{0t}(10), V_{0t}(01), V_{0t}(00)\} > V_{0t}(11)$$

or

$$V_{0t}(10) = V_{0t}(01) = V_{0t}(00) > V_{0t}(11).$$

We need to consider two sub-cases. First, suppose the politician chooses $a_t = (10)$. We can then repeat the argument from case 1 to show that there exists a profitable deviation for group 2. Second, suppose the politician chooses $a_t = (01)$. We can then repeat the argument from case 2 to show that there exists a profitable deviation for group 1■

**Lemma 2** *A pair of performance standards* $x_t = (x_{1t}, x_{2t})$ *is a stationary subgame perfect Nash equilibrium at time t if, and only if*

$$(\mathbf{E1}) \quad V_{0t}(11) = \max\{V_{0t}(01), V_{0t}(00)\};$$

$$(\mathbf{E2}) \quad V_{0t}(11) = \max\{V_{0t}(10), V_{0t}(00)\}.$$

**Proof:** Suppose that $p_1 \geq \frac{1}{2}$. The per-period payoff of group 1 is

$$\begin{aligned} u_{1t} &= x_{1t} \quad \text{if } \max\{V_{0t}(11), V_{0t}(10)\} \geq \max\{V_{0t}(01), V_{0t}(00)\}; \\ u_{1t} &= \underline{x}_{1t} \quad \text{otherwise.} \end{aligned}$$

The per-period payoff of group 2 is

$$\begin{aligned} u_{2t} &= x_{2t} \quad \text{if } \left\{ \begin{array}{l} V_{0t}(11) \geq \max\{V_{0t}(10), V_{0t}(00), V_{0t}(01)\} \\ V_{0t}(01) > \max\{V_{0t}(10), V_{0t}(00), V_{0t}(11)\} \\ V_{0t}(01) = V_{0t}(00) > \max\{V_{0t}(10), V_{0t}(11)\} \end{array} \right. ; \\ u_{2t} &= \underline{x}_{2t} \quad \text{otherwise.} \end{aligned}$$

Recall that $C(x_{1t}, x_{2t})$ and $C_i(x_{it})$ are monotonically increasing in their arguments by [**M**]. Suppose that $\widetilde{x}_t$ is a (stationary subgame perfect Nash) equilibrium. Then, by Lemma 1, (**E0**) is satisfied by $\widetilde{x}_t$. It follows that the payoff of group 1 is maximized by the standard, $x_{1t}$, that satisfies **(E1)**, and that the payoff of group 2 is maximized by the standard, $x_{2t}$, that satisfies **(E2).** Finally, notice that if **(E1)** and **(E2)** are satisfied by a set of performance standards at time $t$, then these standards constitute a stationary subgame perfect Nash Equilibrium. This completes the proof for the case with $p_1 \geq \frac{1}{2}$. The proof for the case where $p_1 < \frac{1}{2}$ is similar and is omitted■

The following two Lemmas explore the implications of assumptions ($\mathbf{C}^+$) and ($\mathbf{C}^-$), respectively.

**Lemma 3** *Conditions* (**E1**), (**E2**), *and* ($\mathbf{C}^+$) *hold at time t if, and only if*

$$V_{0t}(11) = V_{0t}(00) \geq \max\{V_{0t}(10), V_{0t}(01)\}.$$

**Proof:** Note that **($\mathbf{C}^+$)** implies that

$$(\mathbf{C0}^+) \quad V_{0t}(11) + V_{0t}(00) \geq V_{0t}(10) + V_{0t}(01)$$

at any $t$. We prove the Lemma by contradiction. Suppose $V_{0t}(11) > V_{0t}(00)$. Condition **(E2)** implies that

$$V_{0t}(11) = V_{0t}(10).$$

Substitute into **($\mathbf{C0}^+$)** to get that

$$V_{0t}(00) \geq V_{0t}(01).$$

Combing this with **(E1)** yields

$$V_{0t}(11) \leq V_{0t}(00).$$

This is a contradiction, so $V_{0t}(11)$ cannot be greater than $V_{0t}(00)$. It follows directly from (**E1**) that $V_{0t}(11)$ cannot be smaller than $V_{0t}(00)$. Finally, $V_{0t}(11) = V_{0t}(00)$ is compatible with **($\mathbf{C0}^+$)**, **(E1),** and **(E2)** only if $V_{0t}(10) \leq V_{0t}(00)$ and $V_{0t}(01) \leq V_{0t}(00)$■

The next Lemma considers the case of super-additive costs.

**Lemma 4** *Conditions* (**E1**), (**E2**), *and* ($\mathbf{C}^-$) *hold at time t if, and only if*

$$V_{0t}(11) = V_{0t}(10) = V_{0t}(01) > V_{0t}(00).$$

**Proof:** Note that **($\mathbf{C}^-$)** implies that

$$(\mathbf{C0}^-) \quad V_{0t}(11) + V_{0t}(00) < V_{0t}(10) + V_{0t}(01)$$

at any $t$. We begin by proving that $V_{0t}(11) = V_{0t}(10)$. This is done by contradiction. First, suppose that $V_{0t}(11) > V_{0t}(10)$. **(E2)** implies that

$$V_{0t}(00) > V_{0t}(10).$$

Combining this with **($\mathbf{C0}^-$)** implies that

$$V_{0t}(11) < V_{0t}(01).$$

However, **(E1)** implies that $V_{0t}(11) \geq V_{0t}(01)$. This is a contradiction, so $V_{0t}(11)$ cannot be greater than $V_{0t}(10)$. Second, suppose that $V_{0t}(10) > V_{0t}(11)$. **(E2)** implies that

$$V_{0t}(11) \geq V_{0t}(10);$$

This is a contradiction, and so $V_{0t}(10)$ cannot be greater than $V_{0t}(11)$. We conclude that $V_{0t}(10) = V_{0t}(11)$. The proof that $V_{0t}(01) = V_{0t}(11)$ is similar and omitted. Finally, $V_{0t}(11) = V_{0t}(10) = V_{0t}(01)$ is compatible with **($\mathbf{C0}^-$)** only if $V_{0t}(11) = V_{0t}(10) = V_{0t}(01) > V_{0t}(00)$■

The last Lemma establishes that a stationary subgame perfect equilibrium exists.

**Lemma 5** *A stationary subgame perfect equilibrium exists for all $\beta \in (0,1)$.*

**Proof:** Suppose first that **(C$^+$)** holds. In this case, a stationary equilibrium $\hat{x} = \{\hat{x}_1, \hat{x}_2\}$ satisfies **(SC$^+$)** at every $t$. This implies

$$\frac{T - C(\hat{x}_1, \hat{x}_2)}{1-\beta} = T; \tag{25}$$

and that

$$T \geq \max[\frac{T - C_1(\hat{x}_1)}{(1-p_1\beta)}, \frac{T - C_2(\hat{x}_2)}{(1-p_2\beta)}].$$

Equation (25) rewrites as

$$C(\hat{x}_1, \hat{x}_2) = \beta T.$$

Equilibrium levels of $\hat{x}$ satisfy

$$C(\hat{x}_1, \hat{x}_2) = \beta T \tag{26}$$

and

$$T \leq \min[\frac{C_1(\hat{x}_1)}{p_1\beta}, \frac{C_2(\hat{x}_2)}{p_2\beta}]. \tag{27}$$

It follows from conditions **(C$^+$), (M)** and (**K**) that there exists a solution to equations (26) and (27).

Suppose instead that **(C$^-$)** holds. In this case, a stationary equilibrium $\bar{x} = \{\bar{x}_1, \bar{x}_2\}$ must satisfy

$$\frac{T - C(\bar{x}_1, \bar{x}_2)}{1-\beta} = \frac{T - C_1(\bar{x}_1)}{1-p_1\beta} \tag{28}$$

and

$$\frac{T - C(\bar{x}_1, \bar{x}_2)}{1-\beta} = \frac{T - C_2(\bar{x}_2)}{1-p_2\beta} \tag{29}$$

along with

$$\frac{T - C(\bar{x}_1, \bar{x}_2)}{1-\beta} > T. \tag{30}$$

Define the quantities $x_{11}, x_{12}, x_{21}, x_{22}$ as solutions to equations (28) and (29) when $x_1 = \underline{x}_1$ and $x_2 = \underline{x}_2$ respectively. Then,

$$\frac{T - C(x_{11}, \underline{x}_2)}{1-\beta} = \frac{T - C(x_{11}, \underline{x}_2)}{1-p_1\beta};$$

$$\frac{T - C(\underline{x}_1, x_{21})}{1-\beta} = \frac{T}{1-p_1\beta}$$

$$\frac{T - C(x_{12}, \underline{x}_2)}{1-\beta} = \frac{T}{1-p_2\beta}$$

$$\frac{T - C(\underline{x}_1, x_{22})}{1-\beta} = \frac{T - C(\underline{x}_1, x_{22})}{1-p_2\beta}.$$

Solving these equations yields

$$T = C(x_{11}, \underline{x}_2) = C(\underline{x}_1, x_{22});$$

in addition,

$$x_{12} \leq x_{11};$$

and

$$x_{21} \leq x_{22}$$

whenever $\beta \in (0, 1)$. It follows that a solution to equations (28) and (29) exists.

Additionally, if $\bar{x}$ satisfies equations (28) and (29), then restriction (30) holds for all $\beta \in (0, 1)$. To show that an equilibrium exists for all $\beta \in (0, 1)$, rewrite equations (28) and (29) as

$$T\theta = (1 + \theta)C(\bar{x}_1, \bar{x}_2) - C(\bar{x}_1, \underline{x}_2);$$

$$T\eta = (1 + \eta)C(\bar{x}_1, \bar{x}_2) - C(\underline{x}_1, \bar{x}_2);$$

where $\theta = \frac{p_2\beta}{1-\beta}$ and $\eta = \frac{p_1\beta}{1-\beta}$. Adding the two equations, we obtain

$$(\theta + \eta)(T - C(\bar{x}_1, \bar{x}_2)) - C(\bar{x}_1, \bar{x}_2) = C(\bar{x}_1, \bar{x}_2) - C(\bar{x}_1, \underline{x}_2) - C(\underline{x}_1, \bar{x}_2) > 0 \tag{31}$$

by **[C-]**. Note that $\theta + \eta = \frac{\beta}{1-\beta}$ and that (31) implies

$$C(\bar{x}_1, \bar{x}_2) < \beta T$$

as assumed■

Based on this fundamental result, it is relatively straight forward to prove Propositions 1 and 2 in the main text.

**Proof of Proposition 1.** The value of reelection starting from any period $t$ is $V_{0t} = \max[V_{0t}(01), V_{0t}(10), V_{0t}(11), V_{0t}(00)]$. We obtain from Lemma 3 and equation (21) that $V_{0t} = V_{0t}(00) = T$. This implies that

$$V_{0t+1} = T.$$

We obtain, from Theorem 1 and equations (21) to (24), that

$$V_{0t}(11) = V_{0t}(00) \Rightarrow C(x_{1t}, x_{2t}) = \beta T;$$

and that

$$V_{0t}(00) \geq V_{0t}(10) \Rightarrow C_1(x_{1t}) \geq \beta p_1 T;$$

$$V_{0t}(00) \geq V_{0t}(01) \Rightarrow C_2(x_{2t}) \geq \beta p_2 T.$$

The politician's per period payoff is $T - C(x_{1t}, x_{2t}) = (1 - \beta)T$. Moreover, suppose $C(x_{1t}, x_{2t}) = C_1(x_{1t}) + C_2(x_{2t})$. Then, there exist a unique stationary equilibrium, $x_{1t} = x_1^*$ and $x_{2t} = x_2^*$, with

$$C_1(x_1^*) = \beta p_1 T;$$
$$C_2(x_2^*) = \beta p_2 T.$$

**Proof of Proposition 2.** The value of reelection starting from any period $t$ is $V_{0t} = \max[V_{0t}(01), V_{0t}(10), V_{0t}(11), V_{0t}(00)]$. We obtain from Lemma 4 that $V_{0t} = V_{0t}(11)$ for all $t$. Iterative, forward substitution, using equation (24), yields

$$V_{0t} = \sum_{k=0}^{\infty} \beta^k (T - C(x_{1t+k}, x_{2t+k})).$$

For sequences of stationary standards, we get

$$V_{0t} = V_{0t+1} = \frac{T - C(x_1, x_2)}{1 - \beta}.$$

Substituting for $V_{0t+1} = \frac{T - C(x_1, x_2)}{1-\beta}$, we get that

$$V_{0t}(11) = V_{0t}(10) \Rightarrow (\mathbf{SC}_1^-)$$

and

$$V_{0t}(11) = V_{0t}(01) \Rightarrow (\mathbf{SC}_2^-).$$

Finally, $V_{0t} = \frac{T - C(x_1, x_2)}{1-\beta}$ for all $t$ implies that the politician gets $T - C(x_1, x_2)$ per period. This is strictly greater than $(1 - \beta)T$ because $V_{0t}(11) > V_{0t}(00)$ by Lemma 4. For uniqueness, see Proposition 3 in Aidt and Dutta (2004)■

# 9 Supplementary Appendix II

In this appendix, we solves the planner's problem, derive the political cost function under federalism and give sufficient conditions for the case with $s_i > 0$ at the equilibrium.

**The planner's problem** We focus on the case where $s_i > 0$ at the optimum. We assume that the planner maximizes a utilitarian social welfare function with equal weight on each individual. Under [R], the social planner in region $i$ solves (we have omitted subscript $t$ for simplicity)

$$\max_{k_i, s_i} n_i (2k_i^{\frac{1}{2}} - 2\gamma k_{-i}^{\frac{1}{2}} + s_i) \tag{32}$$

subject to $n_i s_i + k_i \leq T_i$. For $s_i > 0$, we can substitute the constraint and find the optimal solution to be $k_i^{R^*} = n_i^2$. We note that $s_i^{R^*} > 0$ if $T_i > n_i^2$. Evaluating the social welfare function at the optima gives

$$SW_i^R = n_i(2n_i - 2\gamma n_{-i}) + T_i - n_i^2. \tag{33}$$

Under [F], the planner solves

$$\max_{k_1,k_2,s_1,s_2} n_1(2k_1^{\frac{1}{2}} - 2\gamma k_2^{\frac{1}{2}} + s_1) + n_2(2k_2^{\frac{1}{2}} - 2\gamma k_1^{\frac{1}{2}} + s_2) \tag{34}$$

subject to $k_1 + k_2 + n_1 s_1 + n_2 s_2 \leq T_1 + T_2$. For $s_i > 0$ all $i$, we can substitute the revenue constraint and find that $k_i^{F^*} = (n_i - n_{-i}\gamma)^2$ under the assumption that $\gamma \in (-1, \frac{n_2}{n_1})$. Since utility is linear in income and transfers, utilitarian social welfare is unaffected by the distribution of the tax revenues not spent on providing local public goods. To ensure that $s_i^{F^*} > 0$ for all $i$, we require $T_1 + T_2 > \sum_i (n_i - n_{-i}\gamma)^2$.

Social welfare under [F] cannot be lower than the sum of regional social welfare under [R] because the federal planner can decide to replicate what the two regional planners do. For $\gamma \neq 0$, it is strictly larger. This implies that it, in principle, is possible to design a transfer policy that will make both regions weakly better off under [F] than under [R]. To see how, let $T^n$ be the net federal tax revenue left over after paying for local public goods, i.e.,

$$T^n \equiv T_1 + T_2 - k_2^{F^*} - k_2^{F^*} = T_1 + T_2 - ((1+\gamma^2)(n_1^2 + n_2^2) - 4\gamma n_1 n_2). \tag{35}$$

Let $\sigma_i$ denote the share of the net surplus allocated to region $i$. For each region, we calculate the difference between regional social welfare under [F] and [R] as

$$\begin{aligned} \Delta_1^*(\sigma_1) &= \sigma_1 T^n + n_1 \left(n_1 - 2\gamma n_2 + 2\gamma^2 n_1\right) - T_1 & (36) \\ \Delta_2^*(\sigma_1) &= (1-\sigma_1) T^n + n_2 \left(n_2 - 2\gamma n_1 + 2\gamma^2 n_2\right) - T_2. & (37) \end{aligned}$$

Define the share of total revenue that makes region 1 indifferent between joining the utilitarian federation ($\Delta_1^* = 0$) as

$$\sigma_1^* = \frac{T_1 - n_1 \left(n_1 - 2\gamma n_2 + 2\gamma^2 n_1\right)}{T_1 + T_2 - ((1+\gamma^2)(n_1^2 + n_2^2) - 4\gamma n_1 n_2)}. \tag{38}$$

This sharing rule makes region 2 weakly better off from joining the federation, i.e.,

$$\Delta_2^*(s_1^*) = \gamma^2 \left(n_1^2 + n_2^2\right) \geq 0 \tag{39}$$

and strictly better off if there is an externality. We need to check that $\sigma_1^* > 0$ such that both regions get a subsidy. That requires that

$$T_1 > n_1 \left(n_1 - 2\gamma n_2 + 2\gamma^2 n_1\right) > k_1^{F^*}. \tag{40}$$

We assume that this is the case and note that this is sufficient to ensure that it is possible to find a transfer policy that will make [F] Pareto superior to [R] in the presence of an externality.

**The political cost function under federalism** Suppose the politician wants to satisfy both regions. He, then, solves the following problem each period (we have omitted subscript $t$ for simplicity):

$$\min_{k_1,k_2,s_1,s_2} k_1 + k_2 + n_1 s_1 + n_2 s_2$$

subject to

$$\begin{aligned} x_1 &\leq 2k_1^{\frac{1}{2}} - 2\gamma k_2^{\frac{1}{2}} + s_1 \\ x_2 &\leq 2k_2^{\frac{1}{2}} - 2\gamma k_1^{\frac{1}{2}} + s_2 \end{aligned}$$

Under the assumption that $\gamma < \frac{n_2}{n_1}$, $k_1$ and $k_2$ are (weakly) positive at the optimum. In the text, we focus on the case in which both groups get transfers, but for completeness, we give the details for all four possible cases:

1. $s_1 > 0$, $s_2 > 0$
2. $s_1 = s_2 = 0$
3. $s_1 = 0$, $s_2 > 0$
4. $s_1 > 0$, $s_2 = 0$

Case 1: Substituting the two constraints, which must be binding at the optimum, into the objective function and taking the first derivatives with respect to $k_1$ and $k_2$ yields:

$$\begin{aligned} 1 - n_1 k_1^{\frac{-1}{2}} + n_2 \gamma k_1^{\frac{-1}{2}} &= 0 \\ 1 - n_2 k_2^{\frac{-1}{2}} + n_1 \gamma k_2^{\frac{-1}{2}} &= 0. \end{aligned}$$

Solving this, we get $k_1 = (n_1 - n_2\gamma)^2$ and $k_2 = (n_2 - n_1\gamma)^2$. The per capita transfers are

$$\begin{aligned} s_1 &= x_1 - 2(n_1 - n_2\gamma) + 2\gamma(n_2 - n_1\gamma) = x_1 - 2n_1\left(1+\gamma^2\right) + 4\gamma n_2 \\ s_2 &= x_2 - 2(n_2 - n_1\gamma) + 2\gamma(n_1 - n_2\gamma) = x_2 - 2n_2\left(1+\gamma^2\right) + 4\gamma n_1 \end{aligned}$$

Notice that $s_i > 0$ requires that $x_i > 2n_i(1+\gamma^2) - 4\gamma n_{-i}$. The political cost function is

$$\begin{aligned} C(x_1, x_2) = &\ (n_1 - n_2\gamma)^2 + (n_2 - n_1\gamma)^2 \\ &+ n_1\left(x_1 - 2n_1\left(1+\gamma^2\right) + 4\gamma n_2\right) \\ &+ n_2\left(x_2 - 2n_2\left(1+\gamma^2\right) + 4\gamma n_1\right) \end{aligned}$$

Case 2: This case applies for $x_1 \leq n_1(1+\gamma^2) - 2\gamma n_2$ and $x_2 \leq n_2(1+\gamma^2) - 2\gamma n_1$. We need to make a distinction between three sub-cases. Firstly, let $\gamma \geq 0$ and $\min\left\{\frac{x_1}{x_2}, \frac{x_2}{x_1}\right\} > -\gamma$

or $\gamma < 0$ and $\min\{x_1, x_2\} > 0$. Then, both constraints are binding and we can solve them to get the lowest spending level on the two local public goods that will generate the required utility levels:

$$\begin{aligned} k_1 &= \left(\frac{x_1 + \gamma x_2}{2(1-\gamma^2)}\right)^2 \\ k_2 &= \left(\frac{x_2 + \gamma x_1}{2(1-\gamma^2)}\right)^2 \end{aligned}$$

and the cost function is

$$C(x_1, x_2) = \left(\frac{x_1 + \gamma x_2}{2(1-\gamma^2)}\right)^2 + \left(\frac{x_2 + \gamma x_1}{2(1-\gamma^2)}\right)^2.$$

Notice that $C(0,0) = 0$. Secondly, suppose that the constraint for group 1 is not binding. First, if $\gamma \geq 0$, then $k_1 = 0$ and $k_2 = \left(\frac{x_2}{2}\right)^2$ and

$$C(x_1, x_2) = \left(\frac{x_2}{2}\right)^2 \quad \text{for } x_1 \leq -\gamma x_2.$$

Second, if $\gamma < 0$, the politician solves

$$\min k_1 + k_2$$

subject to

$$x_2 \leq 2k_2^{\frac{1}{2}} - 2\gamma k_1^{\frac{1}{2}}.$$

Letting $\upsilon$ be the multiplier on the constraint, we can write the first order conditions as

$$\begin{aligned} 1 - k_2^{-\frac{1}{2}}\upsilon &= 0 \\ 1 + \gamma k_1^{-\frac{1}{2}}\upsilon &= 0 \end{aligned}$$

Solving for $k_1$ and $k_2$ and substituting into the constraint yields

$$\upsilon = \frac{x_2}{2(1+\gamma^2)}$$

and we find that $k_2 = \left(\frac{x_2}{2(1+\gamma^2)}\right)^2$ and $k_1 = \left(\frac{\gamma x_2}{2(1+\gamma^2)}\right)^2$ for $x_2 \geq 0$. The political cost function is

$$\begin{aligned} C(x_1, x_2) &= \left(\frac{x_2}{2(1+\gamma^2)}\right)^2 + \left(\frac{\gamma x_2}{2(1+\gamma^2)}\right)^2 \\ &= \frac{x_2^2}{4(1+\gamma^2)} \text{ for } x_1 < 0. \end{aligned}$$

Third, suppose that the constraint for group 2 is not binding. By analogy, we get for $\gamma \geq 0$ that $k_2 = 0$ and $k_1 = \left(\frac{x_1}{2}\right)^2$ and

$$C(x_1, x_2) = \left(\frac{x_1}{2}\right)^2 \quad \text{for } x_2 \leq -\gamma x_1.$$

For $\gamma < 0$, we get

$$C(x_1, x_2) = \frac{x_1^2}{4(1+\gamma^2)} \text{ for } x_2 < 0.$$

Case 3: Substituting $s_2$ out from the beginning, we can write the Lagrangian function as

$$L = k_1 + k_2 + n_2 \left(x_2 - 2k_2^{\frac{1}{2}} + 2\gamma k_1^{\frac{1}{2}}\right) + \psi \left(x_1 - 2k_1^{\frac{1}{2}} + 2\gamma k_2^{\frac{1}{2}}\right)$$

where $\psi$ is a Lagrangian multiplier. We can calculate the first order conditions:

$$1 + n_2 \gamma k_1^{\frac{-1}{2}} - \psi k_1^{\frac{-1}{2}} = 0 \tag{41}$$
$$1 - n_2 k_2^{\frac{-1}{2}} + \psi \gamma k_2^{\frac{-1}{2}} = 0 \tag{42}$$
$$x_1 - 2k_1^{\frac{1}{2}} + 2\psi k_2^{\frac{1}{2}} = 0. \tag{43}$$

Solve equations (41) and (42) to get

$$\begin{aligned} k_1 &= (\psi - n_2\gamma)^2 \\ k_2 &= (n_2 - \gamma\psi)^2 . \end{aligned}$$

Substitute this in equation (43) and solve for $\psi$:

$$\psi = \max \left\{ \frac{x_1 + 4\gamma n_2}{2(1+\gamma^2)}, 0 \right\}.$$

Using this, we get that for $\psi > 0 \Leftrightarrow x_1 > 4\gamma n_2$

$$\begin{aligned} k_1 &= \left(\frac{x_1 + 4\gamma n_2}{2(1+\gamma^2)} - n_2\gamma\right)^2 = \left(\frac{x_1 + 2\gamma n_2(1-\gamma^2)}{2(1+\gamma^2)}\right)^2 \\ k_2 &= \left(n_2 - \gamma \frac{x_1 + 4\gamma n_2}{2(1+\gamma^2)}\right)^2 = \left(\frac{2n_2(1-\gamma^2) - \gamma x_1}{2(1+\gamma^2)}\right)^2 \\ s_2 &= x_2 - 2\left(\frac{2n_2(1-\gamma^2) - \gamma x_1}{2(1+\gamma^2)}\right) + 2\gamma \left(\frac{x_1 + 2\gamma n_2(1-\gamma^2)}{2(1+\gamma^2)}\right) \end{aligned}$$

and the political cost function is

$$\begin{aligned} C(x_1, x_2) = &\left(\frac{x_1 + 2\gamma n_2(1-\gamma^2)}{2(1+\gamma^2)}\right)^2 + \left(\frac{2n_2(1-\gamma^2) - \gamma x_1}{2(1+\gamma^2)}\right)^2 \\ &+ n_2 \left(x_2 - 2\left(\frac{2n_2(1-\gamma^2) - \gamma x_1}{2(1+\gamma^2)}\right) + 2\gamma\left(\frac{x_1 + 2\gamma n_2(1-\gamma^2)}{2(1+\gamma^2)}\right)\right). \end{aligned}$$

We notice that $s_2 > 0$ requires that $x_2 > \frac{2\left(n_2\left(1-2\gamma^2+\gamma^4\right)-\gamma x_1\right)}{(1+\gamma^2)}$. For $x_1 \leq 4\gamma n_2$, $\psi = 0$. The first order conditions are

$$\begin{aligned} 1 + n_2 \gamma k_1^{\frac{-1}{2}} &\geq 0 \\ 1 - n_2 k_2^{\frac{-1}{2}} &\geq 0. \end{aligned}$$

For $\gamma \geq 0$, $k_1 = 0$ and $k_2 = (n_2)^2$ and the cost function is

$$C(x_1, x_2) = (n_2)^2 + n_2 (x_2 - 2n_2).$$

For $\gamma < 0$, $k_2 = (n_2)^2$ and $k_1 = (\gamma n_2)^2$ and the cost function is

$$C(x_1, x_2) = (n_2)^2 (1 + \gamma^2) + n_2 \left(x_2 - 2n_2(1 + \gamma^2)\right).$$

Case 4: This is similar to case 3. For $\psi > 0 \Leftrightarrow x_2 > 4\gamma n_1$, we get

$$\begin{aligned} k_1 &= \left(n_1 - \gamma \frac{x_2 + 4\gamma n_1}{2(1+\gamma^2)}\right)^2 = \left(\frac{2n_1(1-\gamma^2) - \gamma x_2}{2(1+\gamma^2)}\right)^2 \\ k_2 &= \left(\frac{x_2 + 4\gamma n_1}{2(1+\gamma^2)} - n_1 \gamma\right)^2 = \left(\frac{x_1 + 2\gamma n_1(1-\gamma^2)}{2(1+\gamma^2)}\right)^2 \\ s_1 &= x_1 - 2\left(\frac{2n_1(1-\gamma^2) - \gamma x_2}{2(1+\gamma^2)}\right) + 2\gamma\left(\frac{x_2 + 2\gamma n_1(1-\gamma^2)}{2(1+\gamma^2)}\right) \end{aligned}$$

and the political cost function is

$$\begin{aligned} C(x_1, x_2) &= \left(\frac{x_1 + 2\gamma n_1(1-\gamma^2)}{2(1+\gamma^2)}\right)^2 + \left(\frac{2n_1(1-\gamma^2) - \gamma x_2}{2(1+\gamma^2)}\right)^2 \\ &\quad + n_1\left(x_1 - 2\left(\frac{2n_1(1-\gamma^2) - \gamma x_2}{2(1+\gamma^2)}\right) + 2\gamma\left(\frac{x_2 + 2\gamma n_1(1-\gamma^2)}{2(1+\gamma^2)}\right)\right). \end{aligned}$$

We notice that $s_1 > 0$ requires that $x_1 > \frac{2\left(n_1\left(1-2\gamma^2+\gamma^4\right)-\gamma x_2\right)}{(1+\gamma^2)}$. For $x_2 \leq 4\gamma n_1$, $\psi = 0$. For $\gamma \geq 0$, $k_2 = 0$ and $k_1 = (n_1)^2$ and the cost function is

$$C(x_1, x_2) = (n_1)^2 + n_1 (x_1 - 2n_1).$$

For $\gamma < 0$, $k_1 = (n_1)^2$ and $k_2 = (\gamma n_1)^2$ and the cost function is

$$C(x_1, x_2) = (n_1)^2 (1 + \gamma^2) + n_1 \left(x_1 - 2n_1(1 + \gamma^2)\right).$$

Now, suppose that the politician will only try to satisfy the demands of group $i$. Again, we give details for the possible cases, but in the text, we focus on the case with positive transfers:

1. $s_i > 0$.

2. $s_i = 0$.

Case 1: The politician solves

$$\min_{k_i, k_{-i}, s_i} k_i + k_{-i} + n_i s_i$$

subject to $x_i \leq s_i + 2k_i^{\frac{1}{2}} - 2\gamma k_{-i}^{\frac{1}{2}}$. The optimal choice is

$$k_i = (n_i)^2 \text{ and } k_{-i} = 0 \text{ for } \gamma \geq 0$$

and

$$k_i = (n_i)^2 \text{ and } k_{-i} = (n_i \gamma)^2 \text{ for } \gamma < 0.$$

The transfer is

$$s_i = \begin{cases} x_i - 2n_i \text{ for } \gamma \geq 0 \\ x_i - 2n_i(1+\gamma^2) \text{ for } \gamma < 0 \end{cases}$$

The political cost function is

$$C_i(x_i) = \begin{cases} (n_i)^2 + n_i(x_i - 2n_i) \text{ for } \gamma \geq 0 \\ (1+\gamma^2)(n_i)^2 + n_i(x_i - 2n_i(1+\gamma^2)) \text{ for } \gamma < 0 \end{cases}.$$

Notice that for $\gamma \geq 0$, $x_i > 2n_i$ for $s_i > 0$ and for $\gamma < 0$, $x_i > n_i 2(1+\gamma^2)$ for $s_i > 0$.

Case 2: First, if $\gamma \geq 0$, then $k_i = \left(\frac{x_i}{2}\right)^2$ and $k_{-i} = 0$ and

$$C(x_i) = \left(\frac{x_i}{2}\right)^2 \text{ for } x_{-i} \leq -\gamma x_i.$$

Second, if $\gamma < 0$, the politician solves

$$\min_{k_i, k_{-i}} k_i + k_{-i}$$

subject to

$$x_i \leq 2k_i^{\frac{1}{2}} - 2\gamma k_{-i}^{\frac{1}{2}}.$$

Letting $\chi$ be the multiplier on the constraint, we can write the first order conditions as

$$\begin{aligned} 1 - k_i^{-\frac{1}{2}} \chi &= 0; \\ 1 + \gamma k_{-i}^{-\frac{1}{2}} \chi &= 0. \end{aligned}$$

Solving for $k_1$ and $k_2$ and substituting into the constraint yields

$$\chi = \frac{x_i}{2(1+\gamma^2)},$$

and we find that $k_i = \left(\frac{x_i}{2(1+\gamma^2)}\right)^2$ and $k_{-i} = \left(\frac{\gamma x_i}{2(1+\gamma^2)}\right)^2$ for $x_i \geq 0$. The political cost function is

$$\begin{aligned} C(x_i) &= \left(\frac{x_i}{2(1+\gamma^2)}\right)^2 + \left(\frac{\gamma x_i}{2(1+\gamma^2)}\right)^2 \\ &= \frac{x_i^2}{4(1+\gamma^2)} \text{ for } x_i < 0. \end{aligned}$$

In the text, we focus on the case where, at equilibrium, the politician offers local public goods *and* transfers. This requires that tax revenues are sufficiently large. More specifically, it requires the following.

1. Under [R], each regional politician spends $k_i = (n_i)^2$ and $x_{it} - 2n_i + 2\gamma n_{-i}$ on transfers. The equilibrium payoff is

$$x_{it}^R = \frac{\beta T_i}{n_i} - n_i + 2(n_i - \gamma n_{-i}).$$

   Substitute this into the expression for the transfer and note that $s_i > 0$ requires that $T_i > \frac{(n_i)^2}{\beta}$ for $i = 1, 2$.

2. Under [F], two cases can arise. For $\gamma \geq 0$, we can, using Proposition 2, write the payoff to group $i$ at time $t$ as

$$x_{1t}^F = \frac{T\beta p_1 + n_1^2 + \gamma(1-\beta p_1)(\gamma n_2^2 - 4n_1 n_2 + \gamma n_1^2)}{n_1}$$

$$x_{2t}^F = \frac{T\beta p_2 + n_2^2 + \gamma(1-\beta p_2)(\gamma n_1^2 - 4n_1 n_2 + \gamma n_2^2)}{n_2}$$

The transfers are

$$\begin{aligned} s_1 &= x_1 - 2(n_1 - n_2\gamma) + 2\gamma(n_2 - n_1\gamma) = x_1 - 2n_1(1+\gamma^2) + 4\gamma n_2 \\ s_2 &= x_2 - 2(n_2 - n_1\gamma) + 2\gamma(n_1 - n_2\gamma) = x_2 - 2n_2(1+\gamma^2) + 4\gamma n_1 \end{aligned}$$

At equilibrium, they are positive for

$$T > \max_i \left\{ \frac{n_i^2(1+\gamma^2(1+\beta p_i)) - 4\beta\gamma n_1 n_2 p_i + n_{-i}^2\gamma^2(1-\beta p_i)}{\beta p_i} \right\}.$$

For $\gamma < 0$, we can, using Proposition 1, define the minimum equilibrium payoffs as

$$(1+\gamma^2)(n_i)^2 + n_i(x_i - 2n_i(1+\gamma^2)) = \beta p_i T.$$

Solving this yields $x_{it} = \frac{1}{n_i}(T\beta p_i + n_i^2(\gamma^2+1))$. $s_i > 0$, then, requires that

$$x_i > 2n_i(1+\gamma^2) + 4\gamma n_{-i}$$

or

$$T > \max_i \left\{ \frac{n_i}{\beta p_i} \left(4\gamma n_{-i} + n_i \left(\gamma^2 + 1\right)\right) \right\}.$$

So, overall we need

$$T > \max_i \left\{ \frac{n_i}{\beta p_i} \left(4\gamma n_{-i} + n_i \left(\gamma^2 + 1\right)\right); \frac{n_i^2 \left(1 + \gamma^2 \left(1 + \beta p_i\right)\right) - 4\beta\gamma n_1 n_2 p_i + n_{-i}^2 \gamma^2 \left(1 - \beta p_i\right)}{\beta p_i} \right\}$$

and that $T_i > \frac{(n_i)^2}{\beta}$ for $i = 1, 2$.

# 10 Supplementary Appendix III

This appendix studies the incentive of regional politicians to join a federation in the presence of negative externalities. We consider the case with two identical regions: $p_i = \frac{1}{2}$, $n_1 = n_2 = 1$ and $T_1 = T_2 = \frac{1}{2}T$. The rent, $R^F$, extracted by the federal politician is $T - C(x_1^F, x_2^F)$ where $C(x_1^F, x_2^F)$ is given in equation (11) in the main text and

$$x_i^F = \frac{1}{2}\beta T + 1 + 2\gamma(1 - \frac{1}{2}\beta)(\gamma - 2) \text{ for } i = 1, 2. \tag{44}$$

Substitution yields

$$R^F = (1 - \beta) T + 2\gamma (1 - \beta) (2 - \gamma). \tag{45}$$

The politician of region $i$ prefers federalism to regionalism if $q_i R^F > \frac{1}{2}(1 - \beta) T$. A comparison yields that $q_1 R^F > \frac{1}{2}(1 - \beta) T$ if and only if

$$q_1 > \frac{\frac{1}{2}(1 - \beta) T}{R^F} \tag{46}$$

and that $(1 - q_1) R^F > \frac{1}{2}(1 - \beta) T$ if and only if

$$q_1 < 1 - \frac{\frac{1}{2}(1 - \beta) T}{R^F}. \tag{47}$$

Substitution of $R^F$ into equation (47) yields the threshold value $\overline{q}_1$ that makes the regional politician indifferent between joining and not joining the federation:

$$\overline{q}_1 = \frac{\frac{1}{2}(1 - \beta) T + 2\gamma (1 - \beta)(2 - \gamma)}{(1 - \beta) T + 2\gamma (1 - \beta)(2 - \gamma)}. \tag{48}$$

We notice that $\overline{q}_1 > \frac{1}{2}$ because

$$\begin{aligned} & \frac{\frac{1}{2}(1 - \beta) T + 2\gamma (1 - \beta)(2 - \gamma)}{(1 - \beta) T + 2\gamma (1 - \beta)(2 - \gamma)} - \frac{1}{2} \\ = \; & \frac{(2 - \gamma)\gamma}{(T + 4\gamma - 2\gamma^2)} > 0. \end{aligned} \tag{49}$$

Then, for $q_1$ such that

$$\overline{q}_1 > q_1 > 1 - \overline{q}_1 \tag{50}$$

both politicians prefer [F] to [R] This implies that there exist values of $q_1$ such that both regional politicians benefit from centralization. Moreover,

$$\frac{\partial \overline{q}_1}{\partial \gamma} = \frac{2(1-\gamma)T}{(T + 4\gamma - 2\gamma^2)^2} > 0 \tag{51}$$

for $\gamma < \frac{n_2}{n_1} < 1$. Hence, the threshold $\overline{q}_1$ is increasing in the strength of the externality and the two regional politicians are most likely to consent to a federation if externalities are strong.

# 11 Supplementary Appendix IV

Our results regarding the costs and benefits of fiscal centralization derive from the fundamental properties of the political cost functions in the federation. This appendix presents an extension of the baseline model which allows for endogenous distortionary taxation. We show that the political cost function is super-additive with negative externalities ($\gamma > 0$) and sub-additive with positive externalities ($\gamma < 0$), as in the baseline model. As a consequence, the key insight from the baseline model that the direction of the externality creates an asymmetry in the costs and benefits of centralization carries over to the economy with distortionary taxation.

## 11.1 Economic structure

The utility and production structure is as in the baseline model but income in the two regions are subject to a proportional tax $\tau_{it}$ and

$$u_{it} = y_i(1 - \tau_{it}) + g_{it} - \gamma g_{-it}. \tag{52}$$

The parameter $\gamma \in (-1, \frac{n_2}{n_1})$ captures the strength of the externality. Public goods are produced by the technology $g_{it} = 2\sqrt{k_{it}}$. We introduce a leaky bucket assumption and, as in Acemoglu and Robinson (2006, pp. 100-101), the tax distortion is captured by an aggregate cost coming out of the government budget. More specifically, the revenue raised from region $i$ at time $t$ is

$$T_{it} = n_i y_i \tau_{it} - d(\tau_{it}) n_i y_i \tag{53}$$

where the deadweight cost function $d(\tau_{it})$ is increasing in the tax rate at an increasing rate. To get closed form solutions to the political cost functions, we assume that

$$d(\tau_{it}) = \frac{1}{2}\tau_{it}^2. \tag{54}$$

This also implies that the revenue maximizing tax rates are 1. We can write the maximum revenue that can be collected from region $i$ as $T_{it}^{\max} = \frac{1}{2}n_i y_{it}$ and the total amount that can be collected from the federation as $T_t^{\max} = \sum_i \frac{1}{2}n_i y_{it}$. The choice variables of the federal politician are $\{\tau_{it}, g_{it}\}$ for $i = 1, 2$. That is, the tax rates can be region-specific. We omit time subscripts in the following. For simplicity and without any loss of insights, we impose the assumption that the income per capita in the two regions is equal to 1 to simplify some of the welfare differentials.

## 11.2 Political cost functions in the federation

For a federal politician, the political cost of satisfying the two standards announced by the voters in the two regions is the difference between the maximum rent that can be extracted ($T^{\max}$) and the actual rent extracted given compliance with the standards, ($r(x_1, x_2) = \sum_i (T_i - k_i)$):

$$c^F(x_1, x_2) = T^{\max} - r(x_1, x_2) = T^{\max} + k_1 + k_2 - T_1 - T_2. \tag{55}$$

The political cost function is the minimum cost required to meet the standards and is defined as

$$C^F(x_1, x_2) = \min c^F(x_1, x_2) \tag{56}$$

subject to for $i = 1, 2$

$$n_i x_i \leq n_i \left( y_i(1 - \tau_i) + 2\sqrt{k_i} - 2\gamma\sqrt{k_{-i}} \right) \tag{57}$$

$$\tau_i \in [0, 1]. \tag{58}$$

Let the Lagrangian multipliers on the two incentive compatibility constraints be $\lambda_1$ and $\lambda_2$, respectively, and assume that the tax rates are interior (to be verified below). We can then write the first order conditions as

$$1 = \frac{\lambda_i n_i}{\sqrt{k_i}} - \frac{\gamma \lambda_{-i} n_{-i}}{\sqrt{k_{-i}}} \tag{59}$$

$$\lambda_i = 1 - \tau_i \tag{60}$$

$$n_i x_i = n_i \left( y_i(1 - \tau_i) + 2\sqrt{k_i} - 2\gamma\sqrt{k_{-i}} \right). \tag{61}$$

We can solve these equations to get

$$\tau_1^* = \frac{1}{A}(A - x_1 \left(2n_2(1 + \gamma^2) + y_2\right) - 4\gamma n_2 x_2), \tag{62}$$

$$\tau_2^* = \frac{1}{A} \left(A - x_2 \left(2n_1(1 + \gamma^2) + y_1\right) - 4\gamma n_1 x_1\right), \tag{63}$$

where

$$A = 4\left(\gamma^4 - 2\gamma^2 + 1\right) n_1 n_2 + 2y_2 \left(1 + \gamma^2\right) n_1 + 2y_1 \left(1 + \gamma^2\right) n_2 + y_1 y_2 \tag{64}$$

$$k_1^* = (\lambda_1 n_1 - \gamma\lambda_2 n_2)^2 = ((1-\tau_1)\, n_1 - \gamma\,(1-\tau_2)\, n_2)^2 \tag{65}$$

$$k_2^* = (\lambda_2 n_2 - \gamma\lambda_1 n_1)^2 = ((1-\tau_2)\, n_2 - \gamma\,(1-\tau_1)\, n_1)^2 . \tag{66}$$

By inspection of equations (62) and (63), $\tau_i^* \leq 1$ for $\gamma > 0$. For $\gamma < 0$, we require that

$$B_1 x_2 < x_1 < B_2 x_2 \tag{67}$$

where

$$B_1 = \frac{-4\gamma n_2}{2n_2(1+\gamma^2)+y_2} \tag{68}$$

$$B_2 = -\frac{(2n_1(1+\gamma^2)+y_1)}{4\gamma n_1} \tag{69}$$

with $B_1 - B_2 = \frac{A}{4\gamma n_1(2n_2(1+\gamma^2)+y_2)} < 0$. If $x_1 < B_1 x_2$, then $\tau_1^* = 1$ and if $x_1 > B_2 x_2$, then $\tau_2^* = 1$ and the optimal public goods levels are adjusted accordingly. Substitution of the solution into the objective function and imposing the assumption that $y_1 = y_2 = 1$ give the political cost function for $\tau_i \in (0,1)$

$$\begin{aligned} C^F(x_1,x_2) &= \frac{\left(n_1\left(2n_2(1+\gamma^2)+1\right)x_1^2 + 8\gamma n_1 n_2 x_1 x_2 + n_2\left(2n_1(1+\gamma^2)+1\right)x_2^2\right)}{8\left(\gamma^4-2\gamma^2+1\right)n_1 n_2 + 4\left(1+\gamma^2\right)(n_1+n_2)+2} \\ &\quad - \frac{\left(4\left(\gamma^4-2\gamma^2+1\right)n_1 n_2\,(n_1+n_2) + 2\left(\gamma^2+1\right)(n_1+n_2)^2 + n_1+n_2\right)}{8\left(\gamma^4-2\gamma^2+1\right)n_1 n_2 + 4\left(1+\gamma^2\right)(n_1+n_2)+2} \\ &\quad + T^{\max}. \end{aligned} \tag{70}$$

Next, we calculate the political costs of satisfying the voters in just one region. The political cost function is defined as

$$C^F(x_i) = \min T^{\max} + k_1 + k_2 - T_1 - T_2 \tag{71}$$

subject to

$$n_i x_i \leq n_i\left(y_i(1-\tau_i) + 2\sqrt{k_i} - 2\gamma\sqrt{k_{-i}}\right) \tag{72}$$

$$\tau_i \in [0,1]. \tag{73}$$

Clearly $\tau_{-i} = 1$. For a negative externality $\gamma > 0$, we have that $k_{-i}^{**} = 0$ and $\tau_i^{**} = 1 - \frac{x_i}{2n_i+y_i} < 1$ and $k_i^{**} = n_i^2\left(\frac{x_i}{2n_i+y_i}\right)^2$. The political cost function is

$$\begin{aligned} C^+(x_i) &= T^{\max} + n_i^2\,(1-\tau_i^{**})^2 - \left(\tau_i^{**} - \frac{1}{2}\tau_i^{**}\right) n_i y_i \\ &= T_i^{\max} + \frac{x_i^2 - y_i^2 - 2n_i y_i}{4n_i + 2y_i} n_i \end{aligned} \tag{74}$$

for $i=1,2$. For a positive externality $\gamma<0$, we find that $\tau_i^{***}=1-\frac{x_i}{2n_i(1+\gamma^2)+y_i}<1$ and that $k_i^{***}=\left(\frac{n_i x_i}{2n_i(1+\gamma^2)+y_i}\right)^2$ and $k_{-i}^{***}=\gamma^2\left(\frac{n_i x_i}{2n_i(1+\gamma^2)+y_i}\right)^2$. The political cost function is

$$\begin{aligned} C^{-}(x_i) &= T^{\max}+(1+\gamma^2)n_i^2(1-\tau_i^{**})^2-\left(\tau_i^{**}-\frac{1}{2}\tau_i^{**}\right)n_i y_i \\ &= T_i^{\max}+\frac{x_i^2-2n_i y_i(1+\gamma^2)-y_i^2}{4n_i(1+\gamma^2)+2y_i}n_i. \end{aligned} \tag{75}$$

We can now evaluate the additiveness of the political cost functions. To simply the expressions, we impose the assumption that $y_1=y_2=1$. For a negative externality $(\gamma>0)$, we find that

$$\begin{aligned} \Delta C^{+} &= C^F(x_1,x_2)-C^{+}(x_1)-C^{+}(x_2) \\ &= \frac{(4n_2n_1(2n_1+1)(2n_2+1)x_1x_2+n_1^2\gamma(2n_2+1)(2n_2(3-\gamma^2)-1)x_1^2)}{1+2(1+\gamma^2)(n_1+n_2)+4n_1n_2(1+\gamma^4-2\gamma^2)(2n_2+1)(2n_1+1)}\gamma \\ &+\frac{n_2^2\gamma(2n_1+1)(2n_1(3-\gamma^2)-1)x_2^2\gamma}{1+2(1+\gamma^2)(n_1+n_2)+4n_1n_2(1+\gamma^4-2\gamma^2)(2n_2+1)(2n_1+1)}>0 \text{ for } \gamma\in[0,1). \end{aligned} \tag{76}$$

So, the political cost function is super-additive for negative externalities.

For a positive externality $(\gamma<0)$, we find, under the assumption that the tax rates are interior and $y_1=y_2=1$, the following

$$\begin{aligned} \Delta C^{-} &= C^F(x_1,x_2)-C^{-}(x_1)-C^{-}(x_2) \\ &= \frac{\gamma n_1n_2(2n_1\gamma D_2x_1^2+2n_2\gamma D_1x_2^2+D_1D_2x_1x_2)}{4(2(1+\gamma^2)(n_1+n_2)+4n_1n_2(1-2\gamma^2+\gamma^4)+1)D_2D_1} \end{aligned} \tag{77}$$

where $D_i=2n_i(1+\gamma^2)+1$. We requires

$$B_1x_2<x_1<B_2x_2 \tag{78}$$

in order for the tax rates to be interior. We can show that

$$\frac{\partial\Delta C^{-}}{\partial x_1}=\frac{\gamma n_1n_2(D_1x_2+4\gamma n_1x_1)}{4(2(1+\gamma^2)(n_1+n_2)+4n_1n_2(1-2\gamma^2+\gamma^4)+1)D_2D_1}>0$$

for $x_1<B_2x_2$. The sign of $\Delta C^{-}$ is equal to -sign$(2n_1\gamma D_2x_1^2+2n_2\gamma D_1x_2^2+D_1D_2x_1x_2)$. Evaluate this at the upper bound i.e., at $x_1=B_2x_2$ which gives

$$\begin{aligned} &-\frac{4(\gamma^4-2\gamma^2+1)n_1n_2+2(\gamma^2+1)n_1}{8\gamma n_1} \\ &+\frac{2((\gamma^2+1)n_2+1(2n_1(1+\gamma^2)+1)x_2^2}{8\gamma n_1}>0 \text{ for } \gamma\in(-1,0). \end{aligned} \tag{79}$$

This implies that $\Delta C^{-}<0$ and the political cost function is sub-additive.

Unlike in the baseline model, we observe that the political cost functions in the economy which distortionary taxation are quadratic in $x_1$ and $x_2$. This makes it cumbersome to derive closed form solutions which can be used as an input to the welfare analysis of federalism versus regionalism. However, qualitatively the main results will care over to the more complex economy. This is because the key drivers behind them are i) the political cost functions are super-additive with negative externality leading to the rent effect; and ii) the utility allocation is not unique with a sub-additive political cost function. These properties are preserved in the more complex economy with distortionary taxation.

## Extra references

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