

Citizen Preferences and the Architecture of Government

Jean-Marc Bourgeon, Marie-Laure Breuillé

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl

www.cesifo-group.org/wp

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

Citizen Preferences and the Architecture of Government

Abstract

We consider a country whose government provides a bundle of goods and services through a multi-tier administrative organization. We compare the optimal architectures of public governance (i.e. the division of the state into several tiers, the distribution of services among them, their number of jurisdictions and the performance ability of their administrations) of two governments, one centralized and the other decentralized. Under a decentralized government, national and subnational decision-makers only consider the impact of their decisions on the welfare of their constituents, neglecting other tiers' policy. The resulting architecture is generally different from the (first-best) centralized one, and depends on how citizens weight the performance ability of the administrations and the range of goods they provide. If the relative weight on the performance ability is large, the decentralized architecture entails more tiers, less jurisdictions per tier with reduced scope of services than the centralized one, and the reverse if this relative weight is small. We use our results to estimate this weight on U.S. data. We find that the country exhibits two zones (“Northeast & West” and “Midwest & South”) where the estimated values are statistically different.

JEL-Codes: H110, H770.

Keywords: decentralization, fiscal federalism, public governance.

Jean-Marc Bourgeon
INRA, UMR économie Publique
16 rue Claude Bernard
France – 75231 Paris
bourgeon@agroparistech.fr

Marie-Laure Breuille
INRA, UMR1041 CESAER, Université
Bourgogne Franche-Comté, AgroSup Dijon
France – 21000 Dijon
breuille@dijon.inra.fr

September 6, 2018

We are grateful to Julie Le Gallo, Anne-Laure Samson, to seminar participants at the economic department of Ecole Polytechnique, TSE, BETA, the Goethe University Frankfurt, the UCLouvain Belgo-Japanese Public Finance workshop and the CESifo Applied Microeconomics and Political Economy workshops.

1 Introduction

The comprehensive framework of public governance (hereafter CF, which corresponds to the organizational layout and functionality of a country’s public administration in totality, i.e., the number of its tiers and jurisdictions per tier, their geographical distribution, the size of their administrations, and the services they are in charge of providing to their citizenry) differs widely over the world. The number of tiers of sub-national authorities ranges from five, such as in the Philippines, to only one, such as in Kuwait. The number of jurisdictions in each tier also varies widely: focusing on the bottom-most tier of municipal authorities in Europe, 40% are located in a single country: France. Moreover, forms of public governance are regularly changing. As pointed out by the OECD (2014), half of OECD countries (often guided by cost reduction considerations) have planned or completed reforms to redraw the map of CFs in the past 15 years, through municipal-mergers, inter-municipal cooperation, or metropolitan governance. Denmark totally reorganized its CF in 2007; the number of municipalities dropped from 271 to 98 through a series of mergers, and its 13 counties were replaced by 5 regions (Dexia, 2007). Latvia reduced the number of municipalities from 527 to 105 in 2009, and France its number of metropolitan regions from 22 to 13 in 2016. Fewer countries have acted in the opposite direction by increasing the number of municipalities. The United States, however, is one such country; its number of municipalities (including towns and townships) increased from 1952 to 2012 by 5.5%. Although these opposing approaches are admittedly the outcome of country-specific differences in geography, history, and political and social movements, they nevertheless raise the question of the desirable properties of an ideal CF (if any such exist).

In this paper, we analyze this issue in a theoretical framework that allows us to compare the CFs of two governments, one centralized and the other decentralized, in charge of providing the same bundle of public goods. In each case, we determine the number of tiers and jurisdictions per tier, the performance ability of the administrations, and the goods they produce at each level.

Our analytical framework is rooted in the one-tier jurisdiction formation model of Alesina & Spolaore (1997, hereafter AS). In their model, the world is a linear segment populated by a continuum of individuals uniformly distributed and identified by their geographical locations. They consider the emergence of countries, i.e. independent jurisdictions whose governments provide a composite public good. Individuals face two types of costs, namely a “cost of distance” (to reach the location of their jurisdiction’s seat of government) and a “cost of government” (i.e. the production cost of the public

good). We extend this framework to determine the internal organization of a country, detailing the public task into the many goods and services that a government is responsible for, each of them being characterized by an accessibility parameter that determines the citizens' access cost (which, like the "cost of distance" in AS, is also a function of the distance between the citizen and the location of the corresponding public facility) and the provision cost. We allow for as many layers of overlapping jurisdictions as necessary for the government to best serve the citizenry. It is thus a multi-tier, multi-good setup where the linear segment is partitioned at each tier into jurisdictions, each jurisdiction being in charge of providing the tier-specific bundle of public goods.¹ We also detail the provision cost that depends both on the jurisdiction's scope of activities (i.e. the bundle of public goods its administration provides) and on its "capacity" or performance ability (depending on things such as the size of the public facility and the workforce involved), which determines the quality of the public activity. The performance ability of the administration matters for citizens, a greater ability allowing a reduction in congestion effects, but at the expense of higher provision costs. Economies of scope can be achieved by increasing the number of services provided by a jurisdiction, but there is a tradeoff between the quality and the number of services provided by the same administration. The optimal balance depends on citizen preferences and affects the overall organization of the government.

The organization of the country also depends on the leeway given to each jurisdiction. We consider two cases: a centralized country where all jurisdictions are ruled by a benevolent and omniscient planner, which corresponds to the first-best case, and a decentralized country where (elected) decision-makers at each tier decide on the scope of activities and performance ability needed by their administration to maximize the welfare of their constituency without accounting for the decisions made by other tiers. We derive the optimal CF in the latter case by assuming that a social planner decides on the number of tiers and jurisdictions per tier and makes sure that there is no overlap of activities, i.e., no public good is provided by jurisdictions belonging to two different tiers. It is a second-best optimum because the social planner delegates to jurisdictional authorities with narrow objectives the management of their performance ability and the scope of their activities. We assume in both cases that the budget of each

¹The segment in our model is dubbed "a country", the corresponding political entity being either a federation, like the United States or the European Union, or a unitary state with administrative divisions. However, we may consider like AS that the segment is the world since the international governmental organizations that compose the international law system (e.g. the United Nation, the World Bank, the World Trade Organization) correspond to legal entities that form a layer of government (arguably) above nation-states.

jurisdiction is balanced and that its provision cost is financed by a lump-sum tax.²

The optimal CF is driven by three parameters, one related to the valuation of the public activity (the weights citizens assign to the administration's performance ability and scope of activities are given by β and $1 - \beta$ respectively), the provision cost, and the access cost. We first derive the characteristics of the one-tier equivalent of AS in our setup and we show that allowing for varying degrees of performance ability changes the effect of a change in the provision cost: while an increase of the provision cost diminishes the number of jurisdictions in AS, it decreases the performance capacity of each jurisdictional administration but increases their number in the one-tier equivalent. Hence, the performance ability of each administration is reduced, but this is somewhat compensated by a reduction in the citizens' access cost. Allowing for as many layers of administrations as necessary to cover the entire set of public goods leads to the completely opposite effect of an increase in the cost on the number of jurisdictions compared to the one-tier equivalent of AS. Indeed, in the one-tier case an increase in the access or in the provision cost increases the number of jurisdictions, while in the multi-tier setup, although the increased cost induces the government to create another layer of jurisdictions, this actually results in a decrease in the number of jurisdictions (with reduced scopes of activity) in the other layers. The opposite happens when β , the citizens' relative valuation of an administration's performance ability over its scope of activities, increases: in the one-tier case, the number of jurisdictions decreases because of the increase in the provision cost due to the increase in capacity, while in the multi-tier case the number of jurisdictions per tier increases but the resulting cost increase is compensated for by an enlargement of their scope of activities and a reduction in the number of tiers. These general features are shared by centralized and decentralized countries. Hence, if the citizen preference for the administration's performance ability is low, a country adopts a tall and thin CF: the number of tiers is large, the number of jurisdictions is small, and the range of services of those jurisdictions is narrow. Reciprocally, if β is high, the CF is small and fat: the number of tiers is small, the number of jurisdictions is large, and their scope of activities broad. We also obtain that, proceeding downward in the vertical structure of the CF, the range of activities

²AS offers a study of the process of democratization which is different from decentralization: Following Friedman (1977), they compare a democratic world, where the number of nations results from majority votes, to a world of dictators, i.e. rent-maximizing governments who decide on the size and shape of nations such as to maximize their joint potential net revenues. Also note that the democratic world is composed of an excessive number of nations compared to the first-best (that would be reached if the world were ruled by a social planner) unless individuals located far from the seat of government were compensated so that they would not vote in favor of creating a new country.

increases, the administration’s performance ability decreases, and the degree of jurisdictional dispersion (i.e. the ratio of the numbers of jurisdictions of two consecutive tiers) remains constant. As a result, the per capita provision cost, and the amount of satisfaction citizens get from the activity of the government, may decrease or increase as one moves down the CF. We obtain that both increase if β is lower than $1/4$ and decrease otherwise.

Allowing for decentralized decision-making, i.e. jurisdictions that are free to select their range of services and their capacities to perform them, we also obtain that compared to centralization, the difference in the optimal CF depends on how the performance ability and the administration’s scope of services are weighted. If β is larger than a threshold level, the number of tiers is increased but the number of jurisdictions per tier and their scope of activities are reduced under decentralization. This is due to the fact that a decision-maker would like to decrease her jurisdiction’s scope of activities compared to the first-best scenario when the citizens’ relative weight for the jurisdiction’s range of services is low. To compensate for this effect, the social planner reduces the number of jurisdictions per tier but has to increase the number of tiers in order to cover the entire range of public goods provided.

Our structural model leads to close-form results that permit estimation of the citizen preference parameter β using standard panel data procedures. We conducted this empirical investigation using U.S. data on annual federal and state expenditures, share of population, and density (together with data on income per capita and on the partisan composition of the state legislatures) over the period 1977-2015. Assuming the same citizen preferences across the country, we obtain an overall estimate for β equal to 0.18, a value that is neither influenced by the partisan composition of state legislatures nor by other characteristics such as income and population density. However, disaggregating the country according to the four regional divisions used by the Census Bureau (Midwest, Northeast, South, West), we obtain that this weight is lower in the Northeast and West (0.182 and 0.147) than in the Midwest and South (0.279 and 0.238). Gathering these four regions into two zones, we obtain 0.153 for the Northeast & West zone and 0.251 for the Midwest & South zone, estimates that are statistically different. We can thus infer that citizens in the Northeast and West regions value state-provided services more than federally provided services, whereas citizens in the Midwest and South are more likely to be equally satisfied with both levels of government.

Our paper builds a bridge between the literature on the formation of jurisdictions and the literature on fiscal federalism by formalizing for the first time the vertical dimension of the formation of jurisdictions. The literature on the formation of jurisdic-

tions, which analyzes the equilibrium partition of a population into several jurisdictions and their political stability, has been quite extensive over the last twenty years. The breakup or unification of nations has been modeled as the result of a trade-off either between the efficiency gains of unification and the costs in terms of loss of control in political decision making (Bolton & Roland, 1997) or, more commonly, between benefits from economies of scale in the provision of public goods and costs from preferences heterogeneity (AS, Jehiel & Scotchmer, 2001). A cooperative game theory approach is usually used to study the political stability of jurisdictions (Guesnerie & Oddou, 1981; Greenberg & Weber, 1986; Demange, 1994; Casella, 2001; Bogomolnaia et al., 2006, 2008, among others). Although the coexistence of several clubs of different sizes was studied by Hochman et al. (1995), the vertical dimension of the formation of jurisdictions, i.e. the partition of the country into overlapping tiers, is missing in this literature. By contrast, some recent works in the fiscal federalism literature focus on the vertical dimension to determine the optimal level of decentralization of a unique public good (Panizza, 1999) or the partition of a continuum of identical public goods (Wilson & Janeba, 2005). Their framework, however, is restricted to two tiers, i.e., a central government and an exogenous number of same-tier sub-national jurisdictions.

The paper is organized as follows. Section 2 presents the multi-tier framework. Section 3 characterizes the optimal CF. Section 4 investigates the impact of decentralized decision-making. The empirical investigation is detailed in section 5. The last section concludes. All proofs are in the appendix.

2 The model

Following the spatial approach of Hotelling (1929), consider a country with a continuum of citizens uniformly distributed over a segment. The size of the territory and the population mass are both normalized to unity. Each citizen is identified by her geographical location, her point on the segment, supposed to be fixed. We are interested in the vertical organization of the country (the number of jurisdictional tiers and jurisdictions per tier, the size of their administration and the allocation of services provided) to maximize the utility derived by citizens from the public activity. Some of these goods are readily accessible by all, such as the protection provided by the army of the country and the foreign affairs and intelligence services, while others, such as the natural amenities found in a park or the education dispensed in an elementary school, are enjoyable more or less depending on their locations relative to the citizens. To

capture this accessibility concern in a tractable way, we index each public good by an accessibility parameter $x \in [1, \bar{x}]$ and we suppose that the cost borne by a citizen to enjoy good x provided by a public facility located at distance ℓ from her home is given by $\alpha x \ell$ where $\alpha > 0$ is the access cost parameter.^{3,4} To handle this provision problem, the country is composed of jurisdictional tiers, indexed by $t \in \{0, \dots, T\}$, charged to provide a subset of these services. Tier $t = 0$ is the top tier, the central/federal jurisdiction, which provides the most accessible public goods, while $t = T$ is the bottom-most tier, which may correspond to villages or the districts of big cities, in charge of providing the less accessible ones. Each citizen thus belongs to $T + 1$ overlapping jurisdictions, i.e. T sub-national jurisdictions plus the central jurisdiction, each being responsible for delivering a specific bundle of services. The tier t 's range of services is to provide public goods belonging to the subset $(x_{t-1}, x_t]$, with $x_{-1} = 1$ and $x_T = \bar{x}$. At each level t , the territory is divided in n_t jurisdictions ruled by governments indexed by z_t , $z_t \in \{1, \dots, n_t\}$, with $n_0 = 1$. Jurisdiction z_t rules over the area $S_{z_t} \subseteq [0, 1]$, with $S_{z_t} \cap S_{z'_t} = \emptyset$ and $\cup_{z_t=1}^{n_t} S_{z_t} = [0, 1]$, and is geographically defined by three points: its lower and upper borders and the location ρ_{z_t} of its government/public facility which is also where its citizens have to go to enjoy the public goods it provides. Hence, to obtain the whole set of public goods, citizen i incurs a total access cost $\alpha \sum_{t=0}^T \ell_{it} \int_{x_{t-1}}^{x_t} x dx$, where $\ell_{it} \equiv |i - \rho_{z_t}|$ represents the distance to reach the location ρ_{z_t} of the public facility in tier t .

In addition to the citizens' access cost, a jurisdiction incurs a provision cost that depends on the range of the public goods it provides and on the size/performance capacity of the administration that serves citizens. In particular, we suppose that whatever the administration's performance ability, the provision cost of a good is larger the higher its accessibility parameter x ,⁵ but that economies of scope are achieved by providing a wide range of public goods. More precisely, denoting by $c(g, x)$ the cost of providing good x by an administration of performance ability g , we assume that the

³The extent of the rivalry and excludability characteristics of these goods/services may vary and some of them may even be private goods. We do not discuss in the following the overall scope of the services $[1, \bar{x}]$ and we suppose that the exogenous upper bound \bar{x} is the cut-off between the public and the private sectors.

⁴In the examples given above, military protection corresponds to an access cost close to 0 (arguably strictly positive for citizens living on the borders of the territory) and thus we may expect $0 < \alpha < 1$. Observe that at equal distance from home, the accessibility parameter of children's education may be larger than the one associated to natural amenities offered by a park due to inherent constraints, such as the frequency of trips to and from the school and the ease with which parents can conform to the school's operating hours.

⁵For instance, the cost per inhabitant of the army or intelligence services of a country is lower than the cost per inhabitant of teachers in elementary schools.

provision cost of providing the bundle $(x_{t-1}, x_t]$ of services by the same administration of performance ability g_{z_t} is given by

$$C_{z_t} = C(g_{z_t}, (x_{t-1}, x_t]) = \int_{x_{t-1}}^{x_t} c(g_{z_t}, x) dx / \int_{x_{t-1}}^{x_t} dx.$$

Without economies of scope, the provision cost would be given by $\int_{x_{t-1}}^{x_t} c(g_{z_t}, x) dx$. Economies of scope arise from the use of the same administration to produce the whole range of goods and depend on the number of goods provided: $\int_{x_{t-1}}^{x_t} dx = x_t - x_{t-1}$. In addition to a cost that increases with the accessibility parameter, i.e. $c'_x(g, x) > 0$, we suppose that $c'_g(g, x) > 0$ and $c''_{xg}(g, x) > 0$: the larger the performance ability of the administration, the higher the jurisdiction's provision cost.⁶ To ease computations, we consider the case $c(g, x) = kgx$, with $k > 0$, which yields

$$C_{z_t} = kg_{z_t}(x_t + x_{t-1})/2.$$

In this expression, the provision cost is proportional to the product of the administration's performance ability and the average accessibility parameter of the bundle of goods provided. These provision costs are covered by a lump-sum tax τ_{z_t} paid by each citizen resorting to jurisdiction z_t . As tier- t encompasses n_t jurisdictions, the aggregated cost induced by the provision of all public goods nationwide amounts to $\sum_{t=0}^T \sum_{z_t=1}^{n_t} C_{z_t}$.

Each citizen i derives a utility from both the public goods provided and the administration's performance ability of each $T + 1$ overlapping tiers' jurisdictions that she belongs to. Indeed, the larger the administration's performance ability, the lower the congestion citizens endure when accessing the goods and services produced by the administration.⁷ We denote by $u_t = u(g_{z_t}, (x_{t-1}, x_t])$ the utility (expressed in monetary terms) that the citizen derives from the goods provided by her tier- t jurisdiction and we assume that it is given by

$$u(g_{z_t}, (x_{t-1}, x_t]) = g_{z_t}^\beta \left(\int_{x_{t-1}}^{x_t} dx \right)^{1-\beta} = g_{z_t}^\beta (x_t - x_{t-1})^{1-\beta}$$

where β and $1 - \beta$ reflect the importance that the citizen places on the administration's

⁶In addition to the construction/rental cost of the public facility, large administrations accommodating the public necessitate more personnel (administrative, security and maintenance staffs) and equipment than smaller ones.

⁷Alternatively, g_t could be considered as a measure of the overall quality of the administration, which depends on both the public infrastructure and the personnel employed.

performance ability and on its scope of services respectively. Her gross utility from the public activity is thus given by $\sum_{t=0}^T u_t$ which must be larger than the provision and access costs. In our setup, it is shown in the appendix that this is the case at the optimum if $\beta \leq 1/2$.

3 The centralized optimum

Consider a benevolent social planner motivated by the aim of maximizing the country's welfare, who chooses the number of sub-national tiers, the desirable division of the territory in each tier, i.e., the geographical boundaries of each, the location of the seat of their government, the performance ability of their administration and the range of public goods provided at each level. The corresponding optimization problem is given by

$$\max_{T, \{(S_{z_t}, \rho_{z_t}, g_{z_t})_{z_t=1}^{n_t}\}_{t=0}^T} \left\{ \int_0^1 u_i di - \sum_{t=0}^T \sum_{z_t=1}^{n_t} C_{z_t} : n_0 = 1, x_{-1} = 1, x_T = \bar{x} \right\}.$$

where

$$u_i = \sum_{t=0}^T \left[g_{z_t}^\beta (x_t - x_{t-1})^{1-\beta} - \alpha l_{it} \int_{x_{t-1}}^{x_t} x dx \right]$$

is the utility of individual i net of access costs. Because individuals are uniformly located in the territory, we easily show that

Lemma 1 *At each tier t , the territory is divided into n_t jurisdictions of equal size. Their administrations are located at the center of the jurisdiction and have the same performance ability.*

Using lemma 1, the social planner's program can be restated as

$$\max_{T, \{n_t, x_t, g_t\}_{t=0}^T} \left\{ \sum_{t=0}^T W(x_{t-1}, x_t, g_t, n_t) : n_0 = 1, x_{-1} = 1, x_T = \bar{x} \right\} \quad (1)$$

where

$$W(x_{t-1}, x_t, g_t, n_t) \equiv g_t^\beta (x_t - x_{t-1})^{1-\beta} - \frac{\alpha (x_t^2 - x_{t-1}^2)}{8n_t} - \frac{n_t k g_t (x_t + x_{t-1})}{2} \quad (2)$$

corresponds to the net average welfare at the tier- t level. More precisely, the second term in (2) is the per capita average access cost in a tier- t jurisdiction and the last term is the per capita provision cost of its administration. Indeed, since there are

n_t jurisdictions at tier t , the proportion of citizens resorting to one of each of these jurisdictions is $1/n_t$. Hence, to satisfy its budget constraints, the government at tier t must levy a per capita tax equal to $\tau_{z_t} = n_t C_{z_t}$. Also, for the range of good provided by the jurisdiction we have $\int_{x_{t-1}}^{x_t} x dx = (x_t^2 - x_{t-1}^2)/2$ while the average distance to the administration is $2 \int_0^{1/(2n_t)} i di = 1/(4n_t^2)$, hence the expression of the per capita average access cost.

Before solving the social planner's program, it is interesting to analyze the effect of allowing for varying degrees of performance ability in the one-tier case as investigated by AS (with jurisdictions that provide the entire bundle of public goods $[1, \bar{x}]$). The number of jurisdictions and their performance ability are the solutions of the reduced program $\max_{n,g} W(1, \bar{x}, g, n)$. From the first-order condition with respect to n we obtain that⁸

$$n = \sqrt{\frac{\alpha \bar{x} - 1}{4k} \frac{1}{g}} \quad (3)$$

where the term $\sqrt{\alpha/(4k)}$ corresponds to the optimal number of jurisdictions in the analysis of AS. The other term encompasses the effects of the administration's scope of services and performance ability on the number of jurisdictions. The optimal administration's performance ability is given by

$$g = (\bar{x} - 1) \left(\frac{4\beta}{\sqrt{\alpha k} (\bar{x} + 1)} \right)^{2/(1-2\beta)} \quad (4)$$

which shows that absent the economies of scope (captured by the term $\bar{x} + 1$), the performance ability of the administration is proportional to its range of services. Substituting in (3), we arrive at

$$n = \left(\frac{\alpha^{1-\beta} k^\beta}{\beta 2^{3-2\beta}} (\bar{x} + 1) \right)^{1/(1-2\beta)}, \quad (5)$$

hence a number of jurisdictions that increases in both k and α . Indeed, because it is possible to adjust the administration's performance ability, an increase in the access cost or in the provision cost results in a downsizing of the administration. In other words, to maintain an optimal welfare level when costs are exogenously increased, the size of administrations is reduced but the proximity of the government for the citizen is improved by a larger number of jurisdictions. Observe also that an increase in β

⁸Of course, the number of jurisdictions (and the number of tiers) are integers. The following results are thus approximations of the corresponding values.

acts in the opposite direction: the administration's performance ability (4) increases, and from (3), the number of jurisdictions decreases. We show in the following that the opposite effects are at work in the case of several imbricated levels of jurisdictions that are assigned a specific range of the public goods. Indeed, solving the planner's program 1, we obtain the following results

Proposition 1 *The optimal CF entails more than one tier if $\alpha k < 1$, $\alpha/k > (2/e)^2$ and $\beta > \beta_s = \phi^{-1}(1) \in (0, 1/2)$ where $\phi(\beta) = (\alpha/4)^{1-\beta} k^\beta / \beta$. Under these conditions, starring the corresponding variables, it is characterized by*

$$T^* = \frac{\ln \bar{x}}{\ln \lambda} - 1, x_t^* = \lambda^{t+1}, n_t^* = \lambda^{t/(1-2\beta)}, g_t^* = g_0^* \lambda^{-t(1+2\beta)/(1-2\beta)}, u_t^* = u_0^* \lambda^{t(1-4\beta)/(1-2\beta)} \quad (6)$$

where $\lambda > 1$ is given by

$$\lambda = \beta 2^{3-2\beta} / (\alpha^{1-\beta} k^\beta) - 1, \quad (7)$$

$$g_0^* = \frac{\alpha(\lambda - 1)}{4k}, \quad (8)$$

and

$$u_0^* = \frac{\alpha(\lambda^2 - 1)}{8\beta}. \quad (9)$$

The welfare at tier t is given by $W_t^* = (1 - 2\beta)u_t^*$ and the total welfare by

$$W^* = (1 - 2\beta)u_0^* \frac{1 - \lambda^{(T^*+1)(1-4\beta)/(1-2\beta)}}{1 - \lambda^{(1-4\beta)/(1-2\beta)}}$$

if $\beta \neq 1/4$ and $W^* = (1 - 2\beta)u_0^*(T^* + 1)$ otherwise.

The condition $\alpha k < 1$ guarantees that provision costs are sufficiently low to have a multi-tier government. Together with $\alpha/k > (2/e)^2$, it should be the case that $k < e/2 \approx 1.36$. Under these conditions, the optimal CF is characterized by λ , a magnification parameter given by (7), which allows us to determine recursively the number of jurisdictions per tier, their administration's performance ability, and their range of services: we have $n_t^* = n_{t-1}^* \lambda^{1/(1-2\beta)}$, $x_t^* = \lambda x_{t-1}^*$, $g_t^* = g_{t-1}^* \lambda^{-(1+2\beta)/(1-2\beta)}$. The number of tiers is derived from the total range of services of the government using $\bar{x} = \lambda^{T^*+1}$ and, as intuitively expected, it increases with the range of public goods to be provided. The range of services of jurisdictions increases the further away they are from the central government level: $x_{t+1}^* - x_t^* = \lambda(x_t^* - x_{t-1}^*)$. The degree of jurisdictional dispersion in tier t , defined as n_{t+1}/n_t , i.e. the number of jurisdictions in tier $t + 1$ which geographically belong to the same jurisdiction in tier t , is given

by $\lambda^{1/(1-2\beta)}$ for all t . It is thus constant over the CF and can be very large if β is close to $1/2$. The performance ability of the administration is the largest at the central level (tier 0) and decreases at rate $\lambda^{(1+2\beta)/(1-2\beta)}$ as one proceeds down the CF. Hence, the administration at the bottom-most tier, the one the closest to its citizenry (usually, a town), has the smallest administration's performance ability but provides the largest range of public goods, those in the upper range of the cost parameter, i.e. the more costly in terms of provision and accessibility for the citizen (like garbage collection or elementary education). At the other extreme, the central government has the largest administration and provides the smallest range of public services on the lowest end of the cost parameter. Both the fraction of the population belonging to a jurisdiction (given by $1/n_t^*$) and the administration's performance ability decrease with the tier level. Comparing the administration's performance ability per citizen at different levels, we get $g_t^* n_t^* = g_{t-1}^* n_{t-1}^* \lambda^{-2\beta/(1-2\beta)}$ which decreases along the CF. However, since the bundle of public goods increases and includes costlier goods the larger t , the cost per individual may be increasing or decreasing along the CF. This can be inferred from the citizens' satisfaction level u_t^* since at the optimum, provision and average access costs are equalized at each level and represent a total equal to 2β percent of the corresponding utility. In our setup, this ranking depends on the ratio $\beta/(1-\beta)$: if it is lower than $1/3$ (i.e., $\beta < 1/4$) public satisfaction level and public spending are higher the closer the proximity of the jurisdiction with the citizen (the higher t), and the reverse otherwise. The following proposition summarizes these results:

Proposition 2 *Proceeding down the CF,*

- i/ the jurisdiction's scope of services increases while the administration's performance ability decreases,*
- ii/ the per capita administration's performance ability decreases,*
- iii/ the degree of jurisdictional dispersion remains stable,*
- iv/ the utility and the per capita cost increase if $\beta < 1/4$ and decrease otherwise.*

Using (7) allows us to rewrite (5) as

$$n = \left(\frac{\bar{x} + 1}{\lambda + 1} \right)^{1/(1-2\beta)}.$$

As we have $n_t^* = \lambda^{t/(1-2\beta)}$, a striking result coming from Prop. 1, is that the effect of a change in λ (due to the change in one of the parameters) on the number of

jurisdictions in the one-tier case goes in the opposite direction of in the unconstrained $T + 1$ -tier case. More generally, comparative statics exercises allow us to obtain that

Proposition 3 *A decrease in β or an increase in either α or k increases the number of tiers and decreases the number of jurisdictions per tier and their range of services.*

The intuition is the following. Consider first the case where the access cost α is so small that there is only a central state, i.e. a unique administration located at the center of the country that provides the entire bundle of public goods. A large increase in the access cost would induce the country to add a layer of jurisdictions (say, regions) to provide the range of the less accessible services, reducing the services provided by the central state. An even larger increase in α would induce the government to create another layer of jurisdictions (say departments) for the same reason (to reduce the cost of the less accessible goods). However, since the range of services of the regions is reduced, it is optimal to reduce their number too. Hence, it is because another layer of jurisdictions can be added in the unconstrained/multiple tiers case that the number of jurisdictions decreases when α increases contrary to the one-tier situation. The effects at work for the variations of the performance ability cost k is better understood from an initial situation where it is large: a decrease in k then allows the country to increase the administration's performance ability of all jurisdictions so that the scopes of services of the tiers could be increased and thus the number of layers necessary to cover the bundle of public goods decreased. The same reasoning applies for an increase in β : because citizens put more weight on the administration capacities, it acts as a decrease in k .

4 Decentralized organization

So far, we have assumed that jurisdictions are ruled by social planner appointees without leeway to pursue their own political agendas. We now consider (elected) governments that decide on their own jurisdiction's range of services and performance ability. Every jurisdiction thus has the possibility of doing what it perceives as the best for its citizenry. However, jurisdictional authorities may have concerns and priorities that are not perfectly aligned with the social planner's because their actions are limited to their own jurisdictions. This could be because they don't have a say on the decisions of the other tiers' jurisdictions where their constituents also belong, or simply because the elected officials in these jurisdictions make decisions that concern either a larger group

of citizens (the upper tier jurisdiction which is “above” them in the CF), or subsets of their constituents (jurisdictions in a tier “below” them). Hence, the social planner should somehow coordinate the tiers’ decisions in such a way that jurisdictions’ ranges of services do not overlap.⁹ He also has to determine the number of jurisdictions at every level (n_t) and the number of sub-national levels (T) that will maximize the overall welfare, anticipating the decisions of public decision-makers at each tier.

To investigate this problem, we assume that decision-makers at each tier decide on their range of services and the performance ability of their administrations in order to maximize the welfare of their constituencies, given by (2), neglecting the choices of other tiers below or above them. Hence, decentralization in our setup corresponds to forcing the social planner to abide by the choices of decision-makers with narrow objectives. More precisely, decentralization is introduced into the CF problem as a multi-stage delegation game where the social planner first determines the number of jurisdictions in each tier level (n_t for each potential t , with $n_0 = 1$), and then delegates to their decision-makers the choice of their administration’s range of services and performance ability in an order of precedence that corresponds to the citizen representation. Delegation to decision-makers is thus operated successively from tier 0 to the last tier: the planner first asks the decision-maker of tier 0 to choose g_0 and x_0 . This policy maximizes (2) where $t = 0$, $n_0 = 1$, $x_{-1} = 1$. Then, the planner asks the decision-makers of tier $t = 1$ to do the same for their respective constituents under the constraint that the lower bound of their scope of services, x_0 , corresponds to the choice made by tier $t = 0$ to avoid that their services overlap. Given their constituents, they all maximize (2) where $t = 1$, $x_{t-1} = x_0$ with respect to x_1 and g_1 . The planner then asks the decision-makers of tier $t = 2$ to do the same for their respective citizenry under the constraint that the lower bound of their jurisdiction’s range of services corresponds to the choice of the tier $t = 1$ decision-maker that rules the jurisdiction above them. And so on and so on until the whole range of public goods is provided.¹⁰

Given this delegation game, the problem of the planner is to determine

$$\max_{T, \{n_t\}_{t=0}^T} \left\{ \sum_{t=0}^T \max_{x_t, g_t} \{W(x_{t-1}, x_t, g_t, n_t) | x_{t-1}, n_t\} : n_0 = 1, x_{-1} = 1, x_T = \bar{x} \right\}. \quad (10)$$

Before deriving the resulting CF, it is interesting to determine how a jurisdiction

⁹We suppose that the planner cannot discriminate between jurisdictions of the same level: he has to impose the same rules to all jurisdictions belonging to the same tier.

¹⁰We assume that jurisdictions do not coordinate either between themselves (horizontally) or with other tiers (vertically).

with objective (2) would modify the allocation of services and/or the capacity of its administration compared to the first-best levels if it faces different types of constraints.

Proposition 4 *If assigned their scope $(x_{t-1}^*, x_t^*]$, jurisdictions of tier- t would choose the same performance ability as the planner (g_t^*). However, if assigned g_t^* , these jurisdictions have an incentive to carry out the services assigned to upper tiers in the CF, and to lower tiers too if β is lower than $\beta'_s = \Phi^{-1}(1) \in (\max\{1/4, \beta_s\}, 1/2)$ where $\Phi(\beta) = 2\beta\phi(\beta)/(4\beta - 1)$.*

It is not surprising that if decision-makers are constrained in their range of services, they would not change the capacity of their administration compared to the centralized choices since their objective in that case coincide with the one of the social planner. The changes of range of services mentioned in Prop. 4 are also intuitive: because decision-makers consider the result of only their own decisions, ignoring the activity of other jurisdictions, they would like to increase the range of services of their administration. If it were possible, it would always be done by taking on the provision of low-cost public services (the provision cost decreases when moving downward the lower bound of the administration's range of services), and if the citizen's preferences for the range of services are sufficiently large (i.e. β relatively low), by also taking over services costlier than the ones they are charged of providing under a centralized organization.

For the sake of argument, we now investigate partial decentralization, a situation in which the general structure of the CF is not much changed compared to the centralized state. More specifically, we analyze the effects of the delegation process explained above assuming that the social planner implements the same number of jurisdictions per tier as under centralization, given by n_t^* . To be consistent, it should be the case that the range of services resulting from the decentralization process are not too much altered, hence that β is not too different from β'_s , so that the CF entails the same number of tiers. Given the size of their jurisdictions and the lower bound of their range of services, the tier- t decision-makers solve

$$\max_{x_t, g_t} \{W(x_{t-1}, x_t, g_t, n_t^*) | x_{t-1}, n_t^*\} \quad (11)$$

where $n_t^* = \lambda^{t/(1-2\beta)}$, $x_T = \bar{x}$ and $x_{-1} = 1$. It is shown in the appendix that

Proposition 5 *Without reforming the vertical structure of government (same numbers of jurisdictions per tier as under a centralized state), the range of services chosen by decision-makers is reduced (increased) while the performance ability of their administrations is increased (reduced) compared to the centralized levels if β is greater (lower)*

than β'_s . In any case, proceeding down the CF, the range of the tiers' bundle of public goods increases.

Hence, the decentralized public structure is identical to the first-best only in the case where $\beta = \beta'_s$. Otherwise, compared to the centralized levels, scopes are reduced and administration capacities are increased if the citizens' preference for the performance ability of the public administration is large, and the reverse otherwise. This is easily understood from Prop. 4 where threshold β'_s is defined. Indeed, in our delegation game, policy-makers can change the scope of their administration by modifying only the upper-bound of their range of public goods since the lower bound is imposed. From Prop. 4, we know that when their first-best administration's performance ability is imposed upon them, policy-makers are tempted to increase this upper-bound only if the citizens' relative preference for the range of services is large, hence the relative preference for the administration's performance ability is low, a condition given by $\beta < \beta'_s$. Suppose this is the case (a symmetric reasoning applies when $\beta > \beta'_s$). As the increase in range of services increases the provision cost compared to first-best (both x_t and x_{t-1} are increased for all tiers but the state), it is also optimal for policy-makers to save on provision costs by reducing their administration performance ability.

While scopes and capacities are changed at each level, scopes are still larger the lower the jurisdictions in the hierarchy of tiers. Observe that because scopes are reduced when β is larger than β'_s , the number of tiers may not be sufficient to provide the whole range of public goods.

The previous results are derived assuming a partial decentralization process, the number of tiers and the numbers of jurisdictions per tier corresponding to the optimal one under centralization. When the number of tiers and of jurisdictions per tier are also optimally chosen, as in (10), we obtain:

Proposition 6 *The optimal CF under decentralization is characterized by*

$$T^{**} = \frac{\ln \bar{x}}{\ln \lambda_d} - 1, x_t^{**} = \lambda_d^{t+1}, n_t^{**} = \lambda_d^{t/(1-2\beta)}, g_t^{**} = g_0^{**} \lambda_d^{-t(1+2\beta)/(1-2\beta)}, u_t^{**} = u_0^{**} \lambda^{t(1-4\beta)/(1-2\beta)} \quad (12)$$

where λ_d solves

$$\frac{\lambda_d + 1}{(1/\lambda_d + 1 - 2\beta)^{1-\beta}} = \frac{\beta^\beta 2^{2-\beta}}{\alpha^{1-\beta} k^\beta}, \quad (13)$$

$$g_0^{**} = \frac{\alpha\beta(\lambda_d - 1)}{2k(1/\lambda_d + 1 - 2\beta)}, \quad (14)$$

and

$$u_0^{**} = \frac{\alpha(\lambda_d^2 - 1)}{4(1/\lambda_d + 1 - 2\beta)}. \quad (15)$$

The welfare at tier- t is given by $W_t^{**} = (1 - 2\beta)u_t^{**}$ and the total welfare by

$$W^{**} = (1 - 2\beta)u_0^{**} \frac{1 - \lambda^{(T^{**}+1)(1-4\beta)/(1-2\beta)}}{1 - \lambda^{(1-4\beta)/(1-2\beta)}}$$

if $\beta \neq 1/4$ and $W^{**} = (1 - 2\beta)u_0^{**}(T^{**} + 1)$ otherwise.

Observe that the expressions given by (12) and (6) are similar. The only differences come from the magnification parameter λ_d , which is implicitly given by (13), the performance ability of the central government g_0^{**} and the corresponding utility u_0^{**} . Consequently, the optimal decentralized government shares the characteristics of the centralized one, as described in Prop. 2, i.e. a constant jurisdictional dispersion, jurisdiction's scopes that are increasing and administration capacities that are decreasing (overall and per capita) proceeding down the CF, and a threshold value for β equal to $1/4$ for the ranking of the citizen welfare (and of the per capita administration cost) along the CF. However, comparing the expressions (7) and (13), we obtain that $\lambda_d > \lambda$ if β is lower than β'_s and the reverse otherwise. We can thus state that

Proposition 7 *The optimal regional apportionment under decentralization entails fewer (more) jurisdictions per tier, with reduced (increased) range of services and a greater (smaller) number of tiers than under the centralized organization if β is greater (lower) than β'_s .*

This result is a direct consequence of the previous ones. Consider the case $\beta > \beta'_s$: as a decision-maker would like to decrease the range of services of its jurisdiction compared to first-best since the citizens' preferences for the range of services is low relative to the performance ability, it is optimal for the social planner to compensate for this effect by reducing the number of jurisdictions per tier. However, he also has to increase the number of tiers in order to ensure the provision of the entire range of public goods.

5 Eliciting citizen preferences

In this section, we illustrate how the equilibrium conditions derived above can be used to estimate the preference parameter β from a data set with standard econometric

procedures. Using $x_t = x_{t-1}\lambda$ and $g_t = g_{t-1}\lambda^{-(1+2\beta)/(1-2\beta)}$ we obtain

$$\frac{C_{t+1}}{C_t} = \frac{kg_{t+1}(x_{t+1} + x_t)/2}{kg_t(x_t + x_{t-1})/2} = \frac{g_{t+1}x_t}{g_t x_{t-1}} = \frac{g_{t+1}\lambda}{g_t} = \lambda^{-4\beta/(1-2\beta)}$$

while we have $n_t = n_{t-1}\lambda^{1/(1-2\beta)}$. Taking the logarithm of both expressions and eliminating the term involving $\ln \lambda$ on both sides, we get

$$\ln(C_t/C_{t+1}) = 4\beta \ln(n_{t+1}/n_t) \quad (16)$$

which holds for any equilibrium value λ , hence for a centralized or a decentralized organization of the country. Using data on jurisdiction spending, it is possible to compute $Y_{z_{t+1},d} \equiv \ln(C_{z_t,d}/C_{z_{t+1},d})$ for each jurisdiction $z_{t+1} \in \{1, \dots, n_{t+1}\}$ belonging to tier- $t+1$ located below jurisdiction z_t in the CF at each date d . As $1/n_t$ corresponds to the share of the population of a tier- t jurisdiction, $n_{z_{t+1},d}/n_{z_t,d}$ corresponds to the inverse of the share of the z_t citizenry that belongs to jurisdiction z_{t+1} at date d . These individual characteristics, together with other observable characteristics (e.g., disposable income, population density), can be used to estimate the coefficient β using standard panel data procedures.

We apply this methodology on annual federal and state expenditures in the United States over the period 1977-2015 collected by the Tax Policy Center¹¹ to estimate the following relationships

$$Y_{z_1,d} = \beta.P_{z_1,d} + a.X_{z_1,d} + \varphi_{z_1} + \delta_d + \varepsilon_{z_1,d} \quad (17)$$

$$Y_{z_1,d} = \sum_{\tau=1}^4 \beta_\tau \mathbb{1}_{\{z_1 \in \tau\}} P_{z_1,d} + a.X_{z_1,d} + \varphi_{z_1} + \delta_d + \varepsilon_{z_1,d} \quad (18)$$

$$Y_{z_1,d} = \sum_{Z \in \{A,B\}} \beta_Z \mathbb{1}_{\{z_1 \in Z\}} P_{z_1,d} + a.X_{z_1,d} + \varphi_{z_1} + \delta_d + \varepsilon_{z_1,d} \quad (19)$$

where z_1 indexes the states ($z_1 \in \{1, \dots, 50\}$) and $d = 1977, \dots, 2015$. Here, $Y_{z_1,d}$ corresponds to the log of the date- d ratio of the federal expenditures (net of grants to state and local jurisdictions) over state z_1 government spending, $P_{z_1,d} = 4 \ln(pop_{0,d}/pop_{z_1,d})$ accounts for the corresponding population share, $X_{z_1,d}$ is a vector of time-varying control variables composed of the income per capita, state population density and the partisan composition of the state legislatures. The income per capita captures the

¹¹Details on the Tax Policy Center are available on its website <https://www.urban.org/policy-centers/urban-brookings-tax-policy-center>. Queries on its dataset can be performed at <http://sldqs.taxpolicycenter.org/index.cfm>.

Table 1: Summary statistics

Variables	# of Obs.	Mean	SD	Min	Max
ln of expenditure ratio (Y) ¹	1950	4.93	0.95	2.23	6.79
$4 \times$ ln of population ratio (P) ²	1950	17.58	4.03	8.43	25.43
Income per capita ²	1950	25.41	12.58	5.35	68.33
Pop. density ^{2,3}	1950	148.33	190.21	0.6	1024.4
Dem. majority chambers ⁴	1862*	0.49	0.5	0	1
Rep. majority chambers ⁴	1862*	0.3	0.46	0	1

*As Nebraska is a non-partisan, unicameral legislature, it is excluded from the regressions that include partisan composition.

Sources: ¹the Tax Policy Center & the U.S. Government Publishing Office, ²the Bureau of Economic Analysis, ³the Census Bureau, ⁴the National Conference of State Legislatures.

demand for public goods and services, which is expected to increase state expenditures (see for instance Borchering & Deacon, 1972, and Ladd, 1992, for estimations on U.S. data), and therefore to reduce $Y_{z_1,d}$ in our setting. We also consider the partisan composition of state legislatures, as the control of both chambers by democrats is usually shown to lead to significantly higher state expenditures per capita.¹² The variable φ_{z_1} corresponds to state-specific effects that capture time-invariant unobserved heterogeneity, δ_d are year fixed effects accounting for the influence of variables affecting all states identically in year d , and $\varepsilon_{z_1,d}$ is the error term. Table 1 presents the summary statistics of all variables used in our regressions.

Equations (17)–(19) differ by the way β is estimated. In (17), we suppose that citizens preferences are the same whatever the state, while (18) allows for a heterogeneity among the four regional divisions used by the Census Bureau (Midwest, Northeast, South, West), that we denote by $\tau = 1, \dots, 4$. Accordingly, $\mathbb{1}_{\{z_1 \in \tau\}}$ is a dummy variable equal to 1 if the state z_1 belongs to region τ , and 0 otherwise. Therefore, $\hat{\beta}_1$ is the measure of β for the Midwest, $\hat{\beta}_2$ the one corresponding to the Northeast, $\hat{\beta}_3$ for the South and $\hat{\beta}_4$ for the West. The last specification, (19), allows us to test for a coarser partition of the states into two groups, denoted by A (Northeast & West) and B (Midwest & South) with $\mathbb{1}_{\{z_1 \in A\}}$ and $\mathbb{1}_{\{z_1 \in B\}}$ being the corresponding dummy variables. Therefore, $\hat{\beta}_A$ is the measure of β for the West and Northeast regions taken together, and $\hat{\beta}_B$ for the South and Midwest regions.

For the three specifications (17)–(19), we control for the correlation of the error term over time at the state level using cluster-robust standard errors in our panel fixed

¹²See Besley & Case (2003) for a survey of the numerous empirical investigations of the role of political institutions in the United States.

effects estimations. Table 2 gives the estimation results.

Table 2: Estimation results

Model specification	(1)	(2)	(3)	(4)
All states ($\hat{\beta}$)	0.184*** (0.0238)	0.182*** (0.0254)		
Midwest ($\hat{\beta}_1$)			0.279*** (0.0624)	
Northeast ($\hat{\beta}_2$)			0.182** (0.0579)	
South ($\hat{\beta}_3$)			0.238*** (0.0309)	
West ($\hat{\beta}_4$)			0.147*** (0.0255)	
Northeast & West ($\hat{\beta}_A$)				0.153*** (0.0235)
Midwest & South ($\hat{\beta}_B$)				0.251*** (0.0342)
Income per capita	-0.0025 (0.0027)	-0.0029 (0.0029)	-0.0026 (0.0031)	-0.0021 (0.0029)
Pop. density	-0.0006 (0.0005)	-0.0006 (0.0005)	-0.00006 (0.0006)	-0.00004 (0.0006)
Dem. chambers		0.0224 (0.0136)	0.0185 (0.0138)	0.0197 (0.0132)
Rep. chambers		0.0136 (0.0166)	0.0178 (0.0168)	0.0194 (0.0161)
Observations	1950	1862	1862	1862
Hausman's test	23.82***	36.36***	50.33***	65.22***
F-Stat	473.2***	479.8***	845.2***	1021.7***

Notes:*** indicates statistical significance at 1%, ** at 5%, and * at 10% level. Standard errors clustered at the state level are in parentheses. All estimations include year fixed effects. Estimates of the constant term and of the time-dummies are omitted.

The value of β estimated for the whole country is 0.18 (columns (1) and (2); the differences between them being because the state legislature variables are omitted in the first one)¹³, which belongs to the interval $(0,1/2)$ required by our theoretical model to ensure a positive welfare. Furthermore, as β is lower than $1/4$, it suggests that americans are more satisfied by the services provided at the state level than at the federal

¹³As a consequence, Nebraska is included only in regression (1).

Table 3: Partisan views of government

Partisanship	Federal	State	Local
Rep.	13%	57%	63%
Ind.	27%	59%	60%
Dem.	41%	56%	67%
Overall	28%	57%	63%

Source: Pew Research Center (2013).

Answers to the question “Would you say your overall opinion of (government level) is very favorable, mostly favorable, mostly unfavorable, or very unfavorable?” Reported are the proportions corresponding to the sum of “very favorable” and “mostly favorable” by partisanship. Survey conducted March 13-17, 2013 among 1,501 adults living in all 50 U.S. states and the District of Columbia.

one. Observe that none of the estimates corresponding to the partisan composition of state legislatures and the other characteristics (income and density) are statistically significant in the regressions. These results are corroborated by a Pew Research Center’s survey, reported in Table 3 (Pew Research Center, 2013). It shows that whatever their partisanship, individuals express a more favorable view of their local government than their state government, and that the federal government in Washington earns the lowest percentage points of favorable opinions.

Disaggregating β over the four U.S. regions (column (3) of Table 2) reveals a partition of the U.S. territory. Indeed, we obtain that this coefficient is lower in the Northeast and West (0.182 and 0.147) than in the Midwest and South (0.279 and 0.238). Tests for equality of the estimated regions’ β s, presented in Table 4, confirm that the pairwise equalities $\hat{\beta}_1 = \hat{\beta}_2$ (Midwest and Northeast), $\hat{\beta}_1 = \hat{\beta}_4$ (Midwest and West) and $\hat{\beta}_3 = \hat{\beta}_4$ (South and West) are rejected at the 10% significance level (even at the 5% level for the latter equality).

Table 4: Equality tests

H_0	$\hat{\beta}_1 = \hat{\beta}_2$	$\hat{\beta}_1 = \hat{\beta}_3$	$\hat{\beta}_1 = \hat{\beta}_4$	$\hat{\beta}_2 = \hat{\beta}_3$	$\hat{\beta}_2 = \hat{\beta}_4$	$\hat{\beta}_3 = \hat{\beta}_4$	$\hat{\beta}_A = \hat{\beta}_B$
F-test	3.07	0.52	3.28	0.88	0.27	4.76	5.89
p-value	0.0862	0.4726	0.0764	0.3517	0.6041	0.0341	0.0190

Column (4) of Table 2 gives the estimates of this coefficient when the regions are grouped into two zones, Northeast & West (zone *A*), and Midwest & South (zone *B*). They confirm a value much larger for Midwest & South than for Northeast & West

(0.251 versus 0.153). The last column of Table 4 shows that the equality of these estimates is rejected at the 5% confidence level. Hence, from Prop. 2, we may expect that on average, citizens in the Northeast and West regions value state-provided services more than federally provided services, whereas citizens in the Midwest and South are more likely to be equally satisfied with both levels of government.

6 Conclusion

The problem of organizing a country's government entails both horizontal and vertical dimensions and raises the question of the allocation of public services over the tiers. We propose a simple model that allows us to characterize the desirable features of such an undertaking under both a centralized and a decentralized state. Our approach offers a first theoretical foundation of an endogenous multi-tier setting. It highlights the differences between the social optimum and the result of having autonomous jurisdictions. Indeed, decision-makers, whether at the central tier or at sub-national tiers, have an incentive to modify the size of their administration and the range of their services, with respect to the first-best levels. The results we obtain also allow us to perform structural empirical investigations, as illustrated by the elicitation methods of the citizen preference parameter that we detail and apply using U.S. data. This work could be extended on both fronts. From an empirical perspective, organizational choices made by a country that has decided to give more leeway to decision-makers should reveal citizens preferences: if they put more weight on the overall range of services provided by the administrations than on their ability to better perform their tasks, scopes of tiers should increase and the number of bureaucratic layers should decrease compared to the centralized structured. From a theoretical perspective, an avenue of research is to allow for a heterogeneous distribution of the population in order to relax the perfect symmetry of our framework. Also we suppose that the central planner has perfect information about the citizens preferences. Information asymmetry, with sub-national decision-makers being better informed about their constituency than their higher level counterparts, could impact the organization of the government.

References

- Alesina, A. & Spolaore, E. (1997). On the number and size of nations. *The Quarterly Journal of Economics*, 112(4), 1027–1056.
- Besley, T. & Case, A. (2003). Political institutions and policy choices: Evidence from the united states. *Journal of Economic Literature*, 41(1), 7–73.
- Bogomolnaia, A., Le Breton, M., Savvateev, A., & Weber, S. (2006). Stability under unanimous consent, free mobility and core. *Int J Game Theory*, 35(2), 185–204.
- Bogomolnaia, A., Le Breton, M., Savvateev, A., & Weber, S. (2008). Heterogeneity gap in stable jurisdiction structures. *Journal of Public Economic Theory*, 10(3), 455–473.
- Bolton, P. & Roland, G. (1997). The breakup of nations: A political economy analysis. *The Quarterly Journal of Economics*, 112(4), 1057–1090.
- Borcherding, T. E. & Deacon, R. (1972). The demand for the services of non-federal governments. *American Economic Review*, 62(5), 891–901.
- Casella, A. (2001). The role of market size in the formation of jurisdictions. *The Review of Economic Studies*, 68(1), 83–108.
- Demange, G. (1994). Intermediate preferences and stable coalition structures. *Journal of Mathematical Economics*, 23, 45–58.
- Dexia (2007). Subnational public finance in the european union. *Dexia Economic Outlook*, Trends 2000 / 2006.
- Friedman, D. (1977). A theory of the size and shape of nations. *Journal of Political Economy*, 85(1), 59–77.
- Greenberg, J. & Weber, S. (1986). Strong tiebout equilibrium under restricted preferences domain. *Journal of Economic Theory*, 38, 101–117.
- Guesnerie, R. & Oddou, C. (1981). Second best taxation as a game. *Journal of Economic Theory*, 25(1), 67–91.
- Hochman, O., Pines, D., & Thisse, J. F. (1995). On the optimal structure of local governments. *The American Economic Review*, 85(5), 1224–1240.
- Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39(153), 41–57.
- Jehiel, P. & Scotchmer, S. (2001). Constitutional rules of exclusion in jurisdiction formation. *The Review of Economic Studies*, 68(2), 393–413.
- Ladd, H. (1992). Population growth, density and the costs of providing public services. *Urban Studies*, 29(2), 273–295.

- OECD (2014). Regions and cities: Where policies and people meet. *OECD Regional Outlook*.
- Panizza, U. (1999). On the determinants of fiscal centralization: Theory and evidence. *Journal of public economics*, 74(1), 97–139.
- Pew Research Center (April 15, 2013). State governments viewed favorably as federal rating hits new low. Pew Research Center, Washington, D.C. <http://www.people-press.org/category/publications/2013/page/4/> (as at August 2018).
- Wilson, J. D. & Janeba, E. (2005). Decentralization and international tax competition. *Journal of Public Economics*, 89(7), 1211–1229.

Appendix

A Proof of lemma 1

The program of the benevolent planner boils down to

$$\max_{T, s_{z_t}, \bar{\ell}_{z_t}, x_t, g_{z_t}} \sum_{t=0}^T \left[g_{z_t}^\beta (x_t - x_{t-1})^{1-\beta} - \frac{\alpha}{2} (x_t^2 - x_{t-1}^2) \sum_{z_t=1}^{n_t} s_{z_t} \bar{\ell}_{z_t} - k g_{z_t} \frac{x_t + x_{t-1}}{2} \right]$$

where s_{z_t} and $\bar{\ell}_{z_t}$ are the jurisdiction z_t 's size and average distance respectively, with $\sum_{z_t=1}^{n_t} s_{z_t} = 1$ for all t . Because individuals are located uniformly over the territory, $\bar{\ell}_{z_t}$ is minimized when the government —and thus the provision of public goods— is located in the middle of the jurisdiction whatever the tier and the number of jurisdictions in a tier. As a consequence, the average distance to their administration is $\bar{\ell}_{z_t} = \frac{1}{s_{z_t}/2} \int_0^{s_{z_t}/2} \ell d\ell = s_{z_t}/4$. We can thus rewrite the program as

$$\max_{T, s_{z_t}, x_t, g_{z_t}} \sum_{t=0}^T \left[g_{z_t}^\beta (x_t - x_{t-1})^{1-\beta} - \frac{\alpha}{8} (x_t^2 - x_{t-1}^2) \sum_{z_t=1}^{n_t} s_{z_t}^2 - k g_{z_t} \frac{x_t + x_{t-1}}{2} \right] \quad (20)$$

under the condition $\sum_{z_t=1}^{n_t} s_{z_t} = 1$ for all t , where $\sum_{z_t=1}^{n_t} s_{z_t}^2$ is minimized when tier t jurisdictions are of equal size: $s_{z_t} = 1/n_t$ for all t . As jurisdictions are of equal size and their range of services are identical, we also have the same administration's performance ability at each tier level: $g_{z_t} = g_t$ for all t . Replacing in (20) and developing gives (1).

B Proof of Proposition 1

The first-order conditions (FOC) w.r.t. n_t and g_t lead to

$$\frac{\alpha (x_t^2 - x_{t-1}^2)}{4n_t^2} - k g_t (x_t + x_{t-1}) = 0 \quad (21)$$

and

$$\beta u_t - n_t k g_t (x_t + x_{t-1}) / 2 = 0. \quad (22)$$

From (21) and (22) we get

$$\frac{\alpha (x_t^2 - x_{t-1}^2)}{8n_t} = \frac{n_t k g_t (x_t + x_{t-1})}{2} = \beta u_t, \quad (23)$$

and the welfare (1) simplifies to $\sum_{t=0}^T u_t(1-2\beta)$ which is positive if $\beta < 1/2$. The FOC w.r.t. x_t is given by

$$(1-\beta) \left[\frac{u_t}{x_t - x_{t-1}} - \frac{u_{t+1}}{x_{t+1} - x_t} \right] - \frac{\alpha x_t}{4} \left(\frac{1}{n_t} - \frac{1}{n_{t+1}} \right) - \frac{k}{2} (n_t g_t + n_{t+1} g_{t+1}) = 0. \quad (24)$$

Using (22) to substitute for $n_t g_t$ and $n_{t+1} g_{t+1}$ in the last term gives

$$(1-\beta) \left[\frac{u_t}{x_t - x_{t-1}} - \frac{u_{t+1}}{x_{t+1} - x_t} \right] - \frac{\alpha x_t}{4} \left(\frac{1}{n_t} - \frac{1}{n_{t+1}} \right) - \beta \left[\frac{u_t}{x_t + x_{t-1}} + \frac{u_{t+1}}{x_{t+1} + x_t} \right] = 0.$$

Multiplying by $n_{t+1} n_t$ to get

$$(1-\beta) \left(\frac{n_{t+1} n_t u_t}{x_t - x_{t-1}} - \frac{n_{t+1} n_t u_{t+1}}{x_{t+1} - x_t} \right) - \frac{\alpha x_t}{4} (n_{t+1} - n_t) - \beta \left(\frac{n_{t+1} n_t u_t}{x_t + x_{t-1}} + \frac{n_{t+1} n_t u_{t+1}}{x_{t+1} + x_t} \right) = 0,$$

and using (23) to substitute for $n_t u_t$ and $n_{t+1} u_{t+1}$ yields

$$\frac{1-\beta}{\beta} [n_{t+1}(x_t + x_{t-1}) - n_t(x_{t+1} + x_t)] - 2x_t(n_{t+1} - n_t) - [n_{t+1}(x_t - x_{t-1}) + n_t(x_{t+1} - x_t)] = 0.$$

Solutions of this equation are given by $x_t = \lambda x_{t-1} = \lambda^{t+1}$ using $x_{-1} = 1$, and $n_t = \mu n_{t-1} = \mu^t$ using $n_0 = 1$. Indeed, substituting we obtain

$$\frac{1-\beta}{\beta} (\lambda + 1) (\mu - \lambda) \mu^t \lambda^t - 2\lambda^{t+1} (\mu - 1) \mu^t - (\lambda - 1) (\mu + \lambda) \mu^t \lambda^t = 0,$$

which upon factorizing out $\mu^t \lambda^t$ simplifies to

$$\frac{1-\beta}{\beta} (\lambda + 1) (\mu - \lambda) - 2\lambda (\mu - 1) - (\lambda - 1) (\mu + \lambda) = 0.$$

One solution of this equation is $\mu = \lambda = 1$ which implies $T^* = 0$, hence only one tier. More generally, a solution (λ, μ) of this equation can be expressed as (λ, λ^r) , i.e. $\mu = \lambda^r$, where λ and r are deduced from (21) and (23) as follows. Substituting $x_t = \lambda^{t+1}$ and $n_t = \lambda^{rt}$ in (21) yields

$$\frac{\alpha (\lambda^2 - 1) \lambda^{2t}}{4\lambda^{2rt}} - k g_t (\lambda + 1) \lambda^t = 0,$$

hence

$$g_t = \frac{\alpha (\lambda - 1)}{4k\lambda^{(2r-1)t}} = g_0 \lambda^{-(2r-1)t}. \quad (25)$$

As

$$u_t = g_t^\beta (x_t - x_{t-1})^{1-\beta} = \left(\frac{\alpha}{4k}\right)^\beta \frac{\lambda - 1}{\lambda^{(2r\beta-1)t}} \quad (26)$$

we obtain, using (23),

$$\beta \left(\frac{\alpha}{4k}\right)^\beta \frac{\lambda - 1}{\lambda^{(2r\beta-1)t}} = \lambda^{rt} k \frac{\alpha(\lambda - 1)}{8k\lambda^{(2r-1)t}} (\lambda + 1) \lambda^t = \frac{\alpha(\lambda^2 - 1)}{8\lambda^{(r-2)t}}.$$

Collecting terms, we get

$$\frac{8\beta}{\alpha} \left(\frac{\alpha}{4k}\right)^\beta = (\lambda + 1) \lambda^{(2r\beta-r+1)t}$$

that must hold for all t . We thus must have $2r\beta - r + 1 = 0$, hence

$$r = (1 - 2\beta)^{-1} \quad (27)$$

and (7). We have $\lambda > 1$ if

$$1 > \frac{k^\beta}{\beta} \left(\frac{\alpha}{4}\right)^{1-\beta} \equiv \phi(\beta) \quad (28)$$

where $\beta \in (0, 1/2)$. We have $\lim_{\beta \searrow 0} \phi(\beta) = +\infty$ and $\phi'(\beta)/\phi(\beta) = -1/\beta - \ln(\alpha/4k)$ which is negative for all $\beta \leq 1/2$ if $\ln(\alpha/4k) > -2$, i.e. $\alpha/k > (2/e)^2 \approx .54$. Under this condition, we have $\lambda > 1$ if $1 > \phi(1/2) = \sqrt{\alpha k}$ and β greater than $\beta_s \equiv \phi^{-1}(1) < 1/2$. We thus must have $1/k > \alpha > (2/e)^2 k$, which is possible only if $k < e/2$. The number of tiers is deduced from $x_T = \lambda^{T+1}$. From (26) and (27), we get

$$u_t = u_0 \lambda^{-(2r\beta-1)t} = u_0 \lambda^{-(r-2)t}$$

where $r - 2 = (4\beta - 1)/(1 - 2\beta)$ is positive iff $\beta > 1/4$. Using (7), we get

$$u_0 = \left(\frac{\alpha}{4k}\right)^\beta (\lambda - 1) = \frac{\alpha(\lambda^2 - 1)}{8\beta}.$$

As $W^* = (1 - 2\beta) \sum_{t=0}^T u_t^*$, we get

$$W^* = (1 - 2\beta) u_0 \frac{1 - \lambda^{-(r-2)(T+1)}}{1 - \lambda^{-(r-2)}}$$

when $\beta \neq 1/4$, and $W^* = (1 - 2\beta)u_0(T + 1)$ otherwise.

C Proof of Proposition 3

Using (28), $\phi(\beta)$ is decreasing for all $\beta \leq 1/2$ under the condition $\alpha/k > (2/e)^2$, and increases with α and k . Hence, from (7) which can be written as $\lambda = 2/\phi(\beta) - 1$, λ increases with β and decreases with α and k and so does $x_t^* - x_{t-1}^* = \lambda^t(\lambda - 1)$. As $T^* = \ln \bar{x}/\ln \lambda - 1$, T^* decreases with β and increases with α and k . Using $\ln n_t^* = (t \ln \lambda)/(1 - 2\beta)$, we get $(dn_t^*/d\beta)/n_t^* = (2t \ln \lambda)/(1 - 2\beta)^2 + t(d\lambda/d\beta)/(\lambda(1 - 2\beta)) > 0$. As $n_t^* = \lambda^{t/(1-2\beta)}$ is an increasing function of λ , it decreases with α and k .

D Proof of Proposition 4

Denoting $W_t \equiv W(x_{t-1}, x_t, g_t, n_t)$ and differentiating (2) w.r.t. g_t gives

$$\frac{\partial W_t}{\partial g_t} = \beta \frac{u_t}{g_t} - \frac{n_t k (x_t + x_{t-1})}{2} \quad (29)$$

which is null at the social planner's optimum. Hence, if assigned scope (x_{t-1}^*, x_t^*) , the jurisdiction would chose the same administration's performance ability. Differentiating (2) w.r.t. x_{t-1} and x_t gives

$$\frac{\partial W_t}{\partial x_{t-1}} = -\frac{(1 - \beta)u_t}{x_t - x_{t-1}} + \frac{\alpha x_{t-1}}{4n_t} - \frac{n_t k g_t}{2} \quad (30)$$

and

$$\frac{\partial W_t}{\partial x_t} = \frac{(1 - \beta)u_t}{x_t - x_{t-1}} - \frac{\alpha x_t}{4n_t} - \frac{n_t k g_t}{2}. \quad (31)$$

Using (23) to substitute for the first term of (30) evaluated at the social planner's optimum, we get

$$\begin{aligned} \frac{\partial W_t}{\partial x_{t-1}} &= \frac{\alpha x_{t-1}}{4n_t} - \frac{n_t k g_t}{2} \left(\frac{1 - \beta}{\beta} \frac{x_t + x_{t-1}}{x_t - x_{t-1}} + 1 \right) \\ &\leq \frac{\alpha x_{t-1}}{4n_t} - \frac{n_t k g_t}{2} \left(\frac{x_t + x_{t-1}}{x_t - x_{t-1}} + 1 \right) = \frac{\alpha x_{t-1}}{4n_t} - \frac{n_t k g_t x_t}{x_t - x_{t-1}} \\ &= \frac{n_t x_t}{x_t - x_{t-1}} \left(\frac{x_{t-1}}{x_t} \frac{\alpha(x_t - x_{t-1})}{4n_t^2} - k g_t \right) \\ &< \frac{n_t x_t}{x_t - x_{t-1}} \left(\frac{\alpha(x_t - x_{t-1})}{4n_t^2} - k g_t \right) \\ &= 0 \end{aligned}$$

where the first inequality comes from $\beta \leq 1/2$, the second one from $x_{t-1} < x_t$ and the last equality from (21).

Using (23) to substitute for the first term of (31) evaluated at the social planner optimum give

$$\begin{aligned}
\frac{\partial W_t}{\partial x_t} &= \frac{n_t k g_t}{2} \left(\frac{1 - \beta}{\beta} \frac{x_t + x_{t-1}}{x_t - x_{t-1}} - 1 \right) - \frac{\alpha x_t}{4n_t} \\
&= \frac{n_t k g_t}{2} \left(\frac{1 - \beta}{\beta} \frac{x_t + x_{t-1}}{x_t - x_{t-1}} - 1 \right) - \frac{n_t k g_t x_t}{x_t - x_{t-1}} \\
&= \frac{n_t k g_t}{2} \left(\frac{1 - \beta}{\beta} \frac{x_t + x_{t-1}}{x_t - x_{t-1}} - \frac{3x_t - x_{t-1}}{x_t - x_{t-1}} \right) \\
&= \frac{n_t k g_t}{2\beta} \frac{x_t(1 - 4\beta) + x_{t-1}}{x_t - x_{t-1}} \\
&= \frac{n_t k g_t}{2\beta} \frac{x_{t-1}}{x_t - x_{t-1}} [\lambda(1 - 4\beta) + 1]
\end{aligned}$$

which is positive for all $\beta < 1/4$ and more generally if $\lambda(1 - 4\beta) + 1 > 0$. Using (7), this condition becomes

$$1 < \frac{2k^\beta}{4\beta - 1} \left(\frac{\alpha}{4} \right)^{1-\beta} \equiv \Phi(\beta) \quad (32)$$

where $\beta \in (1/4, 1/2)$. Observe that this condition is compatible with (28) since $\Phi(\beta) = 2\beta\phi(\beta)/(4\beta - 1) > \phi(\beta)$ whenever $\beta \in (1/4, 1/2)$. Differentiating $\ln \Phi(\beta)$ we obtain

$$\Phi'(\beta)/\Phi(\beta) = -4/(4\beta - 1) - \ln(\alpha/4k) < -4 - \ln(\alpha/4k) < -2$$

since $\beta \in (1/4, 1/2)$ and $\ln(\alpha/4k) > -2$ by assumption. We thus have $\Phi'(\beta) < 0$ for all $\beta \in (1/4, 1/2)$. As $\lim_{\beta \searrow 1/4} \Phi(\beta) = +\infty$ and $\Phi(1/2) = \sqrt{\alpha k} < 1$, there exists a unique threshold $\beta'_s \equiv \Phi^{-1}(1) \in (\beta_s, 1/2)$ such that $\partial W_t / \partial x_t > 0$ for all $\beta < \beta'_s$ and $\partial W_t / \partial x_t < 0$ for all $\beta \in (\beta'_s, 1/2)$. Static comparative exercises show that both functions Φ and ϕ increase with α and k .

E Proof of Proposition 5

As (29) is equal to 0 at the optimum, we have

$$\beta u_t = n_t k g_t (x_t + x_{t-1}) / 2. \quad (33)$$

Using (33) to substitute for u_t in (31) and equalizing to zero gives

$$0 = \frac{\partial W_t}{\partial x_t} = \frac{1 - \beta}{\beta} \frac{n_t k g_t (x_t + x_{t-1})}{2(x_t - x_{t-1})} - \frac{\alpha x_t}{4n_t} - \frac{n_t k g_t}{2},$$

hence

$$g_t = \frac{\alpha \beta x_t (x_t - x_{t-1})}{2k n_t^2 [(1 - 2\beta)x_t + x_{t-1}]}. \quad (34)$$

Using (33) to substitute for the first term in the decision-maker objective and using (34), we get

$$\begin{aligned} W_t &= \frac{\alpha (x_t^2 - x_{t-1}^2)}{8n_t} \left(\frac{1 - \beta}{\beta} \frac{2\beta x_t}{(1 - 2\beta)x_t + x_{t-1}} - 1 \right) \\ &= \frac{\alpha (x_t + x_{t-1}) (x_t - x_{t-1})^2}{8n_t [(1 - 2\beta)x_t + x_{t-1}]}. \end{aligned}$$

This solution is a maximum for the decision-maker at any given x_{t-1} if W_t is concave in (x_t, g_t) . As

$$\frac{\partial^2 W_t}{\partial x_t^2} = \frac{\partial^2 u_t}{\partial x_t^2} - \frac{\alpha}{4n_t} < 0, \quad \frac{\partial^2 W_t}{\partial g_t^2} = \frac{\partial^2 u_t}{\partial g_t^2} < 0, \quad \text{and} \quad \frac{\partial^2 W_t}{\partial x_t \partial g_t} = \frac{\partial^2 u_t}{\partial x_t \partial g_t} - \frac{n_t k}{2}, \quad (35)$$

we must verify that

$$D \equiv \left(\frac{\partial^2 W_t}{\partial x_t \partial g_t} \right)^2 - \frac{\partial^2 W_t}{\partial g_t^2} \frac{\partial^2 W_t}{\partial x_t^2} \leq 0.$$

Using (35) and

$$\left(\frac{\partial^2 u_t}{\partial x_t \partial g_t} \right)^2 - \frac{\partial^2 u_t}{\partial g_t^2} \frac{\partial^2 u_t}{\partial x_t^2} = 0,$$

we arrive at

$$\begin{aligned} D &= \left(\frac{\partial^2 u_t}{\partial x_t \partial g_t} \right)^2 - \frac{\partial^2 u_t}{\partial g_t^2} \frac{\partial^2 u_t}{\partial x_t^2} + \frac{\partial^2 u_t}{\partial g_t^2} \frac{\alpha}{4n_t} - n_t k \frac{\partial^2 u_t}{\partial x_t \partial g_t} + \left(\frac{n_t k}{2} \right)^2 \\ &= \frac{\partial^2 u_t}{\partial g_t^2} \frac{\alpha}{4n_t} - n_t k \frac{\partial^2 u_t}{\partial x_t \partial g_t} + \left(\frac{n_t k}{2} \right)^2. \end{aligned}$$

From (33), which can re-expressed as

$$\frac{n_t k}{2} = \frac{\beta u_t}{g_t} \frac{1}{x_t + x_{t-1}},$$

we get

$$\begin{aligned} n_t k \frac{\partial^2 u_t}{\partial x_t \partial g_t} &= n_t k \frac{\beta(1-\beta)u_t}{g_t(x_t - x_{t-1})} = \frac{2\beta^2(1-\beta)u_t^2}{g_t^2(x_t^2 - x_{t-1}^2)} > \frac{\beta^2 u_t^2}{g_t^2(x_t^2 - x_{t-1}^2)} \\ &> \frac{\beta^2 u_t^2}{g_t^2(x_t + x_{t-1})^2} = \left(\frac{n_t k}{2}\right)^2, \end{aligned}$$

where we have used $\beta < 1/2$, which implies $2(1-\beta) > 1$, and $(x_t^2 - x_{t-1}^2) < (x_t + x_{t-1})^2$.

We thus have

$$D < \frac{\partial^2 u_t}{\partial g_t^2} \frac{\alpha}{4n_t} < 0.$$

Without changing the regional apportionment, we have $n_t = \lambda^{rt}$. Suppose that first-best scopes are implemented, i.e. $x_t = x_t^* = \lambda^{t+1}$. Using (34) and (25), the resulting performance ability choices are given by

$$g_t = \frac{\alpha\beta\lambda(\lambda-1)}{2k\lambda^{(2r-1)t}[1+(1-2\beta)\lambda]} = \frac{2\beta\lambda g_t^*}{1+(1-2\beta)\lambda}. \quad (36)$$

We have $g_t \geq g_t^*$ iff $\lambda + 1 \leq 4\beta\lambda$ and thus, using the same arguments as those leading to (32), iff $\beta \geq \beta'_s$. As (33) is satisfied when $n_t = \lambda^{rt}$ and $x_t = x_t^*$ only if $g_t = g_t^*$, first-best scopes are implemented only in the case $\beta = \beta'_s$. More generally, plugging (34) into (29) to obtain

$$\frac{\partial W_t}{\partial g_t} = \beta \left(\frac{2kn_t^2[(1-2\beta)x_t + x_{t-1}]}{\alpha\beta\bar{x}_t} \right)^{1-\beta} - \frac{n_t k (x_t + x_{t-1})}{2}$$

where $n_t = \lambda^{rt}$, and denoting $x_t = \omega_t x_{t-1} = \prod_{i=0}^t \omega_i$, we arrive at

$$\lambda^{rt(1-2\beta)} \frac{2^{2-\beta}\beta^\beta}{\alpha^{1-\beta}k^\beta} \left(\frac{(1-2\beta)\omega_t + 1}{\omega_t} \right)^{1-\beta} = (\omega_t + 1) \prod_{i=0}^{t-1} \omega_i$$

which allows to determine the sequence $\{\omega_i\}_{i=0}^T$ recursively. Using (27) and (7), this equation can be written as

$$(\lambda + 1) \lambda^t = h(\omega_t) (\omega_t + 1) \prod_{i=0}^{t-1} \omega_i \quad (37)$$

where $h(\omega)$ is an increasing function defined by

$$h(\omega) \equiv \left(\frac{2\beta\omega}{1 + (1 - 2\beta)\omega} \right)^{1-\beta} \quad (38)$$

verifying $h(\lambda) \geq 1$ iff $\beta \geq \beta'_s$ using the same arguments as those leading to (32). Hence, the sequence $\{\omega_t\}_{t=0}^T$ defined by (37) is given by $\omega_t = \lambda$ for all t iff $\beta = \beta'_s$. Using (37) for $t + 1$ and for t , we get

$$h(\omega_{t+1}) (\omega_{t+1} + 1) \omega_t = h(\omega_t) (\omega_t + 1) \lambda \quad (39)$$

for all $t \geq 0$. As $h(\omega) (\omega + 1)$ is increasing, if $\{\omega_t\}_{t=0}^T$ is increasing (decreasing), all its terms ω_t are bounded above (below) by λ . The monotonicity of the $\{\omega_t\}_{t=0}^T$ for β close to β'_s is derived using a Taylor approximation of (39). We get

$$h(\omega_t) (\omega_t + 1) \lambda \approx h(\omega_t) (\omega_t + 1) \omega_t + \omega_t [h'(\omega_t) (\omega_t + 1) + h(\omega_t)] (\omega_{t+1} - \omega_t)$$

where

$$h'(\omega) = \frac{h(\omega)(1 - \beta)}{\omega[1 + (1 - 2\beta)\omega]}.$$

Substituting yields

$$\begin{aligned} h(\omega_t) (\omega_t + 1) \lambda &\approx h(\omega_t) (\omega_t + 1) \omega_t + \omega_t \left[\frac{h(\omega_t)(1 - \beta)}{\omega_t[1 + (1 - 2\beta)\omega_t]} (\omega_t + 1) + h(\omega_t) \right] (\omega_{t+1} - \omega_t) \\ &= h(\omega_t) (\omega_t + 1) \omega_t + h(\omega_t) (\omega_t + 1) \left[\frac{1 - \beta}{1 + (1 - 2\beta)\omega_t} + \frac{\omega_t}{\omega_t + 1} \right] (\omega_{t+1} - \omega_t) \end{aligned}$$

and after simplifying and rearranging terms,

$$\lambda - \omega_t \approx \left[\frac{1 - \beta}{1 + (1 - 2\beta)\omega_t} + \frac{\omega_t}{1 + \omega_t} \right] (\omega_{t+1} - \omega_t)$$

when $\beta \approx \beta'_s$. Hence, we have $|\lambda - \omega_t| > |\omega_{t+1} - \omega_t|$ if the bracketed term is greater than 1, or equivalently if

$$(1 - \beta)(1 + \omega_t) + \omega_t + (1 - 2\beta)\omega_t^2 > [1 + (1 - 2\beta)\omega_t](1 + \omega_t)$$

that simplifies to $\beta(\omega_t - 1) > 0$, a condition always satisfied since we must have $\omega_t > 1$ to have $x_t > x_{t-1}$. Consequently, if $\omega_t < \lambda$ we have $\lambda - \omega_t > \omega_{t+1} - \omega_t$ and thus $\omega_{t+1} < \lambda$, and if $\lambda < \omega_t$, $\omega_t - \lambda > \omega_t - \omega_{t+1}$ and thus $\lambda < \omega_{t+1}$. Finally, since

$h(\omega)(\omega + 1)$ is increasing, there is a unique ω_0 satisfying (37) evaluated for $t = 0$, i.e. $\lambda + 1 = h(\omega_0)(\omega_0 + 1)$. Moreover, we have $\omega_0 < \lambda$ when $\beta > \beta'_s$ since $h(\lambda) > 1$, and $\omega_0 > \lambda$ when $\beta < \beta'_s$. It thus comes that the sequence $\{\omega_t\}_{t=0}^T$ defined by (39) is increasing and bounded above by λ when $\beta > \beta'_s$, and decreasing and bounded below by λ when $\beta < \beta'_s$.

Comparing the tiers' scopes, we get when $\beta \geq \beta'_s$

$$\begin{aligned} x_t - x_{t-1} &= x_{t-1}(\omega_t - 1) = \omega_{t-1}x_{t-2}(\omega_t - 1) \\ &\geq \omega_{t-1}x_{t-2}(\omega_{t-1} - 1) = \omega_{t-1}(x_{t-1} - x_{t-2}) \\ &> x_{t-1} - x_{t-2}, \end{aligned}$$

since $\{\omega_i\}_{i=0}^T$ is increasing (where the last equality comes from the fact that we must have $\omega_t > 1$ for all t). Because we also have $\omega_t \leq \lambda$ in that case, it comes

$$x_t - x_{t-1} = (\omega_t - 1) \prod_{i=0}^{t-1} \omega_i \leq (\lambda - 1) \lambda^t = x_t^* - x_{t-1}^*$$

and opposite inequalities are derived along the same lines when $\beta \leq \beta'_s$. For the capacities, we get using (34) and (25),

$$g_t = \frac{\alpha \beta \bar{x}_t (\omega_t - 1)}{2k \lambda^{2rt} [1 + (1 - 2\beta)\omega_t]} = h(\omega_t)^{1/(1-\beta)} \frac{(\omega_t - 1) \prod_{i=0}^{t-1} \omega_i}{(\lambda - 1) \lambda^t} g_t^*$$

where $h(\omega_t) = (\lambda + 1)/(\omega_t + 1) \geq 1$ when $\beta \geq \beta'_s$. Hence, the relative capacities verify

$$\frac{g_t}{(\omega_t - 1) \prod_{i=0}^{t-1} \omega_i} = \frac{g_t^*}{(\lambda - 1) \lambda^t} h(\omega_t)^{1/(1-\beta)} \geq \frac{g_t^*}{(\lambda - 1) \lambda^t}$$

when $\beta \geq \beta'_s$ and the reverse otherwise.

F Proof of Proposition 6

We have to solve $\max_{T, \{n_t, x_t, g_t\}_{t=0}^T} \left\{ \sum_{t=0}^T W_t : (33), (34) \right\}$ where constraints (33) and (34) come from the fact that $\partial W_t / \partial g_t = 0$ and $\partial W_t / \partial x_t = 0$ under decentralization. Also, we get from (30) and (31) that

$$\frac{\partial W_t}{\partial x_{t-1}} + \frac{\partial W_t}{\partial x_t} = -\frac{\alpha}{4n_t} (x_t - x_{t-1}) - kn_t g_t < 0$$

and as $\partial W_t / \partial x_t = 0$ that

$$\frac{\partial W_t}{\partial x_{t-1}} = -\frac{\alpha}{4n_t}(x_t - x_{t-1}) - kn_t g_t. \quad (40)$$

The Lagrangian of this program is

$$L = \sum_{t=0}^T \left[W_t + \theta_t \left(g_t - \frac{\alpha \beta x_t (x_t - x_{t-1})}{2kn_t^2(\kappa_1 x_t + x_{t-1})} \right) + \mu_t \left(\beta u_t - \frac{n_t k g_t (x_t + x_{t-1})}{2} \right) \right]$$

where $\kappa_1 = 1 - 2\beta$. Using $\partial W_t / \partial g_t = 0$ and (33), the FOC w.r.t. g_t gives

$$\theta_t = \mu_t \left(\frac{n_t k (x_t + x_{t-1})}{2} - \beta^2 \frac{u_t}{g_t} \right) = \mu_t (1 - \beta) \frac{n_t k (x_t + x_{t-1})}{2}. \quad (41)$$

The FOC w.r.t. n_t yields

$$\frac{\partial W_t}{\partial n_t} + \theta_t \frac{2g_t}{n_t} - \mu_t \frac{g_t k (x_t + x_{t-1})}{2} = 0. \quad (42)$$

Using (41), it simplifies to

$$\frac{\partial W_t}{\partial n_t} = -\mu_t (1 - 2\beta) \frac{g_t k (x_t + x_{t-1})}{2}. \quad (43)$$

Using

$$\frac{\partial W_t}{\partial n_t} = \frac{\alpha (x_t^2 - x_{t-1}^2)}{8n_t^2} - \frac{kg_t (x_t + x_{t-1})}{2}, \quad (44)$$

we get

$$\frac{\alpha (x_t - x_{t-1})}{n_t^2} = 4kg_t(1 - \kappa_1 \mu_t),$$

and, using (34),

$$1 = \frac{2\beta x_t (1 - \mu_t \kappa_1)}{\kappa_1 x_t + x_{t-1}}$$

hence,

$$\mu_t = \frac{(2\beta - \kappa_1)x_t - x_{t-1}}{2\beta \kappa_1 x_t} = \frac{(4\beta - 1)x_t - x_{t-1}}{2\beta(1 - 2\beta)x_t}.$$

Using $\partial W_t/\partial x_t = 0$, the FOC w.r.t. x_t simplifies to

$$0 = \frac{\partial W_{t+1}}{\partial x_t} - \theta_t \frac{dg_t}{dx_t} + \mu_t \left(\frac{(1-\beta)\beta u_t}{x_t - x_{t-1}} - \frac{n_t k g_t}{2} \right) - \theta_{t+1} \frac{dg_{t+1}}{dx_t} + \mu_{t+1} \left(-\frac{(1-\beta)\beta u_{t+1}}{x_{t+1} - x_t} - \frac{n_{t+1} k g_{t+1}}{2} \right) \quad (45)$$

where

$$\begin{aligned} \frac{dg_t}{dx_t} &= \frac{d}{dx_t} \left[\frac{\alpha\beta x_t(x_t - x_{t-1})}{2kn_t^2(\kappa_1 x_t + x_{t-1})} \right] = \frac{g_t[(2x_t - x_{t-1})(\kappa_1 x_t + x_{t-1}) - \kappa_1 x_t(x_t - x_{t-1})]}{x_t(x_t - x_{t-1})(\kappa_1 x_t + x_{t-1})} \\ &= \frac{g_t[2x_t(\kappa_1 x_t + x_{t-1}) - x_{t-1}(\kappa_1 x_t + x_{t-1}) - \kappa_1 x_t(x_t - x_{t-1})]}{x_t(x_t - x_{t-1})(\kappa_1 x_t + x_{t-1})} \\ &= \frac{g_t[x_t(\kappa_1 x_t + 2x_{t-1}) - x_{t-1}^2]}{x_t(x_t - x_{t-1})(\kappa_1 x_t + x_{t-1})} \end{aligned}$$

which gives

$$\theta_t \frac{dg_t}{dx_t} = \mu_t(1-\beta) \frac{n_t k (x_t + x_{t-1})}{2} \frac{g_t[x_t(\kappa_1 x_t + 2x_{t-1}) - x_{t-1}^2]}{x_t(x_t - x_{t-1})(\kappa_1 x_t + x_{t-1})}.$$

Similarly, using

$$\begin{aligned} \frac{dg_{t+1}}{dx_t} &= \frac{d}{dx_t} \left[\frac{\alpha\beta x_{t+1}(x_{t+1} - x_t)}{2kn_{t+1}^2(\kappa_1 x_{t+1} + x_t)} \right] = \frac{\alpha\beta x_{t+1} - (\kappa_1 x_{t+1} + x_t) - (x_{t+1} - x_t)}{2kn_{t+1}^2(\kappa_1 x_{t+1} + x_t)^2} \\ &= \frac{\alpha\beta x_{t+1} - 2(1-\beta)x_{t+1}}{2kn_{t+1}^2(\kappa_1 x_{t+1} + x_t)^2} = \frac{-2(1-\beta)x_{t+1}g_{t+1}}{(\kappa_1 x_{t+1} + x_t)(x_{t+1} - x_t)} \end{aligned}$$

we get

$$\theta_{t+1} \frac{dg_{t+1}}{dx_t} = -\mu_{t+1}(1-\beta)^2 \frac{n_{t+1} k (x_t + x_{t+1})}{(\kappa_1 x_{t+1} + x_t)(x_{t+1} - x_t)} g_{t+1} x_{t+1}.$$

Using (33), we have

$$\begin{aligned} \frac{(1-\beta)\beta u_t}{x_t - x_{t-1}} - \frac{n_t k g_t}{2} &= (1-\beta) \frac{n_t k g_t (x_t + x_{t-1})}{2(x_t - x_{t-1})} - \frac{n_t k g_t}{2} \\ &= \frac{n_t k g_t (1-\beta)(x_t + x_{t-1}) - x_t + x_{t-1}}{2(x_t - x_{t-1})} \\ &= \frac{n_t k g_t (2x_{t-1} - \beta(x_t + x_{t-1}))}{2(x_t - x_{t-1})} \\ &= \frac{n_t k g_t (2-\beta)x_{t-1} - \beta \bar{x}_t}{2(x_t - x_{t-1})} \end{aligned}$$

Similarly

$$\begin{aligned} \frac{(1-\beta)\beta u_{t+1}}{x_{t+1}-x_t} + \frac{n_{t+1}kg_{t+1}}{2} &= (1-\beta)\frac{n_{t+1}kg_{t+1}(x_t+x_{t+1})}{2(x_{t+1}-x_t)} + \frac{n_{t+1}kg_{t+1}}{2} \\ &= \frac{n_{t+1}kg_{t+1}}{2} \frac{(2-\beta)x_{t+1}-\beta x_t}{x_{t+1}-x_t} \end{aligned}$$

Finally, $\partial W_{t+1}/\partial x_t$ can be derived observing that, from (40) and (44) we have

$$\frac{(x_{t+1}+x_t)}{2} \frac{\partial W_{t+1}}{\partial x_t} + n_{t+1} \frac{\partial W_{t+1}}{\partial n_{t+1}} = -kn_{t+1}g_{t+1}(x_{t+1}+x_t)$$

where, from (41) and (42), we have

$$\frac{\partial W_{t+1}}{\partial n_{t+1}} = -\theta_{t+1} \frac{2g_{t+1}}{n_{t+1}} + \mu_{t+1} \frac{g_{t+1}k(x_{t+1}+x_t)}{2} = -\kappa_1 \frac{g_{t+1}k(x_{t+1}+x_t)}{2} \mu_{t+1}$$

yields

$$\frac{\partial W_{t+1}}{\partial x_t} = (\kappa_1 \mu_{t+1} - 2) kn_{t+1}g_{t+1}.$$

We can thus rewrite (45) as

$$\begin{aligned} 0 &= (\kappa_1 \mu_{t+1} - 2) n_{t+1} g_{t+1} \\ &\quad - \mu_t n_t g_t \left((1-\beta) \frac{(x_t+x_{t-1})}{2} \frac{x_t(\kappa_1 x_t + 2x_{t-1}) - x_{t-1}^2}{x_t(x_t - x_{t-1})(\kappa_1 x_t + x_{t-1})} - \frac{(2-\beta)x_{t-1} - \beta x_t}{2(x_t - x_{t-1})} \right) \\ &\quad - n_{t+1} g_{t+1} \mu_{t+1} \left(-\frac{(1-\beta)^2(x_t+x_{t+1})x_{t+1}}{(\kappa_1 x_{t+1} + x_t)(x_{t+1} - x_t)} + \frac{(2-\beta)x_{t+1} - \beta x_t}{2(x_{t+1} - x_t)} \right) \end{aligned}$$

which general solutions are given by $x_t = \lambda \bar{x}_{t-1} = \lambda^{t+1}$ using $x_{-1} = 1$ and $n_t = \eta^t$.

Indeed, in that case we get

$$\mu_t = \frac{(4\beta - 1)\lambda - 1}{2\lambda\beta(1 - 2\beta)} = \mu$$

for all t ,

$$n_t g_t = \frac{\alpha\beta\lambda^{t+1}(\lambda - 1)}{2k\eta^t(\lambda\kappa_1 + 1)},$$

from (34), which yields

$$\frac{n_{t+1}g_{t+1}}{n_t g_t} = \frac{\lambda}{\eta},$$

and thus, replacing and collecting terms,

$$\begin{aligned}
(\kappa_1\mu - 2) \frac{\lambda}{\eta} &= \mu \left((1 - \beta)(\lambda + 1) \frac{\lambda(\kappa_1\lambda + 2) - 1}{2\lambda(\lambda - 1)(\lambda\kappa_1 + 1)} - \frac{(2 - \beta) - \beta\lambda}{2(\lambda - 1)} \right) \\
&\quad + \frac{\lambda\mu}{\eta} \left(-\frac{(1 - \beta)^2(\lambda + 1)\lambda}{(\lambda\kappa_1 + 1)(\lambda - 1)} + \frac{(2 - \beta)\lambda - \beta}{2(\lambda - 1)} \right)
\end{aligned}$$

which does not depend on t .

We can derive λ and η as follows. Using (34) to substitute for g_t in (33) gives

$$2\beta\sigma \left(\frac{2kn_t^2(\kappa_1x_t + x_{t-1})}{\alpha\beta\bar{x}_t} \right)^{1-\beta} = n_t k (x_t + x_{t-1})$$

hence

$$n_t = \left[\frac{k(x_t + x_{t-1})}{2\beta\sigma} \right]^{1/(1-2\beta)} \left(\frac{\alpha\beta\bar{x}_t}{2k(\kappa_1x_t + x_{t-1})} \right)^{(1-\beta)/(1-2\beta)}$$

Using $x_t = \lambda^{t+1}$ and $n_t = \eta^t$, we get

$$\begin{aligned}
\eta^t &= \left[\frac{k\lambda^t(\lambda + 1)}{2\beta\sigma} \right]^{1/(1-2\beta)} \left(\frac{\alpha\beta\lambda}{2k[\lambda(1 - 2\beta) + 1]} \right)^{(1-\beta)/(1-2\beta)} \\
&= \lambda^{t/(1-2\beta)} \left[\frac{k^\beta(\lambda + 1)}{\beta^\beta 2^{2-\beta}\sigma} \right]^{1/(1-2\beta)} \left(\frac{\alpha\lambda}{\lambda(1 - 2\beta) + 1} \right)^{(1-\beta)/(1-2\beta)}
\end{aligned}$$

Using $\eta^0 = 1$, we obtain that $\eta^t = \lambda^{t/(1-2\beta)} = \lambda^{rt}$ and that λ solves

$$\frac{k^\beta(\lambda + 1)}{\beta^\beta 2^{2-\beta}\sigma} \left(\frac{\alpha\lambda}{\lambda(1 - 2\beta) + 1} \right)^{1-\beta} = 1$$

hence

$$\frac{\lambda + 1}{(1/\lambda + 1 - 2\beta)^{1-\beta}} = \frac{2^{2-\beta}\beta^\beta\sigma}{k^\beta\alpha^{1-\beta}} = \frac{4\sigma}{\alpha} \left(\frac{\alpha\beta}{2k} \right)^\beta.$$

Replacing $x_t = \lambda^{t+1}$ and $n_t = \lambda^{rt}$ in (34) we get

$$g_t = \frac{\alpha\beta(\lambda - 1)}{2k(1 - 2\beta + 1/\lambda)} \lambda^{-(2r-1)t} \equiv g_0 \lambda^{-(2r-1)t}. \quad (46)$$

Using (33), we get

$$u_t = \frac{k(x_t + x_{t-1})n_t}{2\beta} g_t = \frac{k(\lambda + 1)\lambda^{(r+1)t}}{2\beta} g_0 \lambda^{-(2r-1)t} = \frac{k(\lambda + 1)g_0\lambda^{t(2-r)}}{2\beta}$$

and thus

$$u_0 = \frac{k(\lambda + 1)g_0}{2\beta} = \frac{\alpha(\lambda^2 - 1)}{4(1 - 2\beta + 1/\lambda)}.$$

G Proof of Proposition 7

Denoting $y \equiv \lambda_d + 1$ and $x \equiv \lambda + 1$, (7) and (13) can be written as

$$x = \beta 2^{3-2\beta} / (\alpha^{1-\beta} k^\beta)$$

and

$$\frac{y}{[y/(y-1) - 2\beta]^{1-\beta}} = \frac{x}{(2\beta)^{1-\beta}}$$

respectively, and give

$$\frac{y}{x} = \left(\frac{y(1-2\beta) + 2\beta}{2\beta(y-1)} \right)^{1-\beta} \equiv f(y).$$

The condition $\lambda_d > \lambda$ is equivalent to $y/x > 1$, hence $f(y) > 1$, or, after reorganizing terms, $y(1-4\beta) > -4\beta$ which is always satisfied if $\beta \leq 1/4$ (as we must have $\beta > \beta_s$ to have $\lambda > 1$, this case is relevant only if $\beta_s < 1/4$). If $\beta > 1/4$, this condition becomes $y < 4\beta/(4\beta - 1)$. As

$$K(y) \equiv \frac{y}{[y/(y-1) - 2\beta]^{1-\beta}}$$

increases with y over $(1, +\infty)$, an equivalent condition is given by

$$K(\lambda_d + 1) = \frac{\beta^\beta 2^{2-\beta}}{\alpha^{1-\beta} k^\beta} < K\left(\frac{4\beta}{4\beta - 1}\right) = \frac{2^{1+\beta} \beta^\beta}{4\beta - 1},$$

hence, after reorganizing terms and using (32), $1 < \Phi(\beta)$. We thus have $\lambda_d > \lambda$ for all $\beta \in (\beta_s, \beta'_s)$ and $\lambda_d < \lambda$ for all $\beta > \beta'_s$.