On the optimal allocation of responsibilities among national and subnational governments

Shani, Ron and Reingewertz, Yaniv

University of Haifa, University of Haifa

30 November 2024

On the optimal allocation of responsibilities among national and subnational governments

Ron Shani^, Yaniv Reingewertz^*

Abstract

This paper introduces a model for optimizing the provision of local public goods across national, regional, and local government tiers. We study how spillovers, heterogeneity of preferences, and economies of scale affect the decision to centralize or decentralize responsibilities among these three tiers of government. We provide three key insights – (a) Adding a regional level to the standard fiscal federalism model creates a subnational solution for the problem of spillovers between local governments, (b) Preference heterogeneity creates an incentive to decentralize the provision of local public goods, (c) Economies of scale create an efficiency gain from centralization. The ultimate distribution of responsibilities is contingent upon the interplay of these three competing forces. Our model integrates insights from previous studies to develop a unified theory encompassing the powers that influence decisions to centralize or decentralize.

Keywords: Fiscal Federalism; subnational governments, decentralization, local public goods

JEL Codes: H77, H11, H41, R13

^ Division of Public Administration and Policy, University of Haifa. * corresponding author. yanivrein@poli.haifa.ac.il . We would like to thank participants of the Regional Inequalities workshop in Bochum, participants in the Israeli Economic Association Annual meeting, as well as Thushy Baskaran for their valuable comments and suggestions.

1. Introduction

Sub national governments in general, and regional governments in particular, are an important yet neglected part of the fiscal federalism literature. An OECD survey finds that out of 101 countries, representing 82% of the world's population, 70 have at least one level of regional governments (OECD/UCLG, 2016).[1] Alas, the theoretical literature on the optimal structure of sub-national governments is very limited, and most of the literature is only providing models with two levels of government (see, e.g. Geys & Konrad, 2010; Hämäläinen & Moisio, 2015; Oates, 1972). This line of literature mostly follows the classic Oates' Decentralization Theorem, which posits that under comparable conditions, "it will always be more efficient for local governments to provide the Pareto-efficient level of output … than for the central government" (Oates, 1972, p. 35). A somewhat more contemporary literature also stresses that centralization is preferred if spillovers between local governments are pervasive (Besley & Coate, 2003; Lockwood, 2002). However, while decentralization and centralization involve the devolution or concentration of power between national and subnational levels of government, research examining the allocation of responsibilities among different tiers of subnational governments is scarce.

While classic fiscal federalism assumes that benevolent officials or politicians make policy decisions, a 'second generation' fiscal federalism literature expands the theory to consider officials who may pursue other interests, primarily political ones (see surveys by Geys & Konrad, 2010; Martinez-Vazquez et al., 2017; Oates, 2005, 2008; Reingewertz, 2014; Slavinskaitė et al., 2019; Weingast, 2009). Such studies delve into various political decision-making procedures and their impact on budgets and project allocation between central and local governments, particularly in the context of spillovers (Besley & Coate, 2003; Cheikbossian, 2016; Dur & Roelfsema, 2005; Feidler & Staal, 2012; Frère & Védrine, 2024; Gregor & Stastna, 2012; Koethenbuerger, 2008; Lockwood, 2002; Lorz & Willmann, 2005). Representatives negotiate under varying conditions, which differ across studies, leading to diverse decisions on budget allocation based on those conditions.

The first contribution of our model to these strands of literature is to add the regional level to a fairly standard model of fiscal federalism. This addition, coupled with the presence of spillovers between local governments, leads to an optimal allocation of responsibilities, with

[1] The OECD sample represents 87.5% of the world GDP. See also Council of European Municipalities and Regions (CEMR) for a description of subnational government structures in Europe (www.ccre.org), and Chatry and Vincent (2019) for subnational governments in Asia.

the regional level overseeing the provision of local public goods. This outcome is further reinforced when the production of these goods benefits from economies of scale, but may be attenuated by increasing preference heterogeneity. By integrating regional governments into the model, the role of central governments in providing local public goods is significantly diminished, thereby extending the traditional decentralization model.

Another contribution of our model is in allowing the preferences of individuals to vary not only between local governments but also within them. This assumption enriches existing models, relaxes the assumption of homogeneity within local governments and allows for a more realistic representation of actual preferences. When spillovers are relatively small and local preference heterogeneity is fairly large, we demonstrate that the aggregate surplus of providing local public goods is greater under local provision. Additionally, we model a political-economy decision making process at the regional level for public products where elected representatives negotiate the number of product units the region provides. We distinguish between pro-public spending representatives who decide to provide more product units than the optimal regional average, and averse public spending representatives who decide to provide less. We show that if they decide to provide more product units they need to subsidize the cost. Yet when they decide to provide less product units, the perceived value of such public product should be low.

Finally, we introduce a forces model to the decision-making process of de(centralization), which integrates findings from previous studies and presents a unified framework of the powers influencing the decision to centralize or decentralize power.

The rest of the paper is organized as follows. Section 2 reviews the theoretical literature on fiscal federalism. Section 3 describes the model and extensions. Section 4 discusses the findings and their policy implications and concludes.

2. Literature review

The classic Fiscal Federalism theory suggests that the lowest tier of government should provide local public goods because of information asymmetries between the local and central levels (Oates, 1972). Roughly speaking, the central government should stabilize the economy and control the distribution-equalization of resources, whereas subnational governments (SNG) should provide local or regional public goods adjusted to the preferences of their residents (Musgrave, 1959; Oates, 1972; Olson, 1969). Oates' classic Decentralization Theorem states that under equal conditions when maximizing the aggregate surplus of residents, local

governments will provide local public goods more efficiently than the central government (Oates, 1972, p. 35). When the public good exhibits economies of scale, or when it spills over jurisdiction borders, the optimal allocation favors central provision. The decentralization theorem assumes that central governments provide equal goods to all residents (equal level of services, same number of product units), regardless of their local preferences.[2] Oates asserts that when the central government possesses the knowledge of individual tastes, it can provide Pareto-efficient levels of consumption in all jurisdictions and adds that in terms of the economic definition of federalism, it is equivalent to decentralization (Oates, 1972, p. 37).

A rich literature evolved from Oates' work, extending his model in various ways. First, many studies analyze the case of spillovers between different jurisdictions (Cheikbossian, 2016; Feidler & Staal, 2012; Gregor & Stastna, 2012; Koethenbuerger, 2008; Lockwood, 2002; Lorz & Willmann, 2005; Seabright, 1996). Second, many studies add the political arena to the model in order to describe a more realistic decision making (Besley & Coate, 2003; Dur & Roelfsema, 2005; Feidler & Staal, 2012; Lockwood, 2002; Lorz & Willmann, 2005; Ponzetto et al., 2016). Third, some studies state that economies of scale lead to centralization (Gregor & Stastna, 2012; Ponzetto et al., 2016); Forth, some add heterogeneity of preferences in each local government (Gregor & Stastna, 2012; Lockwood, 2002; Ponzetto et al., 2016). Finally, only few studies enrich the model by discussing more than one sub-national government level (Geys & Konrad, 2010; Halásková & Halásková, 2015; Seabright, 1996).[3]

Possibly the most important contribution within this literature is Besley and Coate (2003), which add the political point-of-view. They assume that each district requests different local public goods (e.g., parks in the first and roads in the second) and that there are spillovers between districts. A key contribution of Besley and Coate (2003) is the analysis of a political equilibrium at the central level, where minimum winning coalitions dictate the geographical distribution of local public goods. They conclude that, to some extent, their analysis corroborates Oates' insights that decentralization is the more efficient arrangement when spillovers are minor, or regions are heterogeneous, while centralization proves more advantageous when there is a high level of spillovers.

In a complementary paper, Lockwood (2002) studies allocating projects between the central and the local level in a political economy setting, assuming homogenous regional

[2] Ponzetto et al. (2016) further state that they "find that centralization can be welfare-maximizing only if it is accompanied by a uniformity constraint that requires the central government to provide identical public goods to all regions."

[3] Empirical studies confirm that decentralization, which may increase efficiency, positively affects the provision of local public goods (Barankay & Lockwood, 2007; Martinez-Vazquez et al., 2017)

residents. The study concludes that small spillovers and significant heterogeneity between regions still mean that decentralization is preferred, but when spillovers are significant, centralization is more efficient. Dur and Roelfsema (2005) show that the centralization of decision-making results in overspending in some regions and too low in others. They also note that when the central government finances local public goods, the common pool problem may arise.

Focusing on local public goods with inter-region spillovers, Lorz and Willmann (2005) show that centralization is sub-optimally low depending on the elected representatives' approach, as representatives prefer decentralization to reduce their region's cost share through sharing the cost with other regions. With small spillovers, decision-making should be decentralized, and when spillovers are significant, a decision at the center is better.

Several other studies discuss the role of spillovers. Koethenbuerger (2008) concludes that a higher degree of spillovers may reduce the difference in the utility of centralization and decentralization. Gregor and Stastna (2012) study when a spill-in is a complement good and conclude that centralization dominates in most settings, yet preference homogeneity is not a sufficient condition for centralization dominance. Cheikbossian (2016) differentiates between local public goods which are substitutes or complements. He concludes that with a common financing rule, centralization improves welfare if spillovers are significant and the local public goods are complementary. Most studies agree that centralization allows for the internalization of spillovers, but this varies by the level of spillovers.

Several studies add issues of size to the analysis. Feidler and Staal (2012) add that the size of regions influences the decision-making process. Residents of larger local government (LG) favor decentralization, as they can finance the local public good without intergovernmental grants, at least when spillovers are small enough. Smaller districts favor centralization, as other districts' residents participate in the cost of the local public good. They state that decentralization leads to a higher surplus when spillovers are small, and when spillovers are significant, centralization leads to a higher surplus. Frère and Védrine (2024) study voluntary intermunicipal cooperation as an alternative to centralization ("local centralization"). They emphasize the "Zoo effect", where small LGs cannot provide local public goods requiring significant resources (Oates, 1988). Shani et al. (2023) present somewhat similar ideas by discussing a mezzanine tier that agglomerates small rural villages to "locally centralize" and benefit from economies of scale. They find that such structures benefit from economies of scale, yet there are limits to those benefits.

Summing up the literature, we observe that several forces influence the allocation of powers between different tiers of subnational governments. The actual policy decision needs to balance those forces. Table 1 lists the influencing forces and their magnitude towards (de)centralization while referencing the studies that discuss each force. We note that second-generation Fiscal Federalism studies, which add the political economy and decision-making processes, describe the behavior of residents and their elected representatives and the representatives' negotiations in the House. We allocated the political economy behavior into one force in Table 1.

Table 1: Forces influencing centralization - decentralization allocation of powers

Force	Pro-Decentralization	Pro-Centralization	Reference
Information asymmetry	High	Low	Oates (1972)
Spillovers	Small	Significant	Besley and Coate (2003) Feidler and Staal (2012) Lockwood (2002) Lorz and Willmann (2005)
Residents' preferences variations between regions	Heterogenous	Homogenous	Gregor and Stastna (2012) Lockwood (2002) Oates (1972)
Economies of scale, size and scope	Insignificant	Significant	Oates (1972)
Zoo effect: good scale	Small	Large	Frère and Védrine (2024) Oates (1988)
Size of local government	Large	Small	Feidler and Staal (2012)
Complementarity of spillovers	Substitute	Complementary	Cheikbossian (2016) Gregor and Stastna (2012)
Representation	Averse public spending	Pro public spending	Besley and Coate (2003) Lorz and Willmann (2005)

Note: Gregor and Stastna, (2012) add that decentralization may dominate even in homogeneity of preferences.

3. The model

We present a model to describe the division of responsibilities between three tiers of government. The main forces we are concerned with are residents' preferences variations within or between jurisdictions, economies of scale, the magnitude of spillovers, and moderation effects between forces. Assume a country is divided into regions that are subdivided into local governments. C, R, and L denote the central, regional, and local tiers. For simplicity, we assume that all local governments have the same number of residents and that

all regions have the same number of local governments. These governments need to decide on the level of local public goods provision.

Residents of each local government have different preferences for local public goods, where the mean preference and the standard deviation characterize their preferences in each local government. We focus on a local public good, assuming that each government can provide this good, yet with different characteristics, such as the quality of the good or the number of product units each resident prefers to consume. For simplicity, we refer to the number of product units each resident prefers or each government provides as a proxy for the bundle of service quality-quantity.

When an SNG provides Q^x product units to each resident in its jurisdiction, some residents receive more product units than their preferences $Q^x > Q_i$ and some residents receive less $Q^x < Q_i$. Residents assign value to the local public good according to their preferences. If they receive more product units than their preferences, they attribute value only to the number of units that match their preferences, yet they pay for the number of units the government provides (usually through taxation). If, however, they receive fewer units than their preferences, they assign value only to the number of units they receive. Therefore, the utility each resident assumes is $U_i^x = V^x \min(Q_i, Q^x) - Q^x C^x$, where V^x denotes the value that residents attribute to a public product unit, and C^x the cost of a product unit.

The aggregate utility of all residents in a government jurisdiction is

$$U^x = \sum_{i=1}^{N^x} U_i^x = \sum_{i=1}^{N^x} [V^x \min(Q_i, Q^x) - Q^x C^x] = V^x \sum_{i=1}^{N^x} \min(Q_i, Q^x) - N^x Q^x C^x, \quad x \in (C, R, L) \tag{1}$$

where Q^x denotes the per-resident number of product units the government provides, and N^x the number of residents in the government's jurisdiction, $x \in (C, R, L)$. There are *n* regions and *m* local governments in the country.

The level of government which should provide the local public good is the one which maximizes U^x, or $\max(U^C, U^R, U^L)$.

According to the Decentralization Theorem (Oates, 1972, p. 60), each government maximizes its residents' aggregate surplus (or minimizes the welfare loss). The optimal provision of the local public good is the arithmetic mean of the quantity across all residents, which is also Pareto-efficient. Therefore, the average number of product units each government provides is presented in Equation (2).[4]

[4] See Appendix B

$$Q^C = \frac{1}{N^C}\sum_{i=1}^{N^C} Q_i = \frac{1}{n}\sum_{j=1}^{n} Q_j^R = \frac{1}{m}\sum_{k=1}^{m} Q_k^L$$
$$Q_j^R = \frac{1}{N_j^R}\sum_{i=1}^{N_j^R} Q_i : \text{for residents of region j} \quad (2)$$
$$Q_k^L = \frac{1}{N_k^L}\sum_{i=1}^{N_k^L} Q_i : \text{for residents of local government k}$$

In any distribution, there are residents whose preferences are below the average and others who are above. Let $\gamma \in (0,1)$ denote the ratio of residents whose preferences are below the average preferences of the jurisdiction population. Therefore, there are γN^x residents that experience $Q_i < Q^x$, and $(1-\gamma)N^x$ residents which face $Q_i > Q^x$.[5] The left argument of Equation (1) can now be

$$\sum_{i=1}^{N^x} \min(Q_i, Q^x) = \sum_{\substack{i=1 \\ \text{s.t. } Q_i<Q^x}}^{\gamma N^x} Q_i + \sum_{\substack{i=1 \\ \text{s.t. } Q_i>Q^x}}^{(1-\gamma)N^x} Q^x = \sum_{\substack{i=1 \\ \text{s.t. } Q_i<Q^x}}^{\gamma N^x} Q_i + (1-\gamma)N^x Q^x \quad (3)$$

We can express the left sum argument of Equation (3) as the average of the residents whose $Q_i < Q^x$ times the number of those residents. We can further express the average of the residents below the total average using the definition of the Mean Absolute Deviation around a central point (MAD) using Equation A4.[6]

$$\sum_{i=1}^{N^x} \min(Q_i, Q^x) = \gamma N^x\left(Q^x - \frac{1}{2\gamma}\text{MAD}\right) + (1-\gamma)N^x Q^x \quad (4)$$

Substituting Equation (4) into (1) yields

$$U^x = N^x Q^x (V^x - C^x) - \frac{1}{2}V^x N^x \text{MAD} \quad (5)$$

Proposition 1: *the aggregate surplus of providing local public goods under utility maximizing conditions is proportional to the utility from the good minus the loss of utility resulting from the relative dispersion of residents' preferences*.

Proof: using Jensen's inequality, one can show that the MAD is less than or equal to the standard deviation: $MAD: E(|x-\mu|) \leq \sigma$. Embedding it into Equation (5) yields $U^x \geq$

[5] γ is a measure of the skewness of the distribution. In a symmetrical distribution $\gamma=\Phi(0)=½$. In exponential distribution $\gamma=\Phi(1)=0.63$.

[6] MAD is defined as $\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$. For the normal distribution MAD is $\frac{E|x|}{\sqrt{E(x^2)}} = \sigma\sqrt{\frac{2}{\pi}}$. See Appendix A.

$N^x Q^x (V^x - C^x) - \frac{1}{2} V^x N^x \sigma$. The greater the standard deviation, the smaller the aggregate residents' utility. Note that the aggregate surplus is independent of the skewness of the distribution.

Contrary to the Decentralization Theorem (Oates, 1972), the loss of utility which results from the differences of residents' preferences influences the decentralization aggregate surplus. This is somewhat similar to Proposition 3.3.2 of Frère and Védrine (2024), yet it takes the probability distribution function of residents preferences into account, while also differentiating between when the local government provides less product units than the resident prefers, and when it provides more.

Corollary 1a: *The greater the homogeneity of residents within a jurisdiction, the higher the aggregate benefit they derive from local public goods.*

Proof: with Proposition 1, when the residents in a defined jurisdiction have similar preferences, the standard deviation becomes small, resulting in higher aggregate utility.

Corollary 1b: *when the cost of providing a product unit is the same across government tiers, the highest utility is when the lower tier provides the local public good.*[7]

Proof: assume that the perceived value and the cost of a public product unit are equal at all government levels: $V^C = V^R = V^L \equiv V$, and $C^C = C^R = C^L \equiv C$. Assuming even a small Tiebout sorting, the between LGs variance is greater than the within, implying that the standard deviation becomes smaller as the level of government tier is closer to the resident: $\sigma^C > \sigma^R > \sigma^L$, yielding from Equation (5) that $U^C < U^R < U^L$.

Equation (5) describes the aggregate utility for the whole country if the central government provides the local public good for a specific region or local government (*x=c* in Equation (5)). Let us now aggregate all the regions and all the local governments to the country level. According to Equation (5), the specific regional utility of region j is:

$$U_j^R = N_j^R Q_j^R \left(V_j^R - C_j^R\right) - \frac{1}{2} V_j^R N_j^R MAD_j^R, j = 1 \ldots n \quad (6)$$

Let us assume that all regions are similar except for the residents' preferences, i.e., all have the same number of residents, the same good value and cost (N^R, V^R, C^R), but different distributions ($Q_j^R, \gamma_j^R, MAD_j^R$). Aggregating all n regions to the country level when the regions provide the public good and using Equations (2) and (6) yields

$$U^R = \sum_{j=1}^{n} U_j^R = N^C Q^C \left(V^R - C^R\right) - \frac{1}{2} N^C V^R \sum_{j=1}^{n} \frac{MAD_j^R}{n} \quad (7)$$

[7] Another proof of the Decentralization Theorem

Similarly, the expression for the aggregate surplus when local governments provide the local public good is

$$U^{L} = \sum_{k=1}^{m} U_{k}^{L} = N^{C} Q^{C} \left(V^{L} - C^{L}\right) - \frac{1}{2} N^{C} V^{L} \sum_{k=1}^{m} \frac{MAD_{k}^{L}}{m} \quad (8)$$

The left argument of Equations (5), (7), and (8) describes the surplus from the local public good. The right arguments describe the average variability within each government tier. If the local public good has some local merit, we expect the variability (or dispersion) to increase as we move from the lower to the higher tier. When the valuation and the cost are similar at all tiers of government, the tier that better provides the local public good is the one where the average variability is minimal.

Spillovers

Residents of a local government may consume local public goods that an adjacent local government provides if they fit their preferences better than similar goods their local government provides.

Most papers that discuss spillovers use a model of two LGs, each providing a different local public good, where residents have limited resources. Therefore, residents decide about the mix of goods they will consume from the two LGs. We model spillovers based on a different quality of services that each LG provides. Figure 1 illustrates an example of such two LGs. In this sense, our analysis refers to complementary public product that governments at the same level provide, such as education, sanitation, cultural etc.,. Some residents are satisfied with the level of services their local government provides, yet others request better services (or more product units), and may consume the product from another local government. We chose this setting as it complements previous studies, and sheds light on a different type of public products.

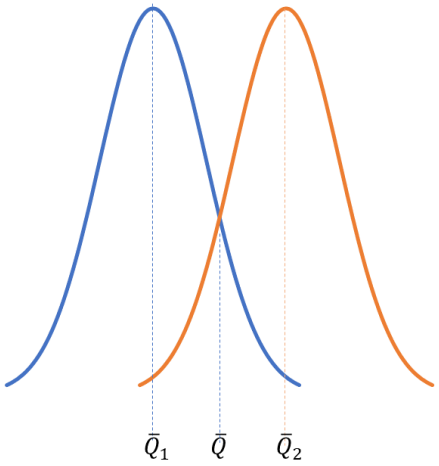

Figure 1: Product competition and spillovers

We assume that there are two local governments, LG_1 and LG_2, that belong to the same region R. Each of the LGs provides the same local public good but with different quantity $\bar{Q}_1 < \bar{Q}_2$. LG_1 and LG_2 are close enough so that residents of LG_1 can consume the local public good LG_2 provides, yet they pay for the good LG_1 provides, most commonly in taxes. There are two options: either each LG provides its residents with the locally average good, or the region provides the good to residents of both LGs with the average number of product units $\bar{N} = \frac{1}{2}(\bar{Q}_1 + \bar{Q}_2)$.

When each LG provides the local public good, there are $\alpha\sigma$ residents of LG_1 who consume the local public good LG_2 provides, as they prefer a more significant number of product units. All residents of LG_2 consume the local good. We also assume that the local public good is free and that the spillover from LG_2 to LG_1 does not burden LG_2.[8] Following Equation (5), the aggregate surplus of the residents of each LG is

$$\begin{aligned} U_1^L &= N^L\bar{Q}_1(V^L - C^L) - \frac{1}{2}V^L N^L MAD_1 + V^L(\bar{Q}_2 - \bar{Q}_1)N^L(1 - \Phi(\alpha)) \\ U_2^L &= N^L\bar{Q}_2(V^L - C^L) - \frac{1}{2}V^L N^L MAD_2 - C^L\bar{Q}_2 N^L(1 - \Phi(\alpha)) \end{aligned} \quad (9)$$

and the aggregate surplus of both LGs is

$$U^L = 2N^L\bar{Q}(V^L - C^L) - \frac{1}{2}V^L N^L(MAD_1 + MAD_2) + [\bar{Q}_2(V^L - C^L) - V^L\bar{Q}_1]N^L(1 - \Phi(\alpha)) \quad (10)$$

where $\Phi(\alpha)$ denotes the cumulative distribution function (CDF) of the residents' preferences for the local public good, and α is the number of standard deviations where residents whose

[8] Upper-tier government may reimburse LG_2 for the spillover through intergovernmental grant.

preferences lie above $\alpha\sigma$ consume goods from LG_2. Figure 2 shows two examples of $1-\Phi(\alpha)$, one for a normal distribution and the second for an exponential distribution.

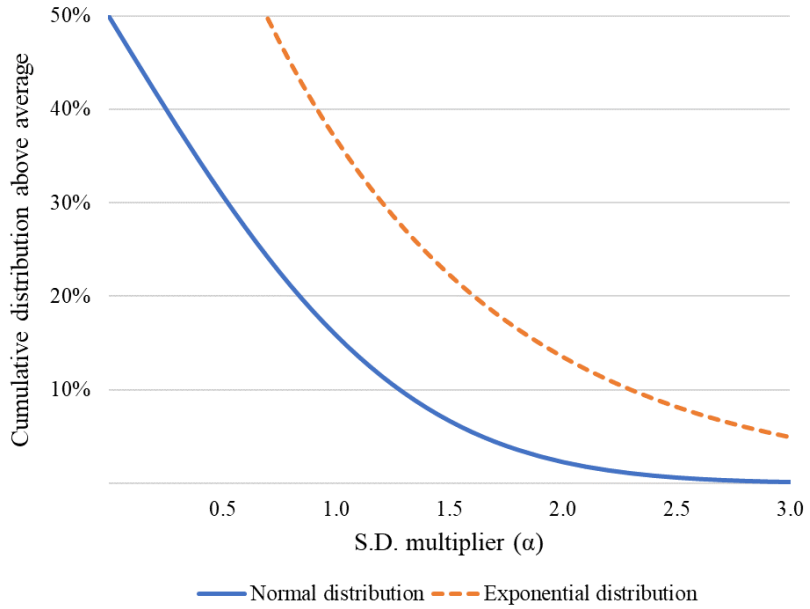

Figure 2: Share of residents of LG_1 requesting more than $\alpha\sigma$ product units and consume spillover from LG_2

When the regional government directly provides the local public good, there will be no spillovers, yet the total number of product units will be the average. For simplicity, we assume that the number of LG_1 residents that their preferences exceed the average $(\bar{Q})$ is equal to the number of LG_2 residents whose preferences are lower than the average (see Figure 1). Therefore, Equation (10) shows the aggregate surplus when the regional government provides the local public good.

To determine which level should provide the local public good, we subtract the aggregate surplus from the regional government's surplus when LGs provide.

$$\Delta U^{L-R} = \frac{U^L}{N^L} - \frac{U^R}{N^L} = \\ = V\left(\frac{1}{2}(\bar{Q}_2 - \bar{Q}_1) - \frac{1}{2}(MAD_1 + MAD_2)\right) + (\bar{Q}_2(V-C) - V\bar{Q}_1)\left(1-\Phi(\alpha)\right) \quad (11)$$

Proposition 2: *Centralization is preferred when spillovers are sufficiently large, and preferences are sufficiently similar.*

Proof: when $\bar{Q}_2 - \bar{Q}_1 < \sigma$, and noting that $MAD \leq \sigma$, the left argument of Equation (11) is negative. We can write the right argument as $(V(\bar{Q}_2 - \bar{Q}_1) - C\bar{Q}_2)\left(1-\Phi(\alpha)\right)$. If $C >$

$V(1-\frac{\overline{Q}_1}{\overline{Q}_2})$, this argument is also negative, resulting in $\Delta U^{L-R}<0$. When spillovers are high, i.e., small α, the right argument becomes even more negative.

Corollary 2a: *when residents' preferences are significantly different, it is more beneficial for the local governments to provide the local public good, regardless of spillovers.*

Proof: when $\overline{Q}_2 \gg \overline{Q}_1$, Equation (11) indicates that $\Delta U^{L-R}>0$.

Corollary 2b: *the difference between residents' preferences moderates the relationship between spillover magnitude and the tier that is more beneficial in providing the local public good.*

Proof: the derivative of Equation (11) with respect to the spillover's magnitude is $\frac{\partial \Delta U^{L-R}}{\partial \alpha}=-(V(\overline{Q}_2-\overline{Q}_1)-C\overline{Q}_2)\frac{\partial \Phi}{\partial \alpha}$. The definition of CDF implies that $\frac{\partial \Phi}{\partial \alpha}>0$. Therefore, the first-order extremum is when $C=V(1-\frac{\overline{Q}_1}{\overline{Q}_2})$. When, on the first hand, the difference in residents' preferences is significant, i.e., $\overline{Q}_2 \gg \overline{Q}_1$, we note that $C<V(1-\frac{\overline{Q}_1}{\overline{Q}_2})$, and therefore, $\frac{\partial \Delta U^{L-R}}{\partial \alpha}<0$: as the magnitude of spillovers increases (α becomes small), ΔU^{L-R} increases, and it will be beneficial for the local governments to provide the local public good, and when the magnitude of spillovers decreases, ΔU^{L-R} decreases, and may reach the point where it will benefit the upper-tier government to provide the local public good.

When, on the other hand, the difference in residents' preferences is small, we note that the first-order extremum may assume $C>V(1-\frac{\overline{Q}_1}{\overline{Q}_2})$, implying that $\frac{\partial \Delta U^{L-R}}{\partial \alpha}>0$: as the magnitude of spillovers increases, ΔU^{L-R} decreases, and may reach the point where it will benefit the upper-tier government to provide the local public good. When the magnitude of spillovers decreases, ΔU^{L-R} increases, and it is more beneficial for the local governments to provide the local public good, and when spillovers increase, ΔU^{L-R} decreases, and it will benefit the upper-tier government to provide the local public good. Table 2 summarizes the moderation effect of residents' preferences on the relations between spillovers and which government is better positioned to provide the local public good.

Table 2: Moderation effect of the difference in residents' preferences

	Differences in residents' preferences	$\frac{\partial \Delta U^{L-R}}{\partial \alpha}$	**Spillovers**	α	ΔU^{L-R}	ΔU^{L-R}	**Tier to provide local public goods**
1	Significant	<0	↓	↑	↓	May reach <0	Regional
2			↑	↓	↑	>0	Local
3	Small	>0	↓	↑	↑	>0	Local
4			↑	↓	↓	May reach <0	Regional

Notes: row 4 is consistent with Proposition 2. The up and down arrows represent the rate of change, increase (↑) or decrease (↓). ΔU^{L-R} is the difference in the aggregate utility when the local government provides the local public good and when the regional government does. α is the number of standard deviations from the average where spillovers start ($\alpha\sigma$).

Economies of scale

So far, we assumed that the cost of providing the local public good is the same across the different government tiers. However, there may be differences in the cost based on economies of scale (see e.g., Andrews & Boyne, 2009). When the central government provides the goods, it produces greater quantities compared to the number of product units a local government provides.[9]

Assuming that the per-unit cost is a declining function of quantity, we obtain that $C^C < C^R < C^L$.

The optimal government tier to provide the public good is the one that maximizes the aggregate surplus, namely $\max(U^C, U^R, U^L)$. To find out which is that tier, we subtract Equation (7) from Equation (5). Assuming that the good value is similar, and adding the economies of scale shows that:[10]

$$\Delta U^{C-R} = \frac{U^C}{N^C} - \frac{U^R}{N^C} = Q^C\left(C^R - C^C\right) - \frac{1}{2}V\left(MAD^C - \sum_{j=1}^{n}\frac{MAD_j^R}{n}\right) \quad (12)$$

Proposition 3: *it is more beneficial for an upper-tier to provide a local public good if the cost saving due to economies of scale is greater than the difference of the average resident preferences variability.*

Proof: The left and right arguments of Equation (12) are positive. For the left argument $C^R > C^C$, and for the right argument variability in the country is greater than the average

[9] Caveat: lower-tier governments which face a smaller number of product units, may form a joint venture to purchase the good at greater quantities, or may outsource, which may result in a lower cost, exhibiting returns to scale. Yet, there can be differences in the management overhead, where lower-tier government, especially smaller local governments exhibit higher per resident costs.

[10] Equation (12) presents the difference between U^C and U^R. The difference between U^R and U^L is analogous.

variability in regions multiplied by -1. Therefore, ΔU^{C-R} will be positive when the number of regions (n, indicating the influence of the economies of scale) is large enough to overcome the difference in the average variability, indicating that under such conditions, it will be more beneficial for the central government to provide the local public good.

Bi-modal distribution

In the previous analysis, we discuss general distributions characterized by mean, standard deviation, and MAD. Here we focus on a specific type of distribution – the bi-modal distribution. This specific case is interesting since there may be jurisdictions where the distribution of the residents' preferences is bi-modal. We may experience a bi-modal distribution in two options. The first is significant differences in residents' preferences in one local government. The second is when two local governments are part of the same regional government, and the regional government provides the goods and faces a bi-modal distribution of preferences (Jones et al., 2022).

We may look at a bi-modal distribution as a joint of two distant-enough uni-modal distributions, each characterized by the mean (μ) and standard deviation (σ), and the mixing ratio (p). Bi-modal distribution requires that $|\mu_1 - \mu_2| >> 2\sigma$. Without loss of generality, let us assume that the mixing ratio is half, the two resident groups have the same number of residents and exhibit the same standard deviation, yet they exhibit different preferences for the number of units of the local public good ($\bar{Q}_1, \bar{Q}_2$). The government faces an average preference $\bar{Q} = p\bar{Q}_1 + (1-p)\bar{Q}_2 = \frac{1}{2}(\bar{Q}_1 + \bar{Q}_2)$. We also assume that $\bar{Q}_1 < \bar{Q} < \bar{Q}_2$.

For clarity, let us denote by R the government which faces the bi-modal distribution and provides the local public good, and by L each of the residents' groups. Each group receives the local public good as a uni-modal per their preferences.[11]

When the R government provides the local public good, all of its residents receive the average number of product units $\bar{Q}$. Half of the population will receive more than their preference, and half will receive less. Figure 3 shows an example of a bi-modal distribution.

[11] We use the second option of bi-modal occurrence, where two LGs are a part of a region. Therefore, we use L to symbolize LGs which exhibit uni-modal distribution, and R the government which face a bi-modal distribution.

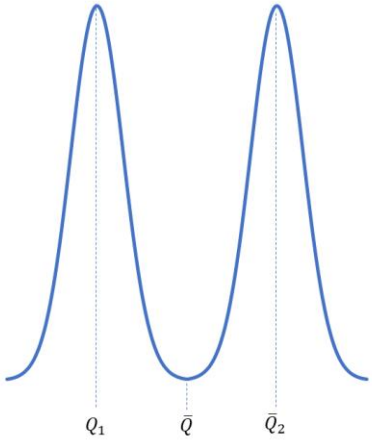

Figure 3: Normal bi-modal distribution
Note: the normal distribution is presented as an example. Our model is agnostic to the distribution function.

The aggregate surplus of the first group is $\frac{1}{2}\mathrm{N}^{\mathrm{R}}\overline{\mathrm{Q}}_1\mathrm{V}^{\mathrm{R}} - \frac{1}{2}\mathrm{N}^{\mathrm{R}}\overline{\mathrm{Q}}\mathrm{C}^{\mathrm{R}}$, and of the second group $\frac{1}{2}\mathrm{N}^{\mathrm{R}}\overline{\mathrm{Q}}(\mathrm{V}^{\mathrm{R}} - \mathrm{C}^{\mathrm{R}})$. The total aggregate surplus of the R government residents is

$$\mathrm{U}^{\mathrm{R}} = \frac{1}{2}\mathrm{N}^{\mathrm{R}}\mathrm{V}^{\mathrm{R}}(\overline{\mathrm{Q}}_1 + \overline{\mathrm{Q}}) - \mathrm{N}^{\mathrm{R}}\overline{\mathrm{Q}}\mathrm{C}^{\mathrm{R}} \tag{13}$$

Now, we calculate the total aggregate surplus when each group is in a different local government (i.e., there is a geographic distinction). In such an arrangement, each group's residents receive the number of product units per the group's average. The surplus of each group will be according to Equation (5)

$$\mathrm{U}^{\mathrm{L}} = 2\mathrm{N}^{\mathrm{L}}\overline{\mathrm{Q}}\left(\mathrm{V}^{\mathrm{L}} - \mathrm{C}^{\mathrm{L}}\right) - \frac{1}{2}\mathrm{V}^{\mathrm{L}}\mathrm{N}^{\mathrm{L}}(\mathrm{MAD}_1 + \mathrm{MAD}_2) \tag{14}$$

Subtracting Equation (13) from Equation (14), assuming that the good value and cost are similar, shows that

$$\Delta\mathrm{U}^{\mathrm{L-R}} = \frac{\mathrm{U}^{\mathrm{L}}}{\mathrm{N}^{\mathrm{L}}} - \frac{\mathrm{U}^{\mathrm{R}}}{\mathrm{N}^{\mathrm{L}}} = \mathrm{V}\left[\overline{\mathrm{Q}} - \overline{\mathrm{Q}}_1 - \frac{1}{2}(\mathrm{MAD}_1 + \mathrm{MAD}_2)\right] \tag{15}$$

Proposition 4: *when residents' preferences exhibit a bi-modal distribution, it is more beneficial to split them into two groups and provide each group with its average preference.*

Proof: in a bi-modal distribution $|\bar{Q}_1 - \bar{Q}_2| >> 2\sigma$. Using the MAD characteristic $E(|x - \mu|) \leq \sigma$ yields that $\Delta\mathrm{U}^{\mathrm{L-R}} > 0$. Proposition 4 is an extension of Corollary 1a.

Political economy analysis

The classic approach to fiscal federalism assumes a benevolent planner, where officials make decisions based on the common good and aim to maximize aggregate welfare.[12] Interestingly, the classic approach shows that the number of public good product units a government provides reflects the average preferences of its residents. When the distribution of preferences is symmetrical, this aligns with the median voter theorem (see Appendix B). The second generation of Fiscal Federalism relaxes this assumption, and incorporates political decision-making processes (Besley & Coate, 2003; Cheikbossian, 2016; Feidler & Staal, 2012; Lockwood, 2002; Oates, 2008). In this section we examine a political decision-making process where a higher-level government may impose a specific number of product units to local governments whose elected representatives form a coalition, thereby complementing previous studies, which show that the number of product units may vary based on political negotiations, yet do not quantify such an influence on residents.

In previous parts of the model, we assumed that local governments decide on the number of public product units based on the average preferences of their residents, which aligns with the median voter in a symmetrical distribution. In contrast, the decision-making process in a regional government occurs in the house of representatives in two stages. In the first stage, residents of each local government elect their representative, reflecting the median voter preferences of that constituency, and in the second stage, representatives negotiate, form a coalition, and decide upon the number of product units to provide based on a majority rule. In this scenario, the regional government will provide Q^R product units based on the perception of the elected representatives, pro or averse public spending, and coalition negotiation results as prior studies describe ((Besley & Coate, 2003; Cheikbossian, 2016; Feidler & Staal, 2012; Lockwood, 2002)).

In any case, there are two options: (a) the region provides $Q^R > \overline{\mathrm{Q}}^{\mathrm{R}}$, or (b) the region provides $Q^R < \overline{\mathrm{Q}}^{\mathrm{R}}$.

The aggregate surplus of residents of region j is[13]

$$\mathrm{U}_{\mathrm{j}}^{\mathrm{R}} = \mathrm{V}\sum_{\mathrm{i}=1}^{\mathrm{N}_{\mathrm{j}}^{\mathrm{R}}} \min\left(\mathrm{Q}_{\mathrm{i}}, \mathrm{Q}^{\mathrm{R}}\right) - \mathrm{N}_{\mathrm{j}}^{\mathrm{R}}\, \mathrm{C}\, \mathrm{Q}^{\mathrm{R}}, \qquad (16)$$

$$\mathrm{i} \in \{\text{residents of region j}\}$$

[12] Oates (2008) indicates that the median voter model provides outcomes which do not deviate much from the benevolent planner model, therefore, the first assumption of the Decentralization Theorem does not require benevolence of public officials.

[13] For convenience and clarity, we omit the subscript j of the region in the following equations.

Let us split Equation (16) into the above two options: $Q^R < \overline{Q}^R$, and $Q^R > \overline{Q}^R$. Let us denote the number of product units the region provides as α multiples of the standard deviation, $Q^R = \overline{Q}^R + \alpha\sigma$. Figure 4 shows the case that α > 0, and Figure 6 when α < 0. Let us further split the distribution into three areas as Figure 4 shows for α > 0, and into two areas as Figure 6 shows for α < 0. We then compute the aggregate surplus for each option.

When $Q^R > \overline{Q}^R$, the aggregate surplus described in Equation (16) becomes[14]

$$U_j^R = V\underbrace{\sum_{i=1}^{\gamma N_j^R} Q_i}_{\text{area I}} + V\underbrace{\sum_{i=\gamma N_j^R+1}^{N_j^R\Phi(\alpha)} Q_i}_{\text{area II}} + V\underbrace{\sum_{i=N_j^R\Phi(\alpha)+1}^{N_j^R} Q^R}_{\text{area III}} - N_j^R C Q^R \tag{17}$$

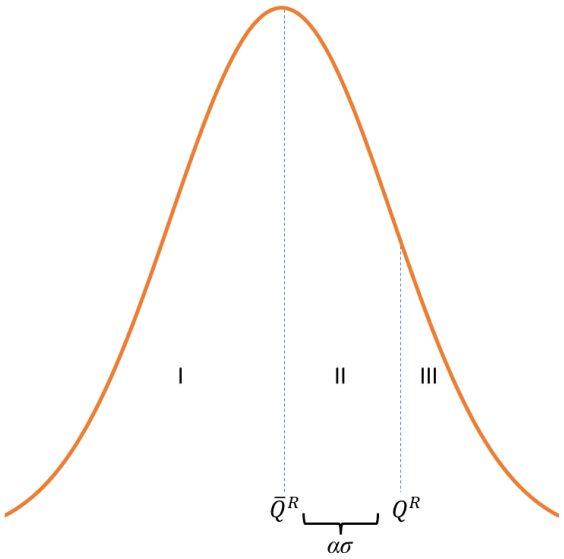

Figure 4: The region provides more product units than the average preferences of its residents
Note: The normal distribution shape is for clarity only

Rearranging Equation (17) gives

$$U^R = \frac{1}{2}VN^R\{(\overline{Q}^R - Q^R)(\gamma + \Phi(\alpha)) + 2Q^R - MAD^R\} - N^R C Q^R \tag{18}$$

The difference between the classic approach, as shown in Equation (5), and the political economy approach shown in Equation (18) is

$$\Delta U^R = N^R\left\{(\overline{Q}^R - Q^R)\left[\frac{1}{2}V(\gamma + \Phi(\alpha) - 2) + C\right]\right\} \tag{19}$$

The coalition elected representatives will benefit their constituencies if $\Delta U^R > 0$, which yields

$$\frac{V}{C} > \frac{2}{2 - \gamma - \Phi(\alpha)} > 1 \tag{20}$$

[14] For simplicity, when α > 0 we approximate the average of area II as the mean between Q^R and $\overline{Q}^R$.

Figure 5 illustrates examples of Equation (20) for both the normal and exponential distributions. It can be observed that, with a normal distribution, if the elected representatives aim to benefit their constituencies by providing higher quality of the public product (e.g., $\alpha > 1$), they need to subsidize the cost to achieve $\frac{V}{C} > 2.5$. They may accomplish this through dedicated subsidies that have price effect.[15] Such policy may reduce the magnitude of the spillovers of that specific public product, as the region provides more product units.

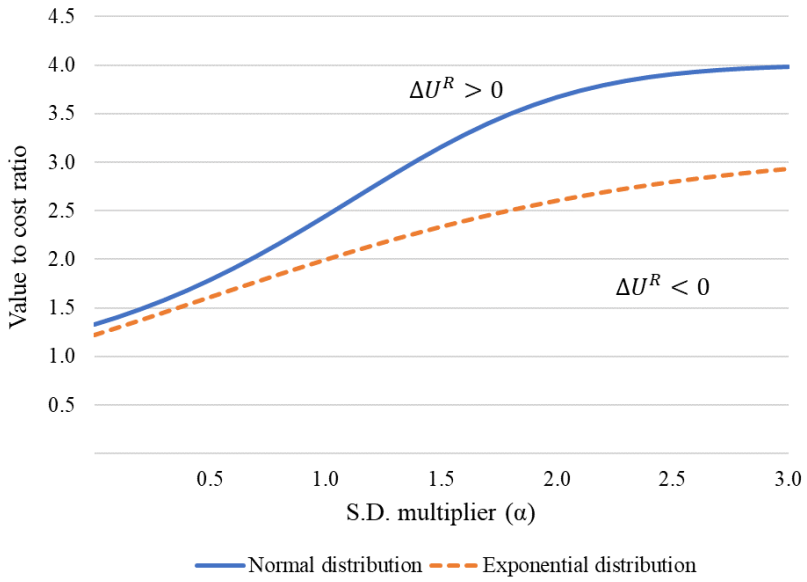

Figure 5: Perceived value to cost ratio frontier
Note: The ratio of perceived value to cost which differentiates between the space where subsidies are economically beneficial and when they are not.

When the elected representatives are averse to public spending, they may decide that the region will provide less product units than the preferences average, namely $Q^R < \overline{Q}^R$. In area I of Figure 6, residents' preferences are lower than Q^R, while in area II residents receive less product units than they prefer. To compute the surplus of residents in area I let us define the ratio between the truncated mean preferences of residents in area I to the number of products the region provides as

$$\int_{-\infty}^{Q^R} Q\, f(Q)\, dQ = \delta Q^R, \delta \in (0,1), f(Q)\ is\ the\ PDF\ of\ Q \qquad (21)$$

The aggregate surplus in this case is

[15] It is assumed that the House cannot discriminate residents and local governments based on different cost, yet they may subsidize LGs differently.

$$U^R = V\underbrace{\sum_{i=1}^{N^R\Phi(\alpha)} Q_i}_{\text{area I}} + V\underbrace{\sum_{i=N^R\Phi(\alpha)+1}^{N^R} Q^R}_{\text{area II}} - N^R C Q^R \tag{22}$$

and rearranging Equation (22) gives

$$U^R = VN^R Q^R\big(1-\Phi(\alpha)(1-\delta)\big) - N^R C Q^R \tag{23}$$

Combining Equations (18) and (23) gives[16]

$$U^R = \begin{cases} VN^R Q^R\big(1-\Phi(\alpha)(1-\delta)\big) - N^R C Q^R, & Q^R < \overline{Q}^R \\ \frac{1}{2}VN^R\big((\overline{Q}^R - Q^R)\big(\gamma+\Phi(\alpha)\big) + 2Q^R - MAD^R\big) - N^R C Q^R, & Q^R > \overline{Q}^R \end{cases} \tag{24}$$

The difference between the classic approach, as shown in Equation (5), and the political economy approach shown in Equation (23) is

$$\Delta U^R = N^R(V-C)\big(Q^R - \overline{Q}^R\big) + VN^R\left(\frac{1}{2}MAD^R - Q^R\Phi(\alpha)(1-\delta)\right) \tag{25}$$

The coalition elected representatives will benefit their constituencies if $\Delta U^R > 0$, which yields

$$\frac{C}{V} > 1 - \frac{\frac{1}{2}MAD^R - Q^R\Phi(\alpha)(1-\delta)}{\overline{Q}^R - Q^R} \tag{26}$$

Figure 7 presents a simulation of Equations (20) and (26) for normal distribution, while indicating the frontier where the representatives' decision benefit residents compared to when the region provides the average of its residents' preferences per the classical approach.

What we deduce from Equations (20) and (26) is: (a) if the elected representatives are pro pubic spending, they'll decide to provide more product units such that $Q^R > \overline{Q}^R$. Residents will appreciate such approach if the benefits outperform the alternative standard approach, meaning that the ratio of perceived value to cost should be higher. As the cost cannot be reduced, the region will have to subsidize at least those LGs which are a part of the coalition through earmarked transfers. (b) if the elected representatives are averse public spending, they may decide to provide less product units, such that $Q^R < \overline{Q}^R$. Similarly, residents will appreciate such policy, provided that the cost will be close to the perceive value. When the perceived product value is so low, we may argue that residents care less about its provision, so the elected representatives may decide to provide fewer product units only to those products that have low perceived value.

[16] When $Q^R = \overline{Q}^R$,$\gamma = \Phi(\alpha)$,and $\delta = 1 - \frac{1}{2\gamma\overline{Q}^R}MAD^R$, we obtain the same equation on both rows, as Equation (6) presents.

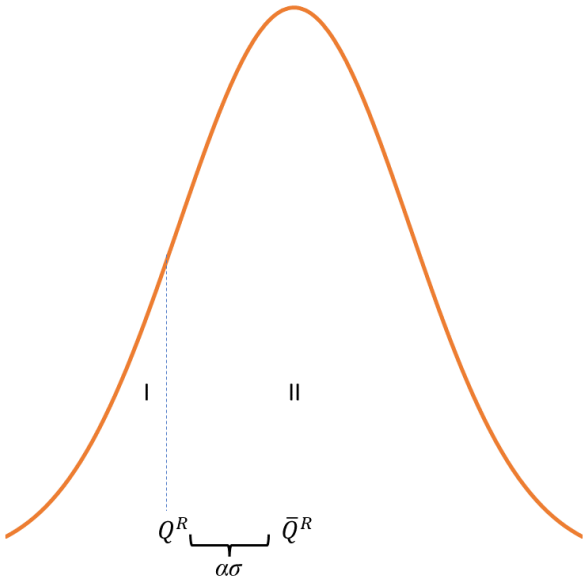

Figure 6: The region provides less product units than the average preferences of its residents

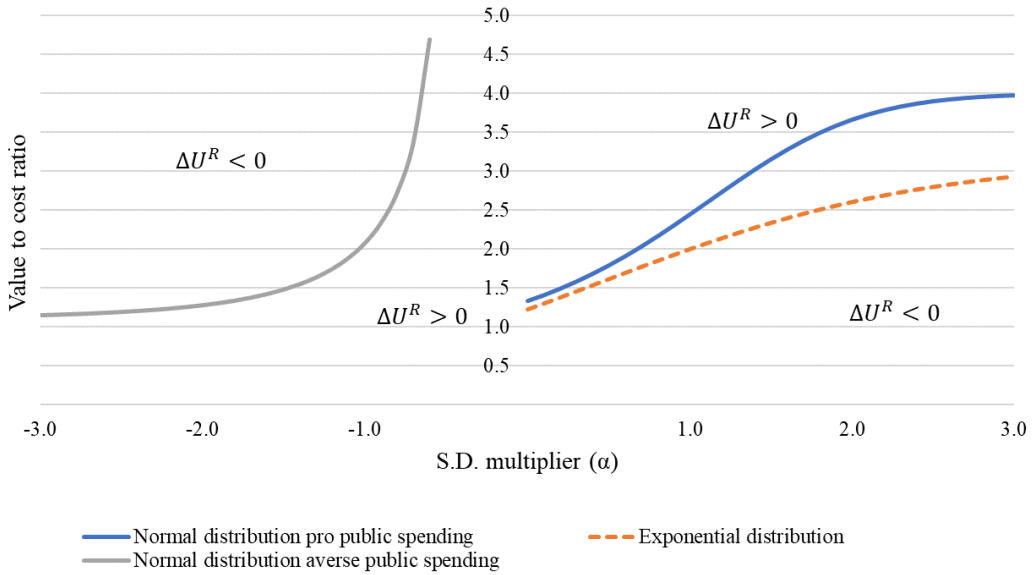

Figure 7: The ratio of perceived value to cost when the region provides more or less product units than the residents' average preferences

4. Discussion and conclusion

The literature on fiscal federalism typically assumes a two-tier government structure and examines the allocation of budgets, projects, and powers between those tiers. Our study introduces a third tier - the regional tier - and demonstrate how this intermediate tier effectively addresses local spillovers.

Building on prior studies, we relax the assumption of homogenous subnational governments and recognize that residents within SNG units may have diverse preferences for local public goods. Residents whose preferences exceed the level of services their SNG provides may consume those goods from other SNGs. In line with the Decentralization Theorem (Oates, 1972) and the Tiebout model (Tiebout, 1956), we show that heterogeneity of residents' preferences reduces the aggregate surplus compared to a homogenous community. Furthermore, we demonstrate that when the cost of providing the good is the same across SNGs, the lower tier should provide the local public good, thereby offering another proof of the Decentralization Theorem.

Most studies also assume that the cost of the public good is the same across SNGs. They note that if the product exhibits economies of scale, the central government should provide the good. Our analysis reveals a more nuanced perspective: economies of scale make it more beneficial for an upper-tier government to provide the local public good if the cost savings resulting from scale outweigh the impact of variations in residents' preferences.

Our approach to handling spillovers diverges from that of previous studies. Earlier research typically assumes two distinct pubic goods, with each SNG provides different goods. Residents of LG_1 consume both the good that their LG provides and the good that LG_2 provides, and vice-versa. In contrast, our model posits that both SNGs provide similar public goods but in different quantities. Each SNG supplies an average number of product units based on residents' preferences. Consequently, residents in LG_1 whose preferences exceed the average can consume the goods that LG_2 provides, provided that they can access them. For example, in education services, some residents may seek high schools offering advanced STEM classes and enroll their children in schools in another SNG that provides such classes. We consider our assumption more realistic and more general because it reflects how elected representatives seek to leverage their region in providing public goods that local residents deem more important. This single-product model is more realistic as it better captures the nuanced variations in service levels across jurisdictions, rather than assuming completely distinct services. It is also more general, as it can accommodate both scenarios where services differ in quality or quantity, thus encompassing a wider range of real-world situations.

We demonstrate that determining which tier is more efficient in providing a specific local public good hinges on the level of heterogeneity between SNGs and the level of homogeneity within each SNG. Moreover, contrary to the findings of most previous studies, we show that it can be advantageous for local governments to provide the public good even in the presence of spillovers. This outcome depends on the nature of the local public good and the

characteristics of the residents. Specifically, differences in residents' preferences (i.e., the level of heterogeneity) moderate the relationship between the significance of spillovers and the more beneficial tier in providing the public good.

To provide a summary of our results and how they complement prior studies we have augmented Table 1 with our findings, adding the new forces and the moderation effect. Table 3 summarizes the additional forces.

Table 3: Additional forces influencing centralization - decentralization allocation of powers

Force	Pro-Decentralization	Pro-Centralization	Comment
Economies of scale	no economies of scale	economies of scale exist	Corollary 1b
Economies of scale vs. preference variability	Cost saving < preference variability	Cost saving > preference variability	Lemma 3
Spillovers High preferences variability	Significant	Insignificant	Corollary 2b
Spillovers Small Preferences variability	Insignificant	Significant	Corollary 2b
Preferences variability between LGs	Large	Small	Proposition 2 Corollary 2a

We can group the forces that influence the decision-making process of the centralization - decentralization allocation of goods, as Tables 1 and 3 show, into five categories: Economic, Residents, Government Structure, Public Good Characteristics, and Politics. Figure 8 presents the allocation of public goods in the multi-tier subnational government forces model inspired from Porter (1985, p. 5).

Explanations for the categories. (1) Economic: This category includes factors such as economies of scale, cost efficiency, and spillovers. (2) Residents: This category encompasses factors related to citizen preferences, heterogeneity of demand, and information asymmetry. (3) Government Structure: This refers to the organization and levels of government (e.g., local, regional, central) and how they interact in the provision of public goods, and the size of the government. (4) Public Good Characteristics: This category includes aspects specific to the nature of the public goods being provided, such as rivalry, excludability, and the presence of spillovers. (5) Politics: This category covers political factors that influence decision-making, such as voting patterns, political ideologies, and governance structures.

These categories provide a comprehensive framework for analyzing the various forces that affect the allocation of public goods across different tiers of government. The reference to

Porter (1985) suggests that this model adapts Porter's Five Forces model to the context of public administration and fiscal federalism.

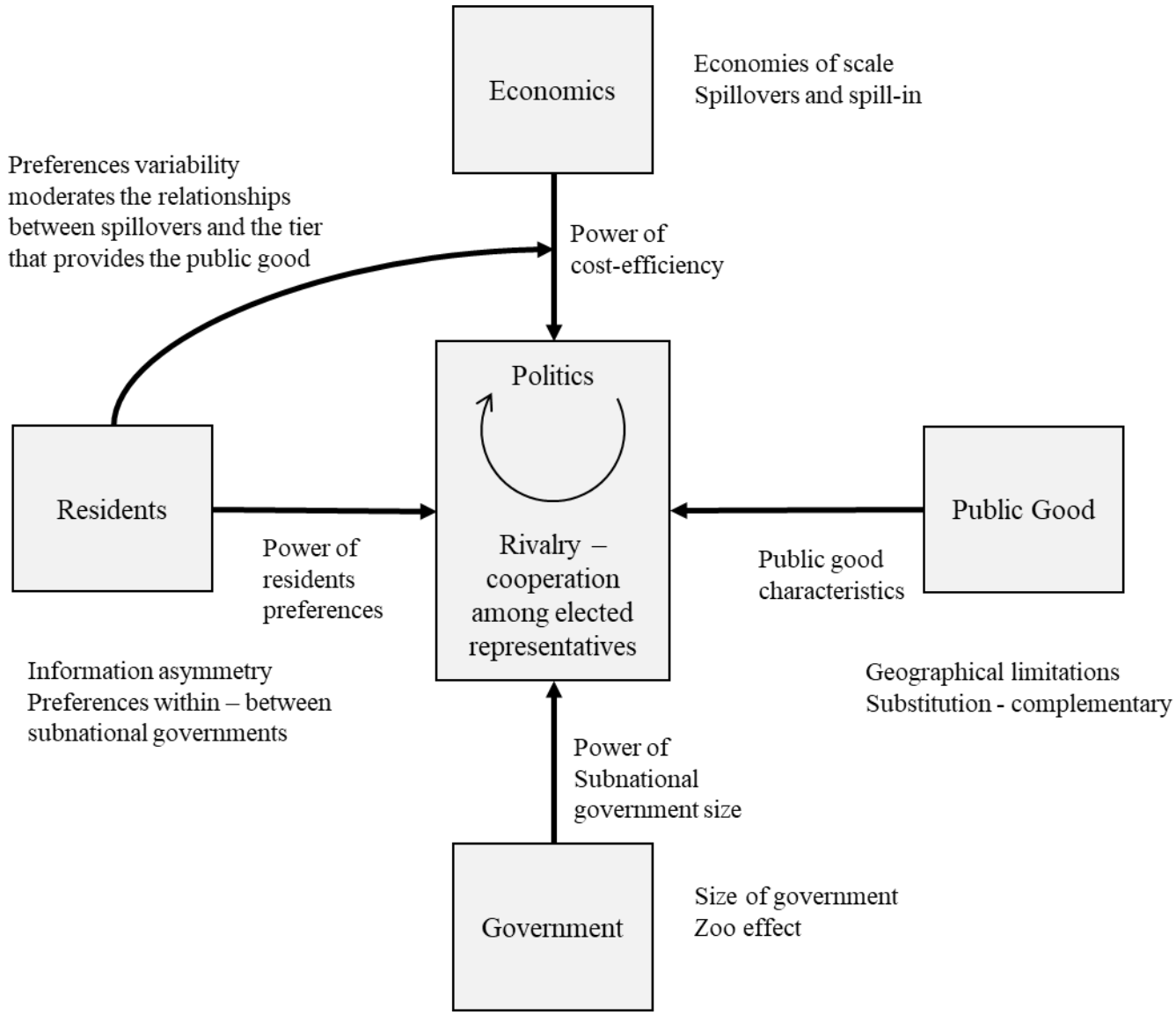

Figure 8: Allocation of public goods in multi-tier subnational governments forces model

The political process at the center of the forces model in Figure 8 integrates all forces and the moderation effect into the decision-making process. Second-generation Fiscal Federalism studies focusing on the political game assume that the decision-making process occurs through negotiations between elected representatives. These negotiations may result with a decision to provide more or fewer product units than the average. We study the consequences of each option, demonstrating that when representatives decide to provide more product units, the region will have to subsidize the cost. Conversely, they'll decide to provide fewer product units if the perceived value is low.

By combining Tables 1 and 3, we create a decision table that can assist policymakers and administrators in determining which public goods should be decentralized and which will be better served by an upper tier. Appendix C and Table C1 provide an example of using such a decision table.

Our research introduces a new model of fiscal federalism that incorporates a third, regional tier of subnational government and addresses local spillovers. This model relaxes the assumption of homogeneous SNGs, allowing for diverse resident preferences for local public goods. The model demonstrates that heterogeneity in preferences can reduce aggregate surplus compared to homogeneous communities. It also shows that when costs are uniform across SNGs, lower tiers should provide local public goods, supporting the Decentralization Theorem. The study explores how economies of scale and spillovers affect the optimal tier for providing public goods, introducing a moderation effect between resident preference heterogeneity and spillover significance. It presents a multi-tier SNG forces model, categorizing influences on centralization-decentralization decisions into economic, resident preference, SNG size, public good characteristics, and political powers. The study concludes with a decision table to assist policymakers in determining the appropriate tier for providing specific public goods. By providing both the theoretical underpinnings of decentralization and a practical tool for applying this theory we aim to contribute not only to academic research but also to applied research and policy making. The decision table we've developed offers policymakers a valuable resource for determining the appropriate tier for providing specific public goods, bridging the gap between theory and practice in fiscal federalism.

Appendix A: The relations between MAD and mean

The mean absolute deviation around the average indicates the average (absolute) distance from the mean. The mean absolute deviation of a set $\{x_1, x_2, \ldots, x_n\}$ is $\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$.

Let us express the average of all the observations which are either less than the average ($\underline{x}^*$) or above it ($\overline{x}^*$) using the mean and MAD for any arbitrary distribution. γ is the ratio of the number of observations less than the average to the total number of observations. For symmetrical distribution γ=½. In general, $\gamma \epsilon (0,1)$, and may be approximated by $\gamma = \frac{mean}{mean+median}$. The averages of the observations, which are either above or below the average, are:

$$\underline{x}^* \equiv \frac{1}{\gamma n}\sum_{\substack{i=1 \\ s.t.\ x_i<\bar{x}}}^{\gamma n} x_i \;;\; \overline{\bar{x}}^* \equiv \frac{1}{(1-\gamma)n}\sum_{\substack{i=\gamma n+1 \\ s.t.\ x_i>\bar{x}}}^{n} x_i \tag{A1}$$

The definition of MAD is

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = \bar{x}(2\gamma - 1) - \gamma\underline{x}^* + (1-\gamma)\overline{\bar{x}}^* \tag{A2}$$

Similarly, the average

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \gamma\underline{x}^* + (1-\gamma)\overline{\bar{x}}^* \tag{A3}$$

Subtracting A2 from A3 yields

$$\underline{x}^* = \bar{x} - \frac{1}{2\gamma}MAD \tag{A4}$$

Adding A2 to A3 yields

$$\overline{\bar{x}}^* = \bar{x} + \frac{1}{2(1-\gamma)}MAD \tag{A5}$$

Appendix B: The optimal allocation of public products – extending the classical approach

Based on Oates Fiscal Federalism, 1972, page 59, which in turn is based on Barzel (1969) “Two propositions on the optimum level of producing collective goods”.

Assume that the cost of additional unit is the same, and the demand curves are linear, at least over the relevant range as Figure B1 shows. C_0 is the product-unit cost to each resident, which results in the demand of Q_i product units. Let us assume two cases: the government provides $\underline{Q}_g < Q_i$, or $\overline{Q}_g > Q_i$. In the first option, residents receive less product units than desired, and their welfare loss is the shaded upper triangle in Figure B1. In the second option residents receive more product units than they prefer, but pay (in taxes) for all the product units their government provides. Their welfare loss is the added cost of the product units above their desired level, which the lower triangle describes.

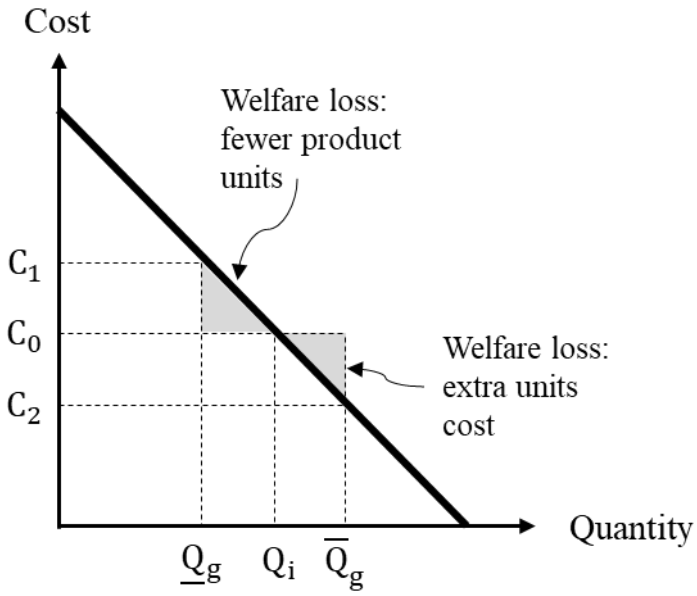

Figure B1: Demand curve of public products

Under the assumption of linear demand curves, the relations between per-product unit cost and the desired quantity is $Q_i = a_i + b_i C_i, b_i < 0$. The welfare loss due to fewer product units is the area of the upper triangle:

$$\Delta \underline{W}_i = \frac{1}{2}\left(Q_i - \underline{Q}_g\right)(C_1 - C_0) = \frac{-1}{2b_i}\left(Q_i - \underline{Q}_g\right)^2 \tag{B1}$$

Further assuming that all residents have the same slope of the demand curve and the same marginal rate of substitution, we can sum the welfare loss for all residents. Let $\gamma \in (0,1)$ be the ratio of the number of residents whose preferences are below the supplied product units, implying that $n(1-\gamma)$ is the number of residents whose preferences are above the supplied product units. Therefore, the total welfare loss of the first case is

$$\Delta \underline{W} = \sum_{\substack{i=\gamma n+1 \\ s.t. Q_i > \underline{Q}_g}}^{n} \Delta \underline{W}_i = \frac{-1}{2b} \sum_{\substack{i=\gamma n+1 \\ s.t. Q_i > \underline{Q}_g}}^{n} \left(Q_i - \underline{Q}_g\right)^2 \tag{B2}$$

Minimizing $\Delta \underline{W}$ results in [17]

$$\underline{Q}_g^* = \frac{1}{(1-\gamma)n} \sum_{\substack{i=\gamma n+1 \\ s.t. Q_i > \underline{Q}_g}}^{n} Q_i \tag{B3}$$

Using the MAD definition,[18] we can write Equation (B3) as

$$\underline{Q}_g^* = \overline{Q} + \frac{1}{2(1-\gamma)} \mathrm{MAD}, \text{where } \overline{Q} = \frac{1}{n}\sum_{i=1}^{n} Q_i \tag{B4}$$

[17] Minimizing ΔW by $\frac{d\Delta W}{dQ_i} = 0, \frac{d^2 \Delta W}{dQ_i{}^2} > 0$

[18] See Appendix A

Similarly, for the case where residents receive more product units than they prefer, the welfare loss will be the area of the lower triangle.

$$\Delta\overline{W}_i = \frac{1}{2}\left(\overline{Q}_g - Q_i\right)(C_0 - C_2) = \frac{-1}{2b_i}\left(Q_i - \overline{Q}_g\right)^2 \quad \text{(B5)}$$

$$\Delta\overline{W} = \sum_{\substack{i=1 \\ s.t.Q_i<\overline{Q}_g}}^{\gamma n} \Delta\overline{W}_i = \frac{-1}{2b}\sum_{\substack{i=1 \\ s.t.Q_i<\overline{Q}_g}}^{\gamma n}\left(Q_i - \overline{Q}_g\right)^2 \quad \text{(B6)}$$

Minimizing $\Delta\overline{W}$

$$\overline{Q}_g^* = \frac{1}{\gamma n}\sum_{\substack{i=1 \\ s.t.Q_i<\overline{Q}_g}}^{\gamma n} Q_i = \overline{Q} - \frac{1}{2\gamma}MAD \quad \text{(B7)}$$

Welfare loss for the whole residents is $\Delta W = \Delta\underline{W} + \Delta\overline{W}$. Minimizing ΔW gives the number of product units that minimizes the welfare loss to be the average over all residents' preferences

$$\overline{Q} = \frac{1}{n}\sum_{i=1}^{n} Q_i = \gamma\overline{Q}_g^* + (1-\gamma)\underline{Q}_g^* \quad \text{(B8)}$$

Appendix C: Using the decision table

Implementing the theoretical model we describe in this study, involves estimating the various forces influencing whether to decentralize powers to sub-national governments. Table C1 **Error! Reference source not found.** presents an application of our model as a tool for a decision-making process in a sub-national structure having three tiers. The planner has to estimate the magnitude of each force between central and regional or regional and local tiers. We provide an example of such a decision table, analyzing which tier will better provide education services. Table C1 lists hypothetical estimates of all forces, where column (5) shows the ranking of local vs. regional tiers, and column (6) ranks regional vs. central governments. Each force ranges between 5 (full force to decentralize), and -5 (full force to centralize).

We describe here the hypothetical numbers given in this theoretical example. For economies of scale the local-regional number equals -3, meaning that centralization to the regional level is moderately preferred to local provision due to economies of scale. When we compare the regional to the central level there is a small advantage to centralize the provision of education services (-1). Moving to spillovers, we assume that these are very small, giving a

small advantage to the local vs. the regional level (+1). The third item is preference variability, which we assume is large, giving an advantage to the level closer to the residents (+4 for local vs. regional, and +4 for regional vs. central). For each additional item in the table, we plug in numbers which correspond to the intensity of the advantage to centralize/decentralize. Finally, we sum all numbers, getting a value of -0.45 for the local-regional comparison, meaning a small advantage to centralize. The value for the regional-central comparison equals 1.09, i.e., there is an advantage to decentralize to the regional level. thus, for education services which exhibit the magnitudes of forces listed in the table there are efficiency gains in providing these services at the regional level.

Table C1: SNG decision table with education service example

Force group	**Force**	**Decentralization**	**Centralization**	**Local - regional**	**Regional - central**
(1)	(2)	(3)	(4)	(5)	(6)
Economics	Economies of scale	insignificant	Significant	-3	-1
Economics	Spillovers	Small	Significant	1	0
Economics	Public good cost vs. preference variability	Cost saving < preference variability	Cost saving > preference variability	3	3
Residence	Information asymmetry	High	Low	4	4
Residents	Residents' preferences between regions	Homogenous	Heterogenous	-3	-1
Residents	Preferences variability between LGs	Large	Small	3	1
Governance	Zoo effect: good scale	Small	Large	-3	3
Governance	Size of local government	Large	Small	-5	2
Public good	Spill-in public goods	Substitute	Complementary	2	2
Public good	Geographic limitations	Bounded	Unlimited	-2	1
Politics	Representation	Averse public spending	Pro public spending	-2	-2
Moderation	Spillovers & High preferences variability	Significant	Insignificant		
Moderation	Spillovers &Small Preferences variability	Insignificant	Significant		
Average				-0.45	1.09

The last row of Table B1 shows the average of each column with equal weights. The result of the local-regional column (5) is a negative average force, meaning it will be better to "locally centralize" the good at the regional tier. The result of the regional-central column (6)

is a positive average force, meaning it will be better to “locally centralize” at the regional tier and not centralize at the central government level.

Future studies may probe further into the question of weights and develop methods to evaluate each force's magnitude.

References

Andrews, R., & Boyne, G. (2009). Size, Structure and Administrative Overheads: An Empirical Analysis of English Local Authorities. *Urban Studies*, *46*(4), 739–759. https://doi.org/10.1177/0042098009102127

Barankay, I., & Lockwood, B. (2007). Decentralization and the productive efficiency of government: Evidence from Swiss cantons. *Journal of Public Economics*, *91*(5–6), 1197–1218. https://doi.org/10.1016/j.jpubeco.2006.11.006

Barzel, Y. (1969). Two propositions on the optimum level of producing collective goods. *Public Choice*, *6*(1), 31–37. https://doi.org/10.1007/BF01718576

Besley, T., & Coate, S. (2003). Centralized versus decentralized provision of local public goods: A political economy approach. *Journal of Public Economics*, *97*, 2611–2637. https://doi.org/10.1016/S0047-2727(02)00141-X

Chatry, I., & Vincent, R. C. (2019). A global view of sub-national governments in Asia: Structure and finance. In J. Kim & S. Dougherty (Eds.), *Fiscal Decentralisation and Inclusive Growth in Asia* (pp. 27–57). OECD Fiscal Federalism Studies. https://doi.org/10.1787/239e95db-en

Cheikbossian, G. (2016). The political economy of (De)centralization with complementary public goods. *Social Choice and Welfare*, *47*(2), 315–348. https://doi.org/10.1007/s00355-016-0962-3

Dur, R., & Roelfsema, H. (2005). Why does centralisation fail to internalise policy externalities? *Public Choice*, *122*(3–4), 395–416. https://doi.org/10.1007/s11127-005-5290-6

Feidler, J., & Staal, K. (2012). Centralized and decentralized provision of public goods. *Economics of Governance*, *13*(1), 73–93. https://doi.org/10.1007/s10101-011-0104-4

Frère, Q., & Védrine, L. (2024). Does decentralisation theorem shape intermunicipal cooperation? *Regional Studies*, *58*(9), 1728–1753. https://doi.org/10.1080/00343404.2023.2282139

Geys, B., & Konrad, K. A. (2010). *Federalism and Optimal Allocation Across Levels of Governance* (SP II 2010-09; WZB Markets and Politics). https://hdl.handle.net/10419/54598

Gregor, M., & Stastna, L. (2012). The decentralization tradeoff for complementary spillovers. *Review of Economic Design*, *16*(1), 41–69. https://doi.org/10.1007/s10058-012-0113-y

Halásková, M., & Halásková, R. (2015). Fiscal decentralisation and provision of local public

services in selected EU countries. *Lex Localis*, *13*(3), 595–613. https://doi.org/10.4335/13.3.595-613(2015)

Hämäläinen, K., & Moisio, A. (2015). *One or Two Tiers of Local Government? – The Cost Effects of a Regional Experiment* (VATT No. 65).

Jones, M. I., Sirianni, A. D., & Fu, F. (2022). Polarization, abstention, and the median voter theorem. *Humanities and Social Sciences Communications*, *9*(1), 1–12. https://doi.org/10.1057/s41599-022-01056-0

Koethenbuerger, M. (2008). Revisiting the "Decentralization Theorem"-On the role of externalities. *Journal of Urban Economics*, *64*(1), 116–122. https://doi.org/10.1016/j.jue.2007.10.001

Lockwood, B. (2002). Distributive politics and the costs of centralization. *Review of Economic Studies*, *69*(2), 313–337. https://doi.org/10.1111/1467-937X.00207

Lorz, O., & Willmann, G. (2005). On the endogenous allocation of decision powers in federal structures. *Journal of Urban Economics*, *57*(2), 242–257. https://doi.org/10.1016/j.jue.2004.11.002

Martinez-Vazquez, J., Lago-Peñas, S., & Sacchi, A. (2017). The Impact of Fiscal Decentralization: a Survey. *Journal of Economic Surveys*, *31*(4), 1095–1129. https://doi.org/10.1111/joes.12182

Musgrave, R. A. (1959). *The theory of public finance : a study in public economy*. McGraw-Hill.

Oates, W. E. (1972). *Fiscal Federalism*. Harcourt Brace Jovanovich.

Oates, W. E. (1988). On the measurement of congestion in the provision of local public goods. *Journal of Urban Economics*, *24*(1), 85–94. https://doi.org/10.1016/0094-1190(88)90048-4

Oates, W. E. (2005). Toward a second-generation theory of fiscal federalism. *International Tax and Public Finance*, *12*, 349–373. https://doi.org/10.1007/s10797-005-1619-9

Oates, W. E. (2008). On the evolution of fiscal federalism: Theory and institutions. *National Tax Journal*, *61*(2), 313–334. https://doi.org/10.17310/ntj.2008.2.08

OECD/UCLG. (2016). *Subnational governments around the world: Structure and finance*. OECD. https://www.oecd.org/regional/regional-policy/Subnational-Governments-Around-the-World- Part-I.pdf

Olson, M. (1969). The Principle of "Fiscal Equivalence": The Division of Responsibilities among Different Levels of Government. *American Economic Review*, *59*(2), 479–487. https://doi.org/10.2307/1823700

Ponzetto, G. A. M., Backus, P., Berry, C., Bustos, P., Carvalho, V., Cicala, S., Donder, P. De, Enikolopov, R., Gancia, G., & Nuevo-, A. (2016). Political Centralization and Government Accountability. *The Quarterly Journal of Economics*, *131*(1), 381–422. https://doi.org/10.1093/qje/qjv035.Advance

Porter, M. E. (1985). *Competitive Advantage: creating and sustaining superior performance*. The Free Press.

Reingewertz, Y. (2014). Fiscal Decentralization - A Survey of the Empirical Literature. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2523335

Seabright, P. (1996). Accountability and decentralisation in government: An incomplete contracts model. *European Economic Review*, *40*, 61–89. https://doi.org/10.1016/0014-2921(95)00055-0

Shani, R., Reingewertz, Y., & Vigoda-gadot, E. (2023). Regional governments in the rural space: the effectiveness of dual-tier regional councils. *Regional Studies*, *57*(9), 1816–1831. https://doi.org/10.1080/00343404.2022.2153825

Slavinskaitė, N., Liučvaitienė, A., & Gedvilaitė, D. (2019). Theoretical Analysis of the Fiscal Federalism. *International Journal of Contemporary Economics and Administrative Sciences*, *9*(2), 250–267. https://doi.org/10.5281/zenodo.3595926

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, *64*(5), 416–424. https://doi.org/10.1086/257839

Weingast, B. R. (2009). Second generation fiscal federalism: The implications of fiscal incentives. *Journal of Urban Economics*, *65*, 279–293. https://doi.org/10.1016/j.jue.2008.12.005