DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 03-09

Can Decentralization Be Beneficial?

Anna Rubinchik-Pessach
Department of Economics, University of Colorado at Boulder
Boulder, Colorado

October 2003

Center for Economic Analysis
Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

Can Decentralization Be Beneficial?*

Anna Rubinchik-Pessach†

Revised: October 2003

Abstract

A conventional justification for government hierarchy in the fiscal federalism literature is based upon asymmetry in policy tools or in information access that is available to different levels of government. This paper demonstrates that even if these asymmetries are eliminated, addition of local (regional) governments to a one-tier central government can be strictly welfare improving.

Keywords: public goods, decentralization
JEL Classification Codes: H41, H72, H73, H11, D72

*This paper is based on my dissertation. I am indebted to Andrew Postlewaite for his guidance, and to Nicola Persico and Steve Matthews for insightful discussions. The paper was revised while the author was visiting CORE, the hospitality of which is greatly appreciated. The research was partially supported by the European Commission Network - HP RN-CT-2000-00072. I would also like to thank Ann M. Carlos, David Cass, Masa Fujita, Hans Gersbach, Robert Inman, George Mailath, Antonio Merlo, Sergio Orioli Parreiras, Ori Pessach, Amir Rabah, Jack Robles, András Simonovits, Jacques-François Thisse and Shlomo Weber as well as participants of the Applied Microeconomic theory seminar at the University of Pennsylvania, Political Economy Speaker Series at the University of Colorado and the Conference on "The Economics of Political Integration and Disintegration" at CORE as well as the anonymous referees who provided useful comments. Any remaining errors are mine.

†Address: University of Colorado at Boulder, Department of Economics, UCB 256, Boulder, Colorado, 80309, USA; e-mail: Anna.Rubinchik@Colorado.edu

1 Introduction

Hierarchical governments, either federations or confederations, are widely spread across the world and are persistent across time. This may suggest that they are outperforming the unitary ones. However, according to a textbook economic argument, a central government can internalize the externalities, hence it has an advantage over a set of local ones, where politicians, representing their constituents, are motivated solely by local interests. So in a world, in which both levels of government have access to the same information, the role for local governments is unclear. Nevertheless, even in such a world maintaining a multi-leveled government may be worthwhile as opposed to a potentially cheaper unitary one. This paper provides an explanation.

Let me briefly describe the set up. Regional representatives, or legislators, striped off their "personal position" (say, identified by the district they will represent),[1] have to write a constitution for their country. "Constitution" in this model is just a set of rules governing acceptance of public projects. As in Harsanyi (1992), the legislators are rational and it is common knowledge that each one of them may occupy any position (represent any locality with equal chance) after the "veil of ignorance" is lifted. Assume their beliefs are consistent, so that there is a common prior, F, with respect to the appearance of public projects, and this prior is shared behind the veil of ignorance. The benefits and costs of a project are expressed in the same (monetary) terms. Thus, by Harsanyi (1992), the social objective is to maximize the expected (net) value of a public project, or simply, the welfare. All the ex-ante identical individuals should agree on this objective, and so they can be thought of as a single (fictitious) individual, named 'constitutional designer.'

The designer will decide to add local governments only if by doing so he can increase the expected welfare. The role of local governments here is reduced to accepting public projects that can be fully financed by one of the regions. In the absence of negative externalities, which is true in this model, the addition of local governments cannot decrease welfare. Then the hierarchy is justified only if generates a strictly positive social benefit that can cover costs of running another tier of government (either explicit or implicit), which are omitted here for simplicity. That is why decentralization is called beneficial in this model only if it is *strictly* welfare improving.

Clearly, the decision of whether or not to decentralize depends on the rules

[1] or behind Rawlsian/Harsanyi's "veil of ignorance."

according to which the central legislature operates and its performance. If the designer is unconstrained, he can condition the constitution on the benefits and costs that become known to the legislators once a project appears on the agenda, and attain the ex-post optimum as in Laffont and Maskin (1982). Then there would be no need for local governments. This result, however, rests on the assumption that the realization of the payoffs is truly exogenous, in other words, it is not subject to strategic manipulations. Unfortunately, in many practical cases the assumption is violated, as estimating benefits and costs of a public project requires specific knowledge, or expertise, that politicians lack. For example, in the U. S. the evaluation of water projects is performed by the U. S. Army Corps of Engineers. Clearly, once it is done, predicted benefits and costs become common knowledge among the legislators.[2] In this case, conditioning the cost sharing arrangements on benefits would be equivalent to delegating *the formulation of the rules* to the "experts," who conduct the evaluation study. As their objectives may not coincide with those of the elected representatives, and, especially, those of the general electorate,[3] this can result in additional informational rents extracted by the experts, which, clearly down-plays the attractiveness of such a constitution. A way to avoid the "garbling of the information structure" (as in Tirole (1999)) resulting from formulating the constitution solely in terms of benefits and costs, is to write it as a "complete contract," fully specifying future contingencies and the corresponding rules of acceptance of public projects. It is commonly known, however, that actual higher order laws are far from being that detailed. Reasons for that are abundant. Apart from computational complexity associated with writing and implementation (see Anderlini and Felli (1999)) or even examining contractual completeness (see

[2]Similar procedures are applied in the other areas of public policy.

[3]Large scale water projects in the U.S. have been evaluated and constructed (if approved by the Congress) by the U. S. Army Corps of Engineers for over a century. From *Washington Post, May 14, 2002; Page A2, "150 Water Projects Halted For Army Corps Review" by Michael Grunwald:*

> ".. In 2000, the Post series detailed how the Corps has justified many projects with skewed assumptions and overly optimistic predictions of barge and ship traffic. E-mails from high-ranking Corps officials revealed that they had manipulated an economic study in order to justify a billion-dollar lock expansion project on the Mississippi River. An internal Pentagon investigation concluded that Corps studies were tainted by an institutional bias toward large-scale construction."

Gilboa, Postlewaite, and Schmeidler (2002)), the enforcement of such a constitution can be problematic, both because some contingencies are impossible to verify and also due to the intricacies of creating a proper motivation for the judges (see Tirole (1999) for the overview). Abstracting from the problem of formulating an "optimally incomplete" contract, which takes into account all the above mentioned considerations, I will follow Aghion and Bolton (2003) in assuming a particular form of incompleteness, so prevalent in practice. A public good has to be provided if at least m out of R legislators vote in its favor. In contrast to Aghion and Bolton (2003), taxes in this model can be specified ex-ante and can be conditioned on the voting behavior.

The core primitive of the model is the distribution, F, over future public projects as perceived by the constitutional designer. It reflects the likelihood of different profiles of gross benefits that will accrue from the projects. Regional variation in the willingness to pay for a project can result from a disparity in tastes, or from technological constraints. For example, the appeal of a publicly broadcasted ballet can vary across regions according to tastes, whereas the benefits of an environmental regulation can vary due to its exact (technological) specifications. The need for a public project can arise from a vast variety of events that are impossible to predict at the time the constitution is created. However, I assume that the designer has a reliable estimate of distribution F over possible realizations of benefits.

In this model the superiority of a hierarchical government is driven by the necessity to use one voting rule and a single tax system for selection of *different* types of projects, i.e., those that generate either equal or unequal distribution of benefits across regions. This creates inefficiency that can be (partially) eliminated by allowing those projects that would be rejected by the central legislature to be accepted on the local level. Indeed, the fact that some of the projects will be approved by the lower tier of government allows a benevolent constitutional designer to set better, or more "specialized" voting rules and taxes for the central legislature. This improves the overall performance of a hierarchical government over its one-tier counterpart. The result is true, even if local governments cannot internalize the externalities at all.

1.1 Related Literature

It would be almost impossible to trace the first discussion of the subject. Already by mid-19th century Guizot (1861),[4] emphasized the tensions between a centrally exercised power, that is "...generally more disinterested, and more capable of taking justice and reason for its sole guide," and local institutions, the preponderance of which was attributed to the "infancy of societies," but which, are, nevertheless needed as the guards against usurpation of power by the center. Under a "truly representative" government described in this model every citizen enjoys sufficient protection from being discriminated against, at least on average, as the expected value of projects accepted by the central government is positive.

Later contributions focused on optimal allocation of responsibilities across levels of government based on the scope of public projects (Oates (1972), Lockwood (2002), Besley and Coate (2000) among others), access to information with respect to benefits and costs of the projects by level of government (Gilbert and Picard (1996), Zantman (2002), see Crémer, Estache, and Seabright (1996) for an overview), or the access to the tax instruments available to the governments (Boadway, Marchand, and Vigneault (1998), Hochman, Pines, and Thisse (1995); see Wellisch (2000) for an overview).

The first two branches of the literature are most closely related to this model, and will be discussed in some more detail. Oates (1972) suggested that central governments make better decisions with respect to "global" public goods, which concern everybody in a country, whereas lower tier is better at providing the "local" public goods, the effect of which is mainly regional. By that argument, which rests on the presumption that central government has to provide the same amount of public good to every region and that the taxes have to be uniform, a country with heterogeneous population needs both levels of government. This is consistent with empirical findings by Panizza (1999), who shows that the degree of centralization is negatively correlated with the differentiation of tastes, where the ethnic fractionalization is used as a proxy for heterogeneity of tastes with respect to public goods.

More recent work by Lockwood (2002) and by Besley and Coate (2000) relax the uniformity of public spending across regions, but still preserve uniform cost sharing for public projects. In these models central government is comprised of regional representatives, who have to allocate tax revenues to public projects. The authors show that even in that environment the

[4] p. 34.

trade-off between centralization and decentralization still remains. A common source of inefficiency on the central level in these models is a budgetary externality created by equal cost sharing. Lockwood (2002), demonstrates that under majority rule with uniform taxation the legislature rejects some desirable projects and accepts some undesirable ones. The author concludes that in the absence of externalities generated by public goods, decentralization is preferable for the country, in which residents are identical within a region. Besley and Coate (2000) reach a similar conclusion by allowing for heterogeneity of tastes within a region. In the presence of equal cost sharing, voters in any region have a motivation to elect to the central legislature a 'biased' representative, who is ready to accept more public projects than a median voter.

Another argument justifying government hierarchy rests on the common wisdom that local governments have superior information about citizens' preferences or technologies to be used in provision of public goods (this asymmetry is assumed away in the current paper). Restriction on *information acquisition* by central governments is questionable both in reality and on theoretical grounds, see Crémer, Estache, and Seabright (1996), although a central government may have fewer *incentives* to acquire the relevant information than the local governments.

The approach used in this paper also borrows from the mechanism design literature. Most recently, Palfrey and Ledyard (2002) rationalized referenda for big populations with independent valuations by comparing them to the optimal procedure that solves the corresponding designer's problem.

In this paper optimal solution is identified for some environments, which allows to rank performance of a hierarchical government versus the unitary one. In other environments, dominance argument is used to make the case for or against decentralization. Note also that this framework is explicitly constructed for a small number of decision makers (legislators) and the dependence in the valuations plays an important role in the analysis.

The rest of the paper is organized as follows. The next section outlines the model and the outcomes of the voting game in the central legislature. These results are used in the following section to formulate the problem of the constitutional designer. It is then demonstrated that an addition of local governments is *not* beneficial if the central legislature has to consider public projects of the same type, with known shape of inter-regional spread of benefits. Environments with very different profiles of benefits accruing from public projects are presented in subsection 3.2 and that is where adding local governments appears worthwhile.

2 The model

There are $R \geq 3$ regions, R is odd. The residents of each region have identical attitudes towards public goods, but the attitudes across regions differ. Each region has one representative in the federal (central) legislature. The legislature has to decide whether to provide a the public project that appears on the agenda.

A project is associated with a profile of benefits, $b = (b_1, .., b_R) \in \mathbb{R}_+^R$, across the regions. Regional benefit, b_r, can be thought of as a sum of (identical) benefits to the residents of the region in case the project is provided.[5] The per-region cost of a project is normalized to unity for simplicity.[6] Therefore, b_r can be thought of as regional willingness to pay per average cost of the project.

As argued in the introduction, the core primitive of the model is probability distribution $F(b)$ over projects. It reflects the frequency with which projects will appear on the legislative agenda. For the rest of the exposition assume F is continuously differentiable and has full support on $[0, B]^R$ with $1 < R < B < \infty$, unless specifically mentioned otherwise.

Next, let us turn to the timing. The **two-tier government** in this model operates in the following fashion:

Stage 0. Given probability distribution over projects, the benevolent designer chooses optimal voting and tax rules for the central legislature.

Stage 1. A new project, b, appears on the agenda of the central legislature and is voted upon according to the rules set by the designer. If the project is accepted, it is implemented and stage 1 repeats. If not, stage 2 takes place.

Stage 2. The rejected project is considered by the local (regional) legislatures. If accepted (by one of the regions), it is implemented. If not, it is never considered again. Stage 1 proceeds.

[5] Assume that regional taxes are uniform. This is optimal from the regional perspective, given identical attitudes towards public goods within each region.

[6] This implies that the projects generating the same ratio of net benefits to the cost receive equal "weight" in the social welfare function. This is unimportant for the following argument, which is based on how spread the benefits are across the regions, and how well different acceptance rules can be adapted to various forms of dependence of benefits across the regions.

The description of a **one-tier government** is the same, apart from the fact that the last stage 2 is omitted.

Throughout the rest of the paper the *addition of the local governments* to a one-tier government means the addition of stage 2 to the legislative process.

Note that local governments can accept the projects rejected by the central government, whereas they are unable to prevent the projects accepted by it from being implemented. This feature of the model reflects the fact that some federal countries (USA, Germany, Russia) have a clause stating the superiority of the federal law in their constitutions. Thus, once a bill (a project) is accepted by the central legislature, it becomes a federal law and local governments have very limited ability to repeal it (see Finer, Bogdanor, and Rudden (1995)). On the other hand, if a bill was rejected by the central government, with no additional restrictions imposed, local governments are free to accept it.[7] Let us impose the following assumption.

Assumption LOCAL A public good is accepted on the local level, iff there is at least one region ready to pay for it.

One could argue that this assumption is somewhat restrictive and unrealistic. In fact, a large portion of public goods provided by the US local and state governments is being partially financed by the federal government (see Inman (1988)). Even so, by invoking this assumption, one "stacks the deck" against justification for local governments. Indeed, if local governments have to bear full financial responsibility for public projects they accept, then fewer projects will be accepted locally. This, in turn, diminishes the importance of the lower tier of government. Implicit here is also an assumption that in case several localities are ready to pay for a project on the local level, the identity of the provider is commonly known (in other words, once the project descents onto the lower level, the local governments play Nash equilibrium of the public good provision game).

Next subsection derives the relationship between voting and tax rules chosen by the designer on one hand and the set of projects accepted by the central legislature in a one-tier government on the other.[8]

[7] In other words, public projects considered in this model are those falling into the area of "concurrent legislation".

[8] See Rubinchik-Pessach (2002) for more details.

2.1 Voting in the Central Legislature

Assume that the constitutional designer is creating the rules of acceptance of projects subject to a set of constraints. So, consider a project b that appears on the agenda.

Assumption VO Project is accepted if and only if at least m legislators vote for it;

Assumption TAX Taxes are imposed only if the project is accepted;

Assumption AN Taxes are anonymous, i.e., a tax can not depend on the name of the region;

Assumption BB The sum of the tax payments should equal to the cost of the project;

Assumption SPM The Supporters of the project Pay (weakly) More than those who oppose the project.

All the assumptions **LOCAL**, **TAX**, **AN**, **BB**, **SPM** will be adopted for the rest of the analysis, and will be discussed in the conclusions.

Specification of the cost sharing rule By anonymity, taxes may differ only on the basis of a legislator's voting decision. So if $k \geq m$ legislators vote in favor of the project, the designer has to set two levels of taxes: $t_k(Y)$ and $t_k(N)$. Let the tax of the supporter be

$$t_k(Y) = \alpha_k, \tag{1}$$

if $k \geq m$ legislators vote for the project. When the taxes are uniform, $\alpha_k = 1$, so that every region pays the average cost of the project, unity. By assumption **SPM**, $\alpha_k \geq 1$. In other words, α_k measures the amount by which a supporter pays more than the average cost. By **BB**, anyone, who opposes the project is required to pay

$$t_k(N) = \frac{R - \alpha_k k}{R - k}. \tag{2}$$

Thus the sharing rule is fully specified by the vector $\alpha = (\alpha_m, \alpha_{m+1}, .., \alpha_R)$. The next subsection demonstrates that we can restrict attention only to the first element, α_m, of this vector in order to determine the set of outcomes of the voting game.

Faced with the cost sharing arrangement, $(\alpha_m, \alpha_{m+1}, .., \alpha_R)$ and given a realization of a project, $(b_1, ..., b_R)$, the legislators play a simultaneous voting game. In this game a legislator i has two actions (votes): $A_i = \{Y, N\}$; and a linear utility :

$$U_i(b_i, a_i, a_{-i}) = \left\{ \begin{array}{l} b_i - t_k(a_i), a_i \in A_i, \text{ if } |\{j : a_j = Y\}| \geq m; \\ 0 \text{ otherwise.} \end{array} \right\}, \quad (3)$$

where $|X|$ denotes cardinality of set X.

Assume the legislators have full information about the benefits that a potential project generates. Consider trembling hand perfect Nash equilibria of this game (Selten (1975)). Denote by M the set of projects that will be accepted, if this equilibrium is being played. The following proposition demonstrates that this set can be fully described by the cost sharing parameter, α_m. The description of this set is simple: the projects that generate m^{th} highest benefit above the threshold α_m will be accepted by the legislature (whereas all the rest of the projects will be rejected).

Proposition 2.1 *The set of projects accepted under a trembling hand perfect Nash equilibrium can be described as follows:*

$$\left\{(b_1, ..., b_R) : b_{[m]} > \alpha_m\right\} = M(\alpha_m, m). \quad (4)$$

Proof is relegated to the appendix. ■

The claim is based on the following result: whenever the project is provided, exactly m people vote for it provided the taxes are non-uniform, i.e., if the supporters pay strictly more than those who oppose the project. The same is true if the taxes are uniform for robust equilibria.[9]

3 Optimal Constitutions Under the Two Regimes

Recall that the benefit profile is normalized so that the per-region cost of a project is unity, thus, the utilitarian welfare from project b is $w(b) \equiv \sum_{i=1}^{R} b_i - R$. A project, generating positive welfare will be referred to as "efficient". Slightly abusing notation, α_m, the crucial component of the cost

[9]See details in the appendix.

sharing vector will be denoted simply by $\alpha \in \mathbb{R}$. Therefore, the optimal constitution should specify the parameters (α, m) that maximize the expected utilitarian welfare calculated over the set of accepted projects. Under the "unitary" state this set is $M(\alpha, m)$, those projects that generate high enough benefit for the pivotal voter (m) to cover his tax bill, α. Under a two-tier government this set also includes the projects generating maximal benefit enough to cover its the costs, $M(R, 1)$. Hence adding another tier adds the set $\hat{M}(\alpha, m) \equiv M(R, 1) \backslash M(\alpha, m)$ to the set of accepted projects.

Therefore, for a one-tier government, the problem faced by the designer is

$$\max_{m,\alpha\geq 1} W_I(\alpha, m), \tag{5}$$
$$W_I(\alpha, m) \equiv \Pr(M(\alpha, m)) E(w(b) \mid b \in M(\alpha, m))$$

whereas for the two-tier government this problem is modified as follows:

$$\max_{m,\alpha\geq 1} W_I(\alpha, m) + Z(\alpha, m), \tag{6}$$
$$Z(\alpha, m) \equiv \Pr\left(\hat{M}(\alpha, m)\right) E\left(w(b) \mid b \in \hat{M}(\alpha, m)\right).$$

Clearly, adding local governments can not reduce welfare in this model, $b_{[1]} > R$ implies $w(b) > 0$, so that $Z(\alpha, m) \geq 0$.[10] It *strictly* increases welfare only if some of the projects rejected in the absence of the lower tier, would be accepted in its presence. In other words, it has to be the case that the set $\hat{M}(\alpha, m) = M(R, 1) \backslash M(\alpha, m)$ is non-empty for (α, m) that solve problem (6). For example, if the optimal constitution in the presence of local governments calls for $m = 1$ and some $\alpha \geq 1$, no projects rejected by the central legislature will be accepted on the local level: $\hat{M}(\alpha, m) = M(R, 1) \backslash M(\alpha, 1) = \varnothing$, provided $\alpha \leq R$.[11] In this case decentralization, or an addition of another tier of government can not be beneficial.

[10] One could add administrative costs of running a local layer of governments, thus the addition of that layer should generate strict benefits to cover the costs.

[11] The last inequality is always true, as for any realization of benefits profile, b, it is optimal to accept the project if $w(b) \geq 0$, or, alternatively, if the average benefit, $\bar{b}$, is bigger than unity, which is equivalent to

$$b_{[m]} \geq \frac{b_{[m]}}{\bar{b}}, \tag{7}$$

i.e., the optimal threshold, α, for a given realization has to be $b_{[m]}/\bar{b}$. The last expression

This leads us to two important conclusions. First, the presence of local projects with few externalities is not sufficient to justify the existence of local governments. Indeed, would these projects be prevalent on the agenda, the rules (constitutions) of the form $(\alpha, 1)$ could have provided an adequate criterion for the central legislature to decide on provision of public projects. Moreover, central provision would dominate local provision, as it can account for externalities, i.e., set $\alpha < R$. The next subsection provides illustration of this idea.

Second, in order to justify the presence of local governments, it is essential to have an environment, in which optimal constitution will not require the minimum winning coalition, m, to be unity, thus leaving a niche for the local governments. As will be shown below this is true if both local and global issues can appear on the agenda of the central legislature.

3.1 Decentralization May Not Be Beneficial

In this section I provide two environments in which adding local governments will not improve welfare.

Let us start with a simple example, in which the (ordered) benefits are perfectly correlated, in other words, the ratios of benefits is constant.[12] It implies that any public project that the legislature considers gives rise to the same spread of benefits *across* the regions. Later this spread will be loosely referred to as "spillovers." Whatever this shape is, as long as it is known *à priori* adding local governments is not beneficial, as they will remain idle, according to the following lemma. Note that it includes both the case of "global" projects with regional benefits being equal (or close) to each other, as well as "local" projects with concentrated benefits.

---

is bounded from above for any realization of the benefits profile:

$$\frac{b_{[m]}}{\bar{b}} \leq \frac{b_{[1]}}{\bar{b}} \leq \frac{b_{[1]}}{\frac{1}{R}b_{[1]}} = R, \text{ provided } b_{[m]} > 0. \tag{8}$$

In the event when all the benefits are zero, i.e., $b_{[1]} = 0$, any choice of $\alpha \geq 1$ is optimal. Thus, without loss of generality we can require $\alpha \in [1, R]$. In the view of this remark, continuity of F and acknowledging that $m \in \{1, 2, ..R\}$, both problems (5) and (6) have solutions.

[12] Clearly, the full support (of F) assumption is dropped for this example.

Lemma 3.1 *Let $b_{[1]}$ be distributed with some continuous (non-degenerate) probability density function S on $[0, B]$, $B > R$. Let*

$$\beta_k \equiv \frac{b_{[k]}}{b_{[1]}}, k \in \{2, 3, ..R\} \tag{9}$$

be constant for any realization b of benefits.[13] *Let $\bar{\beta} = \frac{1}{R}\sum_{k=1}^{R}\beta_k$. Then $(\alpha(1), 1)$ with $\alpha(1) = \frac{\beta_1}{\bar{\beta}}$ solves problem (6), so that $Z(\alpha(1), 1) = 0$.*

Proof Start with problem (5), which in this case simplifies to

$$\max_{\alpha \geq 1, m \in \{1,2,..,R\}} \int_{\alpha/\beta_m}^{B} \left(\bar{\beta}b_1 - 1\right) dSb_1, \tag{10}$$

admitting (multiple) solutions of the form,[14]

$$(\alpha(m), m), \alpha(m) = \frac{\beta_m}{\bar{\beta}} = \frac{b_{[m]}}{\bar{b}} \geq 1. \tag{11}$$

Note that the last constraint is satisfied for at least one of the solutions, as $\alpha(1) = \frac{\beta_1}{\bar{\beta}} \geq 1$. Clearly, also, $\alpha(1) \leq R$. Interestingly, none of these solutions can be improved upon, as they are also optimal conditional on any realization of benefits. Obviously, then $(\alpha(1), 1)$ also solves problem (6). ■

It is true that the possibility of attaining ex-post efficiency is an artifact of this example. Its aim, though, is to illustrate that inefficiency of provision of "local" public goods by the central government can be eliminated without adding local governments. This conclusion is valid in the environments, in which public goods generate "similar" spillovers, or, which give rise to constant benefits' ratios, in terms of the example.

Let me stress that the majority rule ($m = (R + 1)/2$) with uniform taxation ($\alpha = 1$) may not always lead to the efficient provision of public goods

[13] If $b_{[1]} = 0$, let $\beta_k = 0$.

[14] This solution bears a close relationship to the Lindahl taxes. Assume that a beneficiary $r \in \{1, .., R\}$ has to pay a (Lindahl) tax $\alpha(r) = b_r/\bar{b}$, if the project is provided. Allow $\alpha(r)$ to admit any values (fall below unity). Require unanimity. Then all the efficient projects will be accepted, while all the inefficient projects will be rejected. See Myles (1995), p.272-273 for an additional interpretation of the tax.

by the central legislature (and it will not, if $\alpha\left(\left(R+1\right)/2\right) \neq 1$ in this case). It does not follow, however, that decentralization is *necessarily* needed to restore efficiency. An optimal voting rule and cost sharing arrangements for the upper level legislature can achieve the goal without an addition of another layer of government.

The last assertion will also be true for another environment, without the "knife-edge" assumptions, as in the previous example. To stress the comparison with the accepted view, assume that legislature has to deal with (primarily) local projects, thus, the benefits are concentrated. In particular, assume that the second highest benefit is most likely to be below the average cost. In addition, assume that a public project often generates some spillovers. The following lemma demonstrates that no matter how small are the spillovers, decentralization is not beneficial, if the benefits are sufficiently concentrated.

The idea behind the proof is quite simple. First, observe that the problem for the two-tier government (6) can be reduced to the first problem by restricting the benefits to the subset of the support, $[0,R]^R \subset [0,B]^R$. Indeed, any project with at least one benefit realization above R will be accepted in the presence of local government for any pair (α, m). So, the "constitution" affects acceptance of only those projects for which $b_{[1]} \leq R$. Thus, to prove the statement, one needs to show that the solution to the first problem with restricted support does not call for $m > 1$. In this case it is done by making the switch from (α, m') to $(\alpha_\varepsilon, 1)$, $m' > 1$ worthwhile: the expected value of good accepted projects under (α, m'), can be made arbitrarily small, while the expected value of good rejected projects, generating relatively high maximal benefit and some spillovers, is bounded away from zero.

Lemma 3.2 *Restrict the support of F to $[0,R]^R$. Let $\Pr\left(\varepsilon < b_{[2]} < 1\right) = 1-\delta$, $\alpha_\varepsilon = R - \varepsilon$. Assume $\Pr\left(M\left(\alpha_\varepsilon, 1\right)\right) = p$ is bounded away from zero. Then for any $\varepsilon > 0$ there is $\delta > 0$ such that the rule $(\alpha_\varepsilon, 1)$ dominates (α, m) for any $m > 1$ for the two-tier government (problem (6)), therefore a solution to problem (6) is of the form $(\alpha, 1)$.*[15]

[15] As an example consider independently distributed benefits: let b_1 be distributed with some continuous density function H non-degenerate on $[0,R]$, and for $k > 1$ let $b_k \sim S_{\theta,n}$ on $[0,R]$, where $S_{\theta,n}$ is a two sided power distribution: $S_{\theta,n}(b) = \begin{cases} \theta\left(\frac{b}{\theta R}\right)^n & 0 \leq b \leq \theta R \\ 1-(1-\theta)\left(\frac{1-(b/R)}{1-\theta}\right)^n & \theta R < b \leq R \end{cases}$. By van Dorp and Kotz (2002) for big enough n the hypothesis of the lemma will be satisfied.

Proof First, note that for any $m' > 1$ expected value of the projects, W_I, under any rule (α, m') is bounded by $\delta R(R-1)$. Next, let $E(w(b)|M(\alpha_\varepsilon, 1)) = K$. As $\alpha_\varepsilon = R - \varepsilon$, K can be made strictly positive by choosing δ small enough, reducing the chance that $b_{[2]}$ will fall below ε. Pick δ such that

$$\delta < \frac{pK}{R(R-1)}, \tag{12}$$

and $K > 0$. Therefore $W_I(\alpha_\varepsilon, 1) = pK > W_I(\alpha, m')$, which proves the assertion. ■

Clearly, in this case decentralization is useless.

The same is true for another class of examples, which will be mentioned here briefly, mainly, to provide an adequate reference to the literature mentioned in the introduction.

If the valuations (b_i) are identically and independently distributed, say, H, and the population is sufficiently large, then one of the following "degenerate" voting rules can be approximately efficient: accept any project $(m = 0)$ in case the average benefit is above the average cost (unity) and reject any project $(m = \infty)$ otherwise. Indeed, if the distribution H is known ex-ante, so is its expected value, which will be close enough to the realized average of benefits by the law large numbers, and so, the degenerate rule should be close enough to an efficient one.

Similarly, uniform taxation and the simple majority $(m = (R+1)/2)$ are approximately efficient (given R is sufficiently big), if the benefits (b_i) can be decomposed into a common (national) and idiosyncratic (regional) component, so that $b_i = v + \varepsilon_i$, where v is distributed G and ε_i are i.i.d. with zero mean, so that $E(b_i|v) = v$.[16]

3.2 Decentralization Can Be Beneficial

The factor that contributes to the superiority of a hierarchical government in this model is the existence of *different types* of public goods. This will be illustrated with the help of two environments dissimilar in formalizing this idea.

First consider an example demonstrating the necessity of local governments in the environment with two types of projects, one of which is "global" (b^H), with the second highest benefit being always above the average. The

[16] See proposition 4.1. in Rubinchik-Pessach (2002) for more details.

other project is local $\left(b^L\right)$ with benefits accruing to just one region.[17] More precisely,

Proposition 3.3 *Assume the vectors of benefits, b^L and b^H generate constant benefit ratios*

$$\beta_r^i = b_r^i / b_1^i, \; i = H, L; r = 1, .., R \tag{13}$$

and let b_1^H, b_1^L be identically and independently distributed distributed S on $[0, B]$, with bounded density s. Assume $\beta_L = (1, 0, .., 0)$ and $\beta_2^H > \bar{\beta}^H$. Let the projects L and H, appear with frequencies $p, 1-p$ accordingly. Then for $p \in (0,1)$, i.e., when both local (type L) and global (type H) projects appear on the agenda with positive probability the two-tier government is strictly welfare improving over the one-tier government.[18]

Proof First, let us show that it is possible to attain ex-post efficiency with the two tier-government. Next, we have to verify that this is impossible with a one-tier government, which will conclude the proof.

Indeed, denote by $V^*(p)$ the highest attainable welfare. If it were possible to condition the constitution on realization of benefits, the designer would have required $(\alpha, m) = (R, 1)$ for local projects (that generate benefits to one region only) and $(\alpha(m), m)$ with $\alpha(m) = \frac{\beta_m}{\bar{\beta}}$ for the rest (following lemma 3.1). This rule would have generated the following expected welfare

$$V^*(p) = p \int_R^B \left(\frac{1}{R} b_1 - 1 \right) dS(b_1) + (1-p) \int_{1/\bar{\beta}^H}^B \left(\bar{\beta}^H b_1 - 1 \right) dS(b_1). \tag{14}$$

But this value can also be attained ex-ante by the two tier government under constitution, say, $(\alpha(2), 2), \alpha(2) = \frac{\beta_2}{\bar{\beta}} \geq 1$ by assumption.

Next, let us show that the expected value with a one-tier government falls short of $V^*(p)$.

Assume the designer chooses $m = 1$. Then then the projects accepted are those with the highest benefit above α, $b_{[1]} > \alpha$. This condition is the

[17]For this proposition, again, the full support assumption is dropped, as the existence of the solution is verified by construction.

[18]Clearly by the maximum theorem, the result is still true if, the purely local project L, is substituted by a project $L' : \beta_{L'}$ is in the small enough neighborhood of $(1, 0, .., 0)$. In this sense, the result is "robust."

same for both types of projects: H and L. Therefore, the highest attainable welfare in this case is

$$\max_{\alpha\in[1,R]} (1-p) \int_{\alpha}^{B} \left(\bar{\beta}^{H} b_1 - 1\right) dS\left(b_1\right) + p \int_{\alpha}^{B} \left(\frac{1}{R} b_1 - 1\right) dS\left(b_1\right) \tag{15}$$

where $\bar{\beta}^{H} \equiv \bar{b}^{H}/b_1^{H}$ is constant for any realization of b^{H}. Note that the average benefit of a local project is $1/R$. The solution to this problem is

$$\alpha^{*} = \frac{1}{(1-p)\,\bar{\beta}^{H} + p\bar{\beta}^{L_0}} \tag{16}$$

Note that this solution lies between $1/\bar{\beta}^{H}$ (the optimal cost sharing arrangement if only H type projects appear on the agenda) and $1/\bar{\beta}^{L} = R$, (the total cost of the project, an optimal "share" to impose on the single region with the positive benefit, provided only L projects appear on the agenda). Whenever both issues appear on the agenda $(0 < p < 1)$, goods of type H are under-provided: $\alpha^{*} > 1/\bar{\beta}^{H}$, whereas type L goods are over-provided: $\alpha^{*} < R$. Therefore, the expected welfare attained under $(\alpha^{*}, 1)$ – the highest one for $m = 1$ – is strictly below $V^{*}(p)$.

In the complementary case, in which designer chooses $m > 1$, all local projects will be rejected for any α, thus, the attained welfare should be smaller than $V^{*}(p)$ by at least $p\int_{R}^{B}\left(\frac{1}{R}b_1 - 1\right) dS\left(b_1\right) > 0$. In the view of the initial remark, this completes the proof. ■

The reason decentralization is beneficial here is transparent: under the hierarchy each tier filters correctly the corresponding type of public project as the design of acceptance rules for the central government is independent of the presence of good local projects. This not true for the unitary government that is urged to choose a rule that is optimal "on average" for both local and global projects.

The same general idea stands behind the next proposition, demonstrating benefits to decentralization for a wider class of distributions. There, the ability to condition the acceptance rule on the second highest benefit, $b_{[2]}$, as opposed to the first one is very valuable to the designer. For example, $b_{[2]}$ being sufficiently high may indicate that the rest of the benefits are relatively high as well, i.e., it is more likely that a global project is on the agenda. Two-tier governments can require the size of the minimum winning coalition to be at least two without the fear of loosing good "local" projects, that can be picked up by the lower tier. This option is, clearly, infeasible for the unitary governments.

The proof of the next proposition is conceptually dissimilar to the previous one though. Without being able to ensure (as in the previous example) that the "constitution" maximizes ex-post welfare for $m > 1$, and, thus, can not be improved upon, one has to verify that the choice $m = 1$ is dominated. This is done by making the ability to condition on $b_{[2]}$ sufficiently attractive. I have introduced a parameter, θ, which could be interpreted as "news," or "general state of affairs" that the rules may depend upon. As θ decreases (meaning "bad news"), the ex-ante value of a project reduces (becomes more negative), whereas the ex-ante value of a project conditional on the second benefit being above the average cost, $b_{[2]} > 1$, stays positive. Another interpretation for θ is the degree of concentration of the local projects. Thus, when the news are bad enough, in other words, if it really pays to discriminate across projects (based on $b_{[2]}$), decentralization is beneficial, according to the proposition that follows.[19]

Before stating the next result, let us define affiliation as in Milgrom and Weber (1982), who also provide examples of distributions satisfying this property.

Definition 3.4 *The random variables $\{b_1, .., b_R\}$ are affiliated if the joint density $f(b)$ is such that for any b, b'*

$$f(b \wedge b\prime) f(b \vee b\prime) \geq f(b) f(b\prime),$$

where $b \vee b'$ is component-wise maximum,

$$b \vee b' = \left(\max\left\{b_1, b_1^{'}\right\}, .., \max\left\{b_R, b_R^{'}\right\}\right)$$

and $b \wedge b\prime$ is component-wise minimum,

$$b \wedge b' = \left(\min\left\{b_1, b_1^{'}\right\}, .., \min\left\{b_R, b_R^{'}\right\}\right).$$

Affiliation implies that if one of the components of the vector is high (low), it is more likely that the others are high (low) as well. Assuming

[19]No doubt it is possible to construct similar scenarios under which conditioning on $b_{[m]}$, $m > 1$ is very valuable, especially if good local projects are being (almost) adequately accepted on the lower level.

that regional benefits are affiliated is consistent with the idea that a public project has an 'objective' (common) value perceived differently by regions. Note that it is still possible to have projects with *very* different regional valuations. Affiliation just requires these occurrences to be less frequent. Moreover, it does not rule out independently distributed regional valuations.

Proposition 3.5 *Assume*

1. *F is symmetric in all its arguments with the corresponding marginal distribution f, f is continuous on $[0, B]^R$.*

2. *Assume random variables b_k, $k = 1, .., R$ are affiliated;*

3. *Let $\varphi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$; $\varphi(\theta, x) = E\left(w(b)\,|b_{[1]} = x; \theta\right)$, assume $\varphi(\theta, x) < \infty$ is strictly increasing in $\theta \in [0, 1]$ for all $x \in [0, R]$;*

4. *Let $z(\theta) = E(w(b)\,|\theta)$, $t(\theta) = E\left(w(b)\,|1 \leq b_{[2]}; \theta\right)$, assume $-\infty < z(\theta) < 0 < t(\theta) < \infty$ for all θ.*

Then there exist a threshold $-\infty <\underline{\theta}$, such that for any $\theta < \underline{\theta}$, welfare attained under decentralization (in problem (6)*) is strictly higher than that under unitary government (in problem* (5)*).*

Proof is relegated to the appendix. ∎

4 Conclusions

This model provides an economic rationale for the existence of a hierarchical government, which may outperform the unitary one even in the absence of the vertical asymmetric information within the hierarchy and even though the absolute advantage of the (upper) central government in making decisions with respect to public goods is evident. Somewhat contrary to the accepted view, if the issues to be tackled by the government are predominantly of local importance, albeit generating some spillovers, central government alone is sufficient, i.e., there is no need to decentralize. On the other hand, it is the prevalence of valuable global issues *along* with the projects of local importance that generates the need for the local governments. The hierarchy induces specialization of each tier on the corresponding issues, thus, enhancing the overall welfare, which can justify possible costs associated with the additional level of government. Thus, the main argument does not stem from an assumed deficiency of a central government, but rather, rests on the idea of specialization.

Let me briefly discuss the assumptions that underlie the conclusion. First, the restriction to the yes-no voting is crucial. Clearly, departing from voting (extending the set of the possible messages) can restore efficiency, see Laffont and Maskin (1982) for the revelation mechanisms that can achieve this objective. Nevertheless, extreme popularity of voting as a decision mechanism in practice boosts the value of identifying benefits of decentralization under this assumption. Next, recall **SPM** stating that the supporters have to pay more than the 'no' voters in the central legislature. If this restriction is relaxed, multiple outcomes can occur. There are equilibria under which *all* projects are accepted, independent of the benefits they generate. Other equilibria are those under which the projects generating the m^{th} highest benefit above the tax of the supporter, α, are accepted unanimously. Note that even if we allow the designer to set $\alpha < 1$ and ignore the first type of equilibria, it is still impossible to eliminate the inefficiency generated under a one-tier government that has to consider both global and local projects. Therefore, **SPM** is just a simplifying assumption.

Recall that the rules governing acceptance of public projects are fixed, whether because it is too costly to describe all the relevant contingencies, or to verify those. Thus, addition of local governments partially resolves the welfare loss associated with the incompleteness of the contract between the constitutional designer and the future legislators. This implies, in particular, that in the areas in which it is easy to condition the rules of acceptance on the type of public project (either implicitly or explicitly) the role of local governments will be diminished. Observe, however, that more detailed constitutions (laws) do not necessarily undermine the "executive" function of the local governments, that can still be responsible for implementing public projects, rather, it may decrease the "legislative" role of the lower tier.

Finally, in this model the appearance of public projects was driven by a probability distribution F, where all the draws were independent. If the probability, with which the next project appears on the agenda, is *not* independent from the previous decisions of the legislature, decentralization may introduce a new *intergovernmental free rider problem.*

Example 4.1 *Let $R = 3$. Assume that if the project $b^H = (4, 2, 1)$ is rejected, then the project $b^L = (4, 2, 0)$ will be put to voting. The central legislature op-*

erates under the majority rule with uniform taxation $(m = 2, \alpha = 1)$. *Clearly, the second project,* L, *will be accepted by the first region (on the local level), as* $b_1^L = 4 > 3$. *Given this fact, the pivotal voter from region* 2, *prefers to free ride on region* $1 : b_2^H - \alpha = 2 - 1 < 2 = b_2^L$. *Therefore the first project,* H, *will be rejected and, instead, an inferior project* L *is accepted by a hierarchical government.*

This free rider problem can be viewed as another cost associated with a hierarchical government. Modelling the exact relationship between the passage of the bills by the central legislature and the subsequent decisions made on the local level is an insightful investigation left for future research. It can shed further light on the trade-offs associated with decentralization, for which current model can be used as a reliable workhorse.

References

Aghion, P. and P. Bolton (2003). Incomplete social contracts. *Journal of the European Economic Association 1*(1), 38–67.

Anderlini, L. and L. Felli (1999). Incomplete Contracts and Complexity Costs. *Theory and Decision 46*, 23–50.

Besley, T. and S. Coate (2000). Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Analysis. CEPR Discussion Paper No 2495.

Boadway, R., M. Marchand, and M. Vigneault (1998). The Consequences of Overlapping Tax Bases for Redistribution and Public Spending in a Federation. CORE/UCL Discussion Paper No 9803.

Crémer, J., A. Estache, and P. Seabright (1996). Decentralizing public services: What can we learn from the theory of the firm? *Revue D'Economique Politique 106*, 37–60.

Finer, S. E., V. Bogdanor, and B. Rudden (1995). *Comparing Constitutions*. Oxford: Clarendon Press.

Gilbert, G. and P. Picard (1996). Incentives and optimal size of local jurisdictions. *European Economic Review 40*, 19–41.

Gilboa, I., A. Postlewaite, and D. Schmeidler (2002). Incomplete Contracts: A Syntactic Viewpoint. mimeo.

Guizot, F. (1861). *History of the Origin[s] of Representative Government in Europe*. London: Henry G. Bohn. Translated by Andrew R. Scoble; last edition: 2002 by Liberty Fund.

Harsanyi, J. (1992). Game and decision theoretic models in ethics. In R. J. Aumann and S. Hart (Eds.), *Handbook of Game Theory With Economic Applications*, Volume 1. Amsterdam ; New York: North Holland.

Hochman, O., D. Pines, and J.-F. Thisse (1995). On the Optimal Structure of Local Governments. *The American Economic Review 85*(5), 1224–1240.

Inman, R. P. (1988). Federal assistance and local services in the united states: The evolution of a new federalist order. In H. Rosen (Ed.), *Fiscal Federalism*, pp. 33–74. Chicago: University of Chicago Press.

Laffont, J. J. and E. Maskin (1982). The theory of incentives: An overview. In W. Hildenbrand (Ed.), *Advances in economic theory : invited papers for the Fourth World Congress of the Econometric Society at Aix-en-Provence*, pp. 31–94. Cambridge [UK], New York: Cambridge University Press.

Lockwood, B. (2002, April). Distributive Politics and the Costs of Centralisation. *Review of Economic Studies 69*(2), 313–338.

Milgrom, P. and R. Weber (1982). A Theory of Auctions and Competitive Bidding. *Econometrica 50*, 1089–1122.

Myles, G. D. (1995). *Public Economics*. Cambridge, New York, Melbourne: Cambridge University Press.

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

Palfrey, T. R. and J. Ledyard (2002). The Approximation of Efficient Public Good Mechanisms by Simple Voting Schemes. *Jounal of Public Economics 83*, 153–171.

Panizza, U. (1999). On the determinants of fiscal centralization: Theory and evidence. *Journal of Public Economics 74*, 97–139.

Rubinchik-Pessach, A. (2002). Why are some taxes 'more equal than others'? CORE/UCL Discussion Paper 2002/43.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4*, 25–55.

Tirole, J. (1999). Incomplete Contracts: Where Do We Stand? *Econometrica 67*(4), 741–781.

van Dorp, J. and S. Kotz (2002). The Standard Two Sided Power Distribution and its Properties: with Applications in Financial Engineering. *The American Statistician 56*, 90–99.

Wellisch, D. (2000). *Theory of Public Finance in a Federal State*. Cambridge University Press.

Zantman, W. (2002). Constitutional Design and Regional Favoritism. *Journal of Public Economic Theory 4*, 71–93.

A Appendix

Proof of proposition 2.1. Assume first a stronger version of (**SPM**), so that $\alpha_k > 1$ for all $k \geq m$.

Remark A.1 *There is no Nash equilibrium in which strictly more that m members vote for the project. Indeed, suppose $R > k \geq m$ individuals vote for the proposal. By assumption $\alpha_k > 1$, so that $\alpha_{k+1}(R-k) > R-k > R-\alpha_k k$, thus $t_{k+1}(Y) > t_k(N)$. It follows that it is strictly dominant strategy for any player to vote against the project (to save on his tax bill).*

Case 1. Consider projects b such that $b_{[m]} < \alpha_m$. Then not more than $m-1$ people are voting for the project while the rest oppose it in an equilibrium. Clearly in no Nash equilibrium m or more individuals vote for the project. Therefore in this case the project is rejected under any equilibrium .

Case 2. Consider projects b such that $b_{[m]} > \alpha_m$. Exactly m individuals voting in favour of the project is an equilibrium. Indeed, it is a strict Nash equilibrium. We are left to show that there is no equilibria under which the project is rejected. It is obvious that any strategy profile supporting $m-1$ individuals voting in favour of the project is not a Nash equilibrium. A standard trembling hand perfection argument rules out equilibria supporting $m-k$ votes in favor of the project with $k>1$.

Thus we can conclude that

$$M(\alpha_m)=\left\{y\in Y : b_{[m]}>\alpha_m\right\}.$$

Now we can relax the assumption made at the beginning of the proof by allowing weak inequality $\alpha_k\geq 1$. Note that the only time we used the assumption about the strictness of inequality was to assure that $t_{k+1}(Y)>t_k(N)$. The only way to get equality $t_{k+1}(Y)=t_k(N)$ is to have a vector of sharing rules, $\alpha^*=(\alpha_m,\alpha_{m+1},..,\alpha_{k^*},\alpha_{k+1^*},..,\alpha_{R-1})$, such that $\alpha_{k^*+1}=\alpha_{k^*}=1$ for some $k^*\geq m$. Then, there will be equilibria in which k^*+1 legislators vote for the project given there are at least k^*+1 of them have valuations above the average cost. In addition, the equilibria of the second type, in which exactly m people vote for the project if $b_{[m]}>\alpha_m$, still remain. Note that if we take a sequence of vectors α that converge to α^* from above, then, only the latter equilibria will be present, while former will not, as along the sequence α_k and α_{k+1} are above 1 with at least one of them strictly above 1. So, the equilibria of the first type are not robust to the "trembling hand of the designer". Thus we are left with the equilibria supporting $M(\alpha_m)$ as the set of accepted projects. ■

Lemma A.2 *Assume F is symmetric in all its arguments with the corresponding marginal distribution f. Assume random variables b_k, $k=1,..,R$ are affiliated. Then the first derivative of the objective function with respect to α for fixed R,m can be represented as*

$$-m\begin{pmatrix}R\\m\end{pmatrix}A(\alpha,m), \tag{17}$$

where

$$\begin{aligned}
A(\alpha,m) &\equiv \underbrace{\int_{\alpha}^{\bar{B}}..\int_{\alpha}^{\bar{B}}}_{m-1\ times}\underbrace{\int_{\underline{B}}^{\alpha}...\int_{\underline{B}}^{\alpha}}_{R-m\ times}w(b_{-m},\alpha)\,dF(\alpha,b_{-m}) && (18)\\
w(b_{-m},\alpha) &= \sum_{i\neq m}b_i+\alpha-R, && (19)\\
b_{-m} &= (b_1,b_2,..,b_{m-1},b_{m+1},..,b_R). && (20)
\end{aligned}$$

Proof By symmetry of F, objective function $W_I(\alpha, m)$ for a given R, $m \geq 1$, can be represented as $\Phi(\alpha, m)$,

$$\Phi(\alpha, m) \equiv \sum_{k=m}^{R} \binom{R}{k} \underbrace{\int_{\alpha}^{\bar{B}} .. \int_{\alpha}^{\bar{B}}}_{k \text{ times}} \underbrace{\int_{\underline{B}}^{\alpha} ... \int_{\underline{B}}^{\alpha}}_{R-k \text{ times}} \left(\sum_{i=1}^{R} b_i - R \right) dF(b) \qquad (21)$$

Clearly, function $\Phi(\alpha, m)$ is differentiable with respect to the first argument. Taking first derivative of $\Phi(\alpha, m)$ with respect to α and using symmetry of the distribution F,

$$\Phi_\alpha(\alpha, m) = \sum_{k=m}^{R-1} \binom{R}{k} \left\{ -k * A(\alpha, k) + (R-k) * A(\alpha, k+1) \right\}.$$

Note that

$$\binom{R}{k} (R-k) = \frac{R!}{k!\,(R-k-1)!} = \binom{R}{k+1} (k+1).$$

Therefore, after the cancellation, and observing that $A(\alpha, R) = 0$, we get

$$\Phi_\alpha(\alpha, m) = -m \binom{R}{m} A(\alpha, m).$$

■

Lemma A.3 *Let the benefits be affiliated (definition 3.4). Then the first derivative of the objective function $\Phi(\alpha, m)$ can be represented as*

$$\Phi_\alpha(\alpha, m) = -m \binom{R}{m} P(\alpha, m) V(\alpha, m), \qquad (22)$$

where $P(\alpha, m) > 0$ and $V(\alpha, m)$ is strictly increasing in its first argument and is non-decreasing in the second argument, hence $\Phi(\alpha, m)$ is strictly (differentiably) quasiconcave in its first argument on $(0, R)$.

Proof Let

$$V(\alpha, m) \equiv \frac{1}{P(\alpha, m)} \underbrace{\int_{\alpha}^{B} .. \int_{\alpha}^{B}}_{m-1} \underbrace{\int_{0}^{\alpha} ... \int_{0}^{\alpha}}_{R-m} [w(b_{-m}, \alpha)]\, dF(\alpha, b_{-m}), \quad (23)$$

$$w(b_{-m}, \alpha) = \sum_{i \neq m} b_i + \alpha - R, \quad (24)$$

$$P(\alpha, m) \equiv \underbrace{\int_{\alpha}^{B} .. \int_{\alpha}^{B}}_{m-1} \underbrace{\int_{0}^{\alpha} ... \int_{0}^{\alpha}}_{R-m} dF(\alpha, b_{-m}), \quad (25)$$

$$b_{-m} = (b_1, b_2, .., b_{m-1}, b_{m+1}, .., b_R). \quad (26)$$

Clearly, then

$$P(\alpha, m)\, V(\alpha, m) = A(\alpha, m),$$

where $A(\alpha, m)$ is as defined in 18, which justifies 22. It is obvious that $P(\alpha, m) > 0$ for $\alpha \in [1, R)$.

Therefore, it is left to show that $V(\alpha, m)$ is strictly increasing in the first argument. Indeed, in the view of definition 23,

$$V(\alpha, m) = E\left(w(b_{-m}, \alpha)\, | b_{[R]} \leq ... \leq b_{[m+1]} \leq \alpha = b_{[m]} \leq .. \leq b_{[1]}\right).$$

Then, in the view of assumed affiliation of b_k, by Milgrom and Weber (1982), theorem 5, $V(\alpha, m)$ is strictly increasing in α. Therefore,

$$V(\alpha, m) \leq E\left(w(b_{-m}, \alpha)\, | b_{[R]} \leq ... \leq b_{[m+1]} = \alpha = b_{[m]} \leq .. \leq b_{[1]}\right) \leq V(\alpha, m+1), \quad (27)$$

so the statement follows. ∎

Proof of proposition 3.5. To start, let us show that setting $m' > 1$ along with some α' for a unitary government is always strictly dominated by the same rules chosen under a two-tier government, in other words, $Z(\alpha', m') > 0$ for $m' > 1$. Indeed, let $K \equiv E\left(w(b)\, | M(R, 1) \cap b_{[2]} < 1\right)$, clearly, $K > 0$. Moreover, by the full support assumption and as $B > R$, $\Pr\left(M(R, 1) \cap b_{[2]} < 1\right) = q > 0$. Then, if the designer chooses $m' > 1$ for a one tier government, $Z(\alpha', m') \geq qK > 0$, so that he looses compared to the two tier one at least $qK > 0$ in expected terms for any choice of α'.

It is then left to show that the choice of $m = 1$ for a unitary government is dominated as well. Let us demonstrate that there is $\alpha_{II} \in [1, R]$ such that

$$W_I(\alpha_{II}, 2) + Z(\alpha_{II}, 2) - W_I(\alpha', 1) > 0 \quad (28)$$

for any feasible α'. Provided $Z(\alpha, 2) > 0$ for any α, it is sufficient to show that

$$W_I(\alpha_{II}, 2) - W_I(\alpha', 1) \geq 0. \tag{29}$$

conditional on $b_{[1]} \leq R$,[20] in other words, $(\alpha', 1)$ can not be a solution to problem (5) for any α'.

First by lemmata ($A.2$ and $A.3$) the problem admits a unique solution for any fixed m.[21] Let α_{II} be a solution to the problem for $m = 2$ and let α_I be a solution to the problem for $m = 1$. Then, it is sufficient to show that

$$\Delta \equiv W_I(\alpha_{II}, 2) - W_I(\alpha_I, 1) \geq 0. \tag{30}$$

By lemma ($A.3$), $\alpha_{II} \leq \alpha_I$. Let

$$T(\alpha_{II}, \alpha_I) = \left\{ b \in [0, R]^R \,| \alpha_{II} \leq b_{[2]} \leq b_{[1]} \leq \alpha_I \right\}, \tag{31}$$

$$S(\alpha_{II}, \alpha_I) = \left\{ b \in [0, R]^R \,| b_{[2]} \leq \alpha_{II} \leq \alpha_I \leq b_{[1]} \right\}. \tag{32}$$

Note that the difference, Δ, is equal to the difference between the expected value of accepted projects under $(\alpha_{II}, 2)$ that were rejected under $(\alpha_I, 1)$ and that of rejected projects under $(\alpha_{II}, 2)$ that were accepted under $(\alpha_I, 1)$,

$$\Delta = E(w(b) \,| T(\alpha_{II}, \alpha_I); \theta) \Pr(T(\alpha_{II}, \alpha_I)) - E(w(b) \,| S(\alpha_{II}, \alpha_I); \theta) \Pr(S(\alpha_{II}, \alpha_I)). \tag{33}$$

For any θ, if $\alpha_{II} = \alpha_I$, then, clearly, $\Pr(T(\alpha_{II}, \alpha_I)) = 0$. Moreover, $E(w(b) \,| S(\alpha_{II}, \alpha_{II})) \leq 0$, as

$$E(w(b) \,| S(\alpha_{II}, \alpha_{II})) = \int_1^{\alpha_{II}} \nu(\alpha, 2)\, dF(b_{-1}, \alpha), \tag{34}$$

$$\nu(\alpha, 2) = E\left(w(b_{-m}, \alpha) \,| b_{[2]} = \alpha \leq \alpha_{II} \leq b_{[1]}\right), \tag{35}$$

$$\nu(\alpha, 2) \leq V(\alpha_{II}, 2) = E\left(w(b_{-m}, \alpha) \,| b_{[2]} = \alpha_{II} \leq b_{[1]}\right) = 0$$

where the last inequality holds for $\alpha \leq \alpha_{II}$ in the view of lemma ($A.3$) and $V(\alpha_{II}, 2) = 0$ by the first order conditions (lemma $A.2$).

Now assume $\alpha_{II} < \alpha_I$. Note that function

$$\tau(x, y; \theta) = E(w(b) \,| T(x, y); \theta) \tag{36}$$

[20] For the rest of the proof, the condition $b_{[1]} \leq R$ will be omitted for simplicity.

[21] Strictly speaking, the lemmata cover the interior solution cases, the treatment of corner solutions is trivial and, thus, is omitted.

is increasing in the first two arguments by Milgrom and Weber (1982), theorem 5. Let us focus on the non-degenerate case of τ being strictly increasing in x and y.[22] Note that τ is also continuous in x and y due to continuity of f.

Let $\tau^* (\theta) = \tau (\alpha_{II}, \alpha_I; \theta)$. Need to show that $\tau^* (\theta) \geq t (\theta) > 0$ for low enough θ. Clearly, $\tau (\alpha_{II}, R; \theta) > t (\theta)$ by definition of $t (\theta)$. Next, it easy to verify that $\tau (\alpha_{II}, \alpha_{II}; \theta) < 0$, therefore, $\tau (\alpha_{II}, \alpha_{II}; \theta) < t (\theta)$, as $t (\theta) > 0$ for all θ by assumption. Fix $\theta = \hat{\theta}$. Then there exists $\alpha_{II} < y_t < R$, such that

$$\tau \left(\alpha_{II}, y_t; \hat{\theta} \right) = t \left(\hat{\theta} \right). \tag{37}$$

If $y_t \leq \alpha_I$, then, clearly, $\tau (\alpha_{II}, \alpha_I) = \tau^* \left(\hat{\theta} \right) \geq t \left(\hat{\theta} \right) > 0$. If $y_t > \alpha_I$, fix $\hat{\alpha} = \alpha_{II}$, and $\hat{t} = t \left(\hat{\theta} \right)$ then decrease θ to θ_0. As a result α_I should increase. Indeed, α_I solves

$$\begin{aligned} V (\alpha_I, 1; \theta) &= 0; & (38) \\ V (\alpha, 1; \theta) &= \varphi (\theta, x) = E \left(w (b_{-1}, \alpha) | \alpha = b_{[1]}; \theta \right), & (39) \end{aligned}$$

and $\varphi (\theta, x)$ is assumed to be strictly increasing in θ. It follows that α_I is decreasing in θ. Therefore, for fixed $t = \hat{t}$ and $\alpha_{II} = \hat{\alpha}$ inequality $y_t \leq \alpha_I$ is easier to satisfy as θ decreases. In addition, given τ is strictly increasing in the first two arguments, and in the view of equation 37, y_t should increase with t and decrease with α_{II}. Observe that by affiliation of b, and given $\varphi (\theta, x)$ is increasing in θ, $V (\alpha, 2; \theta) \equiv E \left(w (b_{-1}, \alpha) | \alpha = b_{[2]}; \theta \right)$ is increasing in θ, and so is $t (\theta)$. Therefore a decrease in θ should lead to a increase in α_{II} that solves $V (\alpha_{II}, 2; \theta) = 0$, and a decrease in $t (\theta)$. Both lower the resulting y_t that solves (37). Therefore, for θ low enough the inequality $y_{t(\theta)} \leq \alpha_I (\theta)$ should be satisfied. This implies then, $\tau^* (\theta) \geq t (\theta) > 0$ for θ low enough, clearly, finite.[23]

It is left to assure the negativity of the second term, $E (w (b) | S (\alpha_{II}, \alpha_I); \theta)$, in the difference Δ. Let $\sigma : \mathbb{R} \times [\alpha_I, R] \to [0, \alpha_I]$ be defined as a solution to

[22] Cumbersome, but similar proof for the weakly increasing function is omitted.

[23] Note that the requirement $t(\theta) > 0$ still has to be satisfied, which implies, that conditional on the second highest benefit being above unity, decrease in θ is not "as bad news" as the same decrease conditional on the complementary event, i.e., $E \left(w (b) | 1 \leq b_{[2]} \leq b_{[1]} = x; \theta \right)$ decreases in θ slower than $E \left(w (b) | 1 \geq b_{[2]}; b_{[1]} = x; \theta \right)$ for any x.

the following equation,[24]

$$E\left(w(b)\,|b_{[2]}\leq\sigma,b_{[1]}=y;\theta\right)=0. \tag{40}$$

Thus, $\sigma(\theta,y)$ is a function, moreover, it is easy to see that it strictly decreases with y for any given θ and strictly decreases with θ for a fixed y. By the previous argument, without of loss of generality $\alpha_I>R/2$ (otherwise one could pick θ low enough to assure that). Next, as $V(R/2,2;\theta)>0$, $\alpha_{II}<R/2$. Observe that for any $\theta\in\mathbb{R}$, $\sigma(\theta,\alpha_I)=\alpha_I\,(>R/2)$ by definition of α_I (or the first order conditions), and also $\sigma(\theta,R)=0\,(<R/2)$. Therefore, for any θ there exists a threshold y_θ such that for $y\in[\alpha_I,y_\theta)$ $\sigma(\theta,y)>R/2$ and for any $y\in[y_\theta,R]$, $\sigma(\theta,y)\leq R/2$. As a decrease in θ will lead to an increase in $\sigma(\theta,y)$ for any y, the threshold, y_θ, should increase. This will assure that for θ low enough, the desired term, $E(w(b)\,|S(\alpha_{II},\alpha_I);\theta)$ is negative. Indeed,

$$\begin{aligned}
& E(w(b)\,|S(\alpha_{II},\alpha_I);\theta)\leq(E(w(b)\,|S(R/2,\alpha_I);\theta))= && (41)\\
= \;& P_\theta\left(E\left(w(b)\,|S(R/2,\alpha_I),\left\{b_{[1]}<y_\theta\right\};\theta\right)\right)+ && (42)\\
& +(1-P_\theta)\left(E\left(w(b)\,|S(R/2,\alpha_I),\left\{b_{[1]}\geq y_\theta\right\};\theta\right)\right), && (43)\\
P_\theta \equiv\;& \Pr\left(\left\{b_{[1]}<y_\theta\right\}|S(R/2,\alpha_I)\right). && (44)
\end{aligned}$$

Clearly, P_θ is increasing with y_θ. Moreover,

$$\begin{aligned}
N_\theta &\equiv E\left(w(b)\,|S(R/2,\alpha_I),\left\{b_{[1]}<y_\theta\right\};\theta\right) && (45)\\
&= E\left(w(b)\,|b_{[2]}\leq R/2\leq\alpha_I\leq b_{[1]}<y_\theta;\theta\right)<0 && (46)
\end{aligned}$$

by construction of y_θ; $E\left(w(b)\,|S(R/2,\alpha_I),\left\{b_{[1]}\geq y_\theta\right\};\theta\right)$ is bounded by $(R-1)\frac{R}{2}$. It follows that if θ is low enough to assure

$$P_\theta>\frac{(R-1)\,R}{(R-1)\,R-2N_\theta}, \tag{47}$$

then $E(w(b)\,|S(\alpha_{II},\alpha_I);\theta)$ will be negative.

Clearly, θ is finite. ■

[24] Here, again, consider a non-degenerate case so that the function $E\left(w(b)\,|b_{[2]}\leq\sigma,b_{[1]}=y;\theta\right)$ strictly increases in both σ and y, moreover, it also strictly increases in θ by assumption; therefore, σ that solves equation (40) is unique for every pair of values θ,y.