Decentralization and Electoral Acountability : Incentives, Separation, and Voter Welfare

J. Hindriks and B. Lockwood

Discussion Paper 2005-38

Département des Sciences Économiques
de l'Université catholique de Louvain

UCL

Decentralization and Electoral Accountability: Incentives, Separation, and Voter Welfare

Jean Hindriks* Ben Lockwood†

First Version: March 2004
This Version: March 2005

Abstract

This paper studies the relationship between fiscal decentralization and electoral accountability, by analyzing how decentralization impacts upon incentive and selection effects, and thus on voter welfare. The model abstracts from features such as public good spillovers or economies of scale, so that absent elections, voters are indifferent about the fiscal regime. The effect of fiscal centralization on voter welfare works through two channels: (i) via its effect on the probability of pooling by the bad incumbent; (ii) conditional on the probability of pooling, the extent to which, with centralization, the incumbent can divert rents in some regions without this being detected by voters in other regions (*selective rent diversion*). Both these effects depend on the information structure; whether voters only observe fiscal policy in their own region, in all regions, or an intermediate case with a uniform tax across all regions. More voter information does not necessarily raise voter welfare, and under some conditions, voter would choose uniform over differentiated taxes ex ante to constrain selective rent diversion.

JEL Classification: D72, D73, H41

Keywords: fiscal federalism; decentralization; elections; accountability

*Department of Economics, CORE, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium. Email: Hindriks@core.ucl.ac.be

†CEPR and Department of Economics, University of Warwick, Coventry CV4 7AL, England; Email: B.Lockwood@warwick.ac.uk

1. Introduction

Fiscal decentralization, the allocation of tax and spending powers to lower levels of government, is now an established policy objective, in many developed and developing countries. For example, nearly all the large Latin American countries have initiated some form of fiscal decentralization in the last decade e.g. Bolivia (Faguet(2004)), as have Indonesia, the Philippines, and Pakistan, to name just a few. China and Russia's transition from socialism involves various aspects of decentralization. Moreover, it is actively promoted as a development strategy by organizations such as the World Bank[1]. There have also been similar reforms in high-income countries, e.g. devolution of tax and spending powers to Scotland in the UK in 1999, and in Italy, starting in 1993 with the introduction of a municipal property tax.

The usual advantages that are claimed for decentralization, that one can find in the literature, include the following.[2] First, decentralization is claimed to improve *allocative efficiency*, in the sense that the goods provided by governments in localities will be better matched to the preferences of the residents of those localities. This is sometimes known as the *preference-matching* argument. There is now a large theoretical literature evaluating the preference-matching argument,[3] and some more recent empirical papers.[4]

Second, decentralization is argued to increase the *accountability* of government. In the literature, this term is used in rather a broad sense, and refers to constraints on the rent-seeking activities of office holders, such as diverting rents from the public purse, taking bribes, favouring of particular interest groups, and insufficient innovation and effort. Interestingly, in this case, the lead has been taken by empirical researchers: there are now a number of cross-sectional and panel studies that show that across countries, measures of fiscal decentralization are generally negatively correlated with low accountability outcomes, such as corruption and poor governance,[5] although there are some dissenting

[1]For more details on country decentralization programs, and the World Bank's view of the costs and benefits, see http://www1.worldbank.org/publicsector/decentralization/, or World Bank (2000).

[2]For recent reviews of the advantages of decentralization, see Azfar et (2001), Oates (1999) and McKinnon and Nechyba (1997).

[3]See for example, Alesina and Spolare(1997), Besley and Coate(2003), Bolton and Roland(1997) and Cremer and Palfrey(1996), Ellingsen(1998), Gilbert and Picard(1996), Lockwood(2002), Oates(1972), Oberholzer-Gee and Strumpf(2002).

[4]For example, Azfar et al.(2001), Faguet(2004), Oberholzer-Gee and Strumpf(2002), Wallis and Oates(1988).

[5]See among others Huther and Shah(1998), Fissman and Gatti(2002), Mello and Barenstein(2001). More recently, Fissman and Gatti(2002a) and Henderson and Kuncoro(2004) have shown, using subnational data for the US and Indonesia respectively, that expenditure decentralization is only effective

views (Treisman (2000),(2002)).

However, accountability is notoriously difficult to pin down precisely, and perhaps reflecting this, there have been rather few attempts to analyze theoretically how the degree of accountability varies with fiscal (de)-centralization. Indeed, we know of only three contributions that have really made progress on this issue: a paper by Seabright(1996), whose model was refined and extended by Persson and Tabellini(2000) in Chapter 9 of their book, and Bardhan and Mookherjee(2003), which measures accountability (negatively) as the degree to which government policy is distorted by the presence of a lobby group.

To assess these contributions, particularly the first two,[6] and explain how our contribution in this paper is distinctive, we can begin by noting that as stressed by Besley and Smart(2003), elections provide accountability in two senses. First, they allow voters to de-select bad incumbents (*selection effects*). Second, the selection effect provides an incentive for incumbents to change their behavior in order to increase the probability of re-election (*incentive or discipline effects*). A key question, therefore, is what effect (de)-centralization will have on these two accountability mechanisms. The contributions of Seabright(1996) and Persson and Tabellini(2000), we argue, focus exclusively on incentive effects.

In his important contribution, Seabright(1996) stressed two incentive effects of centralization, working in different directions. His setting is a two-period model of the political process where there is an agency problem between politicians and voters due to moral hazard. All policy-makers are identical. The incumbent can vary his effort (or equivalently, the amount of rent he diverts from the public purse to his own pocket). The voters only observe the action of the incumbent with some noise. Indeed, the interpretation of his model that is closest to the standard kind of model in the fiscal federalism literature would be where the voters observe the level of a public good provided by him in the first period, and the level of a public good provided is equal to some exogenous tax revenue, minus diverted rents, plus a productivity shock. As is standard in this kind of model (see e.g. the classic paper of Ferejohn(1986)), the voters set a performance standard $\hat{g}$, by voting the incumbent out of office if his production of the public good is lower than $\hat{g}$. This gives him an incentive to restrain rent-diversion in the first period.

in reducing corruption if it is accompanied by increased powers to raise revenue : "unfunded mandates" lead local officials to find other sources of revenue.

[6]We do not focus on Bhardhan and Mookherjee's contribution, important though it is, because it adopts a different model of the political process: politicians (political parties) in their model are Downsian (they only care about re-election, cannot divert rents, and can pre-commit to a policy platform).

Now suppose that the economy is composed of n regions, and with decentralization, there is one policy-maker in each region, and with centralization, a single policy-maker. Suppose also initially that the productivity shocks are region-specific, rather than specific to the policy-maker i.e. all policy-makers are identical. Then, moving from decentralization to centralization, there are two ways in which the incentive for the policy-maker to restrain rent-diversion changes. First, and most obviously, with centralization, if the policy-maker wins the election, he can expect more rent in the second period (in fact, he extract maximum rent in all regions, rather than one, so in the absence of any exogenous ego-rent from office (Persson-Tabellini(2000)), his future rent rises by a factor of n). We call this the *rent scale effect* of centralization.

But there is a second, more subtle effect of centralization, loss of accountability through the reduction in the probability that the voters in any one region are pivotal in determining the outcome of the election (we will call this the *reduced pivot probability effect* of centralization). To illustrate, consider the case of three regions, and suppose that the voter can choose high rent diversion, in which case he wins with probability 0, or low rent diversion, in which case he wins with probability p. With decentralization, the incumbent can raise his probability of winning by p by cutting rent diversion. With centralization, suppose the incumbent raises his rent-diversion in region i, assuming it is already low in the other two regions. Region i is only pivotal if the incumbent wins in one of the other regions and loses in the other, an event which occurs with probability $2p(1-p)$. So, with centralization, the incumbent can raise his probability of winning by $q = p \times 2p(1-p)$ by cutting rent diversion. Obviously, $q < p$, so the reduced pivot probability effect reduces the incentive to limit rents.

A weakness of Seabright's model is that the voters are not following a voting rule that can be easily justified: all policy-makers are identical, and so *whatever* their performance in office, voters are ex post indifferent about voting them out of office or retaining them at the end of the first period. One way of resolving this indeterminacy is to suppose that the productivity "shock" which maps tax revenue minus rent into public good supply is an inherent competence characteristic of the incumbent. Then, voters are not indifferent about a performance cutoff ex post, because the higher $\hat{g}$, the more likely it is that the incumbent who passes it is competent. Persson and Tabellini(2000, Chapter 9.1) present a model of this form, retaining Seabright's assumption that the first-period incumbent does not observe his competence level.[7] An equilibrium of this model is thus described as

[7]For technical reasons, the public good production function is multiplicative in the competence characteristic ϕ_i i.e. $g_i = (\tau - r_i)\phi_i$.

(i) a level of first-period rent diversion by the incumbent, $\hat{r}$, and (ii) a cutoff $\hat{g}$ such that given $\hat{r}$, his competence is judged to be at least as great as the challenger. Persson and Tabellini show how the rent scale effect and the pivot effect work in the determination of $\hat{r}$.

A more important limitation of Seabright(1996) and Persson and Tabellini(2000) is that they say effectively nothing about how centralization impacts on the *selection* effects of elections.[8] In Seabright, there are no selection effects, as all policy-makers are identical. In Persson and Tabellini(2000), by construction, the probability that an incumbent of given competence wins the election is the same with centralization and decentralization: in the equilibrium with both centralization and decentralization, an incumbent with a competence level higher (lower) than the expected competence of the challenger wins (loses) the election. As both competence levels are random draws from the same distribution, the probability that the initial incumbent has a competence level above the expected level of the challenger is simply 0.5. So, with both centralization and decentralization, by construction, the incumbent loses office with probability 0.5.

So, for separation probabilities to be truly endogenous (and thus vary between centralization and decentralization), there must be *asymmetric information*: the incumbent must be better-informed about his own competence (or some other characteristic) than the electorate. This paper provides a comprehensive analysis of such a model. Our main objective in doing this is to see how accountability (as measured by the pivot probability effect in Seabright's moral hazard model) can be formalized in this setting, and how it interacts with an endogenous separation probability to determine voter welfare under the two fiscal arrangements. It turns out that accountability and voter welfare under centralization depend crucially on the amount of information available to a voter in any jurisdiction i about fiscal policy in other jurisdictions, and comparison of possible information structures is also a major theme of this paper.

Our model has two periods and n regions. In the first period, the type of the incumbent policy-maker is determined by random draw: the incumbent may be "good" or "bad". With decentralization, the policy-maker in region i, knowing his type, then chooses a tax and a level of public good provision in their region. Voters observe this choice and then vote for the incumbent or the challenger. The type of the challenger is also determined by

[8] it is remarkable that the complete contracting principal-agent theory also ignores the selection effect to consider only incentives. One notable exception is Banks and Sundaram (1998) who study the optimal retention rule in agency problems, and show that in equilibrium the chosen retention rule disciplines the agents (incentive effect) and the retained agents are more productive on average (selection effect).

random draw. In the second period, the winner again chooses a tax and a level of public good provision in their region. The (probability of) separation here is clearly endogenous because the bad incumbent may choose to pool with (imitate) the good incumbent, or separate (reveal his type by acting opportunistically). With centralization, the only difference is that in each period, there is only one policy-maker who chooses fiscal policy in *all* regions, subject to a common budget constraint. We do not impose (initially) the requirement that the tax be uniform across regions.

Our results are robust to several ways of specifying "good" and "bad" types. For purposes of exposition, we work mainly with the specification of Besley and Smart(2003), where the good type is benevolent i.e. maximizes the welfare of all the voters in his region, and the bad type maximizes rents diverted from tax revenue.[9] But, in Section 6, we show that all the qualitative results carry over to a variant of the model based on Persson and Tabellini(2000) where policy-makers all maximize rents, but differ in their competence i.e. ability to supply the public good from a given amount of tax revenue.[10]

Our main results are as follows. First, we focus on two key features of the equilibrium when comparing centralization and decentralization, *separation probabilities*, and *expected voter welfare*. In either case, the separation probability is the ex ante probability that a bad incumbent decides to separate in equilibrium, which he will do by diverting rent without restraint in all the jurisdictions he is responsible for.

We begin by studying a benchmark (but not particularly realistic) case where voters have "full information" i.e. can observe taxes set, and public goods supplied in *all* regions, not just their own. In this case, we show that these separation probabilities may be higher or lower with centralization than with decentralization.[11] We show that in this case, comparing voter welfare between centralization and decentralization, all that matters is the separation probability. That is, if voters have a preference for more (less) separation, then the fiscal arrangement that gives a higher (lower) separation probability will be preferred ex ante by all voters. Voters have a preference for more (less) separation when the discount factor and the expected quality of the politicians are above (below) a critical

[9]This has two attractions first, the results work out relatively neatly. Second, in this model, the choice of tax is endogenous, whereas in the competence model, it is basically fixed.

[10]In this model, competence matters to voters because the policy-maker cannot divert all tax revenue as rent: the remainder then provides a public good, with the more competent type providing more of the public good.

[11]Note that, in contrast to Baron and Besanko (1992), the opportunity of *information consolidation* with centralization does not necessarily improves voter information about the quality of the incumbent; this is because the incumbent chooses how much information to reveal in equilibrium by pooling or not.

value.

We then turn to the case where voters have "partial information" (i.e. they can observe the tax set, and public good supplied, only in their own region). In this case, the analog of the pivot probability effect arises, which we call *selective rent diversion*. Specifically, with centralization, if the incumbent wishes to win the election and stay in office, he can do so most efficiently by only imitating the good incumbent in a minimum majority of $m = (n+1)/2$ regions, and can take unconstrained rent in the other regions. In this sense, he is less accountable to the electorate with centralization than with decentralization. Selective rent diversion has two implications. First, it tends to decrease the separation probability relative to decentralization, and second, it unambiguously decreases voter welfare with centralization, for a *given* separation probability.

It does *not* follow from this, however, that voter welfare is always lower with centralization and partial information than it is with either centralization and full information, or decentralization. This is because the separation probability is endogenous: using this fact, counterexamples can be found to both of those statements. So, in particular, with centralization, it is *not* generally true that giving voters more information will make them better off. This result is comparable to Proposition 5 of Besley and Smart(2003), who demonstrate that yardstick competition between regions (which can only occur when voters are fully informed in our sense) does not necessarily increase voter welfare. The mechanism at work is quite different, however: in our case, statistical correlation in the cost of producing the public good in each region is not needed.

Then, we study the empirically important case of *uniform taxation* when decision-making is centralized. This is intermediate between partial and full information, as voters only observe public good provision in their own region, but effectively observe all the information they need about the spending in all regions (although they cannot distinguish whether spending in other regions is on public goods or is diverted as rents). In this case, the results are qualitatively similar to the case of partial information. In particular, accountability of the incumbent is limited because he can selectively pool. But, if he chooses to selectively pool, his ability to extract rents in the minority of regions where he does not pool is lower than with partial information: because the same tax is set and observed by the voters in all regions, he cannot set the maximum tax, but only the highest tax that a good incumbent might possibly set. An implication is that if voters have a constitutional choice ex ante between differentiated and uniform taxes under a wide range of conditions, they will choose uniform taxation (unless they can observe fiscal policy ex post in other regions). Thus, our model provides a novel explanation for the widely observed "stylized fact" that centrally set taxes are almost always uniform. The

argument is that uniform taxation is a useful device to transmit information to voters about spending levels in other regions.

Finally, we extend the model to introduce a "*rent-scale parameter*" in a natural way by supposing (following Persson and Tabellini(2000)) that the incumbent also derives some exogenous "ego-rent" from holding office that is independent of the size of his jurisdiction. The larger this ego-rent, then the smaller the factor by which overall rents from office (exogenous plus endogenous) increase following centralization. Thus, the exogenous ego-rent is an inverse rent scale parameter. It is shown that an increase in this ego-rent (along with an offsetting fall in the discount factor of the incumbent which leaves his behavior under decentralization unchanged) will increase probability of separation under centralization. In turn, this may have a positive or negative effect on voter welfare. This is in contrast to both Seabright and Persson and Tabellini, where exactly the same comparative statics exercise unambiguously lowers rent-diversion under centralization, and thus raises voter welfare since there is no selection effect in their models.

The layout of the remainder of the paper is as follows. Section 2 sets up the model. Section 3 studies the case of decentralization for the benevolence model. Sections 4, 5, and 6 study the cases of centralization with full voter information, partial voter information, and uniform taxation respectively for the benevolence model. Section 7 studies rent scale effects. Section 8 makes the case that most of the key results are robust as they also hold for the competence model. Section 9 discusses other extensions, and Section 10 concludes.

2. The Model

2.1. Preliminaries

There are two time periods $t = 1, 2$ and an odd number of regions $i = 1, ..n$, with $n \geq 3$. In each region in each time period, an incumbent politician makes decisions about taxation and public good provision. Moreover, at the end of period 1, there is an election in which voters choose between the incumbent and a challenger, having observed only first-period fiscal policy. With decentralization, there are n incumbents and n challengers: one in each region. With centralization, there is one incumbent and challenger.

In each region, there are a continuum of measure 1 of identical voters who derive utility $u_t^i = H(g_t^i) + x_t^i$ from a regional public good g_t^i and a private good x_t^i in period t. All agents have an endowment of the private good, normalized to unity. The public good is financed by a lump-sum tax τ_t^i, so that utility of the typical voter is $H(g_t^i) + 1 - \tau_t^i$. The tax can also be interpreted as an income tax at rate τ_t^i on income of unity. It is assumed

that $0 \leq \tau_t^i \leq 1$ so the endowment can be fully taxed. The incumbent can also divert tax revenue of amount r_t^i up to a maximum level of $r \leq 1$ per region in period t. Both voters and politicians have the same discount factor, $0 < \delta < 1$.

In each region in each time period, the unit cost c_t^i of producing the public good from the private good can take on one of two values: $c_t^i \in \{c_L, c_H\}$ with $c_L < c_H$. The determination of c_t^i is described in more detail below. With decentralization, there is a separate budget constraint for each region, of the form:

$$c_t^i g_t^i + r_t^i = \tau_t^i$$

where r_t^i are the rents diverted from tax revenue (if any) in region i. With centralization, the policy-maker is assumed to be able to pool tax revenues, and so faces a single budget constraint. So, the budget constraint is

$$\sum_{i=1}^{n} c_t^i g_t^i + r_t = \sum_{i=1}^{n} \tau_t^i$$

where r_t are the rents (if any) diverted from aggregate tax revenue.[12]

It is a widely observed "stylized fact" that centrally set tax rates are uniform across regions, and consequently, almost all the literature on fiscal centralization assumes that the tax rate is uniform with centralization i.e. $\tau_t^i = \tau_t$. We do not wish to impose the assumption ex ante, for reasons discussed at the end of this section.

Politicians may be of two types, "good" and "bad". In particular, in either region, both the initial incumbent and the challenger at the election are "good" with probability π and "bad" with probability $1 - \pi$. Politicians may differ in competence or benevolence, giving rise to two variants of the model.

Benevolence. A "good" politician derives utility only from the welfare of the voters in his jurisdiction: in particular, he maximizes the sum or average of these utilities. A bad politician cares only about discounted sum of rents diverted. Either type is equally competent in producing the public good. The cost of the public good is high in any region and period with probability $q \geq 0.5$.[13]

Competence. Any politician maximizes the discounted sum of rents diverted; conditional on this, he has a lexicographic secondary preference for supplying the public good at its optimal level. The public good is provided via a technology where the probability q_t^i

[12]Note that as the budget constraint is national, only the aggregate rent matters.

[13]Imposing this constraint on q rules out the "hybrid" equilibrium of Besley and Smart (2003). The reason for this is discussed further in Section 3 below.

that the unit cost is high in region i at time t is (i) uncorrelated across time and regions, and (ii) is conditional on the competency of the incumbent. A "good" politician is more competent than the bad. In particular, if the incumbent is good, then $q_t^i = 0$, and if the incumbent is bad, $q_t^i = q$, with $1 > q > 0$.

Finally, we state our assumptions about the information voters have about fiscal policy in other regions. we study three possible scenarios:

1. *Full voter information;* at time t, the voters in i can observe $(g_t^i, \tau_t^i)_{i=1,..n}$:
2. *Partial voter information;* at time t, the voters in i can observe only (g_t^i, τ_t^i):
3. *Uniform taxation*: at time t, the voters in i can observe only (g_t^i, τ_t^i), but the constraint is imposed that $\tau_t^i = \tau_t$, all i.

The third scenario is of interest because so much of the literature on fiscal decentralization assumes uniform taxation: in this case, voters in one region effectively observe spending in other regions, but they do not know whether spending in other regions is on public goods or rents.

2.2. A Benchmark

Note that in this model, there is an agency problem between voters and the incumbent: the former can only imperfectly control the behavior of the latter through electoral incentives. Note also that in setting up this model, we have abstracted from the usual features that generate a difference between centralization and decentralization in the established literature: there are no economies of scale, there are no spillovers between regions, voters do not differ in tastes for the public good, either within or between regions[14]. So, the difference in outcome between centralization and decentralization is *entirely* due to the difference in the extent to which the voters can control, or hold accountable, the incumbent, in the two cases.

To see this, it is helpful to consider the benchmark in the benevolence model where there is no agency problem i.e. where politicians are "good" with probability 1. In this case, it is clear that there is an equilibrium where the incumbent will always be

[14]Also, by taking a fixed number of incumbents and challengers, we assume away free entry and rule out the district magnitude effect bias in favor of centralization (that larger electoral districts lower barriers to entry and favor competition improving political discipline and selection). The district magnitude effect is related to the idea suggested by Myerson (1993) that electoral rules promoting the entry of many candidates protect voters against corruption in a better way. Myerson (1999) gives an overview of the performance of different electoral systems and Persson et al (2000) give evidence of the district magnitude effect.

re-elected,[15] and in either region and period, the incumbent will provide the public good efficiently, conditional on cost c_L or c_H. This is true whether there is centralization or decentralization. Of course, efficient public good provision, denoted $g_k, k = L, H$ is implicitly defined by the Samuelson rule $H'(g_k) = c_k$. Finally, as the distribution of costs is the same under decentralization and centralization, it follows that public good provision and therefore expected voter welfare must also be the same. In Section 9, we demonstrate that similar equivalence would arise under complete contracting.

2.3. Relation to the Literature

The model of competence is based on the career concerns model of Persson and Tabellini(2000), but with the key difference that in our model, there is initially asymmetric information, as the incumbent is initially informed about his type, as in Rogoff(1990) This means that in equilibrium, the degree of selection is endogenous, as explained in the introduction. The model of benevolence is an $n-$region generalization of Besley and and Smart (2003). We should stress that in their paper, they do not consider centralized decision-making: their benchmark is decentralization without "competition" between regions, and then the impact on selection and incentive effects of introducing either tax or yardstick competition is studied.

3. Decentralization

We solve backward to obtain a unique Bayes-Nash equilibrium in either region.[16] In the second period, the honest policy-maker will provide optimal public good level g_k given the cost realization c_k, and set tax $\tau_k = c_k g_k$. The dishonest policy-maker will just thus take maximum rent by setting a tax of $\tau = r$, and providing no public good i.e. $g = 0$. So, all voters prefer the honest policy-maker.

In the first-period, assume for the moment that good incumbent in either region will be elected with probability 1 if he behaves non-strategically i.e. makes exactly the same policy choices as in the second period. We will shortly verify when this is equilibrium behavior for the voters. In this case, the best strategy for the good incumbent is to behave

[15] There can be other equilibria where the incumbent is no re-elected, as all voters are always indifferent between incumbent and challenger, but these will generate the same outcome as the first one when there is no agency problem.

[16] Obviously, the results in this section recapitulate Section 3 of Besley and Smart (2003): the reader is referred to that paper for deeper discussion of the issues.

non-strategically.

As for the bad type, when cost is high, he always prefers to take maximum rent in the first period, rather than imitate the good type in exchange for re-election: this is because discounting the future, it is better to take maximum rent now, and nothing later, rather than the opposite. When cost is low, the bad type has only two options that may potentially be optimal. First, he can set $(g_H,\, \tau_H)$ and take $\hat{r} = g_H(c_H - c_L)$ in the form of rents: call this the *pooling* strategy. Second, he can set $g = 0$, and take maximal rents, by setting $\tau = r$: call this the *separating* strategy.

We are assuming for the moment that any incumbent who chooses $(g_H,\, \tau_H)$ will be re-elected. So, when cost is low, the payoffs to separating and pooling for the bad incumbent are $r + \delta.0$ and $\hat{r} + \delta.r$ respectively. There is therefore a pooling equilibrium, where the bad politician imitates the good one when the cost of public good provision is low, and is re-elected with probability 1 in that event if $\hat{r} + \delta.r \geq r$, i.e. $\hat{r} \geq (1-\delta)r$, and a separating equilibrium where bad politician does not imitate the good one even when the cost of public good provision is low if $\hat{r} \leq (1-\delta)r$.

To confirm that the pooling equilibrium exists, we only need confirm that voters are willing to re-elect the incumbent if they observe[17] $(g_H,\, \tau_H)$. A voter's posterior belief that the incumbent is good i.e. benevolent, conditional on observing $(g_H,\, \tau_H)$ is

$$q\left(\tau_H, g_H\right) = \frac{\pi q}{\pi q + (1-\pi)(1-q)} \tag{3.1}$$

Note from (3.1) that as $q \geq 0.5$, $q(\tau_H, g_H) \geq \pi$, so the voters are indeed willing to re-elect the incumbent after observing $(g_H,\, \tau_H)$.

So, in any region, the ex ante probability that a bad incumbent separates (the *selection effect*) is

$$s_D = \begin{cases} q & \text{if } \hat{r} > (1-\delta)r \\ 1 & \text{if } \hat{r} < (1-\delta)r \end{cases} \tag{3.2}$$

It is convenient for what follows to show s_D as a function of the discount factor, δ. This is done in Figure 1. It is clear that δ s a key parameter here, as the higher δ, the greater the incentives for pooling, and thus the lower is the separation probability.

Finally, note the role of the assumption that $q \geq 0.5$. This rules out the scenario where the incumbent wants to pool by setting τ_H, g_H, assuming that he can be re-elected, but the voters place a low probability on $c_i = c_H$, and thus will not be willing to re-elect the incumbent if he sets τ_H, g_H. In this case i.e. when $q < 0.5$, and $\hat{r} > (1-\delta)r$, Besley

[17] As a bad incumbent will never set τ_L, g_L, then voters are always willing to re-elect the incumbent having observed τ_L, g_L.

Figure 1: Separation Probabilities,
Decentralisation

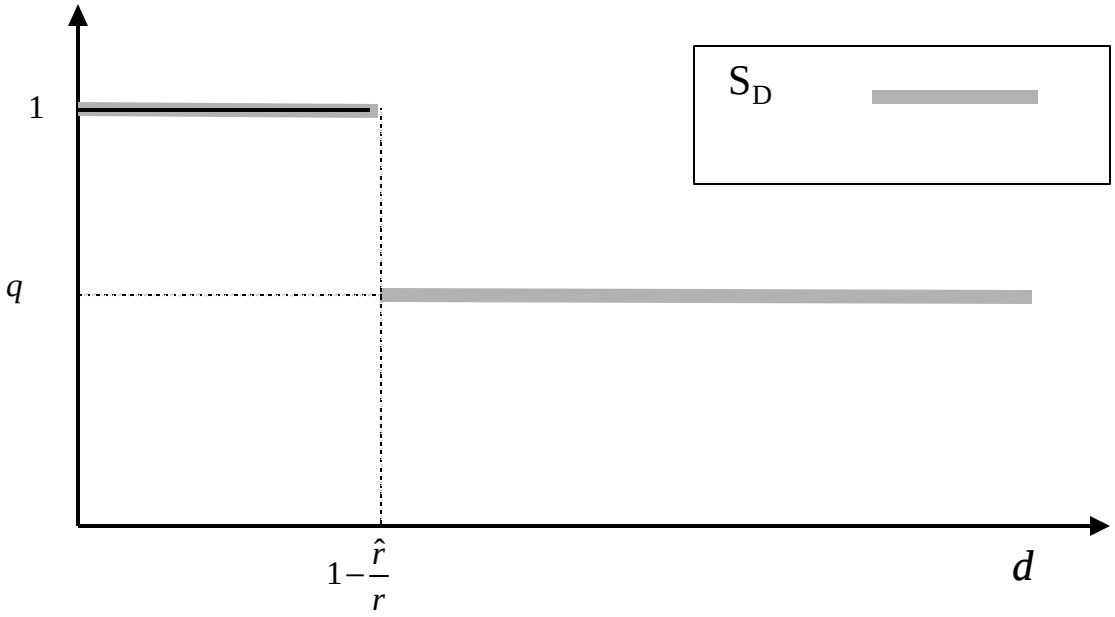

Figure 3.1:

and Smart construct a hybrid equilibrium, where both the bad incumbent and voters randomize. However, for some parameter values this equilibrium does not satisfy the Cho-Kreps stability criterion (Lockwood(2005)). The reason is that the "good " type has an incentive to strategically distort public good provision when cost is high to signal his type to the electorate, in order to avoid being replaced by a (possibly) bad challenger. In this case, a stable fully separating equilibrium can be constructed. We wish to avoid these rather technical issues, and do so by assuming $q \geq 0.5$.

4. Centralization with Full Voter Information

4.1. Equilibrium

We solve backward. In the second-period, the benevolent policy-maker will provide optimal public good level in each region given local costs and charge a tax equal to the cost. The non-benevolent policy-maker will provide no public good and take maximum rent, regardless of the cost configuration. So, all voters prefer the benevolent policy-maker in period 2. In the first-period, the benevolent incumbent behaves non-strategically and so will make exactly the same policy choices as in the second period.

So, it remains to characterize the first-period behavior of the non-benevolent incumbents. At the end of the first period, all voters observe $(g_i, \tau_i)_{i=1,...n}$. Now, if an incumbent extracts maximum rents in one region (by setting $g^i = 0, \tau^i = r$) this will be observable by the voters in the other regions, and the incumbent will thus reveal his type and lose the election.[18] This means that there are only two first-period strategies that are potentially optimal: pooling, which is $(g_i, \tau_i) = (g_H, \tau_H), i = 1, ..n$, and separating, which is $(g_i, \tau_i) = (0, r),\ i = 1, ..n$. Finally, say a region is high-cost (low-cost) if $c_1^i = c_H$ ($c_1^i = c_L$). We then have the following result:

Proposition 1. *Assume that $q \geq (1/2)^{1/n}$. Suppose that $k \in \{0, 1, ..n\}$ of the regions are high cost. If $k = n$, the incumbent always separates. If $k < n$, the incumbent pools if $\hat{r} \geq \frac{n}{n-k}(1-\delta)r = r_k$ and separates otherwise.*

Note the key feature of Proposition 1: the more high-cost regions there are, the higher first-period rent $\hat{r}$ has to be to induce the bad incumbent to pool. Note also that we make an assumption that $q \geq (1/2)^{1/n}$: this plays the role of ruling out a possible hybrid equilibrium, as in the decentralization case.

[18] We call this the "*information consolidation*" effect of centralization.

4.2. Separation Probabilities

Note that r_k is strictly increasing in k, and strictly so when the r_k are strictly positive, with $r_n = +\infty$. So, we can write down a formula for the ex ante probability of separation Let p_k be the probability that k or fewer regions are high-cost[19]. Then:

$$s_F = \begin{cases} 1 - p_k, & r_k \leq \hat{r} < r_{k+1},\ k = 0, 1..n-1 \\ 1, & \hat{r} < r_0 = (1-\delta)r \end{cases} \tag{4.1}$$

The explanation is as follows. If $r_k \leq \hat{r} < r_{k+1}$, the incumbent pools only if there are k or fewer high-cost regions, which occurs with probability p_k so he separates with complementary probability $1 - p_k$. If $\hat{r} < r_0$, the incumbent separates no matter what k is.

How does s_F compare to s_D? It is convenient to use the Figure 1 above to illustrate this. The separation probability s_F as a function of δ, is superimposed on Figure 1 to give Figure 2

. When δ is low, i.e. below $1 - \frac{\hat{r}}{r}$, separation always occurs, even if all regions are low-cost. When δ is high, i.e. above $1 - \frac{\hat{r}}{nr}$, separation never occurs, unless all regions are high-cost, which occurs with probability q^n. Generally, s_F is monotonically decreasing in δ. Note that s_F can be above or below s_D. For low value of δ, separation always occurs in either regime $s_F = s_D = 1$, and for high values of δ, $s_F < s_D$, so that there is more pooling in equilibrium with.centralization If, for example, $\delta \geq 1 - \frac{\hat{r}}{nr}$, the bad incumbent is harder to "detect" than with decentralization, as he only reveals himself when *all* regions are high-cost (whereas the bad incumbent with decentralization reveals himself whenever his own region only is high-cost). But, note that because $q < 1-(1-q)^3 \leq 1-(1-q)^n$, there will always be an intermediate range of values of δ for which $s_F > s_D$. That is, when δ is in the intermediate region, the bad incumbent is easier to "detect" for the voters than with decentralization , as he reveals himself in all cases *except* when all regions are low-cost, whereas with decentralization, the incumbent reveals himself when his own region is high-cost.

The intuition is as follows. Note that the opportunity cost (per region) of pooling with centralization is simply one nth of maximum rent with separation, nk minus the maximum rent with pooling $\hat{r}(n-k)$ i.e. $r - \hat{r}(1 - \frac{k}{n})$ This increases quite smoothly

[19]Note $p_k = \Pr(X \leq k)$, where X is a random variable with a Binomial distribution with parameters q, n.

Figure 2: Separation Probabilities,
Centralization with Full Information

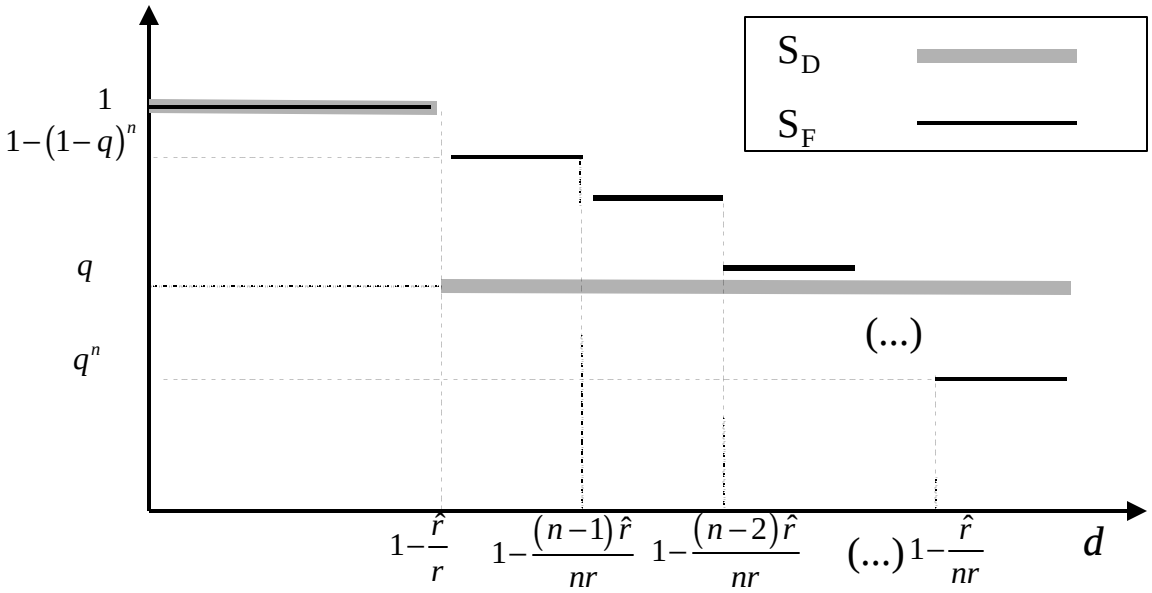

Figure 4.1:

with k, especially when n is large. This is to be contrasted with the decentralization case, where the opportunity cost of pooling, i.e. $r-\hat{r}(1-k)\dot{}\ k \in \{0,1\}$, changes discontinuously when k rises from zero to 1. So, we will call the difference in opportunity costs across the *opportunity cost effect*. With full information, this is the only difference between centralization and decentralization.

Also, note that Figure 2 illustrates nicely that the Baron-Besanko(1992) information consolidation argument, according to which the principal can more easily detect competence when the agent (incumbents) performs in several regions, does not translate immediately to our incomplete contract context. The reason is because the agent (incumbent) decides when to make the information available to the principal. So separation probabilities can go either way.

4.3. Voter Welfare

We now turn to welfare analysis. Let $EW_F,\ EW_D$ be the expected present value of welfare to the voter of any region calculated at the beginning of period 1, before the type of the incumbent and the cost shocks are determined, under full-information centralization and decentralization respectively. .

It is useful to develop the formulae for $EW_F,\ EW_D$ as they will make clear that the welfare ranking of centralization and decentralization depends *entirely* on the separation probabilities. Define

$$W_k = H(g_k) - c_k g_k,\ \hat{W} = qW_H + (1-q)W_L,\ \tilde{W} = \pi\hat{W} - (1-\pi)r$$

where $\hat{W}$ and $\tilde{W}$ denote the expected welfare produced by a good incumbent and by a challenger, respectively. Then, with both centralization and decentralization, second - period expected utility in a region, given that the bad incumbent in that region separates with probability s, is

$$EW_2(s) = \pi\hat{W} + (1-\pi)[s\tilde{W} + (1-s)(-r)] \tag{4.2}$$

The explanation is as follows. With probability π the first-period incumbent is good, in which case he stays in office with probability 1, and delivers expected utility $\tilde{W}$ to the voters in the region. With probability $1-\pi$ the first-period incumbent is bad. If he does not separate, which occurs with probability $1-s$, he will be re-elected and extract maximum rent in the last period. If he separates, he is replaced by a challenger which is good (bad) with probability π $(1-\pi)$. This challenger therefore delivers expected utility of $\tilde{W}$.

Now consider period 1 payoffs, conditional on separation probabilities. With either decentralization or centralization, the first-period expected payoff is

$$EW_1(s) = \pi\hat{W} + (1-\pi)[s(-r) + (1-s)W_H] \tag{4.3}$$

The explanation is the following. With probability π the first-period incumbent is good, in which case he delivers expected utility $\hat{W}$ to the voters in the region. With probability $1-\pi$ the first-period incumbent is bad. If he separates, he extracts maximum rent. If pools, which occurs with probability $1-s$, he always does so by setting g_H, τ_H when the true cost is low.

So, using (4.2),(4.3), the equilibrium welfares with centralization and decentralization are

$$EW_D = EW_1(s_D) + \delta EW_2(s_D),\ EW_F = EW_1(s_F) + \delta EW_2(s_F) \tag{4.4}$$

So, the difference can be decomposed as follows:

$$\begin{aligned} EW_F - EW_D &= [EW_1(s_F) + \delta EW_2(s_F) - (EW_1(s_D) + \delta EW_2(s_D))] \\ &= (s_F - s_D)(1-\pi)[-(W_H + r) + \delta(\tilde{W} + r)] \\ &= (s_F - s_D)(1-\pi)[-(W_H + r) + \delta\pi(\hat{W} + r)] \end{aligned} \tag{4.5}$$

Remembering that W_H, $\hat{W}$ and r are parameters, it follows that the welfare comparison depends *entirely* on whether the separation probability is smaller or larger with decentralization than with centralization. Moreover, note that $\Delta_S \equiv \hat{W} + r$ is the *selection benefit* of separation: if a bad incumbent is replaced by a challenger, the challenger will be good with probability π, in which case he gives the voters $\tilde{W}$ rather than $-r$ in the second period). .Moreover, $\Delta_I \equiv W_H + r$ is the *incentive cost* of separation (the bad incumbent gives the voters $-r$ rather than W_H in the first period). We have therefore proved the following:

Proposition 2. *With either fiscal arrangement, voter welfare is increasing in the separation probability if $\Delta_I/\Delta_S > \delta\pi$. In this case, voter welfare is higher with whichever arrangement gives the higher separation probability. With either fiscal arrangement, voter welfare is decreasing in the separation probability if $\Delta_I/\Delta_S < \delta\pi$. In this case, voter welfare is lower with whichever arrangement gives the higher separation probability. If $\Delta_I/\Delta_S = \delta\pi$, voters are always indifferent between centralization and decentralization.*

The condition determining voter preference over separation is intuitive. The benefits of separation come in the second period, and only occur with probability π. So, δ and π must be sufficiently high for voters to prefer separation.

5. Centralization with Partial Voter Information

5.1. Equilibrium

Second-period behavior of an incumbent of a given type (good or bad) is the same as with full voter information. So, all voters prefer the benevolent policy-maker in period 2. In the first-period, the benevolent incumbent behaves non-strategically and so will make exactly the same policy choices as in the second period.

To analyze the first-period behavior of the non-benevolent incumbents, we introduce the following terminology. The incumbent *separates in region* i if he chooses $g^i, \tau^i \neq (g_L, \tau_L)$ or (g_H, τ_H), and *pools in region* i otherwise. As voters only observe fiscal policy in their own region, all voters in i vote for the incumbent if he pools, and for the challenger otherwise. So, w.l.o.g, we can assume that if the incumbent separates in region i, he sets $(g^i, \tau^i) = (0, r)$. Also, say that an incumbent *separates overall* if he only chooses to pool in a minority of regions, and *pools overall* otherwise. An incumbent wins the election if and only if he pools overall. Thus, the selection effect, i.e. the ex ante probability that he is de-selected if "bad" - which is the focus of our analysis - is the probability that he separates overall.

Proposition 3. *Suppose that k of the regions are high cost. If $k < m = (n+1)/2$, (a majority of low cost regions) the incumbent pools in m low-cost regions, separates in the other regions, and thus pools overall if $\hat{r} \geq \max\{(1 - \frac{n}{m}\delta)r, 0\} = \underline{r}$ and separates in all regions otherwise. If $k \geq m$, (majority of high cost regions) the incumbent will pool overall iff $\hat{r} \geq \max\{(1 - \frac{n}{m}\delta)\left(\frac{m}{n-k}\right)r, 0\} \equiv r_k$. In this case, the incumbent wins the election by pooling in all $n-k$ low-cost regions, and $k-m+1$ randomly selected high-cost regions, and separates in the other regions.*

Note that as in the case of full information, the more high-cost regions there are, the higher first-period rent $\hat{r}$ has to be to induce the bad incumbent to pool. however, the critical value of $\hat{r}$ is lower than in the case of full voter information (for a formal proof, see Section 5.4 below), as the incumbent now has the option (which he takes) of selective rent-diversion. Note now that we only need the condition $q \geq 0.5$ as now the inference problem facing the voter in any region is the same as with decentralization.

5.2. Separation Probabilities

We assume in this section that $\delta < \frac{m}{n}$, because if the opposite is true, the separation probability is always zero, from Proposition 3. If $\delta < \frac{m}{n}$,note that r_k is increasing in k, and strictly so when the r_k are strictly positive, with $r_n = +\infty$. Again, define p_k to be the probability that the number of high-cost regions is less than or equal to k. Then the ex ante probability of separation in any region is :

$$s_P = \begin{cases} 1, & 0 \leq \hat{r} < \underline{r} \\ 1 - p_{m-1}, & \underline{r} \leq \hat{r} < r_m \\ 1 - p_k, & r_k \leq \hat{r} < r_{k+1},\ k \geq m \end{cases} \tag{5.1}$$

The explanation is as follows. If $\hat{r} < \underline{r}$, then the incumbent always separates. If $\underline{r} \leq \hat{r} < r_m$, the incumbent pools only if $k \leq m-1$, which occurs with probability p_{m-1}, so he separates with complementary probability $1 - p_{m-1}$. If $r_k \leq \hat{r} < r_{k+1}$, then the incumbent only separates if the number of high-cost regions is greater than k, which occurs with probability $1 - p_k$.

How does s_P compare to s_D? Again, it is convenient to use Figure 1 above to illustrate this. The separation probability s_P as a function of δ, is superimposed on Figure 1 to give Figure 3.

Here, for clarity, we have assumed $n = 3$. When δ is low, i.e. below $\frac{2}{3} - \frac{\hat{r}}{r}$, separation always occurs. When δ is high, i.e. above $\frac{2}{3} - \frac{\hat{r}}{3r}$, separation does not occur unless all regions are high-cost, which occurs with probability q^3. In between these two values of δ, separation only occurs with partial information if at least two regions are high-cost, an event which occurs with probability $q^3 + 3q^2(1-q)$. In that case, it is possible that $q^3 + 3q^2(1-q) > q$: for example, if $q = \frac{3}{4}$, $q^3 + 3q^2(1-q) = \frac{27}{64}$.

Thus, separation can be more likely with centralization even with the possibility of selective pooling. On the other hand, as we will see below, separation is unambiguously less likely under partial information than it is under full information.

The intuition is simply that there are now two determinants of s_P. As in the case of s_F, the opportunity cost effect is still at work. But now, overlaid on this effect is the selective pooling effect which implies that $s_P < s_F$. But. the opportunity cost can still dominate, implying that we can get $s_P > s_D$, as above.

5.3. Voter Welfare

As before, with both centralization and decentralization, second -period expected utility in a region, given that the bad incumbent in that region separates with probability s,

Figure 3: Separation Probabilities, Centralization with Partial Information

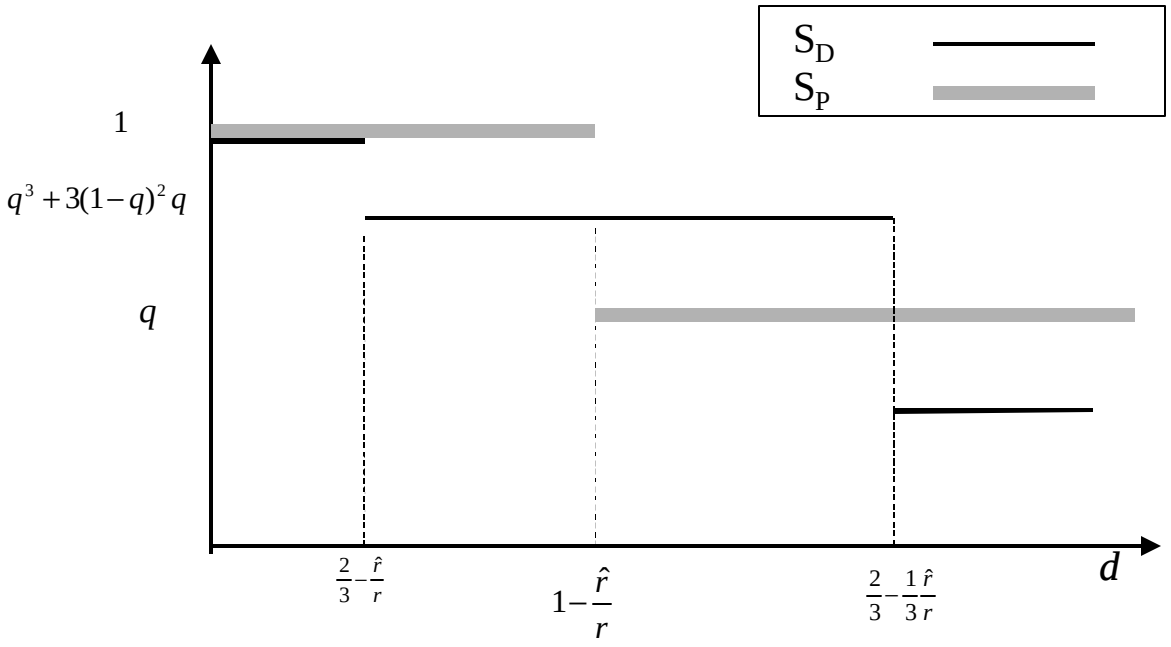

Figure 5.1:

is $EW_2(s)$, defined in (4.2) above. Now consider period 1 payoffs, conditional on separation probabilities. With decentralization, the first-period expected payoff is $EW_1(s)$ as defined above. So, the overall payoff to decentralization is as before.

Now consider the first-period expected payoff with centralization. With partial voter information, there is a distinction between separating (or pooling) in the aggregate, and at the level of the individual region. In particular, as is clear from Proposition 3, when the bad incumbent separates in the aggregate, he does so by separating in each region (i.e. by taking maximum rent), but when he pools, he does so in the minimum number of regions needed to win the election i.e. m regions. That is, in the event of pooling in the aggregate, the expected payoff to a region is $(1-\frac{m}{n})(-r)+\frac{m}{n}W_H$, as the incumbent selects m regions out of n in which to pool (as described in the proof of Proposition 3). As all regions are ex ante identical, the ex ante probability of being selected is therefore $\frac{m}{n}$. So, the expected payoff with separation probability s is

$$\begin{aligned} EW_1^P(s) &= \pi\hat{W}+(1-\pi)[s(-r)+(1-s)((1-\frac{m}{n})(-r)+\frac{m}{n}W_H)] \\ &= EW_1(s)-(1-s)(1-\pi)(1-\frac{m}{n})\Delta_I \end{aligned} \tag{5.2}$$

where the second term is the welfare cost of selective pooling. This is decomposed as follows. With probability $1-\pi$, the incumbent is bad. If this incumbent pools overall ($s=0$), then with probability $(1-\frac{m}{n})$, any region will be chosen to be amongst the unfortunate $n-m$ regions where the incumbent takes maximum rent by setting $g=0,\ \tau=r$, rather than rent $\hat{r}$ by setting $g=g_H, \tau=\tau_H$. The cost to any such region of this is Δ_I.

So, the equilibrium welfares with centralization and decentralization are

$$EW_D(s_D)=EW_1(s_D)+\delta EW_2(s_D),\ \ EW_P(s_P)=EW_1^P(s_P)+\delta EW_2(s_P) \tag{5.3}$$

So, the difference can be decomposed as follows:

$$\begin{aligned} EW_P(s_P)-EW_D(s_D) &= [EW_1(s_P)+\delta EW_2(s_P)-(EW_1(s_D)+\delta EW_2(s_D))] \\ &\quad +\left[EW_1^P(s_P)-EW_1(s_P)\right] \\ &= (s_P-s_D)(1-\pi)[-\Delta_I+\delta\pi\Delta_S] \\ &\quad -(1-s_P)(1-\pi)(1-\frac{m}{n})\Delta_I \end{aligned} \tag{5.4}$$

As is clear from (5.4), there are now *two* effects on welfare of moving to centralization:

1. A change in the separation probability, evaluated using the decentralized welfare criterion:

2. A reduction in welfare at a given separation probability, because limits on rent-diversion are only needed in a majority of regions (instead of all regions) to be reelected - the *selective pooling* effect,

In general, these two effects could go either way. However, we have:

Proposition 4. *If* $\delta > \max\{\frac{\Delta_I}{\pi\Delta_S}, \frac{m}{n}\}$, *then* $s_P = 0$, *and* $EW_D > EW_P$. *If* $\delta \leq \max\{\frac{\Delta_I}{\pi\Delta_S}, \frac{m}{n}\}$, *then examples can be found when* $EW_D < EW_P$.

Proof. (i) If $\delta > \max\{\frac{\Delta_I}{\pi\Delta_S}, \frac{m}{n}\}$, then $\delta > \frac{m}{n}$, so from Proposition 3, $s_P = 0$. Also, $\delta\pi > \frac{\Delta_I}{\Delta_S}$, so $-\Delta_I + \delta\pi\Delta_S > 0$. But then as $s_D \geq s_P$, the result follows from (5.4).

(ii) See Example 1 below. □

The intuition for the general result is as follows. When δ is high, voters prefer a higher separation probability, because the benefits of separation come in the second period. But when δ is high, the incentive to pool with centralization is very strong, as the policy-maker only need sacrifice rent-extraction in $\frac{m}{n}$ of the regions to be elected in the first period, thus gaining second-period rents in all regions. So, voters are worse off with centralization both because the separation probability is lower, and because they prefer decentralization at a given separation probability, due to the selective pooling effect.

To generate an example where centralization is preferred, a necessary condition is that voters dislike separation (i.e. $\delta\pi$ is low enough). But that is not sufficient: we require also that the gain from greater pooling under centralization offsets the loss from selective pooling effect. But this is possible if δ is low enough, as the following example shows.

Example 1. Let $n = 3$, $\delta \leq \min\left\{\frac{2}{3\pi}\frac{\Delta_I}{\Delta_S}, \frac{1}{2}\right\}$. Then, (5.1) gives the relevant separation probabilities. Assume $\hat{r}$ is such that $(1 - \frac{3}{2}\delta)r \leq \hat{r} < (1 - \delta)r$. Then, $s_D = 1$, $s_P = q^3 + 3q^2(1 - q)$. Further, let $q = 0.5$. Then, $s_P = \frac{1}{2}$. Then from (5.4),

$$\begin{aligned} EW_P - EW_D &= (1-\pi)\left[-\frac{1}{2}[-\Delta_I + \delta\pi\Delta_S] - \frac{1}{2}\frac{1}{3}\Delta_I\right] \\ &= (1-\pi)\left[\frac{1}{3}\Delta_I - \frac{1}{2}\delta\pi\Delta_S\right] \\ &> 0 \text{ as } \delta \leq \frac{2}{3\pi}\frac{\Delta_I}{\Delta_S}. \end{aligned}$$

which is the required result. □

5.4. Comparing Partial and Full Voter Information

We are now in a position to ask what the effects on separation probabilities and voter welfares are of switching from partial to full voter information, given centralization. In an incomplete contracting framework such as this, one should not presume that more information is better, and indeed that is not the case. However, it is possible to establish that *conditional* on a fixed separation probability, voter prefer full information. Our results here are:

Proposition 5. *(i) A change from partial to full voter information always increases separation probabilities (ii) A change from partial to full voter information always increases voter welfare, conditional on a fixed s : (iii) A change from partial to full voter information will always increase voter welfare unconditionally if $\delta\pi > \frac{\Delta_I}{\Delta_S}$, but if $\delta\pi < \frac{\Delta_I}{\Delta_S}$, examples can be found where a move from partial to full voter information will decrease voter welfare.*

The reason of part (i) of Proposition 5 is that pooling is more profitable with partial information as the incumbent has to restrain rent diversion in only a majority of regions (instead of all regions). As already remarked, part (iii) of Proposition 5 is comparable to Proposition 5 of Besley and Smart(2003). In the context of decentralized fiscal policy, they demonstrate that yardstick competition between regions (which can only occur when voters are fully informed in our sense) does not necessarily increase voter welfare relative to no yardstick competition. However, in our case, the mechanism at work is quite different.

6. Centralization with Uniform Taxation

6.1. Equilibrium

Again, second-period behavior of an incumbent of a given type (good or bad) is the same as with full voter information. So, all voters prefer the benevolent policy-maker in period 2. In the first-period, the benevolent incumbent behaves non-strategically and so will make exactly the same policy choices as in the second period.

To analyze the first-period behavior of the non-benevolent incumbents, we introduce the following terminology. Define

$$T = \left\{ \tau \left| \tau = \frac{k\tau_H + (n-k)\tau_L}{n},\ k \in \{0, 1, ..n\} \right. \right\}$$

This is the set of uniform tax rates that can possibly be set by a good incumbent in equilibrium. Say that the bad incumbent *pools with respect to taxation* $\tau \in T$. Also, say

that the bad incumbent *pools with respect to expenditure in region i* if $g^i \in \{g_L, g_H\}$. Note that in order to be re-elected, a bad incumbent must (i) pool with respect to taxation, *and* (ii) pool with respect to expenditure in at least m regions. So, (assuming that q is high enough that voters will re-elect the incumbent even if they observe a tax $\tau = \tau_H$) the cheapest way to be re-elected for the bad incumbent is to pool with respect to taxation by setting τ to the maximum value in T i.e. τ_H, and pool with respect to expenditure in exactly m regions by setting $g_i = g_H$ in those regions. Once this observation is made, the next result follows easily;

Proposition 6. *Assume $q \geq (\frac{1}{2})^{1/n}$. Suppose that k of the regions are high cost. If $k < m$, the incumbent sets $\tau = \tau_H$, and $g_i = g_H$ in m low-cost regions and $g_i = 0$ elsewhere, and is thus re-elected if $\frac{m\hat{r}+(n-m)\tau_H}{n} \geq (1-\delta)r$: otherwise, he separates and is not re-elected. If $k \geq m$, the incumbent sets $\tau = \tau_H$, and $g_i = g_H$ in all $n-k$ low-cost regions and $m-(n-k)$ high-cost regions, and $g_i = 0$ elsewhere, and is thus re-elected, if $\frac{(n-k)\hat{r}+(n-m)\tau_H}{n} \geq (1-\delta)r$: otherwise, he separates and is not re-elected.*

6.2. Separation Probabilities

Now we turn to a characterization of separation probabilities. Let $\underline{r}$, r_k solve

$$\frac{m\underline{r}+(n-m)\tau_H}{n} = (1-\delta)r, \ \frac{(n-k)r_k+(n-m)\tau_H}{n} = (1-\delta)r, \ k \geq m \tag{6.1}$$

respectively. Then clearly $\underline{r} < (1-\delta)r$ since $\tau_H > \widehat{r}$. Note also that as $n-k < m$ for $k \geq m$, $r_k > \underline{r}$ and r_k is increasing with $r_n = \infty$. Then the ex-ante probability of separation is given by (5.1) above, but with $\underline{r}$, r_k as defined in (6.1). We can again compare separation probabilities s_D and s_U using the fact that $\underline{r} < (1-\delta)r$. Very much along the lines of the case of partial voter information, it can be shown that "usually" $s_U < s_D$, but cases can be found where $s_U > s_D$.

6.3. Voter Welfare

The computation of voter welfare is similar to the case of partial information, except that first-period welfare is now

$$EW_1^U(s) = \pi\hat{W} + (1-\pi)[s(-r) + (1-s)((1-\frac{m}{n})(-\tau_H) + \frac{m}{n}W_H)] \tag{6.2}$$

So, overall welfare is

$$EW_U(s) = EW_1^U(s) + \delta EW_2(s) \tag{6.3}$$

The reason is as follows. In the event of a bad incumbent pooling with uniform taxation, the expected payoff to a region is $(1-\frac{m}{n})(-\tau_H)+\frac{m}{n}W_H$, as the incumbent selects m regions out of n in which to pool (delivering welfare W_H in each), as the maximal rent the incumbent can extract in the $n-m$ regions is restrained to the uniform tax τ_H.

6.4. Comparing Differentiated and Uniform Taxes

As mentioned in the introduction, our results have some interesting implications for the choice between differentiated and uniform taxes. Assume (as is probably reasonable), that voters only have partial information at the voting stage i.e. they only observe expenditure and the tax rate in their own region, and at a constitutional stage, they have to choose between uniform and differentiated taxes. This amounts to comparing EW_P and EW_U. In our model, a shift from differentiated to uniform taxes will have two effects. First, at a given separation probability, it will increase voter welfare as it constrains the ability of the incumbent to extract rents without being detected. Second, it will raise the equilibrium separation probability, as - for the same reason - pooling is less profitable for the incumbent. This allows us to identify simple conditions under which voters prefer to choose a uniform tax.

Proposition 7. *Assume that voters can only observe expenditure and the tax rate in their own region. Then, with (i) with uniform taxation, the separation probability is at least as high as it is with differentiated taxes, and (ii) if* $\delta\pi \geq \frac{\Delta_I-(1-\frac{m}{n})(W_H+\tau_H)}{\Delta_S}$, *voter welfare is strictly higher with uniform than with differentiated taxes.*

7. The Rent Scale Effect

As emphasized by Seabright(1996), an important determinant of incumbent incentives with centralization (relative to decentralization) is the total size of rents from office as a function of the number of jurisdictions under the incumbent's control, n. Seabright's view is that total rent is an increasing function of n, but may increase less than linearly (diminishing returns). In our model, the total size of rents from office is determined endogenously. So far, to focus on the analytics of the selective pooling effect, we have made enough assumptions so that this is linear in n. In particular, what is crucial for first-period behavior is of course, second period rents accruing to the winner of the election: in our model this is r with decentralization, and nr with centralization.

To allow for diminishing returns in a simple and plausible way, we suppose, following Persson and Tabellini(2000), that there is also an exogenous ego-rent R from office. Then,

second period rents accruing to the winner of the election is now $r + R$ with decentralization, and $nr + R$ with centralization i.e. diminishing returns. Moreover, the larger R, the smaller is $(nr + R)/(r + R)$, so R is an inverse measure of the rent scale effect: the higher R, the lower the effect. We now wish to consider the effect of a change in R on equilibrium under both centralization and decentralization.

By a simple repetition of the argument in Section 3, the condition under which there is pooling under decentralization is $\hat{r} \geq (1-\delta)r - \delta R = r_D$. To make the comparison as clean as possible, consider a change in R, ΔR accompanied by a change in δ that leaves r_D unchanged i.e.

$$\Delta\delta = -\frac{\delta}{r+R}\Delta R \tag{7.1}$$

Then, we have the following:

Proposition 8. *An increase in R, along with an offsetting decrease in δ of (7.1) that leaves r_D, and thus the equilibrium separation probability s_D unchanged, will unambiguously increase s_F, s_P and s_U. That is, whatever the information structure with centralization, a decrease in the rent scale effect will increase the separation probability.*

Proof. We only give the proof for the full-information case: the other cases are similar. Only two strategies can possibly be optimal for the incumbent. The first is to separate, and lose the election which gives him a payoff nr : the second is to pool and win the election, which gives a payoff $(n-k)\hat{r} + \delta(nr + R)$. The second strategy is better iff

$$\hat{r} \geq \frac{1}{n-k}[(1-\delta)nr - \delta R] = r_k \tag{7.2}$$

Now, from (7.1),(7.2), we have

$$\begin{aligned}
\Delta r_k &= -\frac{1}{n-k}[\delta\Delta R + (nr+R)\Delta\delta] \\
&= -\frac{\Delta R}{n-k}[\delta - (nr+R)\frac{\delta}{r+R}] \\
&= \frac{\delta(n-1)r}{(n-k)(r+R)}\Delta R
\end{aligned}$$

So, a fall in R will decrease all r_k. Thus, from (4.1), if s_F changes, it must fall. □

This result is not so surprising: the greater the relative concern for re-election, the greater the incentive to pool with centralization. What *is* interesting is that in our model, with asymmetric information, the effect of a change in rent-scale on voter welfare with centralization relative to decentralization is ambiguous. This is clear from the above analysis. For example, Proposition 2 gives conditions under which an increase in s_F may increase or decrease voter welfare.

This is in contrast to Persson-Tabellini (and implicitly, the analysis of Seabright), where an increase in the rent-scale effect *unambiguously increases voter welfare*. To see this, note that from equations (9.5) and (9.8) of Persson-Tabellini(2000), first-period rents diverted in equilibrium with decentralization and centralization respectively are:

$$r_D = \tau - \xi\delta(R + r),\ r_C = \tau - \frac{\xi}{2}\delta(R + 3r) \tag{7.3}$$

where R, r, δ have the same interpretation as in our model, and τ, ξ are parameters. Then, it is easy to check from (7.3) that a fall in R, along with an offsetting increase in δ of (7.1) that leaves r_D unchanged, will lower r_C. As voter welfare in their model depends (negatively) only on first-period rents diverted (remember that the separation probability is fixed at 0.5) this must increase voter welfare.

8. The Competence Model

Here, we briefly sketch results for the competence model. The purpose of this section is simply to emphasize that for the most part, the results already obtained carry over to the competence case.

8.1. Decentralization

In the second period, both good and bad incumbents will extract maximum rent r. Having done that, they wish to supply the good as close as possible to the efficient level in each region, whatever the cost variable. We will assume that $g_k > (1-r)/c_k = \tilde{g}_k$ i.e. having extracted maximum rent, the maximum possible public good supply -given the maximum tax rate of unity - is less than the efficient level. So, an incumbent with cost c_k supplies $\tilde{g}_k = (1-r)/c_k$, $k = H, L$. So, all voters prefer a competent to an incompetent incumbent, since public good supply is higher.

Now consider the first period. Assume for the moment that good incumbent in either region will be elected with probability 1 if he behaves non-strategically i.e. makes exactly the same policy choices as in the second period. We will shortly verify when this is equilibrium behavior for the voters. In this case, the best strategy for the good incumbent is to behave non-strategically.

Now consider the bad incumbent in i. First, note that by Bayes' rule, the voters' posterior belief that the incumbent is good when $(g_i, \tau_i) = (\tilde{g}_L, 1)$ is observed is

$$\pi' = \frac{\pi}{\pi + (1-\pi)\sigma} \geq \pi$$

where σ is the probability that in equilibrium, a bad incumbent sets $(g_L, 1)$. For the voters to be willing to re-elect the incumbent, we require $\pi' \geq \pi$, which is true for all $\sigma \in [0,1]$ so, the voters are always willing to re-elect in equilibrium the incumbent whenever he pools.

So, if his cost is low $(c_i = c_L)$ he cannot do better than imitate (pool with) the good incumbent, because if he imitates, he will be re-elected while extracting maximum rent. If his cost is high, $(c_i = c_H)$ he has two possible options. The first is to imitate the tax and expenditure of the good incumbent i.e. set $(g_i, \tau_i) = (\tilde{g}_L, 1)$, which leaves him with reduced rent of $\hat{r} < r$, where

$$\frac{(1-\hat{r})}{c_H} = \frac{1-r}{c_L} = \tilde{g}_L$$

giving him payoff $\hat{r} + \delta r$. The second is to separate by taking maximum rent and thus setting $(g_i, \tau_i) = (\tilde{g}_H, 1)$, thus losing the election and giving him payoff $r + \delta.0$. So, will pool if $\hat{r} > (1-\delta)r$, and separate otherwise.

So, in any region, the ex ante probability that a bad incumbent separates (the *selection effect*) is

$$s_D = \begin{cases} 0 & \text{if } \hat{r} > (1-\delta)r \\ q & \text{if } \hat{r} < (1-\delta)r \end{cases}$$

Comparing with the benevolence model, the separation probability is lower in the competence model because when cost is high the incumbents do not separate if $\hat{r} > (1-\delta)r$; and when cost is low the incumbent never separates, no matter $\hat{r}$. In the benevolence model there is always separation when cost is high and separation is also possible when cost is low if $\hat{r} < (1-\delta)r$.

8.2. Centralization

With centralization, again three different possibilities can be analyzed: full information, partial information, and uniform taxation. The qualitative results are very similar to the case of benevolence, with two key exceptions. So, to avoid excessive re-statement of very similar propositions(full details are available on request from the authors), we will just note these exceptions.

First, the partial information case and uniform taxation case are the same in the competence model since both good and bad incumbents always set the same maximum tax of $\tau = 1$ in equilibrium. This leaves only two cases for consideration, full and partial information. In both of these cases, unlike in the benevolence case, voter welfare is *always* decreasing in the separation probability. The reason for this is straightforward.

Let $W_k = H(\tilde{g}_k) - 1$ be the payoff to a voter from an incumbent who behaves as if he has cost c_k, $k = L, H$. Then, in period 1, if the bad incumbent separates rather than pools in a region, he delivers W_H rather than W_L to the voters, so the current loss from separation is $W_L - W_H$. The expected future gain is that next period's incumbent will be high-cost with probability only $(1-\pi)q$, rather than q, implying a gain in voter welfare of $\delta\pi q(W_L - W_H)$. So, it is not surprising that voters are worse off with separation as it involves a cost in the first-period that is certain against an equal-size benefit in the next-period that is uncertain and discounted. All the propositions above hold, appropriately restated, taking into account the differences just noted.

Second, on voter welfare. With full voter information, voter welfare is higher with whichever arrangement gives the lower separation probability: this is the analog of Proposition 2. With partial information, given the same separation probability, decentralization dominates centralization, due to selective pooling effect which is also present in the competence model with partial information. Nevertheless, examples can be found where centralization dominates; these involve a lower separation probability with centralization. This is the analog of Proposition 4. Note in this case, the cost and benefit of decentralization is particularly clear: it prevents selective pooling, to the benefit of voters, but will usually increase the separation probability, to the cost of voters.

9. Some Extensions

9.1. Region-Specific Competence

So far, we have assumed that the unknown characteristic of the incumbent - benevolence or competence - pertains to the incumbent, rather than the region. In the case of benevolence, this is more plausible: it would be hard to explain why an incumbent would be benevolent in one region and rent-seeking in another. On the other hand, in the competence model, it is quite natural to think of the competence characteristic being determined at the regional level, even with centralization. Indeed, this is in some respects a more attractive assumption than the one we have made in the previous section, as it means that the joint distribution of costs of producing the public goods across regions in the first period is unchanged following a move from decentralization to centralization.[20]

In this case, the analysis of decentralization is the same with separation probability s_D (with pooling if $\hat{r} \geq (1-\delta)r$). With centralization, the incumbent is competent in

[20]Under the assumption made above, the unconditional average cost is unchanged following this move, but under centralisation, costs are more highly correlated.

each region with probability π (instead of competent in all regions with probability π). Since competence is uncorrelated across regions, voters only care about performance in their region. Voter information about the fiscal outcomes in other regions is irrelevant. So when the bad incumbent separates in the aggregate, he does so by separating in each region (i.e. by taking maximum rent), but when he pools, he does so in the minimum number of regions needed to win the election (i.e. m regions). That is, in the event of (selective) pooling, the incumbent selects m regions out of n in which to pool and the expected payoff to a region is $(1-\frac{m}{n})W_H + \frac{m}{n}W_L$.

Let k denote the number of high cost regions (i.e., k is a random variable with a Binomial distribution with parameters q', n, where $q' = (1-\pi)q$). If $k < m$, there is a majority of low cost regions and thus the incumbent wins the election. This gives payoff $nr + \delta nr$ to the incumbent. If $k \geq m$, there is a majority of high cost regions and only two strategies can possibly be optimal for the incumbent. The first is to take maximum rent in all regions, thus separating in all regions and lose the election which gives him a payoff $nr + \delta.0$. The second is to limit his rent to $\hat{r}$ in $l = m - (n-k)$ high-cost regions (thus pooling in those regions), and take maximum rent in all other $(n-l)$ regions. This is pooling and thus the incumbent wins the election. This gives payoff $l\hat{r} + (n-l)r + \delta nr$. The second strategy is better iff $\hat{r} \geq (1 - \frac{n}{l}\delta)r = r_k$. Note by inspection that whatever k, the gain to pooling is strictly greater with centralization, as the incumbent can extract maximum rent r from a minority of regions, rather than just $\hat{r}$. So, there is less separation with centralization, $s_C \leq s_D$. Second, on voter welfare. In the competence model, separation is bad for voter welfare (it reduces public good supply in the first period for a possible increase by the same amount in the second period). However, for the same separation probability, decentralization dominates centralization, due to selective pooling effect. Thus, the cost and benefits of decentralization are particularly clear: it prevents selective pooling, to the benefit of voters, but will increase the separation probability, to the cost of voters.

9.2. Multiple Policy-Makers with Centralization

So far, we have assumed that there is a single policy-maker with both decentralization and centralization. This is not the only - or most reasonable - possibility. A related weakness of our approach is that we assume that in the benevolence model, the "good" policy-maker cares equally about welfare in all regions: this is hard to justify. One literature (e.g. Lockwood(2002), Besley and Coate (2003)) takes a legislative bargaining approach to modelling centralization, where decision-making with centralization is assumed to be

made in a legislature composed of one delegate[21] (or equal numbers of delegates) from each region. This approach has two merits: it provides a model of decision-making at the centralized level that captures some features of reality, especially in the US, and it provides a micro-founded account of where the preferences of policy-makers come from. Our analysis could be extended to this more realistic way of modelling centralized decision-making. This is a topic for future work.

9.3. Complete Contracting

In our model, elections only give voters limited control over the incumbent, relative to a complete contract which can also reward the incumbent for performance i.e. all the voters can do is "fire" the incumbent at the end of the first period. How sensitive are our results to this assumed incompleteness? One way to look at this is to write down a "complete contracting" version of our model, where performance-related pay is also an instrument, and see how the equilibrium behavior of the incumbent with the optimal contract differs from the equilibrium in our paper. It should be noted that because the agent (the incumbent) can differ both by preference type and cost, this is not a standard contracting problem. However, in the one-period case, because the cost shocks are i.i.d. across regions, it can be shown (see Appendix B) that with complete contracting, it makes no difference whether the voters contract regionally i.e. with n different regional incumbents or nationally, with one incumbent[22].

However, in the two-period case, which is really the relevant case, the voters have an ex post incentive to fire the bad incumbent if he reveals himself through his first-period choice of contract. In Appendix B, we also consider the two-period contracting problem without precommitment. We show that with either fiscal regime, it is optimal to offer the one-period contract in both periods, with the bad incumbent revealing his type if the cost

[21] This is for convenience only and can be relaxed.

[22] This is contrast to Baron and Besanko(1992), who find that the principal can more easily detect competence when the agent performs several tasks. However, their set-up is rather different. First, in their paper, the asymmetric information pertains to the production process, rather than the agent who runs the process; so even with centralization, there are two pieces of information. Second, the goods produced by the agent(s) are perfect complements: in our setting, they are independent. Third, the results of their paper do not really capture the feature of our paper that centralization allows the principal to have "more observations" on the performance of the agent. In the Baron-Besanko model, centralization can outperform decentralization, but the intuition is to do with internalization of externalities by the agent.

is low[23]. So, again we have equivalence of the two regimes with complete contracting.

10. Related Literature and Conclusions

10.1. Related Literature

Other than the papers already discussed quite extensively already, there are a number of literatures related to the topic of this paper. First, there are a set of political models of corruption assessing the relative performance of different electoral rules (majoritarian. rule versus proportional representation). In Myerson (1993) voting behavior is endogenous to the electoral rule but corruption is assumed to be an exogenous feature of each politician. The ability of voters to hold corrupt incumbents accountable is worse with the majoritarian rule as voters are less willing to switch their vote on a "good" challenger with little chance of winning. Persson and Tabellini (1999) compare political behavior under proportional representation in a nationwide district and the majority rule in a number of local districts. Competition is stronger with local majority rule as politicians are interested to win a majority, not national-wide, but in the marginal districts containing more swing voters. As these voters are more responsive to policy outcomes, politicians are more disciplined and divert less rents. This is a similar finding to ours, but the mechanism at work is quite different.

Second, there is a large body of literature on decentralization within organizations in a complete contracting framework .i.e. where the principal can pay the agent conditional on performance. The basic trade-off is between delegation of decision-making to better informed agents and the associated loss of control and agency costs (since informed agents may not have all the incentives to act in the best interests of the delegating authority). The optimal organizational form emerges from this simple trade off. For a long time it has been a challenge to explain decentralization with this framework. The reason is that from the "revelation principle" any decentralized organization can be replicated by a centralized one in which all agents with different sets of information report their information to a center who then makes all the relevant decision (Myerson(1982)). This revelation principle has for a long time been the cornerstone of the weak superiority of centralization over decentralization. However recent theoretical developments have shown that relaxing some of the underlying assumptions can create a preference for decentralization. For example if the

[23]Thus, if the cost is high, there will be first-period pooling: this is similar to that due to the ratchet effect in dynamic contracts with unobservable ability (Laffont and Tirole(1988)). There, pooling is due to high-ability agents trying to disguise their type in the first period.

center cannot commit not to renegotiate the initial contract after agents have reported their information, then decentralization is an effective commitment device by preventing ex-post opportunistic renegotiations from the center (Beaudry and Poitevin(1995)). When agents can collude when reporting their information to the center, then by creating a conflict of interest between agents, decentralization can prevent any collusion and again decentralization to informed agents may be optimal (Laffont and Martimort(1998)). When communication is costly, decentralization may also be optimal because it reduces communication costs (see Melumad, Mookherjee and Reichelstein(1997)).

However, the model of this paper is rather different to the contractual approach. Models of electoral incentives (as is ours) take an incomplete contracting approach. First delegation of decision making to politicians. The "right" policy choice depends on non-verifiable information about the state of the environment, that cannot be contracted upon ex-ante. Consequently it must be delegated to politicians, thereby creating an incentive problem since giving any residual right to an agent to make collective choices opens up the possibility of opportunistic behavior and abuse of power. Second, the decision maker is selected by election: the politician is not controlled by a contract but by the public opinion. If he wants to remain in office, he must win elections.

10.2. Conclusions

This paper has considered the effects of fiscal decentralization on both incentives and selection of policy-makers. The main message is that (except in the probably unrealistic case where voters have full information about local public good provision across the economy), bad incumbents can pool worth good ones at lower cost to themselves (but at a higher cost to voters) with centralization. This has two consequences. First, at a given separation probability, voter welfare is lower with centralization. But, equilibrium separation probabilities can be higher or lower with centralization: the forces at work on the separation probabilities are quite subtle.

Our model presents a first step toward addressing the question of the expected effect of decentralization on government efficiency in a systematic fashion by studying jointly the incentive and selection effects. One possible direction for future work is to study .empirically separation rates for policy-makers (e.g. legislators or governors) at the national and subnational levels. While there is an existing political science literature on the determinants of job tenure of politicians (see e.g. Finnoccario and Lin(2000)), to our knowledge, there has been no investigation of whether expected tenure is significantly different at different levels of government.

11. Proofs of Propositions

Proof of Proposition 1. Suppose for the moment that the voters are willing to re-elect the incumbent whenever he pools. Only two strategies can possibly be optimal for the incumbent. The first is to separate, and lose the election which gives him a payoff $nr+\delta.0$. The second is to pool and win the election. This gives payoff $(n-k)\hat{r}+\delta nr$. The second strategy is better iff $\hat{r} \geq \frac{n}{n-k}(1-\delta)r = r_k$. As $r_n = \infty$, it is always best to separate when $k = n$.

Now we verify that the voters are willing to re-elect the incumbent whenever he pools. By Bayes' rule, the voters' posterior belief that the incumbent is good when $(g_i, \tau_i) = (g_H, \tau_H), i = 1, ..n$ is

$$\pi' = \frac{q^n \pi}{q^n \pi + (1-\pi)\sigma}$$

where σ is the probability that in equilibrium, a bad incumbent plays the pooling strategy. For the voters to be willing to re-elect the incumbent, we require $\pi' \geq \pi$, or $q^n \geq \sigma$. As the incumbent always separates when $k = n$, which occurs with probability q^n, the maximum possible value of σ is $\sigma = 1 - q^n$ (this occurs when the incumbent pools in all other cost states, which occurs when $\hat{r} \geq n(1-\delta)r$). So, for the voters to be willing to re-elect in equilibrium, we require $q^n \geq 1 - q^n$, or $q \geq (1/2)^{1/n}$. □

Proof of Proposition 3. If $k < m$, there is a majority of low cost regions and only two strategies can possibly be optimal for the incumbent. The first is to take maximum rent in all regions, thus separating in all regions and lose the election which gives him a payoff $nr + \delta.0$. The second is to limit his rent to $\hat{r}$ in m low-cost regions (thus pooling in those regions), and take maximum rent in all other regions. This is pooling overall, and thus the incumbent wins the election. This gives payoff $\hat{r}m + (n-m)r + \delta nr$. The second strategy is better iff $\hat{r} \geq (1 - \frac{n}{m}\delta)r = \underline{r}$.

If $k \geq m$, there is a majority of high cost regions and only two strategies can possibly be optimal for the incumbent. The first is to take maximum rent in all regions, thus separating in all regions and lose the election which gives him a payoff $nr + \delta.0$. The second is to limit his rent to $\hat{r}$ in all $n-k$ low-cost regions (thus pooling in those regions), and limit his rent to zero in $l = m - (n-k)$ high-cost regions, and take maximum rent in all other regions. This is pooling overall, and thus the incumbent wins the election. This gives payoff $\hat{r}(n-k) + l.0 + (n-(n-k)-l)r + \delta nr = \hat{r}(n-k) + (n-m)r + \delta nr$. The second strategy is better iff $\hat{r} \geq r_k$. □

Proof of Proposition 5. (i) We must show that separation probability is higher with full voter information. With full information the separation probability s^F is as follows.

Let $r_k^F = \frac{n}{n-k}(1-\delta)r$; $k \leq n$. If $\hat{r} < r_0^F$, the incumbent separates no matter what k is and $s^F(\hat{r}) = 1$. If $r_k^F \leq \hat{r} < r_{k+1}^F$, the incumbent separates with probability $s^F(\hat{r}) = 1 - F(k)$. With partial information the separation probability s^P is as follows. Let $r_k^P = \max\{(1 - \frac{n}{m}\delta)r, \frac{m}{n-k}(1-\delta)r\}$; with $k \leq n$. If $\hat{r} < (1 - \frac{n}{m}\delta)r = \underline{r}$, the incumbent separates no matter what k is and $s^P(\hat{r}) = 1$. If $r_k^P \leq \hat{r} < r_{k+1}^P$, the incumbent separates with probability $s^P(\hat{r}) = 1 - F(k)$. Thus separation probability is decreasing step function with the same downward jump of $F(k) - F(k-1)$ around $r_k^P$and r_k^F . Since for all $k \leq n$, $r_k^F = \frac{n}{n-k}(1-\delta)r \quad > r_k^P = \max\{(1 - \frac{n}{m}\delta)r, \frac{m}{n-k}(1-\delta)r\}$ then the step function $s^F(\hat{r})$ decreases less rapidly than the step function $s^P(\hat{r})$. Thus given that $s^F(0) = s^P(0) = 1$ we must have $s^F(\hat{r}) \geq s^P(\hat{r})$ for all $\hat{r}$, with strict inequality for $\hat{r} > \underline{r}$. Hence we conclude that separation probability is higher with full information.

(ii) Given s fixed, the fact that voter welfare is higher with full information follows from (5.4) and the fact that $EW^F(s) = EW^D(s)$.

(iii) Let $EW_F(s), EW_P(s)$ be voter expected welfares with full and partial information conditional on a fixed s. Then from (4.4), (5.3) above,

$$EW_F(s) - EW_P(s) = (1-s)(1-\pi)(1-\frac{m}{n})\Delta_I$$

i.e. conditional on a given s, voters prefer full information. So, if $\delta\pi > \frac{\Delta_I}{\Delta_S}$, the voters also prefer a higher separation probability and the result follows immediately.

If $\delta\pi < \frac{\Delta_I}{\Delta_S}$, an example where partial information is preferred to full information is the following. Assume $\hat{r}$ is such that $(1 - \frac{3}{2}\delta)r \leq \hat{r} < (1-\delta)r$. Then, from the formulae (4.1), (5.1), we see that $s_F = 1$, $s_P = q^3 + 3q^2(1-q)$. Then, noting that at a given s, $EW_D = EW_F$, the example is exactly the same as in Example 1 above. □

Proof of Proposition 6. (i) Assume that a voter in i will re-elect the incumbent even if he observes $g_i = g_H$, $\tau = \tau_H$. If $k < m$, only two strategies can possibly be optimal for the incumbent. The first is to take maximum rent in all regions, thus losing the election which gives him a payoff $nr + \delta.0$. The second is pool with respect to taxation, and pool with respect to expenditure in m low-cost regions, thus being re-elected. This gives payoff $m\hat{r} + (n-m)\tau_H + \delta nr$. The second strategy is better iff $\frac{m\hat{r}+(n-m)\tau_H}{n} \geq (1-\delta)r$.

If $k \geq m$, again two strategies can possibly be optimal for the incumbent. The first is to take maximum rent in all regions, thus losing the election which gives him a payoff $nr + \delta.0$. The second is to pool with respect to expenditure in $n-k$ low-cost regions and $m - (n-k)$ high-cost regions, and separate with respect to expenditure elsewhere. This gives payoff $(n-k)\hat{r} + (n-m)\tau_H + \delta nr$. The second strategy is better iff $\frac{(n-k)\hat{r}+(n-m)\tau_H}{n} \geq (1-\delta)r$.

(ii) It remains to verify that a voter in i will re-elect the incumbent even if he observes $g_i = g_H$, $\tau = \tau_H$. But an argument identical to that in the proof of Proposition 1 indicates that this requires $q \geq (\frac{1}{2})^{1/n}$. □

Proof of Proposition 7. (i) First, note by inspection (5.2),(6.2) that whatever k, the gain to pooling is strictly greater with partial information, as the incumbent can extract maximum rent r from a minority of regions, rather than just τ_H. So, $s_U \geq s_P$.

(ii) Now, from (5.2),(5.3), (6.2),

$$\begin{aligned} EW_U(s_U) - EW_P(s_P) &= EW_U(s_U) - EW_U(s_P) + EW_U(s_P) - EW_P(s_P) \qquad (11.1) \\ &= EW_U(s_U) - EW_U(s_P) + (1 - s_P)(1 - \pi)(1 - \frac{m}{n})(r - \tau_H) \end{aligned}$$

But,from (6.3),

$$EW_U(s_U) - EW_U(s_P) = (s_U - s_P)(1 - \pi)[\delta\pi\Delta_S - (r + (1 - \frac{m}{n})(-\tau_H) + \frac{m}{n}W_H)]$$

So, a sufficient condition for $EW_U(s_U) \geq EW_P(s_P)$ is that

$$\delta\pi \geq \frac{\Delta_I - (1 - \frac{m}{n})(W_H + \tau_H)}{\Delta_S}$$

as required. □

12. Complete Contracting

1. *The One-Period Case.* For convenience, we sketch the analysis in the benevolence case since cost is independent of politician type, and assume two regions. We begin with decentralization, so there are two agents (incumbents), one in each region. Consider the contracting problem between the voters and incumbent in one region. This is not a standard problem because *both* the cost of producing the public good and the preferences of the incumbent are unknown. In this type space, the standard single-crossing condition does not hold[24]. Nevertheless, it has relatively simple solution.

The voters contract with an incumbent randomly drawn from the population, so that the probability that the incumbent is good is π. At the time of contracting, the incumbent has also observed the cost realization $c \in \{c_H, c_L\}$. The incumbent makes an announcement of his type $\hat{\sigma} \in \Sigma^D = \{c_H, c_L\} \times \{B, N\}$ where B and N denote good and bad incumbent respectively. Conditional on $\hat{\sigma}$, voters offer a public good and tax combination

[24]That is, the elements of the type space cannot be ordered so that the marginal rate of substitution of the incumbent between g and τ everywhere monotonic in the type variable.

$(g(\hat{\sigma}), \tau(\hat{\sigma}))$ which the incumbent must choose (we assume that if any other g, τ is chosen, the voters can impose a large enough penalty to deter this behavior). The incumbent must pay for the public good out of τ, and there is a limited liability constraint that the remainder $r = \tau - cg$, which can be interpreted as performance-related pay, is non-negative. The voters have a payoff $H(g) - \tau$, the benevolent incumbent gets a payoff equal to the voter's payoff plus performance-related pay, $H(g) - \tau + r = H(g) - cg$, and the non-benevolent incumbent gets $r = \tau - cg$. So, letting $\sigma = (c, \theta)$, where $\theta \in \{B, N\}$, incumbent payoff can be compactly written

$$u(\tau, g, \sigma) = \begin{cases} H(g) - cg, & \theta = B \\ \tau - cg & \theta = N \end{cases}$$

The contract is *incentive-compatible* if

$$u(g(\sigma), \tau(\sigma), \sigma)) \geq u(g(\sigma'), \tau(\sigma'), \sigma), \sigma' \neq \sigma. \tag{12.1}$$

The optimal contract $(g^D(\sigma), \tau^D(\sigma))$ is the one that maximizes expected voter utility[25]

$$\sum_{\sigma \in \Sigma^D} p(\sigma)(H(g(\sigma)) - \tau(\sigma)) \tag{12.2}$$

over the set of incentive-compatible contracts with the limited liability $\tau \geq cg$ constraint also imposed.

The optimal contract is of the following form. Let (g_L^*, τ_L^*), (g_H^*, τ_H^*) solve

$$\begin{aligned} H'(g_H^*) &= c_H + \frac{(1-\pi)(1-q)}{q}[c_H - c_L],\ \tau_H^* = c_H g_H^* \\ H'(g_L^*) &= c_L,\ \tau_L^* = c_L g_L^* + r^*,\ r^* = [c_H - c_L] g_H^* > 0 \end{aligned}$$

Also, we assume that $H(g_H^*) - g_H^* c_H > H(g_L^*) - g_L^* c_H$. Then, the optimal contract has the following form. If $\theta = N$, and $c = c_k$, $(g(\sigma), \tau(\sigma)) = (g_k^*, \tau_k^*), .k = H, L$. If $\theta = B$, $c = c_H$, $(g(\sigma), \tau(\sigma)) = (g_H^*, \tau_H^*)$, and finally[26], if $\theta = B$, $c = c_L$, $(g(\sigma), \tau(\sigma)) = (g_L^*, \tau_L^* - r^*)$

The distortion of the quantity demanded (g_H^*) from the high-cost government trades off the benefit of reducing the rent that must be paid to a bad government with low-cost that occurs with probability $(1 - \pi)(1 - q)$ against the cost of imposing the distortion

[25] Noting also $p(c_H, B) = \pi q$, $p(c_L, B) = \pi(1 - q)$, etc.

[26] Note that we cannot offer a contract with efficient public good provision to the benevolent type with high-cost for otherwise the non-benevolent with low-cost would prefer that contract to the contract (g_L^*, τ_L^*) as it gets higher rent when the quantity demanded of public good to high-cost government is not distorted.

of the quantity on the high-cost government that occurs with probability q.[27]This contract is clearly incentive-compatible. The benevolent type can do no better than pick (g_L^*, τ_L^*) when the cost is c_L, since this is the efficient level of public good (he is indifferent about τ), and (g_H^*, τ_H^*) when cost is c_H, as long as $H(g_H^*) - g_H^* c_H > H(g_L^*) - g_L^* c_H$, which is assumed. When $c = c_L$, the non-benevolent incumbent clearly does better choosing (g_L^*, τ_L^*) rather than $(g_L^*, \tau_L^* - r^*)$, as he gets some informational rent $r^* > 0$. Finally, by construction, (g_H^*, τ_H^*) is preferred to (g_L^*, τ_L^*) by the non-benevolent type when the cost is high and vice-versa[28].

It is also clear that this contract maximizes voter utility subject to the limited liability and incentive-compatibility constraints. First, the revelation is obtained at the least cost by distorting the quantity demanded from the high-cost government. The benevolent type is getting zero pay and choosing either the distorted level of public good provision when cost is high or the efficient level when cost is low. Second, the non-benevolent type is getting the minimum level of pay (informational rent) consistent with incentive-compatibility, and g_H^* is chosen to be second-best efficient given this rent.

Now, turning to centralization, there is one agent (incumbent). So, now the space of types is $\sigma \in \Sigma^C = \{c_H, c_L\}^2 \times \{B, N\}$, and conditional on a declared type $\hat{\sigma} \in \Sigma^C$, the voters choose a pair $(g^i(\hat{\sigma}), \tau^i(\hat{\sigma}))_{i=1,2}$. The payoffs to voters and incumbent are defined analogously to the decentralized case: for example, the voters in i have payoff $H(g^i) - \tau^i$, .the good incumbent gets $\sum_{i=1,2} H(g^i) - c^i g^i$, and the bad incumbent gets $\sum_{i=1,2} \tau^i - c^i g^i$. We assume that the objective of the voters is to maximise $\sum_{i=1,2} H(g^i) - \tau^i$.

For convenience[29], restrict the principal (the voters) to symmetric incentive schemes i.e. $g^i(c^i, c^j, \theta) = g^j(c^j, c^i, \theta)$, etc.. Also, note that because costs are independent across regions, voters cannot gain by conditioning $g^i(c^i, c^j, \theta)$ on c^j and vice versa. [If costs were correlated, then this would not be the case see e.g. Demski and Sappington(1984)].

Then, the relevant type space in each region is simply $\Sigma^D = \{c_H, c_L\} \times \{B, N\}$, so in each region, the problem with centralization reduces to maximising (12.2) subject to (12.1). This is of course, the contract design problem under decentralization, so in the one-period case, the two are equivalent.

2. *The Two-Period Case.* We consider the two-period contracting problem without precommitment. With either fiscal regime, in the second period, the voters offer a contract

[27]The extent of the distortion decreases as the probability that the government is good increases.

[28]In fact, as is standard, the constraint that the low-cost non-benevolent incumbent does not wish to imitate the high-cost non-benevolent incumbent is binding.

[29]This is without much loss of generality - under some conditions, an asymmetric scheme may be better, but this is not the main issue.

conditional on π', there second-period belief that the incumbent is good. Note that the one-period contract is dependent on π,only through the distortion of the quantity demanded of the public good from a high cost government; so voters will offer the one-period contract in the second period modulo the use of the updated belief in the second-period. At the end of the first period, voters will fire the bad incumbent if he reveals himself.

Now suppose that voters offer the optimal one-period contract in the first period. Consider decentralization first. The good incumbent will always choose as in the one-period case. Note that if the bad incumbent chooses (g_L^*, τ_L^*) rather than $(g_L^*, \tau_L^* - r^*)$, he will reveal his type and be fired. If cost is low, this strategy gives him r^*, whereas choosing (g_L^*, τ_L^*) gives him re-election and a payoff of $\delta(1-q)r^*$. So if cost is low, he will choose (g_L^*, τ_L^*) as in the one-period case. Now if cost is high, choosing (g_H^*, τ_H^*) or (g_L^*, τ_L^*) gives re-election; looking only at first-period payoffs, the bad incumbent strictly prefers (g_H^*, τ_H^*),and so this is what he will choose. So, the bad incumbent will choose as in the one-period case, revealing his type when cost is low, and being fired in this event. So, voters cannot do better than offer the one-period contract in the first period.

Now consider centralization. In all four cost states, the good incumbent will choose as in the one-period contract. So, if the bad incumbent chooses (g_L^*, τ_L^*) rather than $(g_L^*, \tau_L^* - r^*)$ in either region he will be fired. This strategy gives him either r^* or $2r^*$ (in states LH/HL or LL respectively), whereas choosing $(g_L^*, \tau_L^* - r^*)$ gives him re-election and a payoff of $2\delta(1-q)r^*$. As $q > 0.5$ by assumption, $2\delta(1-q) < 1$, so that the separating strategy of choosing (g_L^*, τ_L^*) is always best. So if cost is low in a region, he will choose (g_L^*, τ_L^*) as in the one-period case. Again if cost is high, choosing (g_H^*, τ_H^*) is preferred to choosing (g_L^*, τ_L^*) by the above argument. So, the bad incumbent will also choose the same as in the one-period case .So, again, voters cannot do better than offer the one-period contract in the first period. But then it is easy to see that voter welfare is the same under both centralization and decentralization with the optimal two-period contract

Acknowledgements. We are grateful to the participants of the conference on "Political Economy and Institutional Performance" (Brussels, June 11-12, 2004), the Workshop on "Social Choice and Welfare Economics" (Malaga, June 2-4, 2005), the PET 2005 conference (Marseilles, 16-18 June 2005) and seminar participants at Barcelona, Bristol, Essex, Warwick, Nottingham, Milan, Lisbonne and Louvain for their comments. We would also like to thank Massimo Bordignon, Micael Castanheira, Kai Konrad, Nicolas Porteiro, Christian Schultz and Jean Tirole for helful comments on a previous draft.

13. References

References

[1] Alesina, A. and E. Spolare (1997), "On the Number and Size of Nations," *Quarterly Journal of Economics*, **112**, 1027-56.

[2] Azfar, O., S. Kahkonen, and P. Meagher (2001) "Conditions for Effective Decentralised Governance: A Synthesis of Research Findings", IRIS Center, University of Maryland.

[3] Bardhan, P. and D. Mookerjeee (2000), "Capture and Governance at Local and National Levels," *American Economic Review*, **90**(2), 135-139.

[4] Baron, D. and D. Besanko (1992), "Information, Control and Organizational Structure", *Journal of Economics and Management Strategy*

[5] Banks, J.S. and R.K. Sundaram (1998), "Optimal Retention in Agency problems", *Journal of Economic Theory* **82**, 293-323.

[6] Besley, T. and S. Coate (2003), "Centralized versus Decentralized Provision of Local Public Goods: a Political Economy Approach", *Journal of Public Economics*, **87**, 2611-37.

[7] Besley, T. and M. Smart (2003), "Does Tax Competition Raise Voter Welfare?".mimeo, London School of Economics.

[8] Beaudry, P. and M. Poitevin (1995), "Contract Renegotiation: a Simple Framework and Implications for Organization Theory", *Canadian Journal of Economics*, **28**, 302-35.

[9] Bolton, P. and G. Roland (1997), "The Break-Up of Nations: A Political Economy Analysis" *Quarterly Journal of Economics*, **112**, 1057-90.

[10] Cremer. J. and T.Palfrey, (1996), "In or Out? Centralization by Majority Vote", *European Economic Review* **40**, 43-60.

[11] J.S.Demski and M.Sappington (1984) " Optimal Incentive Schemes with Multiple Agents", *Journal of Economic Theory* **33**, 152-171.

[12] Ellingsen, T.(1998), "Externalities vs. Internalities: A Model of Political Integration", *Journal of Public Economics*, **68**, 251-268.

[13] Faguet, J-P, (2004), "Does Decentralization Increase Government Responsiveness to Local Needs? Evidence from Bolivia" *Journal of Public Economics* **88**, 667-893

[14] Ferejohn, J. (1986), Incumbent performance and electoral control, *Public Choice* **50**, 5-25.

[15] Finnoccario, C.J. and T.Lin(2000), "The Hazards of Incumbency: An Event-History Analysis of Congressional Tenure", unpublished paper, University of Michigan

[16] Fissman, R., and R.Gatti (2002) "Decentralization and Corruption: Evidence Across Countries," *Journal of Public Economics,* **83**(3), 325-45.

[17] Fissman,R. and R.Gatti(2002a), "Decentralization and Corruption: Evidence from US Transfer Programs", *Public Choice*, 113, 25-35

[18] Gilbert, G. and P.Picard (1996), "Incentives and the Optimal Size of Local Jurisdictions", *European Economic Review* **40**, 19-43.

[19] Henderson, J.V, and A.Kuncoro (2004) "Corruption in Indonesia", NBER Working Paper W10674

[20] Huther, J. and A.Shah (1998) "Applying a Simple Measure of Good Governance to the Debate on Fiscal decentralization", World Bank Policy Research Working Paper 1894.

[21] Laffont, J.-J., and D. Martimort (1998), "Collusion and Delegation", *Rand Journal of Economics*, **29**, 280-305.

[22] Laffont, J.-J., and J. Tirole (1998), "The Dynamics of Incentive Contracts"", *Econometrica*, **56**, 1153-75.

[23] Lockwood, B., (2002), "Distributive Politics and the Costs of Centralization", *Review of Economic Studies*, **69**(2), 313-337.

[24] Lockwood, B., (2005), "A Note on the Hybrid Equilibrium in the Besley-Smart Model"", Warwick Economics Working Paper 727.

[25] Maskin, E., Y. Quian and C. Xu (2000), "Incentive, Information and Organizational Form", *Review of Economic Studies* **67** (2), 359-78.

[26] McKinnon, R., T. Nechyba (1997), "Competition in Federal Systems: Political and financial Constraint", in J. Ferejohn and B. R. Weingast, eds., *The New federalism: Can the States be Trusted*? Stanford: Hoover Institution Press.

[27] Mello,L. and M.Barenstein (2001), "Fiscal Decentralization and Governance: A Cross-Country Approach", IMF Working Paper 01/71.

[28] Melumad, N., D. Mookherjee, and S. Reichelstein (1997), "Contract Complexity, Incentives, and the Value of Delegation", *Journal of Economics and Management Strategy*, **6**, 257-89.

[29] Myerson, R. (1982), "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems" *Journal of Mathematical Economics* **10**, 67-81.

[30] Myerson, R. (1993), "Effectiveness of Electoral Systems for Reducing Government Corruption: a Game Theoretic Analysis", *Games and Economic Behavior* 5, 118-32.

[31] Myerson, R. (1999), "Theoretical Comparisons of Electoral Systems" *European Economic Review* 43, 671-97.

[32] Oates, W. (1972), *Fiscal Federalism,* Harcourt-Brace, New York.

[33] Oates, W (1999), "An Essay on Fiscal Federalism," *Journal of Economic Literature,* 37(3), 1120-49.

[34] Oberholzer-Gee, Felix and Koleman S. Strumpf (2002): "Endogeneous Policy Decentralization: Testing the Central Tenet of Economic Federalism," *Journal of Political Economy, 110, 1-36.*

[35] Persson, T. and G. Tabellini (1999), The Size and the Scope of Government: Comparative Politics with Rational Politicians, *European Economic Review* 43, 699-735.

[36] Persson, T. and G. Tabellini (2000), *Political Economics: Explaining Economic Policy*, MIT Press, Cambridge, MA..

[37] Persson, T., G. Tabellini, and F. Trebbi (2000), Electoral Rules and Corruption" NBER Working Paper No. 8154.

[38] Rogoff, K. (1990), Equilibrium political budget cycles, *American Economic Review* 80: 21-37.

[39] Seabright, P. (1996), "Accountability and Decentralization in Government:an Incomplete Contracts Model", *European Economic Review* **40**, 61-91.

[40] Treisman, D, (2000), "The Causes of Corruption: A Cross-National Study", *Journal of Public Economics*, 76: 399-457.

[41] Treisman, Daniel (2002), "Decentralization and the Quality of Government", unpublished paper, Department of Political Science, UCLA.

[42] Wallis, J. and W.E.Oates (1988), "Decentralization in the Public Sector: an Empirical Study of State and Local Government", in Rosen, H.J. (ed.), in *Fiscal Federalism: Quantitative Studies*, University of Chicago Press.

[43] World Bank (2000), *Entering the 21st Century*, World Development Report.

Département des Sciences Économiques
de l'Université catholique de Louvain
Institut de Recherches Économiques et Sociales

Place Montesquieu, 3
1348 Louvain-la-Neuve, Belgique

ISSN 1379-244X D/2005/3082/038