Tax Bloc Formation among Leviathans

S. Bucovetsky

York University

Toronto Ontario M3J 1P3 CANADA

sam@dept.econ.yorku.ca

November 2005

Introduction

If countries must use source–based taxes on an internationally mobile tax base, then the Nash equilibrium is inefficient when they set tax rates non–cooperatively. [1] One country's tax increase tends to benefit all other countries (by increasing their tax bases), so that there is an externality. The Nash equilibrium will be Pareto–dominated by some arrangement in which countries coordinate policy, taking into account these externalities.

Why then do countries not better coordinate their tax policies? There are, of course, many difficulties in policy coordination. Individual governments will have incentives to deviate unilaterally from agreements signed by sovereign nations. Enforcement of these agreements may be difficult, because a country's compliance requires outsiders monitoring its own tax collection.

The costs of monitoring may themselves be affected by countries' policies. Sharing of information by the tax authorities in different countries makes monitoring easier. That does not mean all countries will agree to share such information, or that they all will

[1] Wilson (1986) and Zodrow and Mieszkowski (1986) are two important papers presenting this sort of model.

honour such agreements in practice. The European Union's difficulties in forcing member countries' banks to share information are perhaps typical.

There are other reasons why countries may not agree to make coordination easier. It may be that simple models of tax competition have misspecified the countries' payoff functions. The countries' policy makers may be influenced by owners of the mobile tax base, who are interested in keeping tax competition fierce. [2] Politicians may seize part of the tax revenue for themselves, so that voters may want tax competition, in order to discipline such behaviour. [3]

It may be that side payments are difficult to make among countries. As Wilson (1991) demonstrates, small countries may do better under unfettered tax competition than they would under complete harmonization. [4] Despite the enthusiasm of the larger, more powerful countries in the European Union for tax harmonization, statutory corporate income tax rates in the EU have not been harmonized completely.

Of course there also does not seem to be much compelling evidence of a race to the bottom in effective corporate tax rates. The work of Devereux et al (2002), for example, suggests that simple models of capital tax competition do not describe very well the recent European experience. Nonetheless, significant differences across countries in effective source–based tax rates on capital income do seem to persist.

In this paper, I examine the stability of cooperative agreements among countries, in a very simple model of tax competition. One of the assumptions underlying the model is

---

[2] Such fierce competition may be efficient in the long run, as Kehoe (1989) demonstrates.

[3] See, for example, Edwards and Keen (1996), Janeba and Schjelderup (2002), or Wilson and Wildasin (2004)

[4] See also Peralta and van Ypersele (2003) for an analysis of gains and losses from harmonization without side payments.

that countries actually do want to tax the return for capital. That is, I assume a payoff function which implies that countries' governments will benefit from cooperation to reduce tax competition. The problem will be how to agree to cooperate.

The main question addressed here is what sort of coalitions will form. This is a question addressed by Burbidge et al (1997), who developed a fairly general model of tax competition to illustrate the incentives for coalition formation. Here, I use very restrictive functional forms to get a much more specific model, one which can be solved in closed form for the Nash equilibrium to the (final stage of the) tax competition game. Burbidge et al looked seriously at the negotiation process involved in coalition formation. Here, in contrast, several axiomatic notions of stability are applied.

As well, Burbidge et al allowed for side payments among countries in a coalition. Here, instead, if a group of countries choose a common tax rate then each country gets exactly the tax revenue collected within its borders. For that reason, I refer to these groups of countries as "blocs", since "coalitions", in much of the literature, are able to divide their aggregate payoffs arbitrarily among the members. The fixed sharing rule imposed here may be more realistic, if commitment is difficult. In some sense it makes coordination more difficult, since the small countries which benefit most from tax competition cannot be bribed to join a bloc. But flexible sharing rules also create the potential for holdout problems.

Burbidge et al examined under what circumstances a coalition of the whole could be sustained through negotiation. By using a much more simple and restrictive model, I can extend their analysis here. All possible stable bloc structures will be derived. Sometimes the coalition of the whole will not be stable, but some other bloc structure will.

In other words, the gains from free riding may be high enough that a single small country cannot be induced to coordinate taxes with the rest of the countries in the world. I will show that the only possible stable structure in such a case consists of two blocs : one large bloc, and one small bloc which benefits from the high tax rates set by the larger bloc.

Reference to the literature on coalition formation in other areas of economics may be useful. Consider first the very basic Cournot model, in which firms have identical constant–return–to–scale production technologies, and in which the aggregate demand function for the homogeneous product is linear. Suppose the industry starts out with N ($N > 2$) of these identical firms, and that no further entry by new firms is possible. Suppose further that the only way for firms to coordinate policies is to merge. Then there is no incentive for any two firms to merge. [5] A merger of any two of the original N firms will result in increased profits in the new Cournot–Nash equilibrium for each of the other $N - 2$ firms, but the profits of the newly–merged firm will be less than the sum of the profits earned in the original N–firm equilibrium by the two merging firms. Because of this, any number of identical firms (except 2) is a stable industry structure, in that no two firms have the incentive to merge.

The above conclusion is certainly not terribly robust. One of the peculiarities of this simple Cournot model is that there is really no notion of firm size, since there are constant returns to scale. That means there will be no difference between a "big" firm resulting from the merger of 2 of the original identical firms, and any of the remaining $N - 2$ un–merged firms. All produce the same share of industry output in equilibrium. If there were capacity constraints on each firm, then merger may induce an asymmetry : the firm resulting from merger would have twice the capacity of the other firms. If this capacity (of each of the other firms) were less than the per–firm output in the $N - 1$–firm Cournot–Nash equilibrium when capacity constraints were ignored, then it is possible merger may be profitable. [6]

---

[5] See Salant, Switzer and Reynolds (1983)

[6] See Perry and Porter (1985).

Size does seem important in tax competition. First of all, it appears that many tax havens are very small countries. Hansen and Kessler (2001) and Wilson (1991) provide explanations of this phenomenon. Second, the notion of market power seems to make some sense in tax competition. If a bloc of countries contains a non–trivial fraction of the world's stock of some tax base, then the bloc should take into account the effect of its own policies on the world return to that base. In other words, unlike the basic Cournot model, in tax competition the number of countries is not the only measure of the competitiveness of the situation. What matters as well is the share of each country (or bloc of countries) in the world's stock of the tax base.

Of course, the whole point of the tax competition literature is that a country's tax base is endogenous. So a more appropriate measure of the relative size of a country might be the share it would have of the tax base, if all countries levied the same tax rate on that tax base.

In this paper, countries are assumed to have identical technologies. The only difference among them is in their size. Size here is identified as the stock of an immobile input, which is combined with the mobile tax base to produce output, under constant returns to scale. [7] Thus if all countries taxed the mobile factor at the same rate, then its gross return would be the same in every country, so that the constant returns imply a country's share of the mobile tax base would be proportional to its share of the immobile input, its relative size.

There are three purposes of this paper. One is to examine how the size distribution of countries affects the equilibrium tax rates. The specificity of the simple model used

[7] Capital movement in this model thus should be interpreted as movement of physical capital. Although I talk about small, low–tax countries as "tax havens", it is essential in the model that small countries be severely constrained in how much capital inflow they can absorb. It is not obvious that the flow of financial capital to offshore tax havens really is that constrained in aggregate by the lack of complementary resources in the tax havens.

leads to a fairly precise answer to this question. Here an index can be calculated for how much tax competition is going on, based entirely on the size distribution. This index is analogous to concentration indices used in industrial organization, and shares their general properties. The net return to the mobile factor in equilibrium is a monotonic function of this index.

The second purpose is to examine the effects of policy coordination among countries, or among blocs of countries. I consider a very crude form of policy coordination, equivalent to a merger of the countries (or blocs of countries). The simple Cournot model is an example in which merger by two firms makes the two firms worse off, and makes all other firms better off. The theory of customs unions provides examples in which policy coordination among two countries may benefit those two countries while harming the rest of the world. In the model developed here for tax competition, merger will benefit all other countries, but it also must benefit at least one of the merging countries.

The third purpose is to examine countries' incentives to coordinate policy, as in Burbidge et al. Since side payments between parties to a merger are not allowed here, for any structure of blocs of the different countries, the payoff to a single country is determined completely by the Nash equilibrium to the non–cooperative tax–setting game played by the blocs. I assume that countries, and blocs of countries, will merge only if the merger benefits all parties. A coalition structure is considered stable only if there are no further mutually–beneficial bilateral mergers. This is the notion of stability underlying the literature on mergers in Cournot competition. Under fairly weak assumptions, the only stable bloc structures are those containing one or two blocs.

This is a somewhat different result than that obtained for customs unions. There the question has been asked whether division of the world's countries into a few large trading blocs can be considered an intermediate step on the way to global free trade. [8] The

[8] See Kemp and Wan (1976), or Riezman (1985), for example.

answer need not be yes. But in this model of tax competition, the incentives to merge are quite strong. As well, there may be incentives for countries to secede from large blocs. Nonetheless, the "likely" outcome requires the majority of the population to coalesce, either into a single bloc, or into two not–too–similar blocs.

2. The Model

As in Burbidge et al, tax rates are assumed to be set non–cooperatively in a game played by groups of countries. It is assumed that the world has been divided into N countries, and that this initial division is an exogenous datum. What I will refer to as a "bloc structure" is a partition $\mathcal{P}$ of the countries. The players in the non–cooperative tax–setting game are the elements of the partition. So, for example, if $N = 5$, then a possible bloc structure is a partition $\mathcal{P} = (\{1, 2\}, \{3, 4\}, \{5\})$ of the 5 countries into three blocs. For this bloc structure, there would be three players in the tax–setting game. I will denote by $M \leq N$ the number of blocs. Of course M depends on the particular partition which countries have chosen to form.

The technologies and residents' preferences are the same in each of the N countries. The only aspect in which countries — or blocs of countries — may differ from each other is in their populations. I will denote by s_i the share of the world's population living in bloc i. This share is just the sum of the population shares of the individual countries in the bloc.

In each country, capital and labour are used to produce homogeneous output, under constant returns to scale. It is assumed that capital is perfectly mobile among countries, and that labour is perfectly immobile. Labour is in fixed supply, one unit per person.[9]

Let k_i denote the capital–labour ratio in bloc i. The total world capital supply is assumed fixed. The capital–labour ratio for the world as a whole is denoted $\bar{k}$. Although the total capital supply is assumed fixed, I also assume that capital owners can withhold

---

[9] The fact that labour is immobile within countries comprising a bloc does not affect production within the bloc. Since all countries within a bloc are required to levy the same capital tax rate, they all have the same gross return to capital. Capital will move among the countries of a bloc so that the capital–labour ratio is the same in all countries within the bloc.

their supply if the net return is negative. [10] The perfect mobility of capital means that there is one net return to capital, denoted ρ, for the whole world. Then the allocation of capital to the different blocs must satisfy

$$\sum_i s_i k_i \leq \bar{k} \qquad \rho \geq 0 \tag{1}$$

with complementary slackness between the two inequalities.

Each bloc i has one fiscal instrument, a tax on the capital employed within the bloc, at the rate of t_i per unit of capital employed. Thus the tax revenue per person collected in bloc i is

$$\pi_i \equiv t_i k_i$$

One crucial assumption is that π_i is the payoff to each resident of each country in the bloc.

ASSUMPTION 1 : Each resident seeks to maximize the (source–based) capital tax revenue per person in the country in which she resides.

This assumption has been used by Kanbur & Keen (1993), among others. They suggest two motivations for it. One is that local government is controlled by Leviathan, bureaucrats who seize the entirety of the public treasury as rents. This first interpretation of assumption 1 would apply to blocs of countries if it is understood that in a federation of several countries, bureaucrats from each country receive rents in proportion to their countries' share in the federation's population. The second interpretation suggested by Kanbur & Keen is that the source–based capital tax is the only source of public sector revenue, and that the shadow price of public output is very high relative to the shadow price residents place on private consumption. In the limit, as the relative shadow price

[10] so that the world supply curve for capital is "backwards–L" shaped, with a horizontal portion at a net rate of return of 0. Laussel and Le Breton (1998), for example, make this sort of assumption.

approaches infinity, total tax revenue is all that concerns residents. For this interpretation to fit here, it must be assumed that there are constant returns to population in provision of the public output. [11]

Each bloc's strategic variable is its source–based capital tax rate t_i. Each bloc of countries acts non–cooperatively, taking as given the tax rates levied in the other blocs. The net return to capital — which is equalized across countries — is the gross return to capital in the country's bloc, minus its tax rate. The second strong simplifying assumption is that the gross rate of return to capital R_i is a linear function of the capital–labour ratio.

ASSUMPTION 2 : $R_i = a - bk_i \qquad a, b > 0 \qquad a > 2b\bar{k}$

(The assumption that $a > 2b\bar{k}$ serves to ensure that no capital will be withheld in equilibrium.)

Assumption 2, and the assumption of perfect mobility of capital, imply that

$$\rho = a - bk_i - t_i \qquad i = 1, \cdots, M \tag{2}$$

For any given set of tax rates, equation (2) and inequality (1) determine the endogenous allocation of capital to blocs of countries. If no capital is withheld (that is, if the first inequality in (1) is not strict), then

$$k_i = \bar{k} + \frac{1}{b}[\sum_{j=1}^{N} s_j t_j - t_i] \tag{3}$$

[11] Many of the results would continue to hold even if the Leviathan assumption is relaxed somewhat. For example, each country's government might maximize a weighted sum of the wage income of residents, and of the tax revenue. The relative weight on the tax revenue is the shadow price of public expenditure. As long as that shadow price is constant, and is the same across all countries, then the costs and benefits of policy coordination are similar to those obtained in the paper.

From equation (3),

$$\frac{\partial k_i}{\partial t_i} = -\frac{1 - s_i}{b} \qquad \text{if} \quad \rho > 0 \tag{4}$$

If the first inequality in (1) were strict, so that there were excess supply of capital, then each bloc's k_i would satisfy

$$a - bk_i - t_i = 0 \quad i = 1, \cdots, M \tag{5}$$

so that each bloc's capital supply would be independent of other blocs' tax rates. In this regime

$$\frac{\partial k_i}{\partial t_i} = -\frac{1}{b} \qquad \text{if} \quad \sum s_i k_i < \bar{k} \tag{6}$$

The weighted average of all blocs' tax rates plays a major role here, so it will be convenient to denote

$$\bar{t} \equiv \sum_{i=1}^{M} s_i t_i$$

(Note that, unlike $\bar{k}$, here $\bar{t}$ is endogenous.)

Each bloc chooses its t_i so as to maximize its payoff $\pi_i = t_i k_i$ subject to the rules (inequality (1) and equation (2)) for the allocation of capital. If the net return to capital is positive, so that equation (4) holds, then bloc i's best reaction to the tax rates chosen by the other blocs obeys

$$t_i = \frac{1}{2 - s_i}[b\bar{k} + \bar{t}] \tag{7}$$

The system of equations (7) can also be written

$$A\mathbf{t} = \beta$$

where A is an M–by–M matrix with elements $2(1 - s_i)$ along the diagonal, and $-s_j$ in off–diagonal element ij, where $\mathbf{t}$ is the vector of equilibrium tax rates, and where β is a vector with $b\bar{k}$ in every element. If there are at least two blocs in the partition, each with a positive share of the population, then $s_i < 1$ for each bloc i. That means that the matrix A has a dominant diagonal. Since dominant diagonal matrices have non–negative–valued inverses, then the equation $A\mathbf{t} = \beta$ can be solved uniquely for the vector $\mathbf{t}$ of tax rates. Therefore, the following lemma holds.

LEMMA 1: For any partition with at least two blocs, there is at most one Nash equilibrium in tax rates in which the net return to capital is positive.

The solution to equations (7) may imply a negative net return to capital ρ. (This will occur if and only if one bloc has a large enough share of the population.) However, it can be shown that if the solution to (7) implies a non–negative net return to capital, then that solution represents the unique Nash equilibrium (for the given size distribution of blocs). That is,

LEMMA 2: For a given bloc structure, if there is a Nash equilibrium with a positive net return to capital, then there is no other Nash equilibrium.

PROOF : see the appendix •

When one bloc has a large enough share of the population, [12] the solution to (7) may not constitute a Nash equilibrium, since it implies a negative net return to capital. In that case, there will be Nash equilibria with a zero net return to capital. The equilibria need not be unique in this case. However, they all imply no withholding of capital.

---

[12] Or when a is not very large.

LEMMA 3: For any bloc structure, there exists at least one Nash equilibrium. In any Nash equilibrium capital is fully allocated, so that $\sum_{i=1}^{M} k_i = \bar{k}$.

PROOF : see the appendix •

Given a bloc structure, it is straightforward to check whether there is an equilibrium in which the net return to capital is positive.

LEMMA 4 : For a given bloc structure, the net return to capital will be positive in the Nash equilibrium if and only if

$$(1-\zeta)a > b\bar{k}$$

where

$$\zeta \equiv \sum_{i=1}^{M} \frac{s_i}{2-s_i}$$

PROOF : Suppose that there is a Nash equilibrium in which the net return to capital is positive. Equation (7) implies that

$$s_i t_i = \frac{s_i}{2-s_i}[b\bar{k} + \bar{t}] \tag{8}$$

Adding up (8) over all M blocs yields

$$\bar{t} = \frac{\zeta}{1-\zeta} b\bar{k} \tag{9}$$

From equations (2) and (3), if the net return to capital is positive, then

$$\rho = a - b\bar{k} - \bar{t} \tag{10}$$

Whenever $\rho > 0$ in equilibrium, then equations (9) and (10) must hold, implying that $(1-\zeta)a > b\bar{k}$.

Conversely, if $(1-\zeta)a > b\bar{k}$, then a Nash equilibrium exists in which $\bar{t} = \frac{\zeta}{1-\zeta} b\bar{k}$, $t_i = \frac{b\bar{k}+\bar{t}}{2-s_i}$ for which $\rho = a - b\bar{k} - \bar{t} > 0$. •

If $(1-\zeta)a < b\bar{k}$, then there will be Nash equilibria in which the net return to capital is driven down to zero. Suppose that this were the case, for some bloc structure with $M \geq 2$ blocs. In some sense, all the gains from cooperation will have been exhausted even without a coalition of all the countries. In fact, if $M > 1$, and if $\rho = 0$ in equilibrium, then any bloc levying a below–average tax would have a higher payoff than it would earn if all countries merged into a single bloc.

Worse, whenever $(1-\zeta)a < b\bar{k}$, there will be a continuum of equilibria. Equations (4) and (6) show that $\partial\pi_i/\partial t_i$ is discontinuous at $\rho = 0$. In particular, any set of t_i's such that $a/2 \leq t_i \leq a/(2-s_i)$ will constitute an equilibrium, provided that the tax rates average to $a - b\bar{k}$.

So the equilibrium payoff to a bloc is not uniquely defined when $(1-\zeta)a < b\bar{k}$. The range of equilibrium payoffs to a bloc may include payoffs which are less than the payoff $(a - b\bar{k})\bar{k}$ attainable by the coalition of the whole, and payoffs which are greater than this. [13]

Therefore, I would like to restrict attention to bloc structures for which the net return is not driven down to 0. Of course, this cannot be the case if $M = 1$: then $\zeta = 1$. (A monopoly Leviathan would always seize all the rents from absentee capital owners.)

The following assumption ensures that the net return is positive for any bloc structure with two or more blocs :

---

[13] For instance, this will be the case for either bloc, if $a = 2.4$, $b = 1$, $\bar{k} = 1$, $M = 2$, and $s_1 = 0.6$.

ASSUMPTION 3 : The population share s_N of the smallest country is large enough so that

$$\frac{3s_N(1-s_N)}{(2-s_N)(1+s_N)}a \geq b\bar{k}$$

The left side of the above inequality equals 0 when $s_N = 0$, and is an increasing function of s_N (when $0 \leq s_N \leq 0.5$), so that, if a is sufficiently large, there is a unique positive minimal s_N defined by the above inequality.

The result that Assumption 3 ensures that $a(1-\zeta) > b\bar{k}$ whenever $M > 1$ is proved in the corollary to Proposition 3, in the following section.

3. An Index of Tax Competition

A consequence of equation (7) is the following proposition

PROPOSITION 1: In any Nash equilibrium in which the net return to capital is positive, then the equilibrium tax rates are monotonically increasing with the size of the bloc levying the tax.

This result, that small jurisdictions tend to be tax havens, has been derived elsewhere in a variety of contexts.[14]. It is also true here[15] that smaller jurisdictions are better off.

To see that the payoff is a declining size of a bloc's population share s, first define the function $\alpha(s_i)$ as

$$\alpha(s_i) \equiv \frac{1}{2 - s_i}$$

(Where it does not lead to confusion, $\alpha(s_i)$ will be shortened to α_i.) When the return to capital is positive, equations (3) and (7) imply that

$$\pi_i = \frac{(b\bar{k} + \bar{t})^2}{b} \alpha_i (1 - \alpha_i) \tag{11}$$

It then follows that

PROPOSITION 2: In any Nash equilibrium in which the net return to capital is positive, the payoff (tax revenue per capita) is a monotonically decreasing function of bloc size.

PROOF : The expression $\alpha(s_i)$ is a monotonically increasing function of s_i. Suppose there is some given size distribution of blocs of countries, resulting in a Nash equilibrium

[14] for example, Wilson (1991), Kanbur and Keen (1993) and Hansen and Kessler (2001)

[15] as in Wilson (1991) and in Kanbur and Keen (1993)

with some value of $\bar{t}$, in which the net return to capital is positive. Equilibrium payoffs will increase with a bloc's size in this equilibrium, if and only if π_i, as defined in equation (11), is an increasing function of α_i. But the partial derivative of π_i with respect to α_i is

$$\left[1 - 2\alpha_i\right] \frac{(b\bar{k} + \bar{t})^2}{b}$$

and $\alpha(s_i)$ ranges from 0.5 to 1 as s_i ranges between 0 and 1. The partial derivative is negative, so that payoffs decrease with a country's size. •

Therefore, in any Nash equilibrium in which the net return to capital is positive, it is the smaller blocs which do best, due to the greater increase in their payoffs induced by tax reductions (as shown by equation (4)).

The index ζ was defined in the previous section :

$$\zeta \equiv \sum_{i=1}^{M} \frac{s_i}{2 - s_i}$$

where it was also shown, in equation (9), that the average tax rate in the Nash equilibrium was an increasing, convex function of ζ.

From equations (9) and (11), the payoff to a bloc, in the Nash equilibrium, is proportional to

$$\frac{\alpha_i(1 - \alpha_i)}{(1 - \zeta)^2}$$

The constant of proportionality depends on b and $\bar{k}$, which will not vary with the bloc structure. Therefore, a country's decision whether to join a particular coalition will be determined entirely on the value of that coalition's payoff $\alpha_i(1 - \alpha_i)/[(1 - \zeta)^2]$ from the tax–setting game.

The index ζ depends only on the size distribution of blocs. It has all the properties of a good measure of concentration. It reaches its maximum value under monopoly : if countries coalesced into a grand coalition then ζ would equal 1. The index reaches its minimum under perfect competition : as the number of equally–sized blocs grows arbitrarily large, ζ approaches 0.5. If there are M equally–sized blocs, then $s_i = 1/M$, so that the index ζ would decrease with M.

Most important, any merger will increase ζ.

PROPOSITION 3 : Take any partition $\mathcal{P}$ containing $M \geq 3$ coalitions. Then replace it by a new, coarser partition $\mathcal{P}'$ with $M-1$ coalitions, formed by merging two blocs i and j in $\mathcal{P}$, and leaving all the other blocs in $\mathcal{P}$ unchanged. Then the index ζ, and the weighted average $\bar{t}$ of the tax rates, must be higher for $\mathcal{P}'$ than for $\mathcal{P}$.

PROOF : By definition

$$\zeta = \sum_{i=1}^{M} s_i \alpha(s_i)$$

The expression $\alpha(s_i)$ is an increasing function of s_i. Therefore, if $s_j \geq s_i$, then moving ϵ people from bloc i to bloc j ($\epsilon \leq s_i$) must increase ζ, by

$$\alpha(s_j + \epsilon) - \alpha(s_i - \epsilon)$$

A merger of two blocs i and j (with $s_i \leq s_j$) is equivalent to reducing bloc i's population share by s_i, and increasing bloc j's by s_i. $\bullet$

COROLLARY : Assumption 3 ensures that $\rho > 0$ in the Nash equilibrium for any bloc structure in which there are at least two distinct blocs.

PROOF : Proposition 3 implies that ζ takes its highest value, among all bloc structures with at least 2 blocs, in some structure with exactly 2 blocs.

Suppose then that there were two blocs, so that

$$\zeta = \frac{s}{2-s} + \frac{1-s}{2-(1-s)}$$

where s is the share of population of the smaller of the two blocs. Then $\partial\zeta/\partial s$ is proportional to

$$(2s-1)[(2-s)(1+s)+(s^2-s+1)]$$

which is negative whenever $0 < s < 0.5$.

Therefore, the largest possible ζ (other than for the coalition of the whole) occurs when the bloc structure is $(\{1,2,\dots,N-1\},\{N\})$, where countries are numbered in decreasing order of size.

For this bloc structure,

$$\zeta = \frac{s_N}{2-s_N} + \frac{1-s_N}{1+s_N}$$

implying that

$$1-\zeta = \frac{3s_N(1-s_N)}{(1+s_N)(2-s_N)}$$

Therefore, $a(1-\zeta) > b\bar{k}$ whenever assumption 3 holds. •

4. The Incentives to Merge

By assumption, side payments are not allowed among participants in a merger. The payoff to each resident of a given bloc is assumed the same, the per capita tax revenue obtained in the Nash equilibrium when the bloc competes with all the other blocs. This of course is a very strong simplifying assumption, not made by Burbidge et al, nor in most analyses of cooperative coalition formation. However, it is the appropriate assumption if it is very difficult to make explicit constitutional commitments to some residents of a federation. If coordination of tax policy requires some loss of sovereignty by participating nations, then the majority may be able to renege on any explicit side payments made to the minority. [16] Ruling out side payments is also the usual procedure in the literature on merger in Cournot oligopoly, and is frequently done in the analysis of trading blocs.

Were side payments allowed, then any merger of blocs could be made beneficial to both parties [17]. That is, if blocs i and j merged, and then played as a single bloc in the non–cooperative tax–setting game, then the total payoff of this new merged bloc, in the new Nash equilibrium, would be higher than the sum of the total payoffs to blocs i and j in the Nash equilibrium to the tax–setting game played with each bloc acting as a separate player.

Formally, if $\mathcal{P}$ is some partition of countries into blocs, and if $\mathcal{P}'$ is the partition resulting from a merger of blocs i and j, then

LEMMA 5 : $(s_i + s_j)\pi_{i\cup j}(\mathcal{P}') > s_i\pi_i(\mathcal{P}) + s_j\pi_j(\mathcal{P})$, where $\pi_i(\mathcal{P})$ denotes the Nash equilibrium payoff to player i in the tax–setting game in which the players are elements of the partition $\mathcal{P}$.

---

[16] In this model, it will only be the smaller party to any bilateral merger which may require side payments.

[17] provided the merger did not drive the net return to capital down to zero

PROOF : see the appendix $\bullet$

But since side payments are not allowed, Lemma 5 does not necessarily imply that both parties to any merger must be made better off.

Proposition 3 shows that there are positive externalities from any merger. It shows that the only possible opposition to a merger of blocs i and j would come from those blocs themselves. Equation (11), and Proposition 3, show that to these blocs there are two effects of a merger, one positive and one negative. The merger raises the overall average tax rate $\bar{t}$, which raises overall tax collections. But it also raises the share of the population in the bloc (from s_i to $s_i + s_j$), which tends to lower the payoff.

However, this first effect will always dominate for the larger partner to a merger.

THEOREM 1 : If $s_i \leq s_j$, then bloc j will benefit from a merger with bloc i.

PROOF : Proposition 3 shows that the merger will increase the weighted average $\bar{t}$ in the new Nash equilibrium. From equation (7), then, the equilibrium tax rate t_k in every bloc k not party to the merger will increase.

Now consider the best reaction of the new merged entity of blocs i and j. One feasible strategy for this bloc would be for it to simply choose a tax rate equal to the old t_j. What would be the level of capital per worker in the merged bloc? The tax rate in (the former) bloc j has been left the same. The tax rate in bloc i is now t_j, because merger implies $t_i = t_j$. Proposition 1 shows that $t_i \leq t_j$. The first paragraph of this proof shows that each other t_k will have increased. Therefore equation (3) indicates that the merged bloc would have a higher level of capital per worker (were it to choose the old t_j), than bloc j did in the old equilibrium.

That means choosing $t_{i \cup j}$ equal to the old t_j — in response to the other blocs' new equilibrium strategies — yields a higher payoff to residents of bloc j than they received in the old equilibrium. Since the old t_j is a feasible strategy for the merged bloc, then the merged bloc's best response must give it at least that high a payoff. [18] •

It may not be the case that the smaller party to the merger will be made better off. Note the problem for this bloc is not that the larger bloc can somehow treat the minority worse than the majority in the new merged entity. Once the merger has occurred, a resident's original bloc of origin is irrelevant ; every resident's payoff is the same. The problem is that, in a sense, size brings responsibility. Residents of a big bloc cannot attract capital as easily through tax cuts, since they recognize that it must be attracted from a relatively small rest of the world.

As an example, if $s_1 = 0.55$, $s_2 = 0.4$, and $s_3 = 0.05$, then a merger of blocs 2 and 3 lowers the payoff to residents of the former bloc 3. In fact, if $s_i = 0.05$ and if $s_j = 0.4$, then bloc i will lose from a merger with bloc j, regardless of how the remaining 55 percent of the world's population is partitioned.

It turns out to be the case that "merger leads to merger" in the following sense. Suppose a merger of blocs i and j would increase both blocs' payoff, starting from some partition $\mathcal{P}$. Suppose then that there were some mergers of other blocs (not involving i or j). It must still be the case, after these mergers, that blocs i and j both gain from their own merger.

THEOREM 2 : A merger between blocs k and m will make a merger between blocs i and j more attractive to i and j (where i, j, k and m are all disjoint).

---

[18] This argument also can be used to prove Proposition 3. It shows that the payoff to bloc i, were it to choose a tax rate of t_i and were every bloc other than i to stick with their (pre–merger) equilibrium tax rates, is at least high as the payoff that bloc j gets in the (pre–merger) equilibrium. Since changing t_i to t_j is a feasible policy change for bloc i, its actual equilibrium payoff must be at least as high as the payoff from this deviation.

PROOF : Consider the index of competition ζ, before and after a merger of blocs i and j.

Assume that $s_i \leq s_j$.

If I denote by $\tilde{\zeta}$ the terms in the index of competition from all blocs other than i and j, that is

$$\tilde{\zeta} \equiv \sum_{k \neq i,j} \alpha_k s_k$$

then the index of competition before the merger is

$$\zeta_b \equiv \alpha_i s_i + \alpha_j s_j + \tilde{\zeta}$$

and the index of competition after the merger is

$$\zeta_a \equiv \alpha_{i \cup j}(s_i + s_j) + \tilde{\zeta}$$

The payoff to bloc i before the merger is (up to a constant of proportionality)

$$\pi_b = \alpha_i(1 - \alpha_i)(1 - \zeta_b)^{-2}$$

and the payoff after the merger is

$$\pi_a = \alpha_{i \cup j}(1 - \alpha_{i \cup j})(1 - \zeta_a)^{-2}$$

so that

$$\frac{\partial \pi_a}{\partial \tilde{\zeta}} = 2\pi_a(1 - \alpha_{i \cup j}(s_i + s_j) - \tilde{\zeta})^{-1}$$

and

$$\frac{\partial \pi_b}{\partial \tilde{\zeta}} = 2\pi_b(1 - \alpha_i s_i - \alpha_j s_j - \tilde{\zeta})^{-1}$$

Now

$$\alpha_{i \cup j}(s_i + s_j) = (s_i + s_j)\alpha(s_i + s_j) > (s_i + s_j)\alpha(s_j)$$

due to the fact that $\alpha(\cdot)$ is a monotonically increasing function.

Therefore, if $\pi_a > \pi_b$, then $\pi_a - \pi_b$ is an increasing function of $\tilde{\zeta}$.

Blocs i and j are willing to merge if and only if $\pi_a > \pi_b$ (where bloc i is the smaller of the two). A merger of any other two blocs, other than i or j, must raise $\tilde{\zeta}$. Therefore, if blocs i and j were willing to merge, given some bloc structure, then they will still be willing to merge after any merger of two other blocs k and m. •

Consider now the possibility of merger of two blocs, of given size s_i and s_j. Theorem 2 says that a necessary condition for the blocs to agree to the merger is that they agree under the coarsest possible partition of all the other blocs. It also says that a sufficient condition is that they agree under the finest possible partition.

The coarsest possible partition of the other blocs is that they all have merged into one big bloc, so that there are three blocs altogether : bloc i, bloc j, and a single bloc of size $1 - s_i - s_j$, in which case

$$\tilde{\zeta} = \frac{1 - s_i - s_j}{1 + s_i + s_j}$$

The finest possible partition of the other blocs is the limiting case, of an infinite number of arbitrarily small other blocs. In such a case,

$$\tilde{\zeta} = \frac{1 - s_i - s_j}{2}$$

The example above indicates that the smaller party may lose from a merger. However, in that example, although blocs 2 and 3 would not agree to a merger, a merger of blocs 1 and 2 would benefit both of the merging blocs.

In fact, for any partition, there must be some merger that benefits both parties to the merger — provided that return to capital stays strictly positive. While this result may not seem particularly intuitive, it has strong implications for the nature of stable bloc structures.

THEOREM 3 : If $M > 2$, then there is some merger between blocs which will increase the payoff to both parties to the merger.

PROOF : see the appendix $\bullet$

The main conclusion of this section is that there will always be a mutually beneficial pairwise merger, if there are 3 or more blocs, provided that the parameter a is suitably high.

If countries, or blocs of countries, are able to discuss coordination of a very simple sort — agreeing to levy the same tax rate, with no side payments — then this result seems to imply that the only stable outcome may be a considerable amount of coordination. The following section examines this implication, using a few simple notions of stability.

5. Stable Bloc Structures

While there are already a considerable number of solution concepts for the problem of coalition formation, I will add yet more to this proliferation. Since the main concern of this section is the stability of a bloc structure, a bloc structure $\mathcal{P}'$ will be described as **upsetting** a bloc structure $\mathcal{P}$ using the following definitions.

A bloc structure $\mathcal{P}'$ is defined as **following** a bloc structure $\mathcal{P}$ if one of the blocs in $\mathcal{P}'$ is the union of 2 or more of the blocs in $\mathcal{P}$, and if all the other blocs in $\mathcal{P}'$ are blocs in $\mathcal{P}$.

In other words, $\mathcal{P}'$ follows $\mathcal{P}$ if there is one bloc S in $\mathcal{P}'$, and a partition $\{S_1, S_2\}$ of the elements of S, such that $\mathcal{P}$ is obtained by replacing S with the refinement $\{S_1, S_2\}$. $\mathcal{P}'$ follows $\mathcal{P}$ if it results from a single bilateral merger of blocs in $\mathcal{P}$.

Analogously, $\mathcal{P}$ **precedes** $\mathcal{P}'$ if and only if $\mathcal{P}'$ follows $\mathcal{P}$.

Then there are two ways of upsetting some bloc structure $\mathcal{P}$:

UM : $\mathcal{P}'$ upsets $\mathcal{P}$ **by bilateral merger** if $\mathcal{P}'$ follows from $\mathcal{P}$, and if each country in each of the blocs that merge gets at least as high a payoff in $\mathcal{P}'$ as in $\mathcal{P}$.

US : $\mathcal{P}'$ upsets $\mathcal{P}$ **by secession** if $\mathcal{P}'$ precedes $\mathcal{P}$, and if each country in **one** of the blocs in $\mathcal{P}$ which splits to form $\mathcal{P}'$ gets a higher payoff in $\mathcal{P}'$ than in $\mathcal{P}$.

The notion of upsetting has some similarities with that of blocking in cooperative game theory. But there are some differences. The complement to a coalition here does not behave cooperatively. But the composition of the complement matters for the payoff to a group of countries choosing to form a bloc. Here if the countries in blocs S and T in $\mathcal{P}$ contemplate a merger, they assume that all the remaining countries stay in the same blocs as they were in $\mathcal{P}$. Moreover, utility is not transferable. Countries planning a deviation (either by merger or by secession) do not get to devise an imputation for the coalition

: it is required that each country in a bloc get the same payoff, namely the payoff to the non–cooperative tax–setting game played by the blocs.

A bloc structure which is susceptible to being upset seems somewhat unstable. It seems natural to define a coalition as "strongly stable" if it cannot be upset. It follows, virtually by definition, that bloc structures can only be strongly stable if there are only 1 or 2 blocs in the partition.

DEFINITION : A bloc structure $\mathcal{P}$ is **strongly stable** if there is no other bloc structure $\mathcal{P}'$ which upsets it by bilateral merger or by secession.

THEOREM 4 : The only strongly stable bloc structures have 1 or 2 blocs.

PROOF : Theorem 3 implies that if $M > 2$, then there is some merger of two blocs in the partition which will increase the payoff to both parties to the merger.

So if $\mathcal{P}$ has three or more blocs, then there is some bilateral merger, which benefits all countries. That means that there is some $\mathcal{P}'$ which follows $\mathcal{P}$, which is preferred by all countries, which means that $\mathcal{P}'$ upsets $\mathcal{P}$ by merger. $\bullet$

In the definition above of upsetting, only bilateral mergers were considered. However, it would not matter for the definition of strong stability, if multilateral mergers were allowed.

LEMMA 6 : A bloc structure can be upset by bilateral merger if and only if there is some multilateral merger of $m \geq 2$ blocs which benefits all m parties to the merger.

PROOF : The "only if" direction is straightforward, since a two–party merger is a special case of a multi–party merger. For the "if" direction, note that, under assumption 3, the only bloc structures for which there is no mutually beneficial bilateral merger are those with 1 or 2 blocs. $\bullet$

Also, upsetting by secession allowed for an entire "sub–bloc" of countries to secede en masse. It turns out to be no more restrictive to consider only secession by a single country.

LEMMA 7 : A bloc structure can be upset by secession if and only a single country can profitably secede from its bloc.

PROOF : Suppose some set $T \subset S$ of countries can secede profitably from coalition S. Theorem 1 implies that the countries in T must have, in aggregate, less than half the population of all the countries in S. Now consider the payoff to country $i \in T$ from seceding unilaterally from S instead. A bloc's payoff is decreasing in its share of the population, and increasing in the overall index ζ. Splitting S into $\{i\}$ and its complement results in at least as high a value for ζ as splitting into T and its complement : it involves a transfer of population from the smaller T to the larger $S \setminus T$. It results in at least as low a value for the share of population in country i's bloc. Therefore, unilateral secession by $i \in T$ must be at least as profitable for country i as secession by the whole group of countries in T. •

There are, of course, many other notions of coalitional stability. Most of them are defined for situations in which the payoff of a coalition may be divided arbitrarily among its members. However, many of these notions can be applied in a straightforward manner to bloc structures, in which a coalition's payoff is divided among its members using a fixed rule. [19]

For example, Hart and Kurz (1983) provide two versions of stability under group deviation, varying in the specification of the conjectured behaviour of countries not party to the deviation. Start first with γ stability. A bloc structure $\mathcal{P}$ is γ stable if it cannot be upset in the following fashion by some other bloc structure $\mathcal{P}'$: $\mathcal{P}'$ is formed from $\mathcal{P}$ by a bunch of countries forming a bloc T in $\mathcal{P}'$; if $i \in T$ in bloc structure $\mathcal{P}'$, and if $i \in S_i$ in

[19] Bloch (1996) does exactly this.

$\mathcal{P}$, then each j in $S_i \setminus T \cap S_i$ forms a singleton bloc $\{j\}$ in $\mathcal{P}'$; if S is a bloc in $\mathcal{P}$ such that $i \notin T$ for every $i \in S$, then S is a bloc in $\mathcal{P}'$ as well ; every country in T does better in $\mathcal{P}'$ than in $\mathcal{P}$. In other words, under the "γ" definition of stability, if a group of countries plan to deviate, they conjecture that the remainder of every bloc from which they take a member country will dissolve completely into singleton countries.

A bloc structure $\mathcal{P}$ is δ stable if it cannot be upset in the following manner : $\mathcal{P}'$ is formed from $\mathcal{P}$ by a bunch of countries forming a bloc T in $\mathcal{P}'$; the remaining blocs in $\mathcal{P}'$, other than T are the blocs in $\mathcal{P}$ minus the countries which seceded to form T ; every country in T does better in $\mathcal{P}'$ than in $\mathcal{P}$. In other words, under the "δ" definition of stability, if a group of countries plan to deviate, they conjecture that the remainder of every bloc from which they take a country will stay together.

In this model, a fine partition of the other countries is bad for a coalition. Therefore, it is straightforward to show that any δ stable bloc structure is also γ stable.

It also turns out that the set of δ stable bloc structures is the same as the set of strongly stable bloc structures.

THEOREM 5 : A bloc structure is strongly stable if and only if it is δ stable.

PROOF : Any deviation which "blocks" a bloc structure under the definition of strong stability will "block" under the definition of δ stability. That is, if $\mathcal{P} = (S_1, S_2, \ldots, S_M)$ is a bloc structure, it will be "blocked" under δ stability if there are some $T_1 \subset S_1, T_2 \subset S_2, \ldots T_M \subset S_M$, with at least one of the T_i's non–null, such that each non–null T_i increases its payoff in the new partition $\mathcal{P}' = (T, S_1 \setminus T_1, S_2 \setminus T_2, \ldots, S_M \setminus T_M)$, where T is the union of all the T_i's. Both upsetting by secession and upsetting by merger fit this definition, so that any δ stable bloc structure cannot be upset by secession or by merger, and so must be strongly stable.

So suppose now that some bloc structure is strongly stable. From Theorem 4, the bloc structure has either 1 or 2 blocs. If it has only 1 bloc, than the only possible deviation is by secession, so that strong stability implies δ stability.

Suppose then that $\mathcal{P} = (S_1, S_2)$ is strongly stable. It must be shown that it cannot be upset by the formation of a new bloc, comprising of some countries from S_1 and some from S_2. So suppose that there were some non–null $T_1 \subset S_1$ and $T_2 \subset S_2$, such that both T_1 and T_2 did better in $\mathcal{P}' = T_1 \cup T_2, S_1 \setminus T_1, S_2 \setminus T_2)$ than in $\mathcal{P}$.

What if $T_1 \cup T_2$ were at least as large as $S_1 \setminus T_1$ and $S_2 \setminus T_2$. Then T's payoff would (from Theorem 1) increase still further from a merger of $T_1 \cup T_2$ with $S_1 \setminus T_1$, and then with a subsequent merger of the remaining two blocs. So each country in T would be better off in a coalition of the whole than in $\mathcal{P}$. But since T_1 and T_2 are each non–null, then each country in S_1, and each country in S_2 would do better in a coalition of the whole than in $\mathcal{P}$, contradicting the strong stability of $\mathcal{P}$.

The remaining case is that $T_1 \cup T_2$ is smaller than one of the other two blocs. Without loss of generality, assume that $S_1 \setminus T_1$ has a larger share of the population than $T_1 \cup T_2$. Consider then a new partition $\mathcal{P}''$ formed by removing T_1 from $T_1 \cup T_2$ and merging it with $S_1 \setminus T_1$. This new bloc structure has a higher value of the index ζ than $\mathcal{P}'$, since countries have moved from a smaller bloc to a larger bloc. T_2 alone has a smaller share of the population than $T_1 \cup T_2$. So the payoff to T_2 must have increased. Thus, T_2 prefers $\mathcal{P}''$ to $\mathcal{P}'$ to $\mathcal{P}$. But $\mathcal{P}''$ could be formed more simply by the secession of T_2 from S_2. $\mathcal{P}$ is upset by secession, contradicting the hypothesis that it was strongly stable,and completing the proof of the theorem. •

However, there might not be any strongly stable equilibrium. The following example illustrates. Suppose that there are three countries : countries 2 and 3 each have the same population, and country 1's population is 10 times larger than the population of either country 2 or country 3. With three countries, there are 5 possible bloc structures. If $a = 20$, $b = 1$ and $\bar{k} = 1$, then the payoffs to the 3 countries in each of the 5 structures are depicted in the table below.

partition	Π_1	Π_2	Π_2	$\bar{t}$
$(\{1,2,3\})$	19.00	19.00	19.00	19.00
$(\{1\},\{2,3\})$	3.23	6.53	6.53	4.13
$(\{1,3\},\{2\})$	5.82	20.48	5.82	8.06
$(\{1,2\},\{3\})$	5.82	5.82	20.48	8.06
$(\{1\},\{2\},\{3\})$	3.10	6.32	6.32	4.03

The coalition of the whole is upset by secession of country 2 (or of country 3). A bloc structure in which the large country 1 is in the same bloc as one of the smaller countries is upset by secession of the smaller country in the bloc. The other two bloc structures are Pareto dominated by the coalition of the whole, and so are upset by merger.

This example shows that the concept of γ stability is indeed a strictly weaker concept than δ stability here. The coalition of the whole is γ stable in the example, since unilateral secession by one country will not be attractive if that country conjectures that the remaining two countries will split in response to this secession.

The coalition of the whole is not always γ stable : the smallest country may do better off in the finest possible partition than in the coalition of the whole. But there always is some γ stable bloc structure.

THEOREM 6 : For any set of N countries, there is at least one γ stable bloc structure, containing 1 or 2 blocs.

PROOF : see the appendix

THEOREM 7 : Any γ stable bloc structure is Pareto optimal, in the sense that no other bloc structure could yield a higher payoff to all countries.

PROOF : The coalition of the whole is always Pareto optimal. Let (S, T) be some bloc structure, with both S and T non–null, and with T at least as large as S. Suppose that (S, T) were Pareto dominated by some other bloc structure, containing at least two blocs. Theorem 3 then implies (S, T) would be Pareto dominated by a bloc structure with exactly two blocs, say (S', T'), with T' at least as large as S'.

Since (S', T') can Pareto dominate (S, T) only if it has a higher value for ζ, then S' must be strictly smaller than S. So some country i must be in S, and in T'. Since T' has at least half the total population, then countries in T' must prefer the coalition of the whole to (S', T'). That means i prefers the coalition of the whole to (S', T'). Since (S', T') was assumed to Pareto dominate (S, T), then every country in S must prefer the coalition of the whole to (S, T), so that the coalition of the whole Pareto dominates (S, T).

So if (S, T) is Pareto dominated, then it is Pareto dominated by the coalition of the whole. And if it is Pareto dominated by the coalition of the whole, then it cannot be γ stable. •

Bloch (1996) proposed another notion of stability for coalition formation games without side payments. The procedure shares some characteristics with a non–cooperative bargaining game. He considered a sequential game, in which individual countries proposed coalitions, and in which the other countries in the proposed coalitions then got to approve or disapprove. If a country "vetoes" the formation of a coalition in which it would be a member, it gets to counter–propose an alternative coalition. The game ends when all countries have been partitioned. Bloch's basic equilibrium concept is a stationary equilibrium to the sequential game just described. He showed that such an equilibrium must always exist.[20] If I refer to a bloc structure as "Bloch stable" if it can be achieved as a

[20] Bloch considers only deviations by individual countries, unlike Burbidge et al (1997) who require stability against deviation by groups of countries.

stationary equilibrium to the sequential game he described, then

THEOREM 8 : Any γ stable bloc structure must be Bloch stable.

PROOF : It is (fairly) straightforward to show that, if (S, T) is a γ stable bloc structure, then the following strategies constitute a stationary subgame perfect Nash equilibrium to the sequential coalition proposal game : if $i \in S$, and no coalitions have formed, or only T has formed, propose S ; if $i \in T$ and no coalitions have formed, or only S has formed, propose T ; if any coalition other than S or T has formed, i proposes the singleton coalition $\{i\}$; if a coalition U has been proposed, with $i \in U$, and with U different from S or T, then i should veto the coalition and counter–propose $\{i\}$. •

Finally, Burbidge et al model coalition formation by defining a game of coalition proposal. Country i's strategy σ_i is a bloc containing country i. Payoffs to the game are defined through an exogenously–specified rule ψ mapping profiles $(\sigma_1, \sigma_2, \ldots, \sigma_N)$ of proposed coalitions into bloc structures. Their notion of an equilibrium structure is the bloc structure $\psi(\sigma)$ resulting from profiles $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ which are coalition proof Nash equilibrium strategies for the game defined by the rule.

Among the examples they give of reasonable rules ψ is the strict unanimity rule, whereby i's bloc in $\psi(\sigma)$ will be σ_i if and only if $\sigma_j = \sigma_i$ for every $j \in \sigma_i$, and i's bloc in $\psi(\sigma)$ is $\{i\}$ otherwise. They also consider a more "cohesive" rule, the similarity rule. Under similarity, if a group of countries has the same proposed coalition, then that group of countries forms a coalition (which may be strictly smaller than the coalition they each proposed).

These two rules correspond to the notions of γ and δ stability respectively. Suppose that $\psi(\sigma) = \mathcal{P}$ for some profile σ of proposed coalitions, under the strict unanimity rule. Then S is a bloc in $\mathcal{P}$ if and only if $\sigma_i = S$ for every $i \in S$. If a group U plans a deviation from the strategy profile σ, then they expect that every $j \notin U$ to stick to its original profile

σ_j. Under the strict unanimity rule, then, country j will wind up a singleton after U's deviation, whenever $j \in S$ in $\mathcal{P}$, $j \notin U$ and $U \cap S \neq \emptyset$. Thus any deviating coalition expects its complement in each existing coalition to dissolve into singletons.

By the same token, under the similarity rule, any coalition planning a group deviation from strategy profile must expect its complement in any existing coalition to stay together as a single coalition after the deviation.

It then follows that any γ stable bloc structure will be an equilibrium coalition under the strict unanimity rule, and any δ stable bloc structure will be an equilibrium coalition under the similarity rule. In each case, the converse is not true : "blocking" is more difficult under the notion of coalition–proof Nash equilibrium, since a deviation will "block" some strategy profile only if the deviation itself is not blocked by a further deviation. In the three–country example above, the bloc structure $(\{1\}, \{2, 3\})$ is an equilibrium structure under the similarity rule, although it is not δ stable.

6. Identical Countries

If all countries have the same share $1/N$ of the population, then the outcomes can be described fairly precisely.

First of all, when there are N identically–sized countries, assumption 3 can be written

$$a \geq b\bar{k}\frac{(2N-1)(N+1)}{3(N-1)} \equiv a_{min}(N)$$

Figure 1 shows a_{min} for the case of $b = \bar{k} = 1$.

Secondly, the highest payoff to a country i will occur either when the bloc structure is the coalition of the whole, or when the bloc structure is $(\{i\}, \{1, 2, \dots, N\} \setminus \{i\})$. In the former case, the payoff to the country is (from equation (10)) $(a - b\bar{k})\bar{k}$. In the latter case, the payoff is

$$\frac{(1+N)^2 N}{[9(N-1)]b\bar{k}^2}$$

Therefore, the coalition of the whole will Pareto dominate any other bloc structure if (and only if)

$$a \geq b\bar{k} + \frac{(1+N)^2 N}{9(N-1)b}\bar{k} \equiv a_{max}(N)$$

Figure 1 also shows a_{max} as a function of N, when $b\bar{k} = 1$.

So if a is above $a_{max}(N)$, then there is a unique strongly stable outcome, the coalition of the whole. If a is between $a_{min}(N)$ and $a_{max}(N)$, then the only possible strongly stable outcome is a bloc structure with two blocs.

To proceed further, consider the incentive of a small country of size t to secede from a bloc of size $s > t$, when there are only two blocs. The payoff from secession is

$$\Gamma(s,t) = \frac{(1-t)}{(2-t)^2(1-\zeta_1)^2} - \frac{(1-s)}{(2-s)^2(1-\zeta_0)^2}$$

where

$$\zeta_0 = \frac{s}{2-s} + \frac{1-s}{1+s}$$

$$\zeta_1 = \frac{t}{2-t} + \frac{s-t}{2-s+t} + \frac{1-s}{1+s}$$

If $t \to 0$, then $\zeta_0 \to \zeta_1$, and $\partial\zeta_0/\partial s \to \partial\zeta_1/\partial s$. So Γ is an increasing function of s if t is sufficiently small. Γ is also (always) a decreasing function of t. Hence, at least when N is sufficiently large, there is a cut–off bloc size $s^*(N)$, so that a country will want to secede from a bloc if and only if the bloc size is greater than $s^*(N)$.

Straightforward computation yields the result that

$$\Gamma(M/N, 1/N) < 0 \qquad \text{for} \quad \text{all} \quad M < N \qquad \text{if} \quad N \leq 10$$

For $11 \leq N \leq 17$, $N/2 < s^*(N) < (N-1)/N$, and $s^*(N) < N/2$ for $N \geq 18$. [21]

Now if $s^*(N) \leq 1/2$, there can be no strongly stable bloc structure with two blocs : the larger of the two blocs must half at least half the population, implying any country in that bloc wishes to secede unilaterally.

So there is no strongly stable bloc structure when there are more than 17 identical countries — unless the coalition of the whole Pareto dominates other bloc structures.

If $1/2 < s^*(N) < (N-1)/N$, could there be a strongly stable bloc structure with 2 blocs? Let M be the number of countries in the larger bloc. The structure could be stable only if $M < s^*(N)$, since otherwise it would be upset by secession. But now consider the incentives for merger of the two blocs. The larger bloc will always gain from the merger. The smaller is M, the greater is the gain to merger for the smaller bloc. (If there are two blocs, one with share of the population $s < 1/2$, then the payoff to the smaller bloc is a decreasing function of s.) The larger is a, the greater is the gain to merger (since the merger results in the coalition of the whole). So if a merger between the two blocs is profitable for the smaller bloc, of size $N - M$, when $a = a_{min}(N)$ and when M is

[21] Since $t = 1/N$ is not that small when N is small, I did check that $\Gamma(M/N, 1/N) > 0$ whenever $\Gamma(M', N, 1/N)$ for some $M' > M$, for N less than or equal to 17.

the largest integer less than $s^*(N)N$, it will profitable for all larger a and all smaller M. Table 1 shows the payoff from the coalition of the whole, and the payoff to the smaller bloc, of size $N - M$, when M is the largest integer less than or equal to $s^*(N)N$ and when $a = a_{min}(N)$, as N runs from 11 to 17. In each case, the coalition of the whole is preferred.[22]

Therefore, if $11 \leq N \leq 17$, then there can be no strongly stable bloc structure with 2 blocs. If the larger bloc has a share greater than $s^*(N)$, then it is upset by secession, and if it has a share less than or equal to $s^*(N)$ then it is upset by merger.

Table 1

N	π_{whole}	π_{2-bloc}
11	7.4	3.3
12	8.1	2.7
13	8.7	2.3
14	9.4	2.5
15	10.0	2.3
16	10.7	2.1
17	11.4	2.0

So a complete taxonomy of strong stability can be provided when countries are identical : [23]

if $a > a_{max}(N)$, the coalition of the whole is the only strongly stable bloc structure

[22] In table 1, b and $\bar{k}$ have both been set equal to 1. But since both payoffs — that from the coalition of the whole when $a = a_{min}(N)$, and that from the smaller bloc in the 2–bloc coalition — are proportional to $b\bar{k}^2$, the result holds for any values of b or $\bar{k}$.

[23] If $a = a_{max}(N)$, and $N > 10$, then the coalition of the whole is the unique strongly stable bloc structure.

if $a \leq a_{max}(N)$ and $N \leq 10$, there is a strongly stable bloc structure in which one bloc consists of a single country, and the other bloc consists of all the other countries

if $a < a_{max}(N)$ and $N \geq 11$, there is no strongly stable bloc structure

What happens if there is no strongly stable bloc structure? When countries are equal–sized, then the coalition of the whole is always a γ stable bloc structure : for any value of N, and for any value of $a \geq a_{min}(N)$. While this may seem to make γ stability a very weak notion, the previous section demonstrated that any γ stable structure (including the coalition of the whole) can be sustained as a coalition proof Nash equilibrium in the explicit coalition formation game proposed by Burbidge et al, under the strict unanimity coalition formation rule. It also can be sustained as a subgame perfect Nash equilibrium in the sequential coalition proposal game proposed by Bloch.

The threat to stability here is secession, rather than merger. γ stability is easier to achieve since conjectures about the response to secession are so pessimistic.

When all countries are the same size, the coalition of a whole seems a better outcome than a two–bloc structure, even if it does not Pareto dominate. The sum of payoffs must be higher in the coalition of the whole, so that each country's expected payoff is higher, if all are assumed to have the same chance of being in the smaller bloc in a two–bloc structure.

If secession is not allowed after a bloc has formed, then the coalition of the whole can always be achieved by a sequence of mutually–beneficial mergers. For example, if $N = 2^m$ for some integer m, then each country could first pair up with another country. Then two–country blocs would pair up, and so on. Each merger occurs between equal–sized blocs, and therefore must, from Theorem 1, benefit both parties.

Of course having $N = 2^m$ is a very special case. But it turns out that a merger between two blocs is always mutually beneficial whenever the smaller bloc is at least 20 percent of the size of the larger bloc. [24] This result can be used to construct a sequence of beneficial mergers for an arbitrary number of equal–sized countries : first merge pairs of countries ; if N is odd, then merger the "left out" country with one of the pairs ; at each subsequent stage, if the number of blocs is even, merge the largest with the smallest, the second–largest with the second–smallest, and so on ;if the number of blocs at some stage is odd, merge the second–largest with the smallest, the third–largest with the second–smallest, and so on ; at each stage of this process the largest blocjk is no more than twice the size of the smallest, so that each merger is mutually beneficial.

So the order of mergers matters — if secession were not allowed. Not allowing secession may seem an arbitrary restriction. But the coalition of the whole can be sustained as an equilibrium when countries anticipate the consequences of future moves, realizing that assent to mergers is irrevocable. Bloch models coalition formation as an explicit sequential game of this sort, and the coalition of the whole will always be a subgame perfect Nash equilibrium of this game.

The main obstacle to total elimination of tax competition, then, is the persistence of size disparities. A bloc structure with only two blocs may be one step away from the coalition as a whole. But the revenue forfeited by countries in a two bloc structure may be large. Consider the total tax revenue collected in a two bloc structure, a a fraction of the total tax revenue conifiscated by the coalition as a whole. This fraction can be made arbitrarily small, with the two–bloc structure still an equilibrium. That is, for any $\epsilon > 0$, there exists a number of countries N, and a value of $a > a_{min}(N)$ such that $(\{1\}, \{2, 3, \ldots, N\})$ is a γ stable bloc structure, and the total taxes collected from this bloc structure is less than ϵ times the tax revenue confiscated by the coalition of the whole.

The role of small countries in thwarting collusion by tax authorities stands in contrast to some of the results in the theory of customs unions. There it is often the case that

[24] The (brute–force) computation for this result is shown in an appendix, available from the author.

there are no mutually beneficial arrangements between a few large blocs. Small countries coalescing into regional trading blocs may not lead to world free trade. But here, small countries coalescing into regional tax harmonization blocs definitely would lead to world tax harmonization.

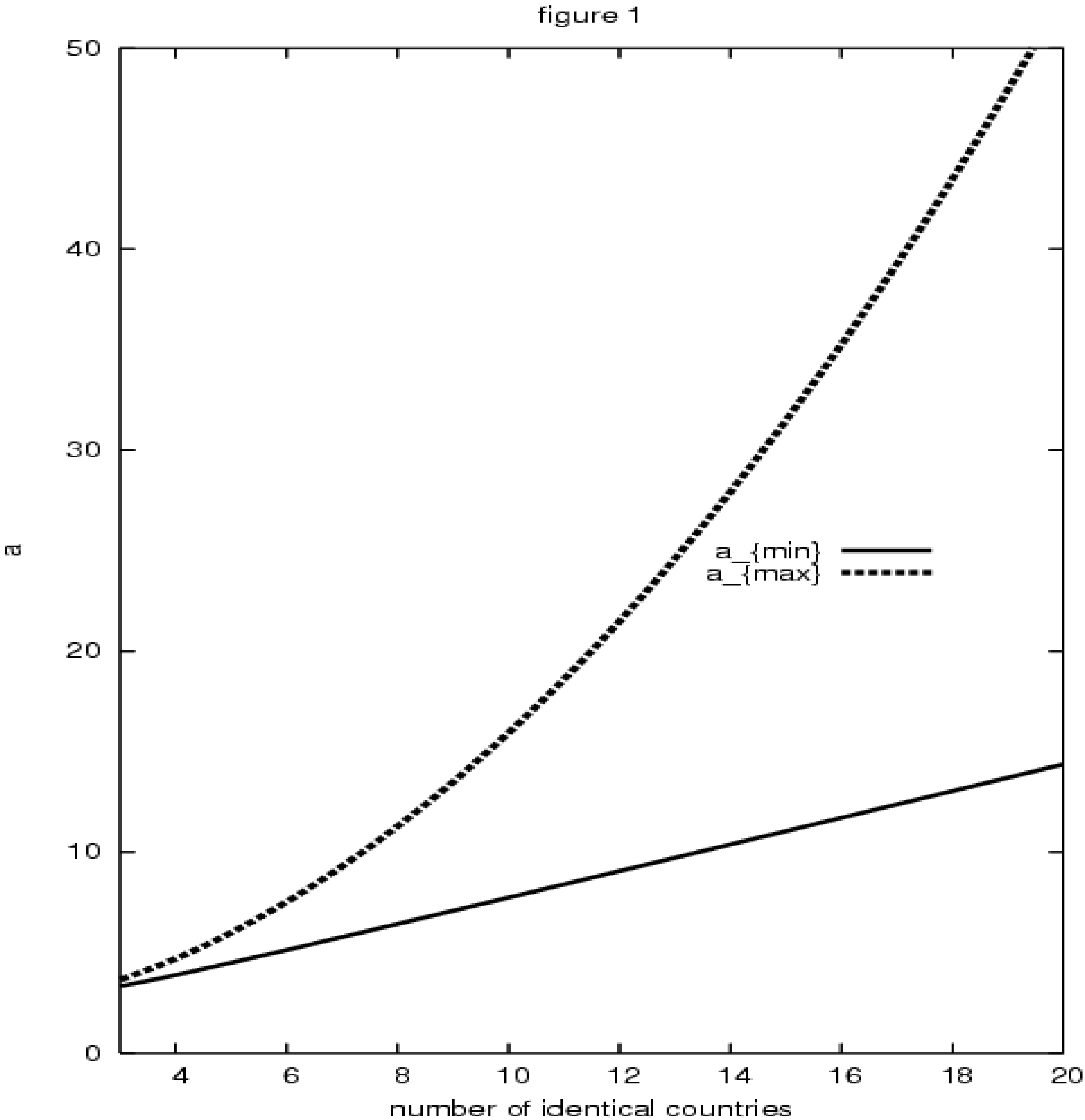
figure 1
50
40
30
20
10
0
a
a_{min}
a_{max}
4
6
8
10
12
14
16
18
20
number of identical countries

Figure 1

7. Concluding Remarks

Tax setting by sovereign countries is usually treated as a non–cooperative game. Given the difficulty of using supranational entities to enforce treaties, this is a natural approach. But in many contexts, the incentives for coordination among countries are very strong. This is certainly the case when they levy sourced–based taxes on a mobile tax base. The central message of the tax competition literature is the inefficiency of the resulting Nash equilibrium tax rates.

Even if cooperation among all the countries in the world cannot be enforced, Burbidge et al showed the importance of considering alliances among countries. They also showed that the possibility of cooperation need not guarantee an efficient result.

In contrast with their general approach, here I use a very special and simple tax competition model. In this model it is almost tautological that there will be some cooperative outcomes which are better for all countries that the Nash equilibrium of a non– cooperative game. Such outcomes should involve increases in the countries' tax rates, and perhaps some side payments among countries. In this paper, in recognition of the difficulty of enforcing cooperation, I have allowed only very simple deals among countries : the only way they have of policy coordination is to agree to set exactly the same tax rates. In this respect, the paper biases the case against cooperation, since small tax havens cannot be bribed to increase their tax rates.

Nonetheless, what the model suggests is that, even clumsy cooperation will virtually eliminate tax competition. Most tax havens will be willing to join tax coordination blocs, even though they do well be free–riding. In a stable outcome, the world's population should be divided into two blocs, if it is divided at all.

These results are derived under very strong assumptions about the production technology (quadratic), and about factor supply (fixed). Nonetheless the assumptions which are really essential, and which drive the results, concern the policy–makers' goals. If each country's payoff really is the tax revenue it collects (or if this revenue has a very high weight in each country's welfare measure), then the problems of non–cooperative tax competition probably are relatively easy to solve. If countries want to tax the return to capital, then increased capital mobility will not prevent them from doing so.

References.

[1] F. Bloch. Sequential formation of coalitions in games with externalities and fixed payoff division. *Games and Economic Behavior*, 14:90–123, 1996.

[2] J. Burbidge, J. de Pater, G. Myers, and A. Sengupta. A coalition–formation approach to equilibrium federations and trading blocs. *American Economic Review*, 87:940–956, 1997.

[3] M. Devereux, R. Griffith, and A. Klemm. Corporate income tax reforms and international tax competition. *Economic Policy*, 35:451–495, 2002.

[4] J. Edwards and M. Keen. Tax competition and Leviathan. *European Economic Review*, 44:113–134, 1996.

[5] N. Hansen and A. Kessler. The political geography of tax h(e)avens and tax hells. *American Economic Review*, 91:1103–1115, 2001.

[6] S. Hart and M. Kurz. Endogenous formation of coalitions. *Econometrica*, 51:1047–1064, 1983.

[7] E. Janeba and G. Schjelderup. Why Europe should love tax competition — and the U.S. even more so. working paper 9334, NBER, 2002.

[8] R. Kanbur and M.Keen. Jeux sans frontieres : Tax competition and tax coordination when countries differ in size. *American Economic Review*, 83:877–892, 1993.

[9] P. Kehoe. Policy cooperation among benevolent governments may be undesirable. *Review of Economic Studies*, 56:289–296, 1989.

[10] M. Kemp and H. Wan. An elementary proposition concerning the formation of customs unions. *Journal of International Economics*, 6:95–97, 1976.

[11] D. Laussel and M. Le Breton. Existence of Nash equilibria in fiscal competition models. *Regional Science and Urban Economics*, 28:283–296, 1998.

[12] S. Peralta and T. van Ypersele. Coordination of capital taxation among asymmetric countries. Technical Report 3695, CEPR, 2003.

[13] M. Perry and R. Porter. Oligopoly and the incentive for horizontal merger. *American Economic Review*, 75:219–227, 1985.

[14] D. Ray and R. Vohra. Equilibrium binding agreements. *Journal of Economic Theory*, 73:30–78, 1997.

[15] R. Riezman. Customs unions and the core. *Journal of International Economics*, 19:355–365, 1985.

[16] S. Salant, S. Switzer, and R. Reynolds. Losses from horizontal merger : The effects of an exogenous change in industry structure on Cournot–Nash equilibrium. *Quarterly Journal of Economics*, 98:185–199, 1983.

[17] J. Wilson. A theory of interregional tax competition. *Journal of Urban Economics*, 19:296–315, 1986.

[18] J. Wilson. Tax competition with interregional differences in factor endowments. *Regional Science and Urban Economics*, 21:423–451, 1991.

[19] J. Wilson and D. Wildasin. Capital tax competition : Bane or boon? *Journal of Public Economics*, 88:1065 – 1093, 2004.

[20] G Zodrow and P. Mieszkowski. Pigou, Tiebout, property taxation and the underprovision of local public goods. *Journal of Urban Economics*, 19:356–370, 1986.

Appendix

LEMMA $A1$: There can be no Nash equilibrium in which $\sum s_i k_i < k$.

PROOF : When there is excess supply of capital, then $\partial k_i / \partial t_i$ is defined by equation 6, so that

$$\frac{\partial \pi_i}{\partial t_i} = k_i - \frac{t_i}{b}$$

Substituting for t_i from equation 5,

$$\frac{\partial \pi_i}{\partial t_i} = 2k_i - \frac{a}{b}$$

One bloc at least must have $k_i \leq \bar{k}$, since the weighted average of the k_i's is less than $\bar{k}$. But then the assumption that $a - 2\bar{k} > 0$ implies that at least one bloc has

$$\frac{\partial \pi_i}{\partial t_i} < 0$$

which means that the vector of tax rates cannot be a Nash equilibrium. •

LEMMA 2 : For any given bloc structure, if there is a Nash equilibrium with a positive net return to capital, then there is no other Nash equilibrium.

PROOF : Suppose that t is a vector of Nash equilibrium tax rates for some bloc structure. From lemmata 1 and A1, the only other possible Nash equilibrium would be some tax vector t' which resulted in $\rho = 0$, and which resulted in no excess supply of capital. When there is no excess supply of capital

$$\rho = a - b\bar{k} - \bar{t}$$

so that the fact that $\rho' < \rho$ means that $\bar{t}' > \bar{t}$.

Now let bloc j be a bloc for which

$$t'_j - \bar{t}' \geq t_j - \bar{t}$$

The fact that $\bar{t}$ and $\bar{t}$ are weighted averages of tax rates imply such a j must exist. Since $\bar{t}' > \bar{t}$, therefore $t'_j > t_j$.

Since $k_i = \bar{k} + \frac{1}{b}[\bar{t} - t_i]$, therefore $k'_j \leq k_j$.

Therefore,

$$\partial\pi'_j \partial t'_j|_- = k'_j - \frac{1 - s_j}{b} t'_j < k_j - \frac{1 - s_j}{b} t_j = \frac{\partial\pi_j}{\partial t_j} = 0$$

which means that a small decrease in t'_j would increase the payoff to bloc j, contradicting the assumption that t' constituted a Nash equilibrium. •

LEMMA 3: For any given bloc structure, there exists at least one Nash equilibrium. In any Nash equilibrium, capital is fully allocated, so that $\sum_{i=1}^{N} s_i k_i = \bar{k}$.

PROOF : The second part of the lemma follows immediately from lemma A1.

To prove the first part, first take the unique solution to equation (7), that is the matrix equation $At = \beta$. If the net return implied by that solution is non–negative, then the solution is a Nash equilibrium.

If not, then the Nash equilibrium will be a t on the border between regimes, so that $\rho = 0$ and $\sum s_i t_i = \bar{k}$. For such a t to be a Nash equilibrium, it must be the case that

$$b\bar{k} + \bar{t} - (2 - s_i)t_i \geq 0 \qquad i = 1, \cdots, N \tag{a1}$$

$$b\bar{k} + \bar{t} - 2t_i \leq 0 \qquad i = 1, \cdots, N \tag{a2}$$

where these inequalities represent the left– and right–hand derivatives of π_i with respect to t_i. For the net return to capital to be 0, it must also be true that

$$\bar{t} = a - b\bar{k} \tag{a3}$$

To construct a Nash equilibrium, start with the solution to $At = \beta$, which involves the tax rates being "too high". Scale down all the tax rates by the same factor of proportionality, so that now $\bar{t} = a - b\bar{k}$. Since all the t_i's and $\bar{t}$ have been reduced by the same factor, and since originally equation $(a1)$ held with equality, at this new tax vector, inequality $(a1)$ is satisfied for each bloc i.

It also will be the case that the larger the bloc, the higher its tax rates. The potential problem is that inequality $(a2)$ may not hold for some smaller blocs. Take the tax rate in each of these smaller blocs, and raise the tax rates in each of these blocs to some t' such that

$$b\bar{k} + \bar{t} - 2t' = 0$$

The facts that $a > 2\bar{k}$ and $\bar{t} = a - b\bar{k}$ imply that $\bar{t} > t'$. To keep the weighted average of the tax rates (and the net return to capital) constant, reduce all tax rates greater than $\bar{t}$ by some $c(t_i - \bar{t})$ where the constant c is chosen so that the weighted average of the tax rates remains at $\bar{t}$. At the new tax vector, there are two types of bloc. Those for which $t_i < t'$ originally now have $t_i = t'$. For these blocs, inequality $(a2)$ now holds with equality. Since the left side of inequality $(a1)$ is larger than the left side of inequality $(a2)$, this inequality is satisfied as well.

Those blocs with $t_i > t'$ now have had their tax rates reduced — but to something greater than (or equal to) t'. The fact that $t_i \geq t'$ for these blocs means that inequality $(a2)$ is satisfied. The facts that $\bar{t}$ is unchanged, t_i has decreased or remained the same, and that inequality $(a1)$ held initially means that inequality $(a1)$ still holds for these blocs.

Hence the new tax vector is a Nash equilibrium. In general, it will not be the only Nash equilibrium : increasing the smaller blocs' tax rates a little above t' (and adjusting the larger blocs' taxes downward to keep $\bar{t}$ constant) will yield another tax vector which satisfies $(a1)$ and $(a2)$.

The existence of a Nash equilibrium could also be demonstrated less constructively by noting that each bloc's payoff is a concave function of its own tax rate, and a continuous function of every other bloc's. •

LEMMA 5 : $(s_i + s_j)\pi_{i\cup j}(\mathcal{P}') > s_i\pi_i(\mathcal{P}) + s_j\pi_j(\mathcal{P})$, where $\pi_i(\mathcal{P})$ denotes the Nash equilibrium payoff to player i in the tax–setting game in which the players are elements of the partition $\mathcal{P}$.

PROOF : Suppose that $t_i \leq t_j$, so that $k_i \geq k_j$. Let the merged bloc $i \cup j$ choose a tax rate

$$t = \frac{s_i t_i + s_j t_j}{s_i + s_j}$$

This leaves the world average tax rate $\bar{t}$ unchanged. From equation (7), no other bloc would want to change its tax rate if $i \cup j$ levied the tax rate t. From equation (3), then, the capital employed per capita in $i \cup j$ would be

$$k = \frac{s_i k_i + s_j k_j}{s_i + s_j}$$

where k_i and k_j were the quantities of capital employed per capita in blocs i and j when they were separate.

The payoff to the new merged entity — should it choose the tax rate t — would be tk. Since $t_i \leq t_j$ and $k_i \geq k_j$, it then follows that

$$(s_i + s_j)^2 tk = [s_i t_i + s_j t_j][s_i k_i + s_j k_j] \geq (s_i + s_j)^2[t_i k_i + t_j k_j]$$

so that the merged bloc can achieve a total payoff at least as great as the sum of the payoffs of the original blocs. •

Proof of Theorem 3

First, number the blocs in order of their population share, so that $s_1 \geq s_2 \geq \cdots s_M$. Next, denote the ratio s_i/s_{i+1} by z_i. Let z denote the minimum of the z_i's. Then

LEMMA $A2$: $1/(1-\zeta) > (z+1)/2$.

PROOF : Since $s_i \leq s_1 z^{-(i-1)}$, then

$$\sum_{i=1}^{M} s_i \leq s_1(1 + (1/z) + (1/z)^2 + \cdot + (1/z)^{M-1}) \leq s_1 \frac{z-1}{z}$$

Since the sum of the population shares must equal 1, therefore

$$s_1 \geq \frac{z-1}{z}$$

Since

$$\zeta = \sum_{i=1}^{M} \alpha_i s_i$$

therefore

$$\zeta \geq \frac{z-1}{z} \frac{1}{2 - ((z-1)/z)} = \frac{z-1}{z+1}$$

so that

$$1 - \zeta \leq \frac{2}{z+1}$$

proving the result. •

Now suppose that blocs i and $i+1$ are two of the blocs for which the ratio z_i is at a minimum. Let $s' \equiv s_i + s_{i+1}$, $\alpha' \equiv 1/(2 - s_i - s_{i+1})$. Further let ζ refer to the original index of tax competition, and let ζ' denote the value of this index should blocs i and $i+1$ merge. Therefore, the gain to bloc $i+1$ from merging with bloc i is

$$\frac{\alpha'(1-\alpha')}{(1-\zeta')^2} - \frac{\alpha(1-\alpha)}{(1-\zeta)^2}$$

LEMMA *A*3: If s_j/s_{j+1} is minimized at $j = i$, and if $s_i + s_{i+1} \leq 4/7$, then residents of blocs i and $i+1$ will benefit from a merger of the two blocs.

PROOF : If the merger occurs, then the overall increase in the index of competition is

$$\zeta' - \zeta = s_i s_{i+1} \alpha'(\alpha_i + \alpha_{i+1})$$

The overall increase in the share of population in the smaller bloc is just $s' - s_{i+1} = s_i$.

Now continuously change the index from ζ to ζ', and continuously change the population share of residents of the smaller bloc from s_{i+1} to s'. That is, let

$$\zeta(x) \equiv \zeta + s_{i+1}\alpha'(\alpha_i + \alpha_{i+1})x$$

$$s(x) \equiv s_{i+1} + x$$

Then $\zeta(0) = \zeta$, $\zeta(s_i) = \zeta'$, $s(0) = s_{i+1}$ and $s(s_i) = s'$. If I denote by

$$\pi(x) \equiv \frac{\alpha(s(x))(1-\alpha(s(x)))}{(1-\zeta(x))^2}$$

then the gain to bloc $i+1$ from merging with the larger bloc i is just

$$\int_0^{s_i} \pi'(x)dx$$

It will be slightly more convenient to use the logarithm of the bloc's payoff. The derivative of $\ln(\pi(x))$ with respect to x is

$$[\frac{1}{\alpha} - \frac{1}{1-\alpha}]\alpha'(x) + \frac{2}{1-\zeta}\zeta'(x)$$

From the definition of α, $\alpha'(x) = \alpha^2$, so that this derivative of the logarithm of the payoff can be written

$$\frac{2}{1-\zeta}s_{i+1}\alpha'(\alpha_i + \alpha_{i+1}) - \frac{\alpha^2 s}{1-\alpha}$$

where I have used the fact that $2\alpha - 1 = \alpha s$. Therefore,the smaller bloc will gain from the merger if

$$\int_0^{s_i} \frac{2}{1-\zeta(x)}s_{i+1}\alpha'(\alpha_i + \alpha_{i+1}) - \frac{\alpha^2(x)s(x)}{1-\alpha(x)}dx \geq 0$$

Lemma $A2$ implies that this integral is greater than

$$\int_0^{s_i} (k+1)s_{i+1}\alpha'(\alpha_i + \alpha_{i+1}) - \frac{\alpha^2(x)s(x)}{1-\alpha(x)}dx$$

Since $\alpha^s/(1-\alpha)$ is an increasing function of x, the integral of the negative second term in the above expression is less than or equal to

$$\frac{1}{2}s_i[\frac{(\alpha')^2 s'}{1-\alpha'} - \frac{\alpha_{i+1}^2 s_{i+1}}{1-\alpha_{i+1}}$$

Since $(k+1)s_{i+1} = s'$, then the integral will be positive if

$$2(\alpha_i + \alpha_{i+1}) \geq \frac{\alpha'}{1-\alpha'}$$

Given the convexity of $\alpha(s)$, $(\alpha[s_i] + \alpha[s_{i+1}])/\alpha[s_i + s_{i+1}] \geq 2\alpha[s'/2]/\alpha(s')$.

Since $2\alpha[s'/2]/\alpha[s'] = 4(1-s'/2)/(2-s'/2)$, and since $1-\alpha' = 1-s'/2-s'$, therefore, the value of the integral is non–negative if

$$8\frac{1-s'}{4-s'} \geq 1$$

which will be true whenever $s' \leq 4/7$, completing the proof of the lemma. $\bullet$

Next, suppose that the share of total population of the two largest blocs is some s'. Consider now varying the share s_1 of population in the largest bloc, changing s_2 so as to keep $s' = s_1 + s_2$ constant. If $\Psi(s_1; s')$ denotes the gain to bloc 2 from merging with bloc 1, then

LEMMA $A4$: $\frac{\partial \Psi}{\partial s_1} < 0$

PROOF : The post–merger payoff depends on s', and on the distribution of population among the other blocs, and so is unaffected by an increase in s_1 and a decrease in s_2 which leave constant s'.

The pre–merger payoff to bloc 2 is $\alpha_2(1-\alpha_2)/(1-\zeta)^2)$. Increasing s_1 and decreasing s_2 must increase ζ. (The derivative of ζ with respect to s_1, when s_2 changes to keep s' constant, is $2([\alpha_1]^2 - [\alpha_2]^2) > 0$.) Since $\alpha(1-\alpha)$ is a decreasing function of s, therefore this pre–merger payoff increases as s_1 rises and s_2 falls, proving the lemma. $\bullet$

Next, define $\gamma_2(s)$ in the following fashion. Suppose that $s < 1$, and that blocs 1 and 2 have a combined share s of the overall population. Suppose further that the remainder of the population is divided as finely as possible. Then $\gamma_2(s) < s/2$ is the share of population in bloc 2, such that bloc 2's residents are just willing to merge with bloc 1.

LEMMA $A5$: If $s_1 = s - s_2 > s_2$, and if $s_2 > \gamma_2(s)$, then residents of bloc 2 gain from a merger with bloc 1.

PROOF : From Theorem 2, the gain to bloc 2 from merging with bloc 1 is larger, the coarser the partition of the remaining countries. So if bloc 2 is willing to merge with bloc 1 when the remaining share $1 - s_1 - s_2$ of the population is divided as finely as possible, then they will be willing to merge for any other possible partition of the remaining share

$1 - s_1 - s_2$.

Lemma $A4$ shows that $\Psi(s - s_2, s) > 0$ whenever $s_2 > \gamma_2(s)$, completing proof of the lemma. •

Figure 2 shows the graph of $\gamma_2(s)$, and of $s_1 = s - \gamma_2(s)$. Notice first that $\gamma_2(s) < 2/7$ for all values of s. Therefore

LEMMA $A6$: If $M > 2$ and if s_i/s_{i+1} is not minimized at $i = 1$, then there is some merger between two blocs which is increases the payoff to both parties to the merger.

PROOF : Figure 2 shows that $\gamma_2(s) < 2/7$ for all $s \in (0, 1)$. That means that $s_i + s_{i+1} < 4/7$ for any $i > 1$. The hypothesis of the lemma is that s_i/s_{i+1} is minimized at some $i > 1$. Therefore, lemma $A3$ implies that blocs i and $i + 1$ will both gain from a merger, if s_j/s_{j+1} is minimized at $j = i$. •

LEMMA $A7$: If $s_1/s_2 \geq s_i/s_{i+1}$ for all $i = 2, \cdots, M - 1$, then

$$\frac{(s_1)^2}{s_1 - s_2} > 1$$

PROOF : The population shares must sum to 1. By hypothesis, $s_1 = zs_2$, and $s_i < s_2 z^{-(i-2)}$ for $i > 2$.

Therefore, the sum of the population shares are less than or equal to

$$s_1 + s_2 + \sum_{j=3}^{M} z^{-(j-2)} s_2$$

In turn, this expression is less than or equal to

$$s_1 + s_2 \frac{z - 1}{z}$$

which (given that $z = s - 1/s_2$) equals

$$s_1 + \frac{s_1 s_2}{s_1 - s_2} = \frac{s_1^2}{s - 1 - s_2} \bullet$$

If $s_2 = \gamma_2(s)$, then $(s-1)^2/(s_1 - s_2) = s + (\gamma_2(s))^2/(s - 2\gamma_2(s))$. This curve is shown in figure 2.

LEMMA $A8$: If $s_2 < \gamma_2(s)$, then

$$\frac{(s_1)^2}{s_1 - s_2} < s + \frac{[\gamma_2(s)]^2}{s - 2\gamma_2(s)}$$

PROOF : If $s_1 + s_2 = s$, then $(s_1)^2/(s_1 - s_2) = (s_1)^2/(2s - 1 - s)$. Differentiation shows that this expression is a decreasing function of s_1. $\bullet$

The graph of $s + (\gamma_2(s))^2/(s - 2\gamma_2(s))$ in figure 2 (labelled "sum of all shares") is everywhere below 1. Therefore, the following result has been demonstrated :

LEMMA $A9$: If $s_2 \geq \gamma_2(s)$, then $s_1/s_2 > \min_j s_j/s_{j+1}$.

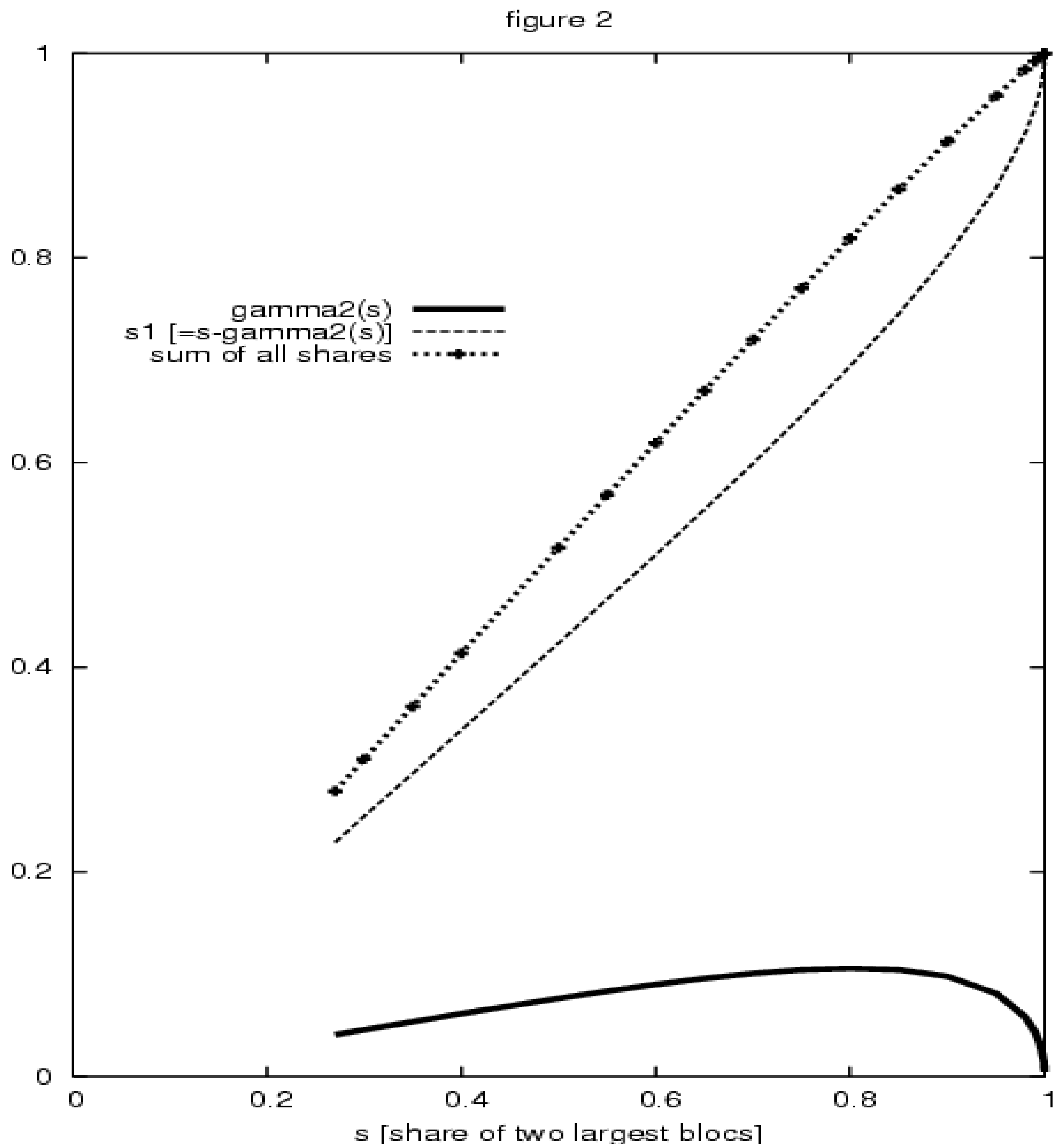
figure 2
gamma2(s)
s1 [=s-gamma2(s)]
sum of all shares
0
0.2
0.4
0.6
0.8
1
s [share of two largest blocs]

The lemmata together then imply

THEOREM 3 : If $M \geq 3$, then there exists some merger of two blocs which would increase the payoff of both merging blocs.

PROOF : Let $s = s_1 + s_2$.

If $s_2 \geq \gamma_2(s)$, then a merger of blocs 1 and 2 would increase their payoffs, as proved in Lemma $A5$.

If $s_2 < \gamma_2(s)$, then lemma $A9$ shows that s_j/s_{j+1} is not minimized at $j = 1$. Lemma $A6$ then shows that there is some other pair of blocs, i and $i+1$, with $i > 1$, for which a merger would increase the payoff to both parties. $\bullet$

THEOREM 6 : For any set of N countries, there is at least one γ stable bloc structure, containing 1 or 2 blocs.

PROOF : The proof proceeds by induction on the number N of countries.

If $N = 2$, the coalition of the whole is γ stable if both countries prefer it to the only other bloc structure, a split into two singletons. If the smaller country prefers the bloc structure $(\{1\}, \{2\})$ to the coalition of the whole, then $(\{1\}, \{2\})$ is γ stable.

So suppose that the induction hypothesis holds for $N - 1$. Take any set of $N \geq 3$ countries, and now consider what would happen if the two largest countries, numbers 1 and 2, were merged into a single country. By the induction hypothesis, there is some γ stable bloc structure for this $N-1$–country world. Now let S' denote the bloc containing the merged entity $1 \cup 2$ in this $N-1$–country world, and then let S denote this bloc, with 1 and 2 "un–merged".

i Suppose first that S is the coalition of the whole. Then S' is the coalition of the whole for the $N - 1$ country world. Since S' is γ stable, then country i did not wish to secede unilaterally from the coalition of the whole, if it conjectured that its secession would break the coalition into singletons. Separating $1 \cup 2$ into countries 1 and 2 makes the payoff from this unilateral secession even smaller (and does not affect the payoff to the coalition of the whole), so that country i will not wish to secede from S (if $i \geq 3$). Since the payoff from a given bloc structure decreases with the size of a country, countries 1 or 2 will not want to secede either. No bloc containing both 1 and 2 will want to secede, since the bloc did not want to secede from S' when 1 and 2 were merged. The remaining possibility is that a bloc containing country 1 and some others (but not country 2), or a bloc containing #2 but not #1 might want to secede. But since #1 is the largest country, the payoff to any bloc T containing #1 would be increased by sequentially adding each of the other countries not in T. If the bloc contained #2, then sequentially adding all countries other than #1 would increase the payoff to T. Then adding #1 would increase the bloc's payoff still further, if country 1 had less than half the total population. So the only remaining possible deviation is a secession of a bloc T containing country 2, but not country 1, in a case in which country 1 had more than half the population. If that were profitable for members of T, then a deviation of $T \setminus \{2\}$ from S' would have been profitable, contradicting the γ stability of S'.

Next, suppose that the S is a proper subset of $\{1, 2, \ldots, N\}$, and that T is its complement.

ii Since country 1 is the largest country in S, it will not want to secede unilaterally. If S contains some country $i \notin \{1, 2\}$, then i would not to secede unilaterally, since it did not want to secede unilaterally from S', and since splitting of $1 \cup 2$ into 1 and 2 makes secession by others less attractive. But the payoff to $i > 2$ from secession is at least as high as the payoff to country #2 (from Proposition 2). Therefore the only possible profitable unilateral secession from S would occur if $S = \{1, 2\}$, in which case country 2 might wish to secede. This case will be examined in part *iv* of the proof below.

iii Under the definition of γ stability, another possibility is a merger of some non–null subset $U \subset S$ with some non–null subset $V \subset T$, with their complements in S and T splitting into singletons. If U contained both 1 and 2, then the γ stability of the original bloc structure in the $N-1$–country world implies the merger would not be profitable. Similarly, merger would not be profitable if $S \setminus U$ contained both 1 and 2. If U contained the largest country 1, then the payoff of $U \cup V$ would be higher yet if country #2 were added to $U \cup V$. Since this deviation was possible in the $N-1$–country world, then the merger of U and V cannot be profitable in the N–country world.

Now if $U \cup V$ were larger than country 1, then sequential merger with all the other countries not in $U \cup V$ (starting with country 1) would raise the payoff to $U \cup V$, implying the coalition of the whole was preferable to (S, T), which contradicts the γ stability of (S', T) in the $N-1$–country world.

So if U and V can secede from S and T, and merge, then it must be the case that $2 \in U$, and $U \cup V$ is smaller than country 1. If that were the case, then a merger of $U \setminus \{2\}$ with V in the $N-1$–country world would have been profitable : this merger would have resulted in a smaller bloc, and a higher value for ζ than the merger of U with V in the N–country world.

The remaining case of merger between subsets of S and T is then country 2 merging with some non–null subset V of T. This would only be profitable if $S = \{1, 2\}$: otherwise a merger of some other country $i \in S'$ with V would have been profitable in the $N-1$–country world.

Since country 2 is the second–largest country, then if it could profitably merge with some proper subset $V \subset T$, its payoff would increase further by sequential merger with all the countries in $T \setminus V$.

Hence, if a merger of $U \subset S$ with $V \subset T$ is profitable for all parties to the merger (under the conjecture that their complements in S and T dissolve into singletons after the merger), then it must be the case that *a* $S = \{1, 2\}$ and *b* a merger of country 2 with all of T is profitable for both 2 and the members of T. This case will be considered in part v of the proof below.

iv Suppose $S = \{1, 2\}$ and country 2 gains by unilateral secession. Consider now the bloc structure $(\{1\}, \{2\}, T)$. Theorem 3 says that there must be some mutually beneficial pairwise merger in this partition with 3 blocs. The fact that country 2 wanted to secede from $\{1, 2\}$ means that the merger of the first two blocs is not mutually beneficial. If blocs 1 and T benefitted from merger, then they would benefit further from adding in country 2. That would imply that the coalition of the whole was better for all countries in the $N-1$ country world than (S', T) contradicting the γ stability of that bloc structure in the $N-1$ country world. Therefore, a merger of 2 with T must be mutually beneficial, a case considered in point v immediately below.

v What has been shown is that if (S, T) is not γ stable, then $S = \{1, 2\}$ and that country 2 and bloc T would gain if 2 seceded from S and joined T.

Therefore, the proof of the theorem will be complete if it can be demonstrated that, in this case, $(\{1\}, \{2\} \cup T)$ is γ stable.

First, the two blocs do not wish to merge, since $(\{1, 2\}, T)$ was γ stable in the $N-1$ country world. Since country 2 is the largest country in the bloc $\{2\} \cup T$, it will not wish to secede unilaterally ; moreover, no bloc containing country 2 would want to secede unilaterally from $\{2\} \cup T$ (since then merging sequentially with the remaining members of T would raise the payoff). Third, suppose that some subset $V \subset T$ were to secede from $\{2\} \cup T$. Its payoff from such a secession would be less than the payoff V would have received from seceding from T in the $N-1$–country bloc structure (S', T) — since this latter secession kept 1 and 2 together. Therefore, the γ stability of (S', T) in the $N-1$ country world shows that no unilateral secession will be made.

Finally, consider the merger of some $V \subset \{2\} \cup T$ with country 1. Since 1 is the largest country, the payoff to the merged group in this bloc structure would be less than the payoff under the coalition of the whole, so that the merger would not be profitable for all parties, completing the proof of the theorem. •