Chapter 2

Taxes and Public Infrastructure Goods

in

THE NEW SYSTEMS COMPETITION

Hans-Werner Sinn

CESifo

Ifo Institute for Economic Research
&
University of Munich
Center for Economic Studies

Yrjö Jahnsson Lectures January 1999
Revised Version, January 2002

March 11th, 2003

Chapter 2

Taxes and Public Infrastructure Goods

Competition between Fiscal Systems

There is no better example of the way systems competition works than tax competition. Everywhere we look there are tax reductions which have been triggered off by this competition. The examples of the falling effective tax rates for the European subsidiaries of US parent companies and the increasing share of labour taxes cited in chapter 1 speak volumes. Globalization has meant that a great many countries are now facing tax competition. The corporate tax rates of all the G7 countries except Italy have fallen in the last 20 years, most of them by more than 10 percentage points. Every country is trying to become an attractive location for investment and the offers being made to willing investors are continually rising. Ireland is using special free trade zones to attract investment, Sweden and Austria have given up the principle of synthetic income taxation, burdening interest income with tax rates that are substantially lower than those on per-sonal income, and Germany has reduced its corporate tax rate from 40% to 25%. It is said that Alabama made Mercedes a tax gift whose value equals 30 years of tax payments to persuade it to locate there. Even Luxembourg believes that it must reduce its already low taxes still further to defend itself against this competition.

A country's attractiveness does not, of course, only depend on its tax rates. The stability of its legal system, protection of property, social harmony, and, not least, its public infrastructure all have a considerable influence on the returns a foreign investor can expect to earn there. This indicates the existence of a natural barrier which weakens the erosive powers of tax competition

– when government services cease to be affordable it is no longer possible to attract more capital by reducing taxes still further.

Nevertheless, one should not be too quick to jump to firm conclusions because, regardless of how useful government services are for mobile capital, the suppliers of immobile factors could always choose to finance these services themselves. Whether and to what extent it is in the national interest to tax mobile capital, and whether the efficient provision of public services is endangered, are not trivial questions. These are the questions which will be discussed in this chapter. To analyse the problem, the level of abstraction is successively reduced. First, a model without public goods will be treated, then a public infrastructure will be added, whose costs are proportional to the capital employed, and finally a realistic type of infrastructure good will be introduced that is used by many firms jointly, that involves congestion externalities, and whose supply is only loosely related to the capital employed.

The Standard Argument

The argument on which practically all the literature on tax competition is based comes from MacDougall (1960) and Richman (1963). Both these authors have pointed out that a small open economy can have no interest in putting a source tax on internationally mobile capital because capital will always be able to shift the tax burden. The source tax drives away mobile capital, and the domestic product and the marginal productivity of the complementary immobile factors both fall. The income of these factors declines, and does so by more than it would if the factors were to pay the tax themselves.

The argument can be shown in a simple way in figure 2.1, which relates to the decision situation of a single country. The country produces a homogeneous output using labour L and capital K, where $f(L,K)$ is a linearly homogeneous production function with the usual properties.

The amount of labour employed is fixed and is provided by domestic residents. The amount of capital employed is variable. Capital is internationally mobile and is available in any amount at the net world market return *r*. The downward sloping line shows the marginal product of capital. When there is no tax, firms invest up to the point where $f_K = r$; that is, they choose the amount of capital K_1.

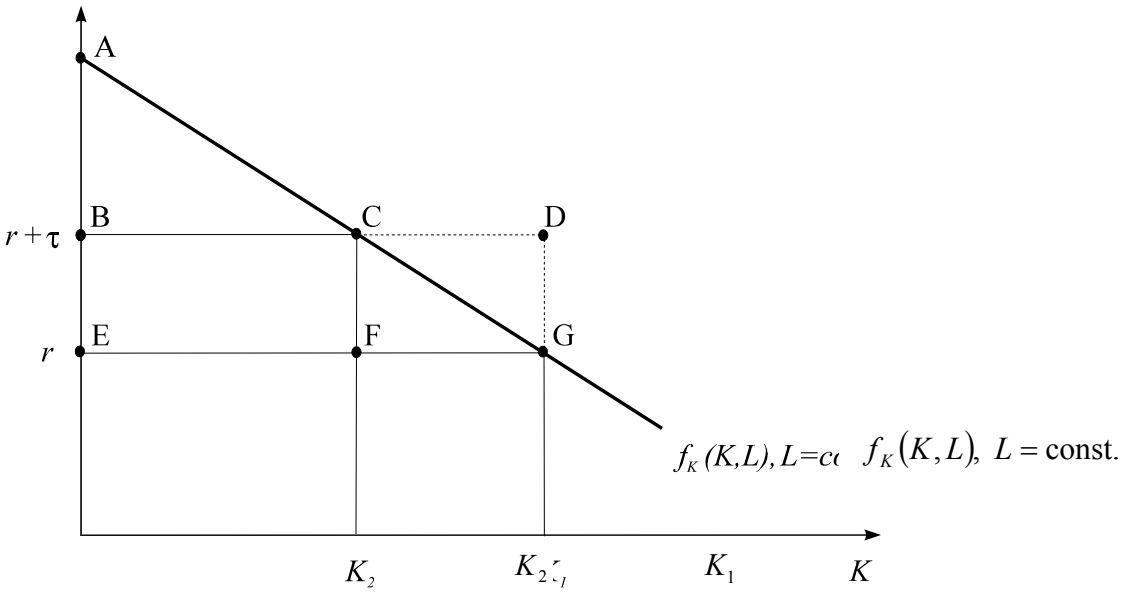

Figure 2.1 The erosion of source taxes in tax competition.

If the government imposes a source tax on capital equal to τ = BE, the amount of capital used will fall to K_2. As the net return *r* is given in the world market, capital leaves the country until its net marginal product after tax is again equal to the given world market level:

$$f_K - \tau = r\,.$$

The tax is shifted completely to the immobile factor. Before taxation the wage income was AGE and when capital leaves it falls to ACB. The tax revenue, which is BCFE, is obviously smaller than the reduction in wages, which is BCGE. Even if the total tax revenue were paid to the wage earners, these would face a loss of CGF. Attempting to tax mobile capital would hurt the owners of the immobile factor themselves. The equilibrium in tax competition between states is therefore K_1. [1]

Whether the mobile capital is domestic or foreign is irrelevant for this reasoning, because, given that the return on the world market is constant, each wealth owner's income from capital is independent of how much capital he invests in the particular country. In a symmetrical international equilibrium it can be assumed without a loss of generality that the wealth $\overline{K}$ of the residents of the country is the same as the equilibrium amount of capital used, K_1.

Competition with Infrastructure Goods

The role of government infrastructure investment is often used as evidence against the pessimistic view of the workability of systems competition based on the results of the previous analysis. It is not only a country's tax burden that is important for location decisions, its infrastructure is important, too. If taxes are seen as the price that must be paid for the publicly provided infrastructure, the investors will accept them. This means that, from the isolated viewpoint of an individual country, it is rational to impose such taxes. A destructive systems competition does not have to be feared.[2]

Figure 2.1 can once again be used to illustrate the infrastructure argument, where r is now to be interpreted as the given net return of capital in countries with a functioning infrastructure.

---

[1] The argument just described has become the basis of the new discussion of tax competition. See, for example, Razin and Sadka (1991).

[2] See Oates and Schwab (1988), Wellisch (1995) or Oates (1995).

Assume that infrastructure costs per unit of capital are BE, so that total infrastructure costs with K_1 capital employed is equal to BDGE. In this case it is not sensible for the country to forego imposing a source tax even when, in principle, it could tax the immobile factor. The reason is easy to see. If the immobile factor pays the tax, its income is AGE - BDGE, that is, ACB - CDG. If the mobile factor pays the tax and τ = BE, the amount of capital used falls to K_2, the infrastructure will be financed by the tax on capital, and the income of the immobile factor, which is no longer being taxed, will be ACB. It is larger by the size of the triangle CDG than if this factor were to pay the tax.

The infrastructure argument is not really convincing in the form used because the infrastructure is modelled as if it were a private intermediate product which has to be used in a fixed proportion with capital. In reality, the public infrastructure is an impure public good with a limited rivalry between competing uses. The state cannot set the tax price for consuming a unit of infrastructure but can only charge for the act of using it jointly with others, and using it does not involve any direct production costs. All that usage imposes are congestion costs which other users have to bear. Pecuniary costs are involved when the infrastructure is improved, but that is a separate government decision.

There is a series of papers which generalize the infrastructure argument to the case of congestion costs. As well as the club goods literature, it includes papers like those of Boadway (1980), Sandler and Tschirhart (1980) or Berglas and Pines (1981). The papers of Gerber and Hewitt (1987), Richter (1994) or Richter, Seitz and Weigard (1996), which follow on from Wildasin (1986), should especially be mentioned. These papers, as a whole, come to an optimistic assessment of systems competition.

However, there are also pessimistic voices. Bewley (1981) argues that increasing returns to scale in the production of infrastructure goods prevents a competitive equilibrium, and Pines

(1991) shows that such an equilibrium probably does not exist when the economies are not ‚replicable'. Zodrow and Mieszkowski (1986) argue that this leads to underprovision of the public infrastructure for private investment.[3]

On the basis of the Selection Principle, this study comes to an ambiguous conclusion about the workability of systems competition between states which take part in infrastructure competition. The study abstracts from the problem of replicability which Pines examined, and it contradicts Zodrow and Mieszkowski's argument. To some extent, it can be interpreted as a generalization of Bewley's analysis to the case of public goods. Bewley examined a publicly provided private good. Here, however, a public infrastructure good will be considered, which is a public good in the sense that it is jointly consumed by all users and cannot be separated among them. To avoid any obvious bias against a market solution, rivalry in use, in the sense of the congestion costs mentioned, will be assumed. This provides an efficiency motive for imposing compensatory taxation, and given that there is a tax revenue it can be used to finance the infrastructure. The approach will be compared with the theory of private clubs, in order to find suitable assumptions for the analysis of systems competition, and used to discuss various policy measures to improve the outcome of systems competition.

Fiscal Competition and Impure Public Goods

An approach is chosen for analysing the infrastructure argument which is familiar from highway congestion models. Using a highway incurs a private unit cost $c(K,W) > 0$ for the user such as time, petrol, car depreciation and the like. The size of this cost depends on the number of usage acts, K, and the capacity of the infrastructure provided by the government, W. Think, for

[3] This argument should not be confused with the view that tax competition leads to an underprovision of public consumption goods. For this see the overview article of Wilson (1999). The issue will be extensively discussed further below in this chapter.

example, of the number of trips over the highway and the width of this highway. It is not by chance that the variable K stands for both the number of usage acts and the amount of capital employed. It is assumed that the two variables are proportional to one another so that, with an appropriate choice of units, they can be taken to be equal numerically. The properties of the usage cost function are given by the derivatives $c_K \geq 0$ and $c_W < 0$. When $c_K = 0$ we have a pure public good in the sense of Lindahl, Musgrave and Samuelson for which there is no rivalry in use. Where $c_K > 0$ we have an impure public good where the users get in each others' way. It is assumed that the function c is homogeneous of degree λ where the sign of λ is still to be determined. The usage cost function was first used by Mohring and Harwitz (1962), and it is well known in the literature on congestion externalities.[4]

The usage cost is quite distinct from the production cost of the public good. Without limiting the generality, it can be assumed that the production cost per unit of capacity has the fixed value $\rho > 0$ so that the total cost for the provision of the public good is $\rho \cdot W$. By contrast, the total cost of using the public good is $c(K,W) \cdot K$.

As before it is assumed that a homogeneous output is produced according to the linearly homogeneous production function $f(K,L)$, where K is the amount of capital employed and L the amount of labour employed. Capital is completely mobile internationally and, for the time being, labour is assumed to be completely immobile – it is inelastically supplied and cannot migrate across borders. The net return r which the capital can earn in other countries after tax and after deducting the usage cost of the infrastructure is constant from the point of view of the (small) individual country. The country only has a source tax on capital and a head tax on labour

[4] Cf. Oakland (1972) and Boadway (1980). For the role of public goods in the production function, see also Pfähler (1995).

available to it and these are raised at the rates τ and ω.[5] The residents of the country have a fixed amount of wealth, $\overline{K}$, which they can invest at home or abroad. Their capital income is $r \cdot \overline{K}$.

Profit maximizing firms invest capital up to the point where the marginal productivity of capital is equal to the sum of the marginal interest, usage and tax costs:

$$f_K(K,L) = r + c(K,W) + \tau . \tag{2.1}$$

It should be noted that $c(K,W)$ measures both the average social usage cost and the marginal private usage cost. The marginal social usage cost on the other hand is $c + c_K \cdot K$, where $c_K \cdot K$ is a marginal congestion externality which the representative firm does not take into account in its planning. In what follows, τ and W are taken to be the choice variables of the government. The lump sum labour tax rate ω is endogenously determined such that the government budget is balanced:

$$\omega L = \rho W - \tau K . \tag{2.2}$$

Note that this equation holds algebraically and is compatible with any sign of the labour tax revenue. If the tax on capital generates more revenue than needed for the provision of the public infrastructure, there will be a subsidy to labour to balance the budget. Taking account of the constraints (2.1) and (2.2), the government's aim is to maximize the rents R of the domestic residents. R is the sum of the gross wage income, which is output minus the return to the factor

[5] For analytical convenience, the source tax is modelled as a periodic levy on the stock of capital rather than on the return of capital.

capital,[6] and the net interest income, minus the labour tax whose revenue is needed to cover a potential deficit in the provision of the infrastructure:

$$R = \left(f - f_K \cdot K\right) + r\overline{K} - \omega L \,. \tag{2.3}$$

Inserting (2.1) and (2.2) into (2.3) gives the expression

$$R = f\left(K, L\right) - r\left(K - \overline{K}\right) - c\left(K, W\right)K - \rho W \,. \tag{2.4}$$

This shows that the total rent can also be expressed as the difference between the output and the sum of the interest cost of the imported capital, the total usage cost, and the total cost of providing the public infrastructure. Equation (2.4) contains an implicit tax shifting result which is due to the fact that, according to (2.1), firms adjust competitively to the given world market rate of interest. Given that the single country can take the total capital income $r \cdot \overline{K}$ as given and the labour tax clears the budget, variations in the usage cost $c \cdot K$ and the cost of providing the infrastructure, ρW, are fully absorbed by the income of the immobile factor (labour).

Knowing this, the government tries to adjust the tax rate τ and the capacity of the public good W in a way that maximizes the rent of domestic citizens. As (2.1) shows that K is a monotonically declining function of τ, this also implies that (2.4) can be maximised by the choice of K and W. The first-order conditions for a national policy optimum are

---

[6] It should be noted that the assumption of a linearly homogeneous production function ensures that $f - f_K \cdot K$ is just sufficient to pay the employees their marginal productivity wage $f_L \cdot L$.

$$f_K = r + c + c_K \cdot K \tag{2.5}$$

and

$$-c_W \cdot K = \rho . \tag{2.6}$$

Equation (2.5) requires that the marginal product of capital equal the marginal social cost of capital which is the sum of the marginal interest cost *r*, the marginal individual usage cost *c*, and the marginal congestion externality $c_K \cdot K$. Equation (2.6) is the Samuelson condition for the optimal provision of public goods. If the capacity of the public good is increased by one unit, the usage cost per trip changes by c_W. The negative of this value is the marginal willingness to pay per trip. The multiplication with *K* is the same as summing over all trips. The condition therefore says that the sum of all users' marginal willingness to pay be equal to the marginal cost of providing the infrastructure.

The optimal benefit tax rate

Comparing (2.1) with (2.5), it can be seen that, in order to achieve the optimum, the government sets a tax rate equal to the marginal congestion externality:

$$\tau = c_K \cdot K . \tag{2.7}$$

This is not the same as the choice of an optimal tax rate equal to the marginal cost of providing a private infrastructure good shown in the previous section, because this cost bears no obvious relationship to the congestion cost.

The choice of the optimal tax rate is illustrated in figure 2.2. Here a symmetrical equilibrium is assumed, in which every country invests as much as it owns, $K = \overline{K}$. The figure shows the level

of the internationally given net return r, the private marginal cost of capital $c + r$, the social marginal cost of capital $c + c_K \cdot K + r$, and the marginal product of capital $f_K(K, L)$. The area under the marginal product curve represents total output, and the area under the social marginal cost curve represents total social costs. The total rent is the difference between the two areas, ACD, plus the fixed capital income $r \cdot \overline{K}$ (i.e. the area FGIH) and minus the cost ρW of providing the public good, which is not shown in the figure. It is obvious that total revenue is maximized at the intersection of the two curves. The resulting tax rate τ is CE, and is exactly equal to the marginal congestion externality. A higher tax rate would drive too big a wedge between f_K and $c + r$, and a smaller rate would drive too small a wedge between them; too little or too much investment would thus be induced.

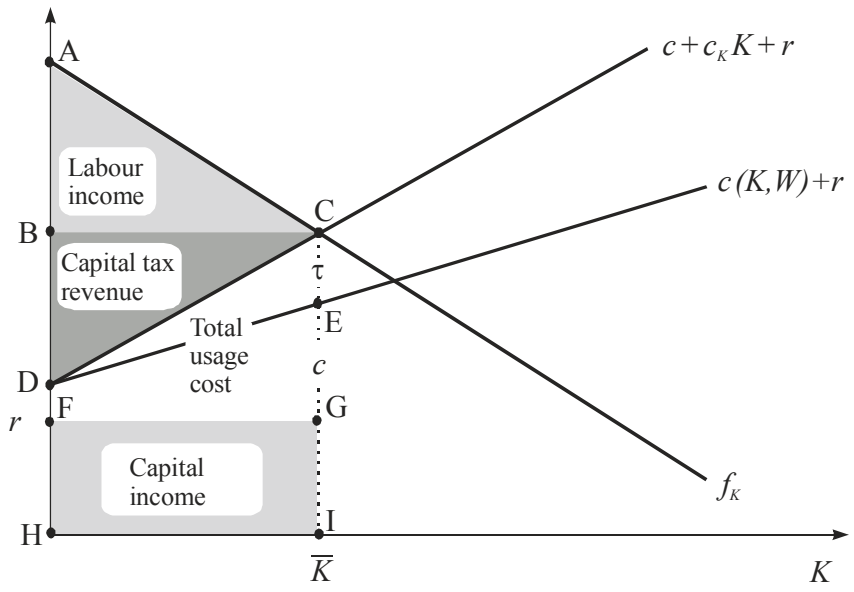

Figure 2.2 The optimal compensatory tax rate.

In accordance with equations (2.2) and (2.3), the total rent can also be shown as the sum of labour income, capital income and capital tax revenue minus the cost of providing the infrastructure. In the optimum, labour income equals ACB, capital income equals FGIH and capital tax revenue equals BCD. The latter follows from the fact that the sum of the interest cost FGIH, the usage cost DCGF and the capital tax revenue is equal to the gross income BCIH that accrues to capital.[7] If K falls below $\overline{K}$, the sum of labour income and capital tax revenue is reduced because a triangle to the left of point C must be taken from the corresponding area in figure 2. The sum of labour income and tax revenue is also reduced by a triangle to the right of C if K is increased beyond $\overline{K}$. Thus, the sum of labour income and tax revenue is maximised at point C.

The Social Optimum

How does the competitive equilibrium thus characterized compare with the international social optimum? Suppose there is a supra-national central planner who chooses the international capital allocation and the respective national provisions of public goods such that the sum of all rents is maximized. As the aggregate stock of capital invested equals the given aggregate stock of wealth, the planner's goal can be taken to be

---

[7] It can also be seen that BCD represents the tax revenue when it is considered that $\tau \cdot K = (\tau + c)K - K \cdot c$ and $K \cdot c = \int_0^K [c(u,W) + c_u(u,W) \cdot u]\,\mathrm{d}u$. If $K = \overline{K}$, ($\tau + c$)K corresponds to the rectangle BCGF and the integral (the usage cost $K \cdot c$) to the white area DCGF. The difference between these areas is the triangle BCD. It is equal to the capital tax burden $\tau \cdot K$.

$$\max_{K_i, W_i} \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} f(K_i, L_i) - c(K_i, W_i)K_i - \rho W_i \qquad \forall\ i = 1,..., n,$$

$$\text{s.t. } \sum_{i=1}^{n} K_i = \sum_{i=1}^{n} \overline{K}_i \ ,$$

where the subscript *i* denotes a particular country and *n* is the total number of countries. It can easily be shown that, in the optimum, it is necessary that

$$f_{K_i} - c(K_i, W_i) - c_{K_i} \cdot K_i = f_{K_j} - c(K_j, W_j) - c_{K_j} \cdot K_j \qquad \forall\ \ i, j = 1,..., n,$$

which says that the marginal product of capital net of the marginal social cost of using the infrastructure be equal in all countries. This is an obvious generalization of the usual requirement for an efficient allocation of a given stock of capital to competing uses. It coincides with national optimality condition (2.5) since there is a common rate of return to capital *r* for all countries. (Firms invest up to the point where their respective marginal product of capital net of the national source tax and net of the private infrastructure usage cost equals the common rate of return and national governments impose source tax rates that equal the marginal congestion externalities.) Similarly, the central planners marginal condition for an optimal provision of public goods,

$$-c_{W_i} \cdot K_i = \rho \qquad \forall\ i = 1,..., n\ ,$$

coincides with national optimality condition (2.6). A proposition summarizes this result.

Proposition 2.1: *The equilibrium in systems competition is efficient. Both the international allocation of capital and the pattern of infrastructure provision are chosen so as to maximize the sum of all rents accruing in all countries.*

Proposition 2.1 sheds a rather favourable light on systems competition from an efficiency perspective. The Invisible Hand does seem to work in the context of systems competition, and the pessimism implied by the Selection Principle seems unfounded. Unfortunately, however, the distributional implications of this equilibrium are less convincing, and there may even be existence problems. These issues will be considered in the next few sections.

Who Pays for the Infrastructure?

As to the distribution problem, the relevant question is: how high is the congestion-driven capital tax revenue the single government collects in systems competition relative to the cost of providing the public infrastructure? Will the competitive government be able to demand from capital a contribution to financing the budget which exceeds the cost of accommodating this capital? Or must the government accept a deficit which would have to be covered by a levy on the fixed factor?

That the latter is a relevant possibility is obvious when the infrastructure is a pure public good. There is no rivalry between the users $(c_K = 0)$ with pure public goods and thus no tax is levied to internalize the marginal congestion cost. The immobile factor bears the burden of paying tax alone. (In figure 2.2 the social and private marginal cost curves become horizontal and the area BCD that represents the tax revenue shrinks to zero.)

The assumption of a pure public good is, however, not realistic. In the more general case of strictly positive congestion externalities, there is a revenue from capital taxation, but it is not

clear whether it is sufficiently large to finance the infrastructure. To see how the capital tax revenue is related to the cost of providing the infrastructure, note that Euler's Theorem implies

$$c_K \cdot K + c_W \cdot W = \lambda c \tag{2.8}$$

where λ is the degree of homogeneity of the usage cost function $c(K,W)$. Inserting the maximization conditions (2.6) and (2.7) into (2.8) gives the expression

$$\tau K = \rho W + \lambda c K. \tag{2.9}$$

The basic insight of this formula, which has already been gained by Mohring and Harwitz (1962, pp. 85-7), can be formulated as follows.

Proposition 2.2: *The optimal congestion charge is sufficient to finance the public infrastructure when* $\lambda \geq 0$*, that is, when the usage cost function does not have a negative degree of homogeneity. If* $\lambda < 0$ *there will be a fiscal deficit that must be covered by taxing the immobile factor.*

To understand the importance of the degree of homogeneity, it can be asked whether the average usage cost of the public infrastructure falls, rises, or remains constant when both the number of usage acts (trips) and the cost of providing the infrastructure double. In the first case $\lambda < 0$, in the second $\lambda > 0$ and in the third $\lambda = 0$. Analogously, it can be asked by how much must the public expenditure on the infrastructure increase in order to keep the average usage cost, and thus the perceived quality of the infrastructure, constant when there are twice as many usage acts.

If, for example, the expenditure must more than double then, $\lambda > 0$ and self-financing of the infrastructure with an efficiently designed congestion charge is warranted. Conversely, the optimal congestion charge is not sufficient to finance the infrastructure if less than doubling the public expenditure suffices to keep the quality of the infrastructure constant when the number of usage acts doubles ($\lambda < 0$).

The problem is broadly equivalent to the problem of returns to scale in the production of a private good. Only when there are falling or constant returns to scale will marginal cost pricing generate enough revenue to cover the total cost of production. However, when there are increasing returns to scale, there is a financing deficit which has to be covered elsewhere, similar to the case of pure public goods. This is the case supporting the pessimistic suspicion of chapter 1 that wage earners are the victims of systems competition.

The Selection Principle, the Theory of Clubs and the Race Below the Bottom

What is now the appropriate assumption about the sign of λ? At first sight $\lambda \approx 0$ would appear to be plausible as constant returns to scale is one of the usual assumptions of microeconomic production theory. But one needs to be careful here. An uncritical use of assumptions from the theory of private markets is not permissible in the light of the Selection Principle introduced in chapter 1 because it neglects the reasons that induced government intervention. If governments do what they ought to do and do not provide public goods which could also be provided by private clubs, arbitrary assumptions about the size λ would not be appropriate.

A look at the theory of clubs as derived by Buchanan (1965), Boadway (1980) and Berglas and Pines (1981) quickly shows which assumptions about λ the Selection Principle requires. It is assumed that there are $i = 1,\ldots, m$ identical private clubs which supply the infrastructure at entrance fees $\tau_1,...,\tau_m$. The qualities of the services supplied by the clubs may differ and they are

inversely related to the usage costs $c(K_i, W_i)$, where K_i is the number of usage acts sold by, and W_i is the capacity of, club *i*. The production cost per unit of capacity is once again ρ. Let *P* be the overall usage price of club services in the sense that it incorporates both the pecuniary entrance fee and the non-pecuniary individual usage cost. In a competitive equilibrium members must be indifferent between all clubs and thus there will be a uniform overall usage price:

$$P \equiv \tau_i + c(K_i,\, W_i) = \tau_j + c(K_j,\, W_j) \qquad \forall \;\; i, j = 1,\, ...,\, m\,.$$

The (small) individual club *i* takes *P* as a magnitude given by the market and chooses K_i and τ_i or, equivalently, K_i and W_i such that its profit is maximized:

$$\max_{K_i, W_i} \; \left[P - c(K_i,\, W_i)\right] K_i - \rho W_i \;.$$

The necessary conditions for an interior optimum are

$$-c_{W_i} K_i = \rho \tag{2.10}$$

and

$$\tau_i = c_{K_i} K_i \,. \tag{2.11}$$

They correspond to conditions (2.6) and (2.7). The private club also provides a capacity which meets the Samuelson condition and chooses an entrance fee which adequately takes account of the internal congestion costs and the resulting reduction in the quality of its services. Analogously to (2.9), the condition

$$\tau_i \cdot K_i = \rho \cdot W_i + \lambda \cdot c(K_i, W_i) \cdot K_i$$

follows from (2.8), (2.10) and (2.11). It shows that private clubs can operate without loss in a competitive equilibrium when $\lambda \geq 0$. If, on the other hand, $\lambda < 0$, the competitive process would be ruinous and would not find an equilibrium. If the government limits itself to its genuine responsibilities and does not take away the legitimate business of the clubs, then it will only step in where clubs would engage in ruinous competition. It follows that the only reasonable assumption for a competition between states is that $\lambda < 0$.

Proposition 2.3: *The Selection Principle implies that the state limits itself to the provision of those public goods for which $\lambda < 0$. An efficient congestion charge for the use of the public infrastructure is therefore not sufficient to finance the cost of providing this infrastructure*.

Whether the state is actually constructed in accordance with the Selection Principle is an empirical question, but the information available about this is less than adequate. According to a study by Borcherding and Deacon (1972) city size and expenditure on public goods are proportional, which implies that $\lambda = 0$.[8] The authors assume, however, that the quality of the public goods provided is independent of the city size, without giving any statistical information that this is actually true. When, contrary to this assumption, the quality of provision increases with city size, Borcherding and Deacon's research implies that λ is negative. A study in which the quality of the public good is measured is that of Brueckner (1981). Brueckner examines the

[8] See also Blankart (1996, p. 89).

provision with public fire brigades, measuring the quality of the protection they offer in terms of the size of private fire insurance premia, and comes to the unambiguous conclusion that there are economies of scale in the provision of public fire protection ($\lambda < 0$). While this is some indication of economies of scale, it is not compelling evidence, since lower insurance premia in big cities could also have other causes than better fire protection including, for example, a more competitive insurance market. In a review article screening the existing literature, Weichenrieder and Reiter (1997) come to the conclusion that the matter is still undecided.[9]

One problem with the existing literature is that all studies relate to local rather than national public goods, which are the subject of this book. Examples of national public goods are the legal system, defence, highways, and the administrative services of the federal government. The empirical studies at the local level cannot be transferred indiscriminately to these examples, because a variant of the Selection Principle may also apply to the allocation of public functions to the different levels of government. If municipalities and provinces have been optimally designed, they will only provide those public goods for which low-scale minima of the user cost functions with regard to the number of usage acts K and the capacity W are attainable and they will operate at these minima. At the margin, the average usage cost function $c(K,W)$ then has a degree of homogeneity equal to zero, and workable governmental competition among provinces or cities is conceivable. However, for the very same reason, public goods with increasing returns to scale will then be assembled at the national level, and the difficulties pointed out above will arise.

It may be that the foundation of a European state will at some stage in history make it possible to apply the Selection Principle to the nation states because they may then have been designed optimally in the sense discussed. However, the current situation in Europe is far removed from that situation. As long as the small European nation states carry out functions which in the United

---

[9] Cf. Holocombe and Sobel (1995) and Walzer (1972).

States are considered federal tasks and as long as no superior level of governments exists that could absorb the activities unsuitable for fiscal competition, increasing returns to scale at the margin seem likely, and under that condition it must be expected that fiscal competition will lead to financing deficits.

The result shows that fiscal competition among states has ruinous aspects when (central) governments step in to avoid ruinous competition in the private sector (or among lower levels of government). However, this does not mean that governments go bankrupt, that there is no equilibrium or that the equilibrium is inefficient. In the case considered, it is still in the interest of domestic residents to bear the deficit arising from the accommodation of the mobile factor capital itself.

As capital income is fixed in the rest of the world, the optimization result also implies that the subsidization of mobile capital by means of a labour tax will maximize net-of-tax labour income. Thus, if we distinguished between capital owners and workers, it would even be in the interest of workers to vote for a tax on labour income helping to finance the public infrastructure used by capital alone.

Nevertheless, the fiscal deficit is a problem. In the absence of tax competition, a fiscal tax on capital is feasible (i.e. a tax burden can be imposed on capital which is greater than the cost of the infrastructure it uses and which allows the government to finance its other responsibilities). When there is tax competition, only benefit taxes on capital are possible. From a distributional perspective this is already a problem. What is worse, however, is that only marginal benefit taxes are possible whose revenue is insufficient to cover the cost of the infrastructure. Capital receives a net subsidy at the expense of immobile taxpayers. Tax competition not only implies a race to the bottom, as is often argued, but in a certain sense it also even implies a race below the bottom.

This gives a deeper meaning to the OECD's (1998) fear that the global economy suffers from ‚harmful tax competition'.

Tax Harmonization and the Overprovision of Public Goods

To avoid the race below the bottom, compensating policy measures could be sought, both at a national and at an international level. Such measures will be discussed next.

One conceivable measure to avoid the distributional consequences is the international harmonization of capital tax rates at a level above the one resulting from fiscal competition. If capital is taxed excessively, it can escape from one country, but not from all countries. The tax authorities therefore gain more power if they commit themselves to jointly determine their tax rates. By fixing the tax rates above the competitive level, the tax authorities can try to collect more revenue from capital and mitigate the distributional consequences.

The problem with tax harmonization, however, is that it eliminates only one of two competition parameters available to the national government. Despite the fixing of the capital tax rate, labour taxation still leaves the free choice of how much public infrastructure to provide, and it is unclear whether the government will continue to chose an infrastructure W compatible with the Samuelson condition.

To see how the rent of the domestic population reacts to an increase in the provision of public infrastructure, given the capital tax rate, differentiate equation (2.4):

$$\left.\frac{\mathrm{d}R}{\mathrm{d}W}\right|_{\tau=\text{const.}} = \left(f_K - r - c_K \cdot K - c\right)\cdot \Phi - c_W \cdot K - \rho\,. \qquad (2.12)$$

Here Φ is the reaction coefficient for capital which results from implicitly differentiating the arbitrage condition (2.1) with given τ,

$$\Phi \equiv \left.\frac{\mathrm{d}K}{\mathrm{d}W}\right|_{\tau=\text{const.}} = \frac{c_W}{f_{KK} - c_K} > 0. \tag{2.13}$$

It is obvious that this coefficient is always strictly positive: an improvement in the infrastructure with a given capital tax rate attracts more capital into the country. Using equation (2.1), it follows from (2.12) that

$$\left.\frac{\mathrm{d}R}{\mathrm{d}W}\right|_{\tau=\text{const.}} = (\tau - c_K \cdot K)\cdot \Phi - c_W \cdot K - \rho\ . \tag{2.14}$$

In the national optimum, $\mathrm{d}R/\mathrm{d}W = 0$ and hence

$$(\tau - c_K \cdot K)\cdot \Phi = \rho + c_W \cdot K \tag{2.15}$$

is the optimality condition. In the unconstrained optimum discussed previously, it follows from (2.7) that $\tau - c_K \cdot K = 0$ and from (2.6) that $\rho + c_W \cdot K = 0$. Thus, of course, condition (2.15) would automatically hold if the harmonization constraint on the national tax rate were not binding.

With an effective constraint, however, which forces τ to obtain a value above the marginal congestion externality $c_K \cdot K$, the left-hand side of the expression is strictly positive, and so the

right-hand side must be positive, too. The marginal willingness to pay, summed over all usage acts, is less than the cost of providing the public infrastructure,

$$-c_W(K,W)\cdot K<\rho, \tag{2.16}$$

indicating an oversupply of the infrastructure as measured by the Samuelson rule. Given that capital is deterred by a tax rate higher than necessary to cover the marginal congestion externality, it pays domestic residents to lure more of it into the country by offering a better infrastructure.

Equation (2.16) describes one country's reaction to an exogenous constraint on the domestic source tax rate, given the world market rate of interest. If this country were the only one to face such a constraint, its capital stock would shrink and (2.16) might be satisfied even at a level of W which is equal to the unconstrained optimum. There would be an oversupply of infrastructure relative to a constrained optimum in which the stock of capital is too small, but not necessarily relative to the laissez faire optimum. However, in fact, all countries face the same constraint. Thus the world market rate of interest has to adjust so as to satisfy equation (2.1), and in a symmetrical equilibrium with a given world stock of capital the same amount of capital will be working in each single country as in the absence of constraints. Expression (2.16) therefore not only implies that the Samuelson rule is violated, but also that more infrastructure is provided than in a laissez faire equilibrium where τ and W are both chosen competitively.[10]

[10] Fuest (1995) offers a useful approach with tax and infrastructure competition where capital and infrastructure inputs are mobile across the borders. He finds that tax harmonization may lower social welfare if infrastructure competition continues. However, in his model there is no overprovision of public infrastructure.

Proposition 2.4: *Tax harmonization intensifies infrastructure competition and leads to an overprovision of the public infrastructure in equilibrium.*

This result is undesirable not only from an efficiency but also from a distributional perspective. At least part of the extra tax revenue resulting from a harmonized tax rate increase will dissipate by financing the excessive amount of infrastructure at the expense of the fixed factor. To avoid this consequence, harmonizing the supply of infrastructure goods between the countries in addition to harmonizing the tax rates could be considered as a policy measure. This is only a theoretical possibility, however. Given that the infrastructure needs and preferences of the different countries are as different as they are, a useful harmonization of infrastructure expenditures is hard to imagine in practice.

Self-Financing Constraints and Average Cost Pricing

As was shown, tax harmonization is a problematic means to avoid the distributional consequences of unbridled tax competition. It induces countries to engage in excessive infrastructure competition which counters the distributional goal and leads to an inefficiently large provision of infrastructure. An alternative examined in this section is the imposition of a self-financing constraint on capital. Suppose instead of harmonizing the tax rates, the states agree not to subsidize capital and to finance the infrastructure investment exclusively with capital charges, leaving the immobile factors tax free (or charging them only with taxes for the public goods they need themselves). Otherwise, the states retain their rights to choose the policies they want. By definition, this policy prevents labour from subsidizing capital, and perhaps it is also a useful corrective device that avoids or mitigates the overprovision of infrastructure goods.

To find the answer, eliminate the labour tax in the government budget constraint (2.2), $\omega = 0$, so that this constraint becomes a self-financing constraint:

$$\rho W = \tau K \,. \tag{2.17}$$

As before, the single competitive government seeks to maximize the sum of all national rents. Differentiating equation (2.3), given the labour tax ω, yields the following first-order condition for a national maximum:

$$\left.\frac{\mathrm{d}R}{\mathrm{d}\tau}\right|_{\tau K=\rho W} = -\varphi \cdot f_{KK} \cdot K = 0 \,. \tag{2.18}$$

Here

$$\varphi \equiv \left.\frac{\mathrm{d}K}{\mathrm{d}\tau}\right|_{\tau K=\rho W} = \frac{1 + c_W \dfrac{K}{\rho}}{f_{KK} - c_K - c_W \dfrac{\tau}{\rho}} \tag{2.19}$$

is a differential quotient for the response of capital to a marginal increase in the tax rate under the condition of self-financing which follows from equations (2.1) and (2.17). The differential quotient combines the deterrent effect of the tax, one divided by the denominator, and the attraction effect of the public infrastructure, the second item of the numerator divided by the denominator. The denominator is negative if, as is assumed, the second-order condition of the

government's optimization problem is met.[11] As $f_{KK} < 0$, this ensures that the sign of $\mathrm{d}R/\mathrm{d}\tau$ is simply the negative of the sign of the numerator of (2.19). In the optimum, the numerator of (2.19) has to be zero, so as to satisfy (2.18). It follows that the condition for an optimal provision of infrastructure with a self-financing constraint is given by the equation

$$\rho = -c_W(K,W) \cdot K \,.$$

As this is the Samuelson condition (2.6), the initial suspicion is confirmed. Indeed, a self-financing constraint will do the job and induce the competitive government to provide an efficient amount of infrastructure.

As the self-financing constraint excludes any subsidies paid by the immobile factor, it is also clear that the income of this factor, and hence also the total sum of rents, is maximized when the tax-infrastructure combination is chosen which meets the Samuelson condition and maximizes the country's capital import.

To interpret the result, suppose the tax rate is gradually raised, starting from zero. Initially, $\rho < -c_W K$, and hence the numerator of (2.19) is negative such that (2.19) itself is positive and, from (2.18), $\mathrm{d}R/\mathrm{d}\tau > 0$. The attraction effect of the infrastructure which can be financed with the additional tax revenue outweighs the deterrent effect of the tax. However, the higher the tax rate

---

[11] The second-order condition is $\mathrm{d}^2R/\mathrm{d}\tau^2 < 0$ or, which comes to the same thing, $\mathrm{d}\varphi/\mathrm{d}\tau = \mathrm{d}^2K/\mathrm{d}\tau^2 < 0$. Since $W = \tau K/\rho$, this condition becomes $\mathrm{d}\varphi/\mathrm{d}\tau = c_{WW} \cdot (K/\rho)^2 / D < 0$, where D is the denominator of the expression on the right-hand side of (2.19): $D = f_{KK} - c_K - c_W \tau/\rho$. Since $c_{WW} > 0$, the second-order condition is met when $D < 0$, which in turn requires $f_{KK} - c_K < c_W \tau/\rho$. Since $f_{KK} < 0$ and $c_K > 0$, $c_W \tau/\rho$ can thus not be too strongly negative. It is assumed that this is the case.

already is, the smaller is the overweight. In the optimum the two effects are balanced at the margin and this is precisely the point where the supply of public goods meets the Samuelson condition.

Since the Samuelson condition still holds true despite the self-financing constraint, the budget deficit must be avoided through an infringement of the condition for an optimal congestion charge. In fact, taking account of (2.6), (2.8) and the self-financing constraint (2.17), one can derive an expression for the optimal tax rate,

$$\tau = c_K \cdot K - \lambda c,$$

which, when $\lambda \neq 0$, is not compatible with a congestion cost charge that covers the marginal crowding externality (see 2.7). In the relevant case where $\lambda < 0$, the tax rate is greater than the marginal crowding externality $c_K \cdot K$.

If only one country imposed the self-financing constraint and the world market rate of interest could therefore be taken as given, less capital would be invested than in the first best optimum and, with the smaller amount of capital employed, the sum of the marginal willingness to pay, $-c_W K$, would be smaller. Accordingly, the Samuelson rule (2.6) would imply a smaller optimal capacity W of the infrastructure, too. However, once again, this is not what was assumed. If all countries are identical and all face the same self-financing constraint then the amount of capital employed in each country must be the same as is the case where there is no self-financing constraint. Hence the world market rate of return will have to fall so as to satisfy equation (2.1) despite the higher tax rate on capital. In this case not only the Samuelson condition holds, but in addition the amount of infrastructure it implies is the same as without the self-financing constraint.

Given that the same infrastructure capacity and the same amount of capital are available, but self-financing is required, it is clear that gross wage income is the same and net-of-tax wage income is higher than in the case without any constraints. The distributional goal aimed at with the self-financing constraint is achieved.

Proposition 2.5: *Suppose that all countries agree to respect a self-financing constraint which obligates them to charge capital with the full cost of the infrastructure it uses despite economies of scale in infrastructure investment. Then the single competitive country will provide an efficient amount of infrastructure satisfying the Samuelson rule, and in a symmetrical equilibrium it will even provide the same amount as in a first best optimum. The imposition of a self-financing constraint increases the net-of-tax wage income at the expense of capital income*.

This result suggests that average cost pricing for public infrastructure may be a more useful rule for the EU countries than was previously thought. In its 1998 White Paper, ‚Fair Payment for Infrastructure Use', which concentrated on road and railway issues, the European Commission argued that public infrastructure needs marginal cost pricing, and it pointed out that crowding or congestion externalities are the dominant component of the marginal cost to be considered in this context. This view is well founded, as the laissez faire version of the above model has shown, but it should be clear now that when the Selection Principle holds and increasing returns to scale are the normal case for state-run operations, marginal cost pricing implies prices below the average cost of the infrastructure that need to be accompanied by a public subsidy from other sources. If distributional goals matter in addition to efficiency considerations in the EU – and they certainly do – then average cost pricing and the exclusion of subsidies can also be legitimated. When constraints force countries to use average cost prices, they can keep full autonomy concerning

their infrastructure decisions and will make the right allocative decisions without incurring income losses for the fixed factors.

This demonstrates the point made in the introductory chapter that it would be a mistake to conclude from a failure in systems competition that a centralized solution must necessarily be sought. Sometimes it is possible to establish rules of conduct for the competing countries which make systems competition workable.

The legal implementation of average cost pricing could be facilitated by extending the existing prohibition on subsidies.[12] According to the present EU rules, the individual countries are forbidden to subsidize firms directly, and it is one of the most visible, and by now accepted, limitations of a country's sovereignty when the European Commission prohibits a national subsidy, as it frequently does. If this prohibition is extended to cover indirect subsidies through reducing the price of using the public infrastructure below the average cost of providing it, then the undesirable distributional implications of the unconstrained competitive equilibrium can be avoided.

A Critique of the Underprovision Hypothesis

The optimism about the efficiency of the supply of public goods in systems competition that emerges from proposition 2.5 contrasts sharply with the underprovision hypothesis of Zodrow and Mieszkowsi (1986) and Wilson (1986), which is one of the most frequently cited results in the tax competition literature.[13] The point this literature makes is that capital mobility increases the supply elasticity of the taxable factors, thereby raises the marginal cost of public funds, and hence reduces the optimal supply of public goods.

---

[12] See Article 92 of the Treaty Establishing the European Community.

[13] See also Oates (1972) and Wilson (1999).

In its basic version the argument was derived by assuming that only a source tax on capital is available to finance a public consumption good and that the fixed factor cannot be taxed. If the fixed factor can be taxed, the cost of a euro of public funds is one euro, because there is no excess burden from taxation. Thus the optimal supply of public consumption goods is determined by the Samuelson rule according to which the sum of all consumers' marginal willingness to pay equals the marginal cost of providing the public good. If, however, only mobile capital can be taxed, the cost of raising one euro of public funds is above one euro, because the excess burden comes in addition to the mere financing cost, and accordingly fewer public goods are supplied than implied by the Samuelson rule.

This can easily be understood by once again looking at figure 2.1 which showed that labour bears more than the full burden of a source tax on capital. In order to raise the source tax revenue $\tau \cdot K$, or BCFE, while keeping capital's net rate of return at par with the given world level, capital must be exported so as to raise its pre-tax rate of return at home. Labour, which is the complementary factor of production, will then experience a lower marginal product and receive a lower income. However, labour income falls not only by the tax revenue imposed on capital. In addition it falls by the amount $-\tau \cdot \Delta K/2$ or CGF where ΔK is the capital export. Thus, the cost of public funds exceeds the revenue raised by this extra decline in labour income, and the underprovision result follows.

The Zodrow-Mieszkowski-Wilson approach has a certain formal similarity with the self-financing constraint introduced in the previous section in lieu of tax harmonization. Why is it that the approach of the previous section implied an efficient provision of public goods while the Zodrow-Mieszkowski-Wilson approach implies an underprovision?

The basic reason lies in the fact that the latter refers to public consumption goods rather than intermediate goods which would benefit private firms, and that countries do not compete in

public consumption goods. While public consumption goods do not benefit the factor capital, the taxes needed to finance them are assumed to be imposed on this factor. This asymmetry makes financing and providing public consumption goods a difficult matter, with a substantial excess burden of taxation being involved. By contrast, public intermediate goods by themselves attract capital. This makes financing and providing the goods easy, and on balance no excess burden occurs. Taxing capital to finance public consumption goods results in underprovision, but taxing capital to finance public intermediate goods results in an efficient provision.

This emphasizes once again the importance of distributional issues. As public consumption goods benefit primarily the recipients of labour income, the imposition of a capital tax for the purpose of financing these goods is a measure which can be justified with a distributional objective in mind. In a sense, therefore, the lesson of the underprovision literature is that the attempt to redistribute from capital to labour is costly and futile if capital is mobile. This is similar to the basic message derived by MacDougall (1960) and Richman (1963), the only difference being that now redistribution in kind instead of a money transfer is considered.

As a counter argument against the view that it is the distinction between public consumption and intermediate goods that matters it could be pointed out that Zodrow and Mieszkowski did not limit their approach to public consumption goods but also extended it to the case of public intermediate goods (1986, pp. 362-6). Did they not show that these goods will also be underprovided in fiscal competition? They did not. Unfortunately, this part of their analysis suffers from a formal mistake that drives their results. The authors *assume* that an increase in the tax rate, despite the improvement in the infrastructure associated with it, has, on balance, a negative effect on the amount of capital employed, without recognizing that such a negative effect is an indication of a too high tax rate and an overprovision of infrastructure. In fact, they derive the alleged underprovision from the implicit assumption of an overprovision of

infrastructure. The Zodrow-Mieszkowski model is presented in the appendix to this chapter where the mistake in their reasoning is explained in detail.

Residence Taxes

Let us return to the question of how the distributional consequences of systems competition, in particular the need to subsidize capital, can be avoided. Are there measures which, unlike the collective imposition of self-financing constraints, can be carried out by a single country alone without finding the consent of others?

One possibility, recommended by Bucovetsky and Wilson (1991), is replacing the principle of taxing at source with the residence tax principle. When the revenue that domestic capital owners earn in foreign countries can be taxed by the domestic authorities, the evasive reaction of transferring the investment location is no longer possible. In the model this means that $r\overline{K}$ will become the tax base and the capital income tax will become a lump sum tax.[14]

In principle the residence principle can be adopted unilaterally, without other countries following or using the same rule. A country that taxes the income of its own residents regardless of where it is earned and leaves the income of foreigners untaxed does not have to fear the kind of tax evasion studied in the previous sections. However, in practice, the residence principle can hardly be implemented without the help of other countries. Thus, in 1977 the OECD countries agreed on a Model Double Taxation Convention, in which the residence principle was foreseen for the taxation of interest income.[15] This agreement has had limited relevance in practice, though, because countries did not assist one another in collecting the revenue, and tax evasion has become a common practice. To fight tax evasion the EU is currently seeking an additional

[14] In an intertemporal allocation model there is, of course, the opportunity of taking evasive action in the form of reducing savings. See Sinn (1985, chapters 7 and 9).

[15] See OECD (1977).

agreement between its member countries to either impose a uniform source tax on interest income or to mutually inform the residence countries about the interest incomes earned on their bank accounts by foreign residents. It remains to be seen how successful this approach will be.

It should be noted that the OECD Model Double Taxation Convention and the EU rules discussed only refer to interest income and do not incorporate business profits. Distributed earnings are taxed according to bilateral double taxation agreements and usually incorporate both residence and source tax elements. Retained earnings are nearly always taxed by the source country alone.[16] It will be a long struggle to introduce the residence principle for business profits, if it ever happens.

Even if the political impediments on the way towards residence taxation could one day be overcome, this principle is not really a safeguard against ruinous tax competition since investors could change their residence instead of the location of their investment. In the EU in particular, this possibility is gaining importance since the freedom of establishment has been granted to all citizens. Emigrating to other countries has become easy. By the arguments given above, tax competition in such a situation will lead to an equilibrium where the tax rates are driven down to the marginal cost of providing a formal residence in a country, and this may not be very much. After all, the marginal cost of hosting a mail box company is close to zero.

Dividend Taxes and Equity Traps

An alternative, unilaterally available taxation approach for preventing the evasive response of capital would be to shift the tax system from income taxation toward cash flow taxation, as was recommended by the Meade Committee (1978).[17] Cash flow taxes are taxes which are

[16] Cf. Sinn (1985, chapter 7) and Weichenrieder (1995).

[17] See Sinn (1985, chapter 11) for an extensive discussion of the theoretical underpinnings.

periodically raised on the real and/or financial surplus of firms. As they allow an immediate write-off of real and/or financial investment projects, they effectively leave marginal new investment projects, whose capital value is zero, tax free. They nevertheless generate a tax revenue since they effectively tax the returns from previous investment, carried out before the cash flow taxes were introduced. Although old projects die out, repeated replacement investments ensure that the cash flow tax generates a permanent and even growing tax revenue whose present value equals the product of the tax rate and the historical stock of capital existing at the time of tax reform.

Cash flow taxes are powerful revenue raisers that cannot be competed away in tax competition. The tax on marginal investment cannot be eroded by tax competition because there is no such tax, and the tax on the returns of existing capital cannot erode because a country can neither hope to attract capital by cutting the tax rate and nor does it have to fear a capital flight by raising it. A flight of capital means that investors do not reinvest but use the funds freed through depreciation for foreign investment. Given that gross investment can immediately be written off under the cash flow tax, the lack of reinvestment increases the tax base and the tax liability. The cash flow tax thus incorporates an exit fee that compensates the state for the foregone tax on the future cash flow that would have been collected had capital stayed in the country.

While the chances for cash flow taxes being implemented in the foreseeable future are small, it is important to realize that the existing taxes on corporate distributions come close to such taxes. In fact, the so-called S-based tax of the Meade Committee, which is a tax on the real and financial cash flow of a firm, is a dividend tax.[18] The S-based tax differs from a dividend tax only in that it involves a subsidy on new share issues (negative dividends) at the rate of the dividend

[18] The financial cash flow is net borrowing minus net interest payments, and the real cash flow is sales revenue minus wage costs, minus the costs of intermediate products and minus gross investment. The sum of the real and financial cash flow is the cash flow accruing to the firms shareholders, i.e. dividends net of new share issues.

tax. The existing dividend taxes, which do not have this feature, discriminate against equity creation in the form of issuing new shares but are neutral with regard to equity creation through profit retentions and hence do not increase the cost of capital when firms react flexibly and finance new investment entirely with retained earnings.[19] They leave marginal investment financed with retained earnings tax free, since the tax on future dividends earned by this investment is outweighed by the dividend tax saved by deciding to retain and invest rather than distribute the earnings. A country does not have to fear that it will drive out existing capital if it raises the dividend tax rate. A dividend tax punishes capital flight in a similar way as a tax on real cash flow does, because a termination of reinvestment within the firm means that more dividends are distributed and more dividend taxes are paid.[20] Capital is trapped in the corporate firm and cannot escape without paying a toll to the government. By contrast, taxes on retained earnings increase the cost of capital and drive capital out of the country.

It follows that tax competition will reduce the taxes on retained earnings but not the dividend taxes.[21] If anything, there is the risk that countries use dividend taxes to exploit foreign shareholders after the investment is made and choose excessively high tax rates. This may explain why the international double taxation agreements between countries refer to the taxes on distributed rather than retained earnings and why they specify upper rather than lower limits on the tax rates. The observation may also explain, for example, why Germanys tax reform in 2000,

---

[19] See Sinn (1985, chapters 4, 5 and 7) and Sinn (1991b).

[20] However, the dividend tax punishes only the withdrawal of previous retentions. When previous retentions are used up, a withdrawal of the capital stemming from new issues is possible without incurring a tax liability.

[21] Dividend taxes can be interpreted as taxes on quasi-rents. Thus, the result explained in the text has a certain resemblance with the finding of Huizinga and Nielsen (1997) that tax competition does not eliminate taxes on pure profits if some of these profits accrue to foreigners. However, the argument developed above holds true even when, unlike the Huizinga-Nielsen model, there are not rents flowing to foreigners, say, because the production function is linear homogeneous and fixed factors are exclusively owned by domestic residents.

which certainly was carried out with an eye on international tax competition, involved a substantial cut in the taxes on retained earnings while dividend taxes were tacitly increased by moving from the previous full-imputation to a partial imputation system with a double taxation of dividends at the level of firms and shareholder households.

This observation partially alleviates the fear that tax competition will benefit capital and hurt immobile labour. Capital caught in the corporate equity trap may also be a victim of tax competition. If all capital owners had the same diversified portfolio of assets and hence shared equally in the losses resulting from high dividend taxes, this could be welcomed as a mitigating distributional effect which reduces the adverse effects on the net-of-tax income of immobile labour. However, this is certainly not the case. Owners of financial assets would benefit from the higher interest rates resulting from the dissipation of taxes on retained earnings and so would new purchasers of shares who come in after the forces of tax competition have become operative: share prices would adjust such that their net rate of return on shares would be the same as the net-of-tax rate of return in financial assets. Moreover, of course, owners of non-corporate capital would also not be hurt by an increase in dividend taxes. Existing owners of corporate shares, on the other hand, would have to bear the increased dividend taxes alone since the additional future dividend taxes would be capitalized in the share prices prevailing at present.

Apart from problematic redistribution effects among the group of capital owners, the empirical importance for Europe of the upward forces on dividend taxes should not be overestimated, though. First, corporations, which are subjected to dividend taxes, play only a limited role on this continent. In Germany, for example, joint stock companies do not represent more than 20% of overall business revenues. Second, the theoretical point made above would only be valid empirically if marginal corporate investment could be entirely financed with retained earnings. While this used to be a valid approximation for long periods of time, the many start-ups

characterizing the New Economy have made new issues of shares an increasingly important marginal source of finance. New firms have to rely on new share issues and, in a dynamic context, dividend taxes raise the cost of capital for such share issues much more than was previously thought (Sinn 1991a, 1993). Thus, a competitive country does, in fact, have a lot to gain if it reduces the tax on corporate distributions, and the difference among the erosive forces working on new share issues and retained earnings may, after all, not be all that large.

Existence Problems

As was shown, the real problem of systems competition is not the underprovision of public goods but rather its distributional implications. Judged against the current tax systems where capital contributes towards financing the general budget, it is a disturbing aspect that uncontrolled tax competition results in labour partially subsidizing the infrastructure used by capital. This is not quite what one can call ruinous tax competition, but it comes close to it.

Private competition is called ruinous if, because of increasing returns to scale, competitive prices do not cover costs and firms try to reach a profitable situation by expanding excessively, so that in the end most firms go bankrupt and a non-competitive situation with one or a few surviving firms emerges. Modelling what goes on in ruinous competition is difficult, but to show that it will emerge is easy, because it is sufficient to demonstrate that due to scale economies a competitive equilibrium fails to exist. This section considers the possibility of truly ruinous systems competition to clarify the issue from a theoretical perspective and to identify the necessary assumptions without necessarily claiming that the world we live in will, in the foreseeable future, come close to this possibility.

In the model set up, only one new assumption has to be introduced: the international mobility of labour in addition to that of capital. When only one factor was mobile, increasing returns to

scale in the production and provision of the public infrastructure was not a fundamental problem for competition because the immobility of the other factor prevented the individual countries from being able to exploit the scale economies at the cost of their competitors. The situation looks very different when labour is also mobile.

A non-existence proof

To show that there is an existence problem for a competitive equilibrium when all factors are mobile, a proof by contradiction will be undertaken. Suppose that there is an equilibrium in which the net-of-tax wage rate in the country considered is equal to the corresponding wage rate in the international labour market. The values w^*, K^*, L^*, and W^* indicate the equilibrium values of the net-of-tax wage rate, the capital stock, employment, and the capacity of the public infrastructure, respectively. The income of the resident population, which consists of capital and wage income, is

$$R^* = r\overline{K} + w^* L^* . \tag{2.20}$$

Since capital income is fixed by the conditions in the world market we only need to consider wages. Solving for w^* and using (2.4) gives the expression

$$w^* = \left[f\left(K^*, L^*\right) - rK^* - c\left(K^*, W^*\right)K^* - \rho W^* \right] / L^*$$

for the net-of-tax wage rate. If an equilibrium would exist, then, with the given world market conditions, the single government would be unable to increase the income of the existing population L^* through its own actions. In fact, however, it can easily be shown that this is always

possible when the government takes action to make the economy grow larger. For example, the government can increase W to αW^*, allow immigration until employment reaches αL^*, and, taking account of (2.1), choose a tax rate τ which ensures that $K = \alpha K^*$ when $L = \alpha L^*$, where α is a parameter larger than one. Because of the homogeneity assumptions, which apply to $f(K,L)$ and $c(K,W)$ with $\lambda < 0$, the net-of-tax wage rate increases to

$$\begin{aligned} w &= \left[f\left(\alpha K^*, \alpha L^*\right) - r\alpha K^* - c\left(\alpha K^*, \alpha W^*\right) \alpha K^* - \rho\alpha W^* \right] / \left(\alpha L^*\right) \\ &= \left[f\left(K^*, L^*\right) - rK^* - \alpha^{\lambda} c\left(K^*, W^*\right) K^* - \rho W^* \right] / L^* > w^* \end{aligned}$$

where $f_{LL} < 0$ implies immigration up to the limit αL^*. The income of the existing population rises correspondingly to

$$R = r \cdot \overline{K} + w \cdot L^* > R^* = r \cdot \overline{K} + w^* L^*.$$

This contradicts the assumption that w^*, K^*, L^* and W^* represent an equilibrium and proves:

Proposition 2.6: *If all factors of production are mobile, the production function $f(K,L)$ is linearly homogeneous, and the Selection Principle is operative in the sense of concentrating activities with increasing returns to scale in the government sector, there is no equilibrium in infrastructure competition.*

Why the Selection Principle implies non-existence

This result is a logical implication of a very abstract model and caution is required in transferring it to the real world. The assumption that all factors of production are perfectly mobile is an

idealization that may well be considered exaggerated given the language barriers that still exist in the European Union. However, it should be taken into account that it depends not so much on the mobility of the existing EU population as on the differential mobility among the EU countries of those people who have already decided to immigrate from third countries. This has already been pointed out in the introductory chapter. Increasing returns to scale in the provision of public infrastructure concentrates the flow of migrants only into a few countries and increases the imbalances in country size that already exist now. For Germany, which, even relative to its size, absorbed the lion's share of European immigrants in recent years, proposition 2.6 is more than a theoretical possibility conceived in an ivory tower. Perhaps it is the non-existence of an immigration equilibrium which provides the material for the distressing sensational reports which the television channels broadcast all over the world. From the fall of the Iron Curtain up to the end of 1997, on balance, around five million people streamed into western Europe, more than three million of them to Germany alone. At the time of writing, the foreign-born percentage of the population in Germany is about as high as the respective percentage in the US, the classical immigration country. And this does not include the internal east-west German migration, which was a net 1.3 million people for western Germany within a decade after the wall had come down. In view of these figures, the destabilizing power of systems competition should not be underestimated.

From a theoretical point of view, the result could be criticized because it abstracts from land as a factor of production, which is, of course, not mobile. If land is important for production, the economy as a whole may not exhibit scale economies with regard to the mobile factors even if the government sector itself operates under increasing returns to scale. However, land is no longer an important factor of production in a modern industrial society. Less than 3% of the

workforce in Germany now works in agriculture, and the returns to land from all economic uses may be in the order of magnitude of not more than 4 to 5% of the German national product.[22]

If land can be neglected, the production of goods and services in the public and private sectors can really only have constant or increasing returns to scale. Constant returns to scale show up when, as is always possible, larger scale production processes can be realized through replication of the processes already available at a lower level, and there are increasing returns to scale when indivisibilities in production mean that there are additional efficient production possibilities at higher levels of scale which were not available at lower levels. In general, we should assume that both types of production processes exist. If the two types were spread uniformly over the private and public sectors, then competitive market economies could not exist because the private sector would exhibit increasing returns to scale. However, if the Selection Principle is operative, production is organized so that the private sector takes over the production processes with constant returns to scale and the public sector takes over the remaining ones with increasing returns to scale, and both sectors coexist. A fundamental problem arises, however, if, in addition to market competition, systems competition is introduced, because then the exceptions from the market process which the government had administered are forced back into a competitive environment. The ruinous consequences which prevented a competitive solution in the first place reappear on the higher level of systems competition.

---

[22] In Sinn and Sinn (1991, appendix II) the value of land in west Germany is calculated for the year 1989. If a yield of 4% is assumed, which is the average obtained for residential property in Germany, the share of the returns from land in GDP is around 3.7%. Based on a study by Bach and Bartholmai (1998) a value of 5.1% is obtained. I have Marcel Thum to thank for these calculations.

Policy Implications

What can be learnt from the analysis of this section for the construction of a new Europe without internal borders? Much depends on the development of labour mobility. If labour mobility stays limited for the time being, systems competition in terms of taxes and infrastructure levels can be expected to work in the sense that an efficient equilibrium among the countries will be found. However, efficiency is not equity. Capital, except possibly for corporate capital trapped by dividend taxes, will be the big winner of systems competition. It no longer has to contribute towards financing the general government budget, and, what is more, it will even enjoy subsidies in the sense that its taxes have to cover only part of the public infrastructure it uses.

It is not a good idea to harmonize tax rates to prevent the distributional consequences, because that would only trigger off an even more intense competition for mobile capital with public infrastructure expenses. If harmonization imposes an effective constraint on tax rates, an overprovision of infrastructure results.

To avoid this consequence, a European agreement to impose self-financing constraints in the form of an extended subsidy prohibition or the requirement to use average cost pricing for public facilities would be useful. A self-financing constraint leads to efficient provision of public goods in systems competition without burdening labour with a tax to subsidize capital. Net-of-tax labour incomes would be higher, and profit incomes lower.

If labour income, too, becomes more flexible, then an equilibrium in systems competition may no longer exist. Instead, scale economies would tend to concentrate the population in large countries at the expense of the smaller ones. Systems competition would be ruinous, since it includes activities which the state originally absorbed to avoid ruinous competition.

Before this happens it may be better to think about creating a new layer of government in Europe whose function would be to provide those public services which exhibit pronounced

supra-national scale economies. Defence, money supply and constitutional jurisdiction are among the first functions that can be shifted to a central government.

Of course, there is the danger of a misuse of tax money by a central government that operates at a distance from its voters. That such a danger cannot be dismissed out of hand is shown by the incident concerning the Brussels Perry Lux Agency, which in 1998 resulted in the European Parliament refusing to exonerate the European Commission. However, as was said before, it makes no sense to judge the political bodies to be created with the Nirvana approach of first best policy decisions. European history is full of examples of successful unification and centralization, a process which Popitz (1927) even saw as a historical law. Professional economists, too, should give some sympathetic, though also critical, attention to the peaceful process of European political integration, which has only just begun.

Appendix to Chapter 2

Criticism of the Zodrow and Mieszkowski Infrastructure Model

Zodrow and Mieskowski (1986) first show that public consumption goods are underprovided in equilibrium if these goods must be financed with a tax on mobile capital and they then extend their argument to the case of public infrastructure or investment goods. This appendix shows why this extension is flawed. [23]

The authors assume that output is produced according to a production function $F(K,W),\ F_{KW} > 0,\ F_W > 0,\ F_{WW} < 0$, where the arguments are capital K and the capacity of infrastructure W,[24] and they postulate analogously to (2.1) and (2.2) (for $\omega = 0$)

$$F_K(K,W) = r + \tau \tag{A2.1}$$

and

$$W = \tau K\,, \tag{A2.2}$$

where the price ρ of the infrastructure is equal to one and crowding externalities are assumed away.[25]

---

[23] An alternative criticism of the Zodrow-Mieszkowski argument can be found in Noiset (1995). See also Sinn (1997).

[24] The symbols are changed here in order to make the coverage as far as possible the same as that in the model used in this paper.

[25] To see that the Zodrow-Mieskowski model is incompatible with crowding externalities assume that $F(K,W) = f(K,L) - c(K,W)\cdot K$ as in this model. Equation (A2.1) then becomes $f_K(K,L) - c(K,W) - c_K(K,W)\cdot K = \mathrm{r} + \tau$ rather than $f_K(K,L) - c(K,W) = \mathrm{r} + \tau$ according to equation (2.1). The comparison shows that the authors implicitly assume pure public goods or private infrastructure goods for which the congestion externality $c_K(K,W)\cdot K$ is not an externality at all, but rather a productivity effect taken fully into account by the individual investor. This difference is important but it is not the decisive criticism of the derivation of the underprovision hypothesis.

Analogously to (2.18) they calculate from (A2.1) and (A2.2) a differential quotient

$$\varphi \equiv \left.\frac{\mathrm{d}K}{\mathrm{d}\tau}\right|_{(\mathrm{A2.1}),(\mathrm{A2.2})} = \frac{1 - K\,F_{KW}}{F_{KK} + \tau F_{KB}}, \tag{A2.3}$$

which measures the net effect on the amount of capital employed of a tax rate increase and the improvement in the infrastructure that this increase makes possible. They also assume that the denominator of the right-hand side of (A2.3) is negative but they assume in addition that the numerator is positive,

$$K \cdot F_{KW} < 1, \tag{A2.4}$$

i.e. that a tax increase will deter capital despite the improvement of the infrastructure which this tax increase makes possible. They base this assumption on the statement, which is not explained further, that this is a ‘stability assumption‘ (ibid., p. 363). ‘We assume that the model is stable in the sense that the marginal cost of diverting a unit of output to public services for firms (which is equal to unity) is greater than the associated increase in output due to the increased marginal productivity of capital $\left(K \cdot F_{KW}\right)$‘. The authors then show that assumption (A2.4) implies an underprovision of infrastructure goods if the tax rate is chosen such that the income of the fixed factor is maximized:

$$F_W > 1. \tag{A2.5}$$

Unfortunately, the reasoning is contradictory. It cannot in any case be applied to all the production functions that the authors allow for. An example can explain this. Suppose that $F_{KWK} = F_W(0,W) = 0$ which, of course, does not have to hold in general, but is perfectly in line with what the authors assume. Then

$$K \cdot F_{KW}(K,W) = \int_0^K F_{KW}(u,W)\mathrm{d}u = F_W(K,W),$$

and equation (A2.4) becomes

$$F_W < 1,$$

which indicates an overprovision of infrastructure goods. This obviously contradicts (A2.5). Apparently the authors assume overprovision of infrastructure goods in order to prove that there is underprovision.

This does not mean that there are no production functions for which (A2.4) and (A2.5) are compatible with one another, but it is quite inadmissible to use assumption (A2.4) to prove that the public goods are underprovided. This assumption lacks an economic basis and is too close to the condition for an overprovision for it to be able to serve as the basis for the underprovision hypothesis.