Chapter 7 Local Public Finance[1]

7.1 Introduction

The theory of local public goods differs from the standard analysis in that goods are assumed to be specific to a particular geographical location, and consumers, in deciding on their location, can exercise choice with respect to the quantity and types of public goods provided. For some public goods there may be no spatial restriction (for example, the benefits from research and development); but for others the benefits, although available at no additional cost to new residents, are construction of sea defenses benefits those protected by the sea wall; the transmission of a television programme benefits those within a certain distance of the transmitter. In this chapter we examine some of the implications of the local nature of such public goods and their provision by local communities. There is of course no necessary reason why they should be provided by local rather than central government; our focus in on the former, but in the final section we consider the fiscal relations between different levels of government.

Local Public Goods and the Market Analogy

The mobility of individuals between communities supplying local public goods has a number of major implications. It is in particular relevant to the problem of the revelation of preferences. Indeed, much of the interest in local public goods was stimulated by the intriguing suggestion of Tiebout (1956) that, if there were enough communities, individuals would reveal their true preference for public goods by the choice of community in which to live (in much the same way as individuals reveal their preferences for private goods by their choices). Where there is a wide range of choice, all those deciding to live in the same community would have essentially the same tastes, and there would be no problem of reconciling conflicting preferences. Moreover, it is often asserted that such a local public goods equilibrium would be Pareto-efficient.

This argument is based largely on the analogy with private goods:

> Just as the consumer may be visualized as walking to a private market place to buy his goods, …we place him in the position of walking to a community where the prices (taxes) of community services are set.

[1] This part draws heavily from Atkinson and Stiglitz (1980, Chapter 17, pp.519-556).

> Both trips take the consumer to the market. There is no way in which the consumer can avoid revealing his preferences in a spatial economy. [Tiebout, 1956, p.422]

This parallel ignores however certain key characteristics of local public goods. One of the most important of these is the essential non-convexity associated with the provision of such goods to individual citizens. In the conventional analysis of markets with only private goods, the assumption of convexity is critical in three ways; (1) as a result of non-convexities, there may exist no competitive equilibrium; (2) non-convexities in practice are likely to be associated with various kinds of non-competitive behaviour; and (3) where there are non-convexities, it is not necessarily the case that every Pareto-efficient allocation can be supported by a competitive equilibrium with appropriate lump-sum redistributions.

In the case of local public goods, non-convexities are inherent in that the cost of supplying a given quantity of a public good (e.g., a local radio programme) to an additional individual is zero (in the pure case). As we shall show, a local public goods equilibrium may not exist. Whether or not it does depends on the precise equilibrium notion employed, and, as we note below, several alternative concepts suggest themselves. Second, when there is a limited number of communities, they may attempt to make themselves more attractive to outsiders, acting in this way analogously to monopolistically competitive firms. On the one hand, this provides a motive for ensuring efficiency in the provision of public services; on the other, the mix and level of public goods provided may not be Pareto-efficient. Finally, not every Pareto-efficient allocation can be sustained by a local public goods equilibrium.

There are therefore reasons to doubt the usefulness of the competitive market analogy when considering the provision of local public goods and the claims that have been made for its efficiency. In a local public goods equilibrium, there may well be fewer communities than different types of individuals (it may indeed be socially optimal to have only one). The person may not therefore be able to find a community of individuals whose tastes are essentially identical to his own, and there may not be an optimum number and mix of people in a community.

Finally, there are the issues raised by redistribution. In the United States at least, it would appear that the pattern of local community formation has much to do with the rich attempting to segregate themselves from the poor, in part because there is a large element of redistiribution involved in the provision of education and other services of the local community. By moving to their own communities, the rich can avoid this redistribution. This phenomenon, which has no direct parallel with private goods, clearly must form part of the analysis.

7.2 Optimum Provision of Local Public Goods

For a pure public good that is not spatially limited (such as the benefits from research and development), the issue of the number and size of communities does not arise. Where however the benefits from a public good are spatially restricted, we have to consider these questions. As far as the public good is concerned, it is indeed natural to ask why there should be more than one community. If the addition of a person does not detract from the benefits enjoyed by others, then – from this point of view – the optimum allocation would involve everyone living in the same community. Against this, however, must be balanced the diminishing returns to labour with a fixed quantity of land, or the declining utility arising from congestion (e.g., as residential density increases). Moreover, for some public goods congestion may set in beyond a certain size of community (as emphasized in the treatment by Tiebout)[2].

In this section we focus on a single pure public goods, considering the balance between the increasing returns inherent in its provision and the decreasing returns to labour as the population within a community is increased. We assume at this stage that all individuals are identical and examine the optimum allocation over a number of identical communities (i.e. with the same quantity and quality of land). One can envisage there being a large number of islands, and we want to know how many islands should be inhabited and what is the optimum level of public goods provision on each island (This is essentially the optimum "club" problem formulatede by Buchanan, 1965).

Basic Framework

The model is a highly simplified one, in which total output, *Y*, in a community can be used either for private consumption (*X* per person) or for the public good, *G*, in that community. It is assumed that output is an increasing, concave function of the number of workers in the community, *N*:

$$Y = f(N) \qquad f' > 0, f'' < 0 \tag{1}$$

where $f \to 0$ as $N \to 0$ and as $N \to \infty$, $f \to \infty$ and $f' \to 0$. On the assumption that everyone in the community is identical and is treated the same, the aggregate production

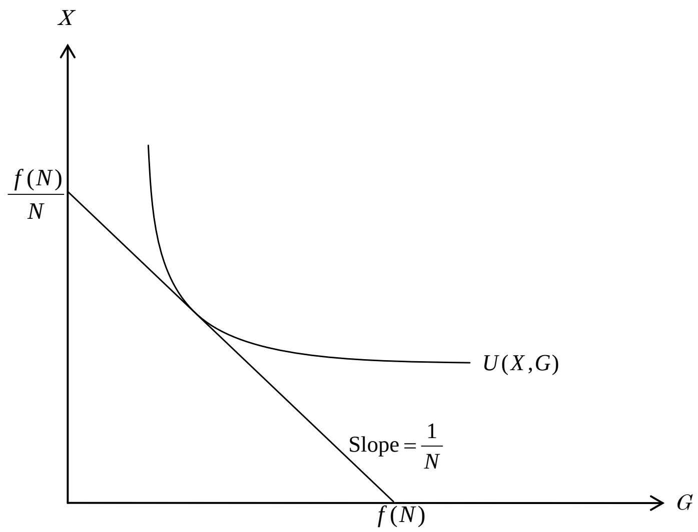
2 An essential aspect of the analysis is that individuals belong to a single community; i.e., their place of residence, work and consumption coincide. Obviously, this is not strictly true, but the assumption seems a useful abstraction.

constraint gives:

$$Y = XN + G = f(N) \qquad (2)$$

For fixed *N*, this defines the consumption opportunity set illustrated in Figure 1.

Figure 1 Opportunity Set for Fixed Population



We assume that individuals have identical preferences, represented by the utility function $U(X,G)$ where *U* is assumed to be quasi-concave. If the government chooses *G* to maximize *U* for a given level of *N*, this gives the point of tangency in Figure 1. The condition for a maximum of *U* is that

$$U_x = NU_G$$

or

$$\frac{NU_G}{U_x} = 1 \qquad (3)$$

which is the conventional result that the sum of the marginal rates of substitution equal the marginal rate of transformation ($\sum MRS = MRT$).

As we increase *N*, output, and hence the maximum level of public goods, increases (since $f'(N)>0$) but the maximum level of consumption *per capita* ($f(N)/N$) decreases. The variable *N* opportunity locus is the outer envelope of the fixed *N* opportunity loci – see Figure 2. This outer envelope may be characterized by taking a fixed value of *G* and then varying *N* to maximize *X*. Since

$$N = \frac{f(N) - G}{N} \tag{4}$$

the first-order condition implies

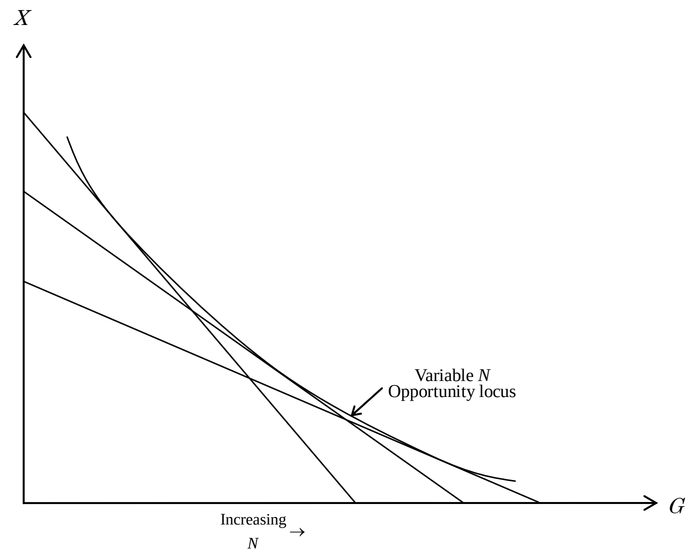
$$f' = \frac{f(N) - G}{N} = X \tag{5-a}$$

or

$$G = f - Nf' \tag{5-b}$$

The second of these conditions has an interesting interpretation. Since f' is the marginal product of labour, $f - Nf'$ is output minus wage payments if workers are paid their marginal product. Thus, if the level of public expenditure is fixed, but the population is variable, the population that maximizes consumption *per capita* is such that *rents equal public goods expenditure*. This has been dubbed the "Henry George" theorem (Stiglitz, 1977), since not only is the land tax non-distortionary, but also it is the "single tax" required to finance the public good.

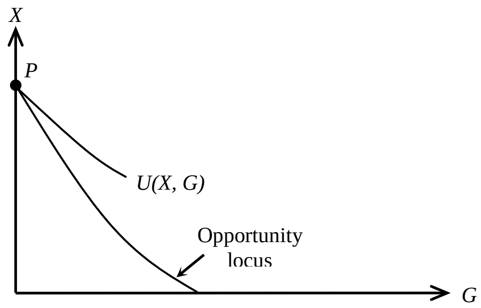
Figure 2 Opportunity Set for Variable Population



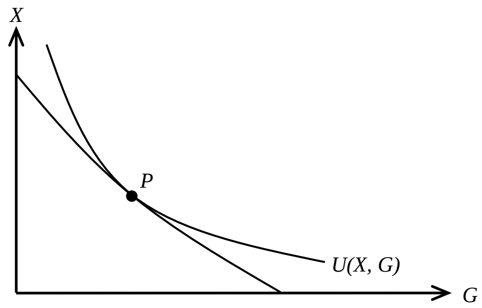
Properties of the Social Optimum

If we now put these two elements together – variation in *G* and variation in *N* – then we immediately have to face the problem that the variable *N* opportunity locus is convex to the origin (rather than concave, as typically assumed with private goods in a conventional model). As a result, the community size that maximizes *per capita* utility may be zero, infinite or finite, as illustrated by Figure 3a – 3c. If the indifference curve is more "curved" than the opportunity locus, then there is an "interior" solution. This is likely to be the case if public and private goods are strong complements so the indifference curve is very curved. Otherwise, utility is maximized with only private goods being produced and a "zero" population, or with only public goods being produced and an "infinite" population (It is assumed that *N* may be treated as a continuous variable, and that we can ignore the problems that may arise if the total population is not a multiple of the optimum *N*; these aspects are discussed below).

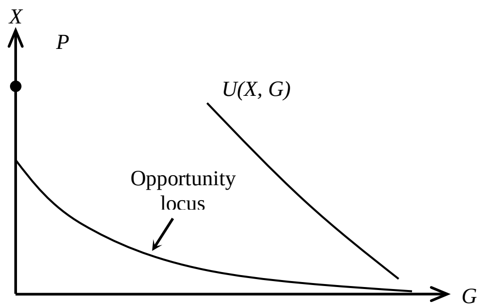
Figure 3 Optimal Population



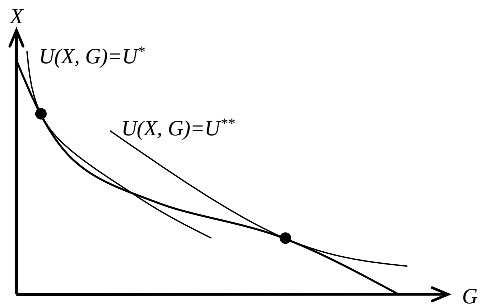
(a) Optimal population is zero



(c) finite optimal population:
single optimal *N*



(b) Optimal population is infinite



(d) finite optimal population:
multiple local optima

These findings may be related to the results on optimum population. If the objective is to maximize per capita utility, then with only private goods consumption is maximized with an infinitesimal population. On the other hand, if there were only public goods, utility would be maximized with the largest possible population – we would have a national public good. If individuals value both private and public goods, then there is a balancing of these two effects.

Even where there is an interior solution, it may not be unique, as is illustrated in Figure 3d, where there are two combinations of *X*, *G* (and hence *N*) that give local maxima. In order to explore this further, let us define the maximum level of utility that can be attained for a given community size, *N*, by *V*(*N*). In other words, it is the value of the maxim and obtained from solving the fixed *N* problem with which we began:

$$V(N) \equiv \max_X [U(X, f(N) - XN)] \tag{6}$$

Differentiating with respect to *N* and using the envelope condition (i.e., that *X* is chosen optimally for any given *N*),

$$\begin{aligned} V'(N) &= U_G(f' - X) \\ &= \frac{U_G}{N}(Nf' - NX) \end{aligned} \tag{7}$$

Using (2) and the first-order condition (3)

$$V'(N) = \frac{U_X}{N^2}[G - (f - Nf')] \tag{8}$$

At an interior optimum for *N*, where $U_X > 0$, the square bracket is zero, which gives equation (5b). If we now take the second derivative and evaluate at $V' = 0$,

$$V''(N)\big|_{V'=0} = \frac{U_X}{N^2}\left(\frac{dG}{dN} + Nf''\right) \tag{9}$$

The second term in the bracket is negative; its magnitude depends on the elasticity of substitution of the production function. From the definition of the elasticity (σ_p),

$$-Nf'' = \left(\frac{1}{\sigma_p}\right)\frac{f'(f - Nf')}{f} \tag{10}$$

If we define γ to be the share of government spending in total output, then at the optimum (from (5b))

$$\gamma = \frac{G}{f} = \frac{f - Nf'}{f} \tag{11}$$

On the other hand, from the indifference map (which we assume for convenience to be homothetic), the elasticity of substitution along an indifference curve is

$$\sigma_c \equiv \frac{d\log(G/X)}{d\log N} = \frac{d[\log\gamma - \log(1-\gamma) + \log N]}{d\log N} \tag{12}$$

So that (if $\gamma \equiv d\gamma / dN$)

$$\sigma_c - 1 = \frac{\gamma' N}{\gamma(1-\gamma)} \tag{13}$$

Hence

$$\frac{dG}{dN} = \frac{d}{dN}(\gamma f) = \gamma f' + \gamma' f = \gamma f' \sigma_c \tag{14}$$

where the last step substitutes from (13) and (11). Assembling the pieces (equation (10), (11) and (14), and substituting into (9),

$$V''\big|_{V'=0} = \frac{U_X}{N^2}\gamma f'\left(\sigma_c - \frac{1}{\sigma_p}\right) \tag{15}$$

If $\sigma_c\sigma_p$ is everywhere less than 1, this rules out a local minimum, and hence cases such as that shown in figure 3(d) (where there is a local minimum between the two maxima). This confirms the earlier suggestion that strong complements in consumption (low σ_c) increase the curvature of the indifference map and tend to lead to a unique interior solution. It also brings out that strong complementarity in production (low σ_p) has the same effect, since this leads to a flat opportunity locus[3].

---

[3] Along the variable *N* opportunity locus $\frac{dX}{dG} = -\frac{1}{N}, \quad \frac{X}{G} = \frac{f'}{f - f'N}$.

Hence $-\frac{d\log X/G}{d\log(-dX/dG)} = \frac{d\log f'/(f - f'N)}{d\log N} = \frac{ff''N}{f'(f - f'N)} = -\frac{1}{\sigma_p}$

In the analysis so far it has been assumed that labour is supplied in elastically, but the results can readily be extended and the Henry George theorem remains valid. This is left for the reader to consider (see Stiglitz, 1977, p.28).

Fixed Population and Fixed Number of Communities

To this point we have assumed that there is no obstacle to the establishment of sufficient local communities of optimum size to accommodate the total population. One problem is that the total number of people may not be an integral multiple of *N*. This has been discussed in the literature on optimal club size (e.g., Pauly, 1967). More serious in the context of local governments is likely to be the limit on the number of potential communities. Although in a frontier society it may be possible to establish new towns, and thus reduce *N*, there is likely to be an end to this process. Settlement in most advanced countries is restricted to a fixed number of locations.

We now consider the implications of this feature of local jurisdictions. For ease of exposition, we assume that there are two communities, denoted by 1 and 2, with identical quantity and quality of land, and that a fixed population, $2N^*$, has to be divided between them. If the social optimum involves equal treatment, then the solution is relatively straightforward. There is however no necessary reason why equal treatment should be implied.

In order to examine the social optimum, let us denote by N_i the number of people in community *i* and by V_i the level of utility where G_i is chosen optimally in each community. Suppose that the government maximizes the Benthamite social welfare function

$$\Psi = N_1 V_1 + N_2 V_2 \tag{16}$$

The first and second derivatives are (substituting $N_2 = 2N^* - N_1$)

$$\frac{d\Psi}{dN_1} = (V_1 + V_2) + N_1 V'_1 - (2N^* - N_1)V'_2 \tag{17}$$

$$\frac{d^2\Psi}{dN_1} = 2(V'_1 + V'_2) + N_1 V''_1 + (2N^* - N_1)V''_2 \tag{18}$$

Evaluating at $N_1 = N_2$, the equal treatment case is clearly a turning point, but there is no guarantee that it is a maximum, since at the solution

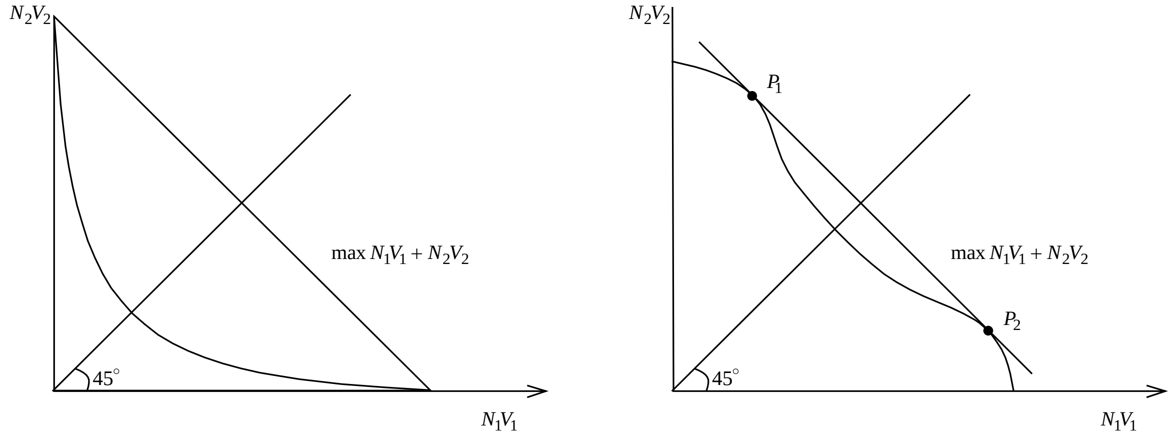
$$\tfrac{1}{2}\left.\frac{d^2\Psi}{dN_1^2}\right|_{N_1=N_2} = V'(N^*) + N^*V''(N^*) \tag{19}$$

Suppose first that N^* coincides, by chance, with a value of *N* that, in the variable number of communities case, gives a local maximum. Then $V'(N^*) = 0, V''(N^*) < 0$, and we have a local maximum of the constrained case. If there is an "excess" population, so that $V'(N^*) < 0$, then it is sufficient for a local maximum that $V''(N^*) \le 0$. On the other hand, it is quite possible that there is a population "shortage", so that $V'(N^*) > 0$. It can then happen that the equal treatment solution is a local minimum. This means that social welfare could be increased by moving to an asymmetric allocation, which in effect allows one community to get closer to the optimum and a larger fraction of the total population to enjoy the consequent higher level of *V*.

Some possible situations are illustrated in Figure 4(a) and (b). In the first case, the social optimum involves all the population being in one jurisdiction; this is horizontally equitable in that all people are treated identically. On the other hand, if the possibility frontier has the shape illustrated in Figure 4(b), then the social optimum involves the asymmetric treatment of identical individuals. This may at first seem surprising, but it is only a further illustration of the conventional wisdom that welfare maximization does not necessarily imply equal treatment of equals. It is quite possible that we may want to constrain the government to choose only between policies that ensure equal utilities, but this must be introduced as a separate principle of horizontal equity.

The solution does of course depend on the instruments at the disposal of the government. We have not, for example, allowed for lump-sum subsidies between communities. It may be seen however that with a utilitarian objective this involves the equalization of the marginal utility of consumption and that this does not necessarily imply equalization of utility if the level of public good provision differs.

Figure 4 Social Optimum



(a) all population in one jurisdiction (b) asymmetric treatment

Differences Among Individuals

The analysis of the case with identical individuals is mainly of interest because it provides the necessary background for the general theory where individuals differ. As we have seen, the hypothesis of Tiebout was that, where there are heterogeneous individuals, they would sort themselves out according to their preferences; communities would thus be homogeneous. We need to ask, however, under what conditions such complete sorting is optimal.

The first point concerns the production side of the economy. Such considerations were in effect assumed away by Tiebout: "restrictions due to employment opportunities are not considered. It may be assumed that all persons are living on dividend income" (1956, p.419). This clearly ignores an important factor leading to mixed communities. If doctors and lawyers are not perfect substitutes, then it may pay to have communities in which there are both. Of course, if doctors and lawyers have the same preferences, then it is still possible that all individuals in the same community have the same tastes. But this seems unlikely. More generally, we would require that the distribution of tastes of lawyers and doctors be identical, and that they have the same incomes; but since the latter depends on their relative supplies, this could not be true in general unless they were perfect substitutes for each other.

Leaving aside the mixing due to interactions in production, it is not the case that individuals are always better off forming homogeneous communities with people of identical tastes. Suppose that there are two communities that could be settled, and equal numbers of two types of person, identical except for their preferences regarding public goods. There are three public goods, and the utility functions of the two types are

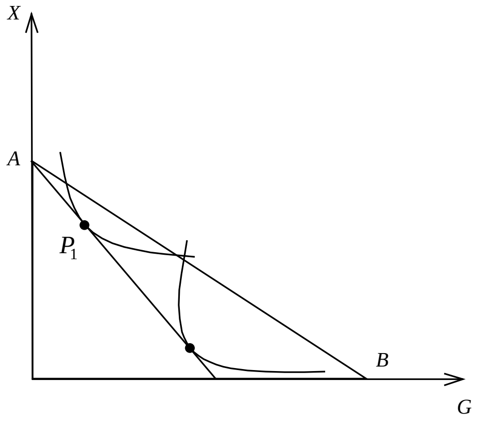
$$U(X, G_1 + \kappa G_3) \text{ and } U(X, G_2 + \kappa G_3) \tag{20}$$

where $0 < \kappa < 1$. In other words, group 1 prefers public good 1 (swimming pools), gets no utility at all out of public good 2 (ski lifts), but enjoys hiking trails (public good 3). Hiking and swimming are perfect substitutes, but at a trade-off of less than 1 to 1. Group 2 has symmetric preferences, preferring public good 2, getting no utility out of public good 1 and limited enjoyment from public good 3.

Clearly, if they form separate communities, each will produce the public good of its own preference: swimming pools in 1 and ski lifts in 2. We need however to compare this with the possibility of a merged community, where – as a compromise – good 3 is produced. In this case, they can enjoy the benefits of the economies of scale associated with public goods: if $\kappa > 1/2$, then with the same tax payments the effective public goods supply to each person goes up. Against this must be balanced the diminishing returns to labour as the size of the community is doubled, but it is clear that there are circumstances in which everyone is better off. This is more likely to be the case, the closer κ is to 1 and the less is the extent of diminishing returns. (For a related discussion, see McGuire, 1974, and Berglas, 1976.)

The desirability of forming homogeneous or heterogeneous communities may depend on the ability to identify different groups. Assume, for instance, that there are two groups in the population, one of which has a low preference for the public good, the other of which has a high preference. Assume there are no diminishing returns to labour. Clearly, if a single community were formed, a supply of public goods equal to the original high level could be provided, and everyone's taxes cut. Hence, such a combination would be Pareto-improving. But if everyone in the mixed community has to be taxed identically, on the grounds that we cannot identify those who prefer the low quantity of public good, there might be no allocation that would improve the position for both types. This is illustrated in Figure 5, where P_1 and P_2 denote the positions chosen when the two groups form separate communities. The mixed community with equal treatment involves a point on the line *AB*, and there is no such point that is preferred by both groups.

Figure 5 Pareto-inefficient Community Formation



This provides one reason why benefit taxation may be desirable, even though it may reduce the consumption of a public good that has no marginal cost of usage (e.g., tolls on uncrowded bridges). Although such taxation may, with perfect information, be sub-optimal, it may be warranted if it allows the assignment of tax burdens in such a way as to permit the formation of larger communities than otherwise would be the case.

7.3 Market Equilibria and Optimality: Identical Individuals

To assess the claims made for the market provision of local public goods, we need to specify the way in which the mechanism is assumed to work and what is meant by a local public equilibrium. We must then ascertain under what conditions such equilibrium exists (recall that in the presence of non-convexities, competitive equilibria often do not exist); finally, we need to determine whether, if equilibrium exists, it is Pareto-efficient.

We proceed in the same way as in the earlier analysis. We first assume, in this section, that people are identical. This means that the critical issue of matching people by communities does not arise, but we can still ask whether communities of the optimal size will be formed, and whether, within each community, the optimal supply of public goods will be provided. We then turn in Section 4 to the more difficult issue of analyzing local public goods equilibria when individuals differ.

Basic Model

As explained in the Introduction, the behaviour of the market process depends on the conditions governing migration and the way in which the decisions are made regarding local public goods.

Here we assume that there is free migration, and that in each community all individuals are treated identically[4]. It is then a condition of equilibrium that all individuals have the same level of utility. As far as local public good decisions are concerned, we assume initially that each community acts to maximize utility for a given population. In other words, decision-makers ignore the effect on migration. Alternatives to this myopic assumption are discussed below.

For ease of analysis, we make the same simplifying assumptions as earlier. There is a single private good and a single public good. There are two potential communities, both identical. The conditions for equilibrium may be given in terms of *V*(*N*), which represents the maximum utility assuming *N* is constant (i.e., ignoring the effect on migration):

$$V(N_1) = V(N_2) \qquad \text{if both communities settled}$$

$$V(2N^*) \geq V(0) \qquad \text{if only one community settled} \qquad (21)$$

Some of the various possibilities are illustrated in Figure 6. We may note that continuity of *V*(*N*) is sufficient to ensure existence of at least one equilibrium[5].

There are in fact quite possibly multiple equilibria. Let us take first the case shown in Figure 6(a), where $V' < 0$ for all *N*. In the market economy, there is an equal-population equilibrium at *E*, and two single-community equilibria at E_1 and E_2. Which of these is attained depends on the adjustment process. Suppose that migration takes place according to the difference in utility levels. If the population is disturbed from the equilibrium *E* in Figure 6(a), it will tend to diverge. If $N_1 = N^* + \varepsilon$, where $\varepsilon > 0$, then $V(N_1) > V(N_2)$ and people will move to community 1. The limit of this process is a locally stable equilibrium at E_1, with only community 1 inhabited. The case in Figure 6(b) also has three equilibria, with the same stability pattern (although different welfare implications – see below). The third case, Figure 6(c), exhibits three interior equilibria, of which E_1 and E_2 are locally stable under the assumed adjustment process. The final case, 6(d), has no fewer than five equilibria. The equal-size equilibrium *E* is locally (but not globally) stable, as are the one-community equilibria.

Turning to the efficiency properties of these equilibria, we can see that the case shown in Figure 6(a) corresponds to that in Figure 4(a), where maximization of $N_1V_1 + N_2V_2$ involved

---

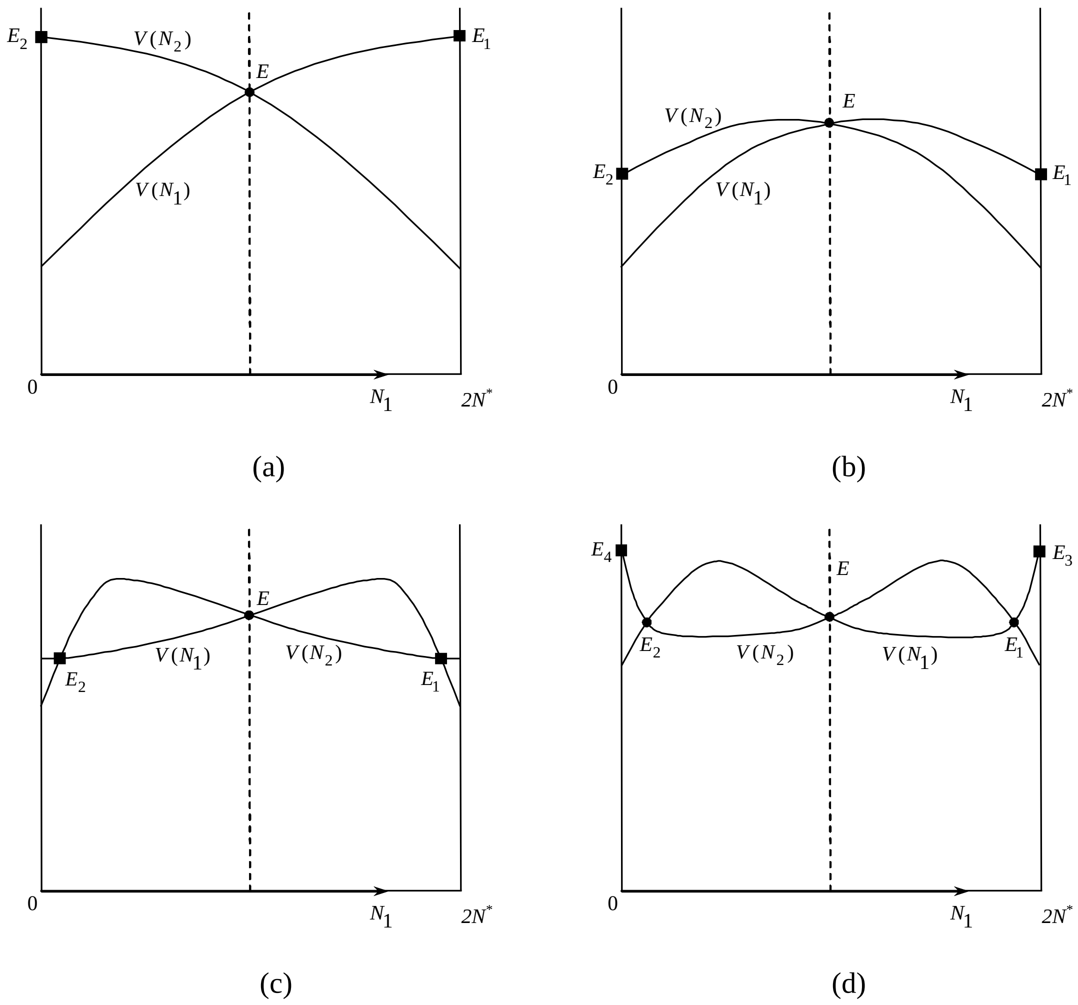
4 That is, we assume there is no differential taxation of immigrants and original occupants. The analysis can be viewed as applying to a "socialist economy" in which all residents share equally in rents.

5 If $N_1 = 0$ is not an equilibrium, then V_1 must be above V_2; conversely V_2 is above V_1 at $N_1 = 2N^*$; hence by continuity there is an intersection.

only one community being populated. As we have seen, the only locally stable equilibria of the market process in this case are those with single communities, so that the migration of individuals does achieve an efficient allocation. On the other hand, there is no guarantee that this will come about. Figure 6(b) shows the case where the single-community equilibria are again locally stable, but there exist allocations at which everyone is strictly better off. For example, if the population were allocated equally (i.e., at *E*), this would make everyone strictly better off. The converse applies in Figure 6(d), where the equal-community equilibrium, which is locally stable, is clearly Pareto-inferior to the one-community equilibria.

This simple model demonstrates the lack of generality of Tiebout's hypothesis. Even in the absence of any problems of sorting individuals according to differences in tastes, the local public goods equilibrium may not be Pareto-efficient. Nor is the problem alleviated if we let the number of communities and the number of individuals increase (in proportion). The analysis has moreover ignored two problems that mean that it is even less likely that the equilibrium be efficient: the effect of migration on land values, and differences among communities.

Figure 6 Market Equilibria



■: Locally stable equilibrium, ●: Unstable equilibrium

Land Values and Capitalization

The pervious analysis assumed that all individuals had identical claims; in effect, we modeled a state in which land is publicly owned and all migrants have equal access to the rents (after paying for the public goods). Equivalently, there is a 100 percent tax, with the deficit or surplus between government revenue and expenditure being made up by lump-sum taxes or subsidies. Assume now, however, that we give all individuals one unit of land but for half of the population we concentrate δ of their ownership claims in one community ($1-\delta$ in the other); for the other half of the population, there is δ in the other community ($1-\delta$ in the first), where $\delta > 1/2$. Moreover, we assume that the government is restricted in its imposition of rent taxes to a rate τ which is less than 100 percent. The difference between government expenditure and the revenue from rent taxes is raised (or distributed) as before as a uniform lump-sum tax, T_i in community *i*.

In this situation, the citizens will take account of the effect of decisions about public goods on the rents they receive, and it is quite possible that this will entail an inefficient level of expenditure on local public goods. There is in effect "capitalization" of the benefits in land values. To see the considerations involved, consider a position where $N_1 > N_2 > 0$. There are some people working in community 1 whose land is more (i.e., δ of it) in community 2; they are not however the majority. Majority voting means that the level of G_1 is chosen to maximize the utility of a person who owns δ of his land in community 1. The consumption of this person is given by

$$X^{11} = f'(N_1) + (1-\tau)[\delta R_1 + (1-\delta)R_2] - T_1 \tag{22}$$

where R_i denotes the rent per unit of land:

$$R_i = \frac{f(N_i) - N_i f'}{N^*} \tag{23}$$

and the tax required per worker in community *i* is

$$T_i = \frac{G_i}{N_i} \frac{\tau N^* R_i}{N_i} \tag{24}$$

In contrast to the earlier analysis, individuals are assumed to act non-myopically to the extent that they allow for the effect of migration. The total derivative of $U(X^{11}, G_1)$ with respect to G_1 is therefore given by

$$\frac{dU(X^{11}, G_i)}{dG_1} = U_G(X^{11}, G_1) - \frac{U_X(X^{11}, G_1)}{N_1}\left(1 - \frac{N_1 dX^{11}}{dN_1}\frac{dN_1}{dG_1}\right) \tag{25}$$

From this we can see that the level of public goods is influenced by two considerations not previously present: the difference in the interests of different community members and the effect of migration. To see the effect of the former, suppose that $dN_1 / dG_1 = 0$. The level of public goods is then determined by equating the sum of the marginal rates of substitution to the MRT, but assuming that everyone places the same value on the public good as does the majority.

The effect of migration on the consumption of a member of the majority can be broken down into several components:

$$\frac{dX^{11}}{dN_1} = -[-f''(N_1)] + \underbrace{\frac{(1-\tau)}{N^*}\{\delta[-N_1 f''(N_1)] + (1-\delta)[N_2 f''(N_2)]\}}_{\text{rent effect}} + \underbrace{\frac{T_1}{N_1}}_{\substack{\text{spreading}\\ \text{tax burden}}} + \underbrace{\frac{\tau}{N_1}[-N_1 f''(N_1)]}_{\text{tax on landowners}} \quad (26)$$

The first term is the reduction in wages caused by the induced migration. This appears as in the earlier analysis, but the effect on rent is different (previously the rent R_1 accrued to all residents in community 1). Migration to 1 raises rents and hence land values in community 1 and lowers land values in community 2. The net effect depends on the pattern of ownership; and the benefit to the individual depends on the extent to which increases in land values are taxed. If $\delta = 1$, so that land holdings are concentrated, then the net result of the first two terms (wage and rent effects) is a rise in X^{11} if $\tau < 1 - N^* / N_1$. The third term arises from the spreading of the tax burden, as before, but the final effect allows for the fact that some of the increased land value is taxed away.

The level of migration depends on the equilibrium condition, and in this sense those whose land is predominantly in 2 but who live in 1 can exercise an influence, even though they are not decisive in the majority vote. In an equilibrium with $N_2 < N^* < N_1$, their utility must equal that of residents in community 2:

$$U(X^{21}, G_1) = U(X^{22}, G_2) \quad (27)$$

where

$$\begin{aligned} X^{21} &= f'(N_1) + (1-\tau)[\delta R_2 + (1-\delta)R_1] - T_1 \\ X^{22} &= f'(N_2) + (1-\tau)[\delta R_2 + (1-\delta)R_1] - T_2 \end{aligned} \quad (28)$$

Differences among Communities

Returning to the basic model, with myopic decisions, we may consider the consequences of differences in land size or quality. The optimum allocation of population with a utilitarian social welfare function between two islands does not in general entail equal utility (see Exercise 2). Yet the market equilibrium always implies that all individuals have the same utility. The utilitarian optimum cannot therefore be achieved by a market solution. We can however go further and show that the market equilibrium is not in general Pareto-efficient.

To bring this out, consider the effect of allowing a lump-sum transfer subsidy from community 1 to community 2 at rate *T*. Can this transfer raise the common level of utility? This may be seen by taking the derivative of *V* at the equilibrium with respect to *T* and evaluating at $T = 0$. In the case where the communities are identical, then at the equal allocation equilibrium $N_1 = N_2$, and the transfer cannot raise utility. In contrast, where they are asymmetric, with (say) $V_1(N) > V_2(N)$ for all *N*, then the market equilibrium does not involve $N_1 = N_2$, and a transfer can raise the common level of utility. The market equilibrium is not then Pareto-efficient. Where there are only two communities, it is reasonable to suppose that each would perceive this, and that the transfers would take place. But when we increase the number of communities and people proportionately, then any community from which a transfer is due will attempt to be a "free-rider". It would prefer all other donor islands to provide the subsidy, while it enjoys the benefits in terms of the allocation of the population. This may lead to arguments for a central authority to enforce transfers. It may also be noted that in the situation where certain communities have "excess" populations, if free mobility were permitted, there may be attempts to restrict migration.

7.4 Market Equilibria and Optimality: Heterogeneous Individuals

This section allows for differences between individuals in tastes and endowments. These do not in themselves mean that the Tiebout argument cannot be employed, and we begin with a model where communities are mixed (by virtue or the assumptions made about production) but where there is unanimity about the level of public goods and this is Pareto-efficient. The model does however assume that individuals take account of the effects on migration (act non-myopically) and that the number of communities is freely variable. Where these assumptions do not hold, there may exist no local public goods equilibrium and there may be inefficiency – both in the level of public goods, and in the matching of types of people in communities.

The Tiebout Hypothesis in Mixed Communities

The model we employ initially is one in which the conditions of production are such that communities must be mixed. There are two groups, who interact in production, and both are essential to produce a strictly positive output. The two types are denoted by *m* and *n*, with number m_j, n_j in community *j*. Output in community *j* is

$$Y_j = f(m_j, n_j) \quad (29)$$

where

$$f(0, n_j) = f(m_j, 0) = 0\,. \quad (30)$$

Members of the two groups may have different tastes, and their utility in community *j* is written $U_j^i(X_j^i, G_j)$, where $i = m, n$. None the less, under certain assumptions it can be shown that, if each group acts in a utility-taking manner (the natural analogue of price-taking), then in an equilibrium, if it exists, there will be unanimity on the allocation of public goods and it will be Pareto-efficient (For a more general statement of this result, see Stiglitz, 1979).

The first condition for a local public goods equilibrium concerns migration. For equilibrium, all people of a given type must have the same utility in all communities in which they live, and must perceive themselves to obtain a lower utility in any other community[6]. Given that all communities contain people of both types,

$$U_i^m = U_*^m \quad \text{all } i\text{, and } \quad U_i^n = U_*^n \quad \text{all } i \quad (31)$$

Now, any community is assumed to act as a *utility-taker*. In other words, it believes that, so long as it offers to people of type *i* a utility level U_*^i, it can attract an arbitrary number of such people. This is a natural extension of price-taking behaviour. There is assumed, for example, to be an international market for doctors, and if any community offers a lower utility level (taking account of both private consumption and local public goods) then it cannot secure their services.

Now consider the characterization of a Pareto-efficient allocation, on the assumption that all people of a given type are treated symmetrically. This may be formulated in terms of maximizing for a given community (where we drop the subscript *j*):

$$U^m(X^m, G) \quad (32)$$

Subject to

---

[6] We use the term "perceived" because the individual must form a conjecture about what his utility would be if there is no one of exactly his type within the community. For instance, if there are no doctors within a community, a doctor would have to conjecture the wages that a doctor would be paid (after tax). We assume that these conjectures are correct.

$$U^n(X^n, G) \geq U_*^n \tag{33}$$

and

$$G + mX^m + nX^n = f(m,n) \tag{34}$$

Forming the Lagrangean

$$\mathfrak{L} = U^m + \lambda_1 U^n + \lambda_2 [f(m,n) - G - mX^m - nX^n] \tag{35}$$

the first-order conditions are

$$U_X^m = \lambda_2 m \quad \text{and} \quad \lambda_1 U_X^m = \lambda_2 n \tag{36a}$$

$$f_m = X^m \quad \text{and} \quad f_n = X^n \tag{36b}$$

$$U_G^m + \lambda_1 U_G^n = \lambda_2 \tag{36c}$$

Dividing by $\lambda_2 (= U_X^m / m$ and $\lambda_1 = U_X^n / n$ from (36a)), condition (36c) gives

$$\frac{mU}{U_X^m} + \frac{nU_G^n}{U_X^n} = 1 \tag{37}$$

This is the conventional $\sum MRS = MRT$ condition. Moreover, from (36b), the marginal product of each group equals its consumption, and

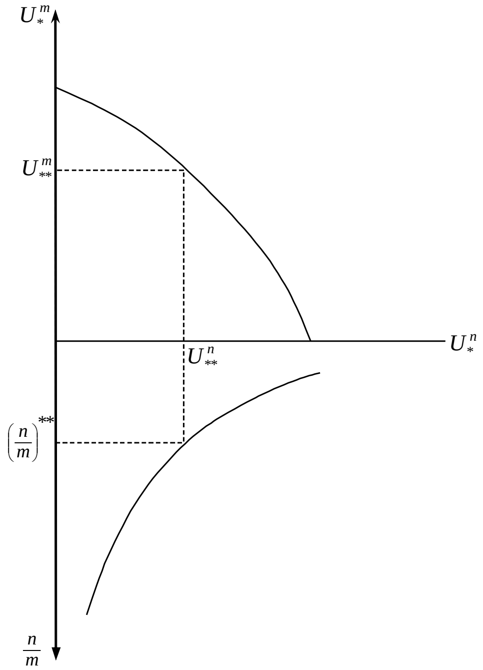
$$f - mf_m - nf_n = G \tag{38}$$

In other words, the Henry George theorem again holds[7].

The nature of the solution may be seen in terms of the utility possibility curve generated by varying U_*^n, and this is shown in Figure 7. For each value of U_*^n, there is a maximum value of U_*^m and the associated ratio of *n* to *m*. This n/m ratio may be thought of as reflecting the relative "demand" for the two types of person. We would normally expect that, as the level of utility we give people of type *n* is increased, the relative demand would be decreased, as shown in the lower part of the diagram.

[7] For a discussion of the conditions under which the theorem is valid, see Arnott and Stiglitz (1980).

Figure 7 Utility-taking Communities



Let us now return to the characterization of the market equilibrium, where one exists. Suppose that the actual relative supplies are $(n/m)^{**}$, as indicated on Figure 7. Then we can show that the Pareto-efficient allocation corresponding to this ratio is a market equilibrium, with utility-taking behaviour. Consider a community in that situation. The supply price of type *n* workers, in terms of utility, is U_{**}^n. If a group of type *m* workers got together, the best they could do is to attain the point on the utility possibility schedule U_{**}^m. They can reach this by forming a community with the optimum population size and supplying the Pareto-efficient level of public goods. Since everyone is then indifferent whether they live in this or another community, the given population size is attainable. Finally, when they all have the given population ratio, with an arbitrarily large number of islands and individuals, everyone will be within a community, and there is no incentive for anyone to move. Thus, under these highly idealized conditions, even though communities are mixed, there is unanimity. Given that each recognizes that there is a utility supply curve for individuals of a particular type, there is no longer any scope for political choice and the market equilibrium generates a Pareto-efficient level of public goods.

The conditions are very strong, however. We have assumed that there is an arbitrary number of communities, and that decisions take account of the effects on migration – the utility-taking assumption. Where these conditions do not hold, there is no guarantee of efficiency, or indeed that a market equilibrium exists.

Non-Existence of Local Public Goods Equilibrium

The possible non-existence of an equilibrium is illustrated by the example of Westhoff (1977), where there is a continuum of consumers with differing preferences and a limited number of communities. In each community the level of public goods is determined by myopic majority voting, this being a most important assumption. Here we give a rather simpler example.

There are three types of local public good, G_1, G_2 and G_3, and three types of person, m, n and o. The preferences of the different types may be written

$$U^m = u(X^m) + v(G_1 + \kappa_m G_3) \tag{39a}$$

$$U^n = u(X^n) + v(G_2 + \kappa_n G_3 + \kappa_N G_1) \tag{39b}$$

$$U^0 = u(X^0) + v(G_3 + \varepsilon G_1) \tag{39c}$$

Where u is strictly increasing, $v(0) = 0$, $0 < \kappa_m < 1$, $0 < \kappa_n < \kappa_N < 1$ and ε is a small positive number. In other words, *m* gets no utility from good 2 and prefers 1 to 3, *n* prefers good 2 to good 1, and slightly prefers good 1 to good 3; *o* gets no utility from good 2, and almost none from good 1[8]. There are assumed to be an odd number P_i of each type, where

$$P_m < P_o < P_n \text{ and } P_m + P_o > P_n \tag{40}$$

Everyone is assumed to have the same income, *I*, and the public good is financed by a uniform poll tax.

The technical conditions of production of the public good are such that it is either produced or not, and it has to be used exclusively for one of the three types (e.g., there can only be one television channel and it has to be used *either* for sport *or* for music *or* for news) [9]. The cost is fixed at unity (independent of the type of use). Finally, we assume that

$$u(I - 1/2) + v(1) > u(I) > u(I - 1) + v(1)$$

This means that a group of two people of the same type would choose to produce the preferred public good, but one person on his own would not.

Within each community, the decisions regarding public goods are made by a majority vote. Voters are myopic and take no account of the effect on migration. Migration takes place

8 It should be noted that the example is not based on cyclical voting. The preferences are (in decreasing order) (1, 3, 2), (2, 1, 3) and (3, 1, 2), so that, unless one of types *n* or *o* was in an absolute majority, public good 1 would always be selected by the population as a whole.

9 This assumption is not essential; indeed, no one would vote (myopically at least) for a mixed programme.

where a person can obtain a higher utility level in a different community, including the possibility of not joining. We consider in turn the possible equilibrium configurations, and indicate how a set of conditions can be derived under which no equilibrium exists:

1. *A single community* (which we denote by *MNO*). Majority voting leads to the choice of good 1, preferred to 2 by *m* and *o* (who form a majority), and preferred to 3 by *m* and *n* (who form a majority). However, if there is a strictly positive level of provision, then for small enough ε the benefit to a person of type *o* is insufficient to outweigh the cost of the poll tax. He therefore migrates to form a new community on his own; hence this is not an equilibrium.
2. *Two Communities* (*MN* and *O*). In the former, type *n* is now in a majority, so that good 2 is produced. Type *m* obtains no utility from good 2, so that its members migrate.
3. *Two Communities* (*N* and *MO*). If the type *m* members join with type *o*, then good 3 is produced (since *o* is in a majority). For a given quantity of the public good, the tax rate in *MO* is lower than in *N*, since the former has a bitter population. Type *n* prefers good 2 to good 3, but if the margin of preference is not too great, then members of type *n* migrate.
4. *Two Communities* (*M* and *NO*). In the latter, good 2 is produced, since type *n* is in a majority. Type *o* migrates, since its members get no utility from the good.
5. *Three Communities* (*M, N* and *O*). A person of type *m* considers joining the community with type *O*. In *O*, good 3 is produced rather than good 1, but the tax from a given quantity is lower (since $P_o > P_m$). If the relative preference for good 1 is sufficiently small, then type *m* migrates.
6. *Equilibria*, where there are people of type *i* in more than one community.

The equilibrium condition is that the level of utility of people of type *i* must be the same in all communities in which they reside. Suppose, for example, that we have (*MN* and *NO*). For this to be an equilibrium, type *n* must be in a minority in both, so that the former produces G_1, and the latter G_3. If the relative preference of type *n* for good 1 is slight, then the total numbers I the two communities must be close (so as to equalize the tax burden). On the other hand, for P_m small, this involves type *n* being a majority in the first community. Hence it cannot be an equilibrium. Other cases of split populations can similarly be ruled out.

Land Values

We saw earlier that, with private land ownership, people voted for public goods not just on the basis of their direct utility but also allowing for the effect of any induced migration on the value of their land. In order to bring out the implications in the context of heterogeneous tastes, we now consider a model where land is owned by people whose only concern is with the effect of the choice of public goods on the land value. We have, for example, a lake in each community that can be used by the residents for two mutually exclusive activities (swimming and boating), and the public decision concerns the proportion, η, of the time for which the lake is devoted to the first of these activities. Individual preferences (which are otherwise independent of location) are given by the distance from their preferred value, η^h:

$$v^h(\eta) = \left|\eta^h - \eta\right| \tag{41}$$

with η^h varying across individuals, with median η^*. This means that, if there are two communities, offering η_1 and η_2 respectively (where $\eta_2 > \eta_1$), then all people with $\eta^h < 1/2(\eta_1 + \eta_2)$ live in community 1 and the remainder live in community 2.

The decision about η is made collectively by those owning land in the community, who are assumed to get no direct enjoyment from the public good, either because the local authority is a land development agency rather than a democratic body, or because decisions are made by a generation who are at a stage of the life cycle when they have acquired assets (land) but lost the taste for water sports. The majority voting outcome is to maximize land values, and this is taken to coincide with maximizing the number of people who wish to live in the community[10]. Moreover, it is assumed that each community takes land usage in the other as given. It can then be seen that the position $\eta_1 = \eta_2 = \eta^*$ is an equilibrium of the model. Where the land use is equal to that preferred by the median, neither community can raise land rents by departing from the median, taking the behaviour of the other community as given.

In this equilibrium, the two communities produce exactly the same public goods, in spite of the heterogeneity of tastes. Obviously, this is not a social welfare optimum; it is only the preferences of the marginal individual that are taken into account. The preferences of the intramarginal individuals – virtually the total population – are completely ignored. Any social welfare function that did not assign all the weight to the median individual in society would have the different communities produce different public goods. Moreover, it is quite possible to construct examples where the equilibrium is not only inconsistent with any social welfare function that does not give all weight to the median, but is actually Pareto-inefficient.

The similarity between this model and the standard theories of product differentiation, in particular that derived from the seminal work of Hotelling (1929), should be clear. Indeed, the issues are closely parallel. The number of communities is limited by the returns to scale

10 This can be obtained from a model of demands derived from utility maximization where the utility functions are Cobb-Douglas and on the basis of certain assumptions about the formation of expectations regarding future land prices.

associated with public goods, while the number of commodities is limited by the returns to scale in production. The market solution involves firms maximizing profits and ignoring the effects of their actions on the profits of others. So too, here, communities do not pursue the correct objective function; they maximize the value of land, rather than social welfare, and ignore the effect on intramarginal individuals and on the other communities[11].

Rich and Poor Communities

Differences in tastes are no doubt significant, but probably much more important are differences in endowments. One of the most striking aspects of local government in the United States and other advanced countries is the marked difference in the wealth of local communities.

If the local public good is in fact a publicly provided private good, then there are clear reasons why the rich would be interested in excluding the poor[12]. With the case of pure public goods, with which we are concerned here, there is no additional cost to supplying a further individual within a geographical area: the consumption of the poor does not detract from that of the rich. On the other hand, the rich may be interested in excluding the poor because of differences in the levels of demand for the public goods and the redistribution implied in the method of financing. The poor may vote for a different combination of public goods and taxation, and if there is a specified method of financing (e.g., a property tax) then the taxes paid may not match benefits received.

In order to illustrate the way in which exclusion may be practiced we assume, for simplicity, that there are only two groups in the population. The rich, referred to by a superscript *R*, have income (*per capita*) of M^R, while the poor, referred to by a superscript *P*, have *per capita* income M^P, where $M^P < M^R$. The income is assumed independent of the number of people living in the community –there are no costs of congestion in terms of diminishing returns. The utility functions are given by $U^P(X^P, G^P)$ and $U^R(X^R, G^R)$. With complete exclusion, the equilibrium for each group is given by maximizing $U^i(X^i, G^i)$ subject to $N^i X^i + G^i = N^i M^i$ where N^i is the number in each group. The solution values are denoted by an asterisk and are shown in Figure 8(a).

Now let us suppose first that direct (costless) exclusion can be practiced by either group. Each community then compares the utility obtained in the exclusionary equilibrium with that

11 We should note that recent work has established the special nature of the Hotelling model, and the problems that arise when there is more than one dimension over which the firms can compete.

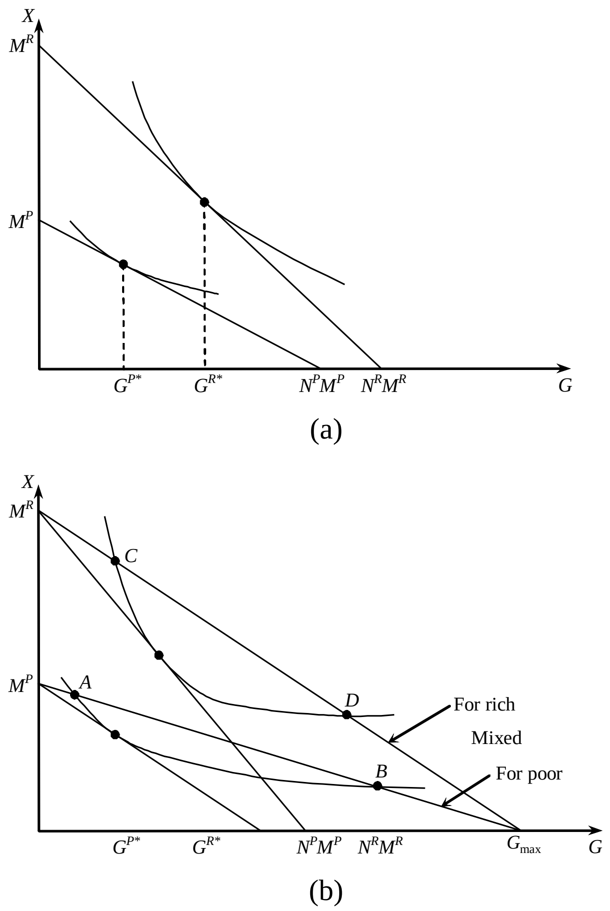
12 This may well be the case with education. The provision of a uniform level of education to all children regardless of the wealth of their parents, financed by a proportional wealth tax, involves in effect considerable redistribution; and this is still greater if taxation is progressive. For analysis of this case, see Stiglitz (1977).

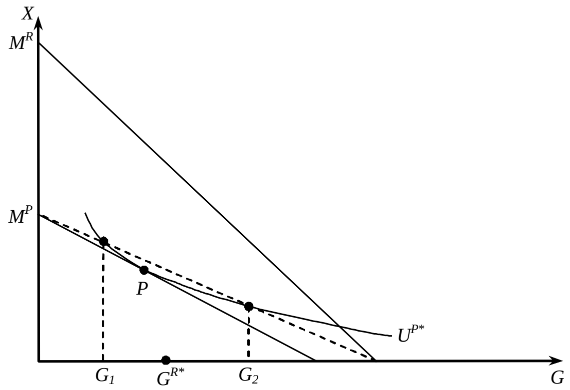
obtainable if they merge. The production possibilities of the merged community are:

$$N^P X^P + N^R X^R + G = N^P M^P + N^R M^R + G_{\max} \qquad (42)$$

If we assume that the tax levied is a proportional income tax at rate t (an assumption that is critical to much of the analysis), then the merged community offers the poor person points along the line joining M^P to $G_{\max}$ in Figure 8(b), and the rich person points on the line joining M^R to $G_{\max}$.

Figure 8 Rich and Poor Communities





(c)

As drawn in Figure 8(b), there is scope for both groups to gain from a merged community. The poor gain if the level of public good is set between that corresponding to the points *A* and *B*, and the rich gain between *C* and *D*. However, the outcome depends on the process by which the conflicting interests of rich and poor in the mixed community are reconciled. Suppose that the poor are in a majority, and that they are able to exercise political control. The decision depends on the level of sophistication exercised in voting, and if they vote myopically, the resulting equilibrium may well be inefficient – as we have seen before. Suppose that the poor maximize U^P without regard to the position of the rich. If the resulting level of U^R is less than that obtainable at the exclusionary equilibrium (e.g., to the right of *D* in Figure 8(b)), the rich will opt out. There will be two separate communities, even though both could be better off with a single, integrated community. Since there are no diminishing returns, the social optimum is that where both groups live in one community – and share as fully as possible in the spill-overs from public goods.

How is this affected if direct exclusion is not possible? Suppose that a poor person can choose to live in a rich community if he wishes. In order to preserve a segregated community, the rich are restricted to choosing a tax rate and a level of public spending that do not attract the poor. We can then put eh problem of the rich community as

$$\max U^R(X^R, G^R)$$

subject to

$$X^R = (1-t^R)M^R$$
$$t^R = G^R / N^R M^R$$
$$U^P[(1-t^R)M^P, G^R] \le U^{P*} \quad (43)$$

where U^{P*} is the level that the poor achieve in the exclusionary case. Figure 8(c) shows the solution to this problem diagrammatically. The possibilities open to a single poor person joining the rich community are indicated by the dashed line. Consequently there are two *exclusionary points*. For levels of G^R below G_1 or above G_2 the poor person will not be attracted to the rich community; for other levels of G^R will be. As shown, the exclusionary constraint is binding. The rich community, in order to exclude the poor, chooses either a higher or a lower level of public expenditure than it would have chosen in the equilibrium where direct exclusion was feasible[13]. Essentially, in the higher equilibrium, the tax rate is so high that the poor cannot afford to live in the community; the amount of private consumption that they are left with is "inadequate". In the lower equilibrium, the government expenditure is very low. The rich can purchase private goods that are "substitutes" for the public good; the poor, however, cannot do this, and thus they prefer to remain in their own communities.

One can observe both extremes of behaviour in rich communities in the United States.

Concluding Comments

The Analogy between local public goods that are competitively supplied by different communities and the conventional competitive equilibrium model for private goods is a suggestive one, but, for reasons, that we noted in the introduction, the analogy is of more limited validity, and the analysis is of far greater complexity than Tiebout's original article suggested. There are certain circumstances in which communities and individuals exist in just the right proportions so that every community is at the optimal size, and where individuals act non-myopically, in which a local public goods equilibrium is Pareto-efficient. But in the more realistic case, where there is a limited number of jurisdictions, or where people act myopically, equilibrium may not exist, and when it does exist it may not be Pareto-efficient. The equilibria displayed inefficiencies in (1) the numbers of individuals within the community; (2) the level of public goods and the choice of public goods supplied within each community; (3) the number of communities formed; and (4) the matching of individuals together to form communities. Even, therefore, without introducing any social judgements as to the desirability of certain types of community (e.g., favouring integrated communities), there may be strong arguments for intervention by a central authority.

[13] We need also to allow for the possibility that the rich do not act collusively and that it may pay individual members of the rich community to join the poor. The relevant budget line joins M^R to $N^P M^P$.

7.5 Optimal Federalism and Grants-in-Aid: Normative Analysis

First-Best Policy Environment

Whether grant-in-aid has any role in an optimal, first-best federalist system of governments depends upon the underlying model used to establish the notion of a social welfare optimum. Recall that in the conventional model of optimal federalism redistributional policy is the sole responsibility of the national government, whereas allocational functions reside in the lowest level governments consistent with pareto optimality. Consequently, only the national government is concerned with social welfare optimization as traditionally defined. The lower level governments care only about efficiency.

Grants-in-aid are unnecessary in this model, as long as the policy environment is truly first best and a perfect correspondence of jurisdictions exists for all allocational problems. The national government satisfies its interpersonal equity conditions with lump-sum taxes and transfers among individuals (and firms, with decreasing cost production), exactly as in the single-government model of the public sector. Similarly, all governments, whether national or "local", interact only with the individual consumers and firms within their jurisdictions when correcting for resource misallocations. Thus, they simply follow the normative decision rules derived under the assumption of a single government. There is no need for the grant-in-aid, because no government need be directly concerned with any other jurisdictions. In our view, this is yet another reason for rejecting the traditional model of optimal federalism. It seems implausible that intergovernmental relations would be of no consequence in a federalist system of governments, even under first-best assumptions.

Our alternative model of federalism, defined the social welfare optimum as an equilibrium in which each government maximized its own dynastic social welfare function, with the restriction that the arguments of each government's social welfare function are the social welfare functions of those governments immediately below it in the fiscal hierarchy. Grants-in-aid are required in this model to resolve the distribution question, since all but the lowest level governments must tax and transfer resources lump sum among the governments immediately below them in the fiscal hierarchy. In the parlance of grants-in-aid, these lump-sum grants would be *unconditional*, *nonmatching*, and *closed-ended*: unconditional, because one government cannot dictate to any other government how to dispose of the funds, the "states' rights" criterion; nonmatching and closed-ended, because the interpersonal equity conditions require straight resource transfers of some finite amount. Notice, too, that the "grants" are negative for those governments that must surrender resources.

Our alternative model shares with the conventional model the attribute that grants-in-aid are not required for allocational purposes in a first-bet policy environment with a perfect correspondence of local functions. Simultaneously with satisfying all possible interpersonal equity conditions, satisfying all necessary pareto-optimal conditions proceeds government-by-government in the usual manner. To develop a further role for grants-in-aid, then, requires introducing some second-best distortion into the policy environment.

Second-Best Policy Environment

Imperfect Correspondence

A second-best restriction commonly analyzed In the literature is a maintained imperfect correspondence for an externality-generating activity, which causes each local government to follow the wrong decision rule. Imagine the following situation[14]. Community A, consisting of H_A individuals, provides a Samuelsonian nonexclusive public good in amount $\overline{X}_G$, the services of which are consumed directly by its own citizens. In determining the amount $\overline{X}_G$, the government of A follows the standard first-best decision rule:

$$\sum_{h_A=1}^{H_A} MRS^{h_A}_{X_G, X_{h_A 1}} = MRT_{X_G, X_1} \quad (44)$$

Suppose that H_B citizens of contiguous community B benefit from the existence of X_G in community A even though they cannot directly consume the services of X_G. For example, X_G may be police protection which has the spillover effect of reducing criminal activity in community B. In effect, then, X_G in community A becomes an aggregate external economy for the citizens of community B, entering into each person's utility function. The aggregate gain to community B's citizens on the margin can be represented as:

$$\sum_{h_B=1}^{H_B} MRS^{h_B}_{X_G, X_{h_B 1}}$$

with each MRS^{h_B} measured positively. The true first-best pareto optimal conditions are, therefore:

$$\sum_{h_A=1}^{H_A} MRS^{h_A}_{X_G, X_{h_A 1}} + \sum_{h_B=1}^{H_B} MRS^{h_B}_{X_G, X_{h_B 1}} = MRS_{X_G, X_1} \quad (45)$$

---

14 A similar example appears in Wallace Oates, *Fiscal Federalism*, Harcourt Brace Jovanovich, New York, 1972, pp.95-104. Oates's Chapter 3 and appendices provide an excellent analysis of the uses of grants-in-aid within the conventional model of fiscal federalism.

Without any intervention from a higher level government in the fiscal hierarchy, X_G will be misallocated (presumably undersupplied), because community A ignores the second set of terms on the left-hand side of Equation (46). The situation exemplifies the notion of an imperfect correspondence, since the jurisdictional boundaries of community A, which makes the allocational decision on X_G, do not encompass all citizens affected by the production and consumption of X_G.

There is no deed for a grant-in-aid in this case. The next highest government in the fiscal hierarchy, one that includes the citizens of both A and B, could provide X_G to the citizens of A in accordance with Equation (46). It does have the option, however, of allowing community A to decide on the level of X_G as before and influencing its decision with an appropriate grant-in-aid. Hence, its choice is fully analogous to that in single-government models of aggregate externalities, in which the government can either dictate the consumption of the good or use a Pigovian subsidy and maintain decentralization.

A society committed to federalism would presumably choose the grant-in-aid since it promotes decentralized local autonomy, much as a single government under capitalism would choose decentralized subsidies for aggregate externalities.

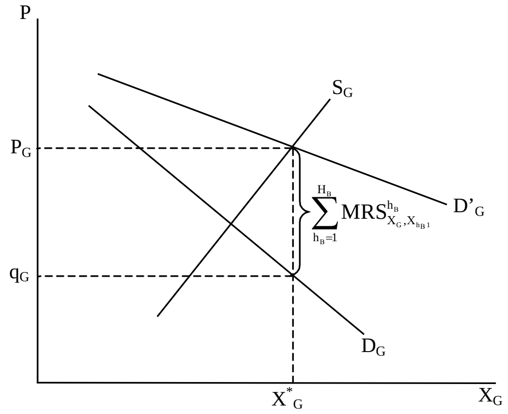
The appropriate subsidy is a per-unit subsidy, equal to the aggregate gain to the citizens of B on the margin, or:

$$s = \sum_{h_B=1}^{H_B} MRS^{h_B}_{X_G, X_{h_B 1}}$$

which, in this case, is a grant-in-aid from the higher level government to community A. The grant, depicted in Figure 9, would be conditional, matching, and open-ended: conditional on expenditures for X_G with a matching rate equal to the ratio s/P_G at the optimum, where P_G is the producer price of X_G (see Figure 9), and open-ended because it is not optimal to limit the size of the grant to any value other than $s \cdot X_G^*$, where X_G^* is determined by the receiving government.

These simple grant-in-aid examples can be quite misleading, however. Localities tend to provide the same kinds of public services, so that the actual pattern of externalities is likely to be far more complex than depicted in our simple story. If community A's police expenditures generate external economies in community B (and, possibly, other neighboring communities), then community B's police expenditures can be expected to generate external economies for all its neighbors, including A. But, if this is so, then the spillover component of the externality is likely to be individualized by community, in which case the required pattern of grants-in-aid becomes extremely complex.

Figure 9



To see the possibilities, define (X_{G_A} , X_{G_B} , …, X_{G_C}) as the vector of the individual community outputs, C in number, with $X_G = X_{G_A} + X_{G_B} + \ldots + X_{G_C}$ the aggregate output of X_G across all communities. If the aggregate X_G enters each person's utility function, then a single matching grant is appropriate, with $s = \sum_{\text{all h}} MRS^h_{X_G, X_{h1}}$. Referring again to police expenditures, the assumption is that the spillover effects on criminal activity within a region depend upon aggregate police expenditures across all communities within the region.

Although expenditures on police may give rise to an aggregate externality, each community is more likely to receive the most benefit from police expenditures in its contiguous communities and increasingly less benefit from police expenditures in ever more distant towns. If so, then the spillover externality remains individualized and pareto optimality requires a complex set of subsidies, one for each town. Moreover, the subsidies are interdependent, with each matching rate dependent upon police expenditures in every community. Thus, the situation is exactly analogous to case of individualized externalities arising from private sector activities.

We have seen that aggregate externalities admit to relatively simple solutions whereas individualized externalities do not. The existence of federalism, with imperfect correspondences, adds nothing to the complexity of the problem. Even if the next highest government in the fiscal hierarchy chose to provide X_G , it would still follow the same decision rules, providing, of course, that the direct services of each individual X_{G_i} are consumed exclusively by members of the corresponding community, as posited in our example. If there is not even a perfect correspondence for the direct consumption of these services, then a set of grants-in-aid is unlikely to be appropriate. In this case, the next highest government should decide upon the level of the aggregate $\overline{X}_G$ and its individual subcomponents. For instance, police services may be exclusive by town because the laws of each town forbid police to cross jurisdictions. But if there really is an imperfect correspondence here then these exclusions are arbitrary and nonoptimal. Fewer, larger police departments with a regional orientation would be the optimal solution, but these would have to be provided by the next highest government in the fiscal hierarchy.

Note finally that the analysis carries through in both the conventional model of federalism in which only the national government has a social welfare function, or in our alternative model in which each government possesses a social welfare function. As long as income is optimally distributed according to the interpersonal equity conditions of each model, allocational issues dichotomize from distributional concerns just as in single-government models.

These same points apply to externalities generated by private sector activity. Unless the direct component of the activity can be localized within a single community (say, a production externality arising at a particular site), grant-in-aid are unlikely to be pareto optimal. And even if pareto optimality could be achieved by the grants-in-aid, it may not be the most direct fiscal tool. Not surprisingly, grants-in-aid are most appropriate for publicly provided services.

Consider the example of a production site located in community A. Suppose its external diseconomies affect both citizens in A and those in other neighboring towns. If town A taxes the producer, it will undoubtedly base the tax on the marginal damages only to its own citizens. The next highest government could design a negative conditional matching grant (i.e., a tax) levied on town A that would optimally adjust for the broadened scope of the external diseconomy, but an additional direct tax on the producer would seem less cumbersome. Other more complex situations, such as the individualized pollution example in which production at multiple sites along a river generates external diseconomies for the other firms, can best be solved by producer taxes established by a higher level government and not by a set of grants-in-aid to a number of localities. There is no compelling reason to involve lower level governments as intermediaries in correcting for private sector externalities.

7.6 Alternative Design Criteria

That actual grant-in-aid bear little relationship to theoretical design criteria is hardly surprising, because the theory is so difficult to apply In this instance. In terms of our alternative model, distributional norms based on social welfare functions can never be more than suggestive to the policymaker. In terms of imperfect correspondences for externality-generating public services, varying matching formulas across “local” governments on the basis of marginal external benefit or harm may be unconstitutional. Faced with these realities, economists have resorted to developing practical design criteria that are at least roughly consistent with the underlying theory.

A surprising feature of the more practical literature is that it has tended to focus on distributional concerns, more in line with our alternative model of optimal distribution under federalism than with the mainstream position. A principal question is how to design grants to correct for perceived resource imbalances either across states (for federal grants) or across localities (for state grants). This focus makes sense at a practical level because many federal and state grants in the Unites States do attempt to direct aid disproportionately toward poorer states and localities. Examples are the federal grants to support states' public assistance payments under Temporary Assistance to Needy Families (TANF) and Medicaid and state grants to support local public school expenditures.

The LeGrand Guidelines

In the mid-1970s, Julian LeGrand suggested three sensible practical guidelines for grant-in-aid programs whose goals are redistributional. First, the grants must be a function of the real income or wealth of the receiving governments, commonly referred to as its fiscal capacity. LeGrand argues that jurisdictions with fiscal capacities below some target level should receive aid and jurisdictions above the target should pay a tax (receive a negative grant). In contrast, existing grant-in-aid programs always give something to all governments. The political motivations behind giving something to everyone are clear, but such grants tend by their very nature to have limited redistributional power. Note, also, that fiscal capacity accounts for differences in prices across communities, the relative expenditures required to achieve comparable levels of public services.

LeGrand's second guideline is that the amount of aid received (tax paid) should be independent of any expenditure decisions made by the receiving government. This guideline honors two principles: Redistributional policy ought properly be concerned with each government's overall initial level of resources, and, consistent with the federalist ideal, the grantor should not attempt to influence the specific spending decisions of lower level governments.

LeGrand's third guideline states that grants should vary directly with the receiving government's fiscal effort, the idea being that governments with less interest in providing public services should receive correspondingly less aid. This criterion is somewhat troublesome because it tends to contradict the second guideline. It implies that the grantor will try to influence the overall level of public services beyond the giving or taking of resources, although not the composition of these services. In any case, it is a commonly accepted principle. The

U.S. Congress has frequently incorporated effort parameters into aid formulas[15].

LeGrand shows that basing grants-in-aid on differences in fiscal capacity automatically incorporates each community's fiscal effort. To see this let:

T_i = total taxes per capita collected by government i.

P_i = a price index of public services provided by government i.

E_i = the effective tax rate in government i, the effort parameter

Y_i = the per capita tax base in government i.

The fiscal capacity of government i is Y_i/P_i. LeGrand defines a purchasing power effort (PPE) ratio as:

$$PPE_i = \frac{T_i}{E_i P_i} \qquad (46)$$

where purchasing power refers to the purchasing power of the taxes. But $T_i = E_i Y_i$. Therefore,

$$PPE_i = \frac{T_i}{E_i P_i} = \frac{E_i Y_i}{E_i P_i} = \frac{Y_i}{P_i} \qquad (47)$$

LeGrand's PPE ratio is the same as fiscal capacity.

Under LeGrand's preferred grant-in-aid formula, the grantor picks a target PPE ratio or fiscal capacity, $PPE_T = Y_T/P_T$. The per-capita grant, G_i, is then designed to put all jurisdictions at that target PPE_T. Thus, G_i is such that:

$$\frac{T_i + G_i}{E_i P_i} = \frac{Y_i}{P_i} = \frac{G_i}{E_i P_i} = \frac{Y_T}{P_T} \qquad (48)$$

or

$$G_i = E_i\left(\frac{P_i}{P_T} Y_T - Y_i\right) \qquad (49)$$

The grant received (tax paid) depends upon a locality's fiscal effort as embodied in the tax rate, and its relative fiscal capacity, defined as the difference between its per-capita tax base and the target per-capita tax base adjusted by the differences in the prices of public services in the

15 When Congress replaced Aid to Families with Dependent Children (AFDC) with TANF it stipulated that the states could not reduce the expenditures on public assistance that they had been making under AFDC.

locality relative to the target community. Hence, all three of LeGrand's criteria are satisfied by this simple formula.

LeGrand's formula would lead to a substantial amount of redistribution, since richer than average towns would actually pay taxes. By including E_i, the formula also addresses a problem with federalism that many people find particularly inequitable; namely, welthy communities can offer better public services than the poorest communities even though their tax rates are only a fraction of the tax rates in the poorest communities. LeGrand's formula doubly rewards the poor communities who have high tax rates. Finally, if one concedes that social welfare rankings may properly be functions of fiscal effort, among other things, this simple formula is reasonably consistent with the redistributional decision rules of our alternative model of fiscal federalism. It bears roughly the same relationship to these norms as the Haig-Simons ability-to-pay criterion does to the interpersonal equity conditions of single-government social welfare maximization. Both substitute income for utility, although the Haig-Simons criterion contains nothing comparable to the fiscal effort term.

Applying LeGrand's Principles: Bradbury *et al*.

LeGrand's grant formula is much too egalitarian to be politically acceptable. A more practical version of his proposal would be to close only a portion of the disparities in fiscal capacity:

$$G_i = kE_i\left(\frac{P_i}{P_T}Y_T - Y_i\right) \quad k < 1 \tag{50}$$

subject to the constraints:

$$G_i \geq 0 \quad \text{all } i \tag{51}$$

and

$$\sum_i G_i N_i = D \tag{52}$$

where D is the budget given to the distributional granting authority for the grants to reduce fiscal disparities (N_i is the population of locality i). Equation (51) ensures that no communities with fiscal capacities greater than Y_T/P_T would be taxed under the formula. The granting authority would maintain the budget constraint by varying k and the reference community Y_T. A high Y_T combined with a low k gives smaller amounts of aid to more communities, and vice versa. Taxes to support the grants would come from general tax revenues, not from levies on the high-fiscal-capacity communities.

Katherine Bradbury *et al*. (1984) were commissioned by the Massachusetts state government to design an equalizing grant program for distributing 5% of the state's grant budget, approximately $110 million, to the cities and towns with low fiscal capacities. They approached the problem in the spirit of LeGrand, but they used a different measure of fiscal disparity in the aid formula. They based their formula on what they termed a community's fiscal gap, equal to:

$$\text{Gap}_i = \overline{E}C_i - \bar{t}B_i \qquad (53)$$

where

$\overline{E}$ = the average per capita expenditures across all communities,
C_i = the cost of providing the average expenditures in community i.
$\bar{t}$ = the average tax rate across all communities
B_i = the per-capita tax base in government i.

In other words, a community's fiscal gap is the difference between what it would have to spend to provide the average local public service bundle and the tax revenues it would raise if it applied the average tax rate across all communities to its tax base.

A reference, or target, fiscal gap is defined in the same way:

$$\text{Gap}_* = \overline{E}C_T - \bar{t}B_T \qquad (54)$$

The grant formula closes a portion of the difference between a community's fiscal gap and the reference fiscal gap,

$$A_i = k(\text{Gap}_i - \text{Gap}_T) = k[\overline{E}(C_i - C_T) - \bar{t}(B_i - B_T)] \qquad A_i \ge 0 \qquad (55)$$

where A_i is the per-capita grant. The first term on the right-hand side (RHS) is the cost disadvantage suffered by community i relative to the reference community, and the second term is community i's tax-base disadvantage. The main deviation from LeGrand's principles is that the Bradbury *et al.* formula does not include an effort term.

Bradbury *et al.* argue that the average expenditure level $\overline{E}$ and the cost of providing the services C_i should be based on regression analysis. They also believe that the relative cost advantages or disadvantages should reflect only environmental factors that are beyond the immediate control of the communities, such as population density, the condition of the housing stock, and the crime rate. They posit a supply of expenditures function:

$$E_i = E_i\left(\vec{S}_i, \vec{P}_i, \vec{C}_i\right) \quad (56)$$

where

$\vec{S}_i$ = the vector of public services offered in community i

$\vec{P}_i$ = the vector of input prices for the factors used to produce the public service vector in community i

$\vec{C}_i$ = the vector of environmental factors that influence the cost of providing the public service in community i

The demand side of the model is a standard median voter model (described later) in which the median household solves an as-if maximization problem in terms of a numeraire private composite commodity and the vector of public services, subject to its individual budget constraints and the overall community budget constraint. The supply relationship, Equation (56), enters as the expenditures in the overall community budget constraint. The analysis leads to a reduced-form equation for overall public expenditures (individual public service outputs are not measurable):

$$E_i = f\left(\overline{V}_i, \vec{A}_i, \vec{P}_i, \vec{D}_i, \vec{C}_i\right) \quad (57)$$

where:

$\overline{V}_i$ = the average (mean) property value in community i

$\vec{A}_i$ = a vector of other resources available to community i, such as other grants-in-aid

$\vec{D}_i$ = a vector of taste parameters "demand" factors

The demand factors Bradbury *et al*. chose were per-capita income and the percentage of the population ≥ 65. The five environmental cost factors were population density, the condition of the housing stock, the ratio of children in the public schools to the entire population, the crime rate, and the poverty rate. They had no data on variation of input prices, $\vec{P}$, across the cities and towns. Equation (57) was estimated on a sample of 300 Massachusetts and towns.

To estimate $\overline{E}$ in the grant formula, Equation (55), they set the values of all the explanatory variables in Equation (57) equal to their average values across all 300 cities and towns. To compute the relative cost term C_i in their grant formula, they estimated $\hat{E}_i$ by setting the values of all the explanatory variables except $\vec{C}$ at their average values, and the values of the variables $\vec{C}$ at their actual values in community i. Then $C_i = \hat{E}_i/\overline{E}$ or $C_i\overline{E} = \hat{E}_i$ in the grant formula.

In applying the Bradbury *et al*. formula, the state:

1. Set the reference $Gap_T = 0$, to maximize the number of communities receiving aid.
2. Set an additional condition that every community receives a grant of at least $5 per capita from the budget set aside for these grants.
3. Defined the fiscal gaps to include existing state aid, $\overline{A}$:

$$Gap_i = \overline{E}C_i - \bar{t}B_i - \overline{A}_i \tag{58}$$

Finally, since all grant, expenditure, and tax-base variables are in per-capita terms, the cities and towns received a proportion of the entire distribution budget equal to the product of their fiscal gaps and population divided by the sum of the fiscal gaps times populations of all the aided localities.

Bradbury *et al*. proposed that the aid be adjusted each year using the same estimating equation for E_i and just adjusting the values of the explanatory variables.

Redistributing Through Matching Grants

As our final example of practical grant design criteria we will consider Martin Feldstein (1975)'s proposal for remedying unequal local public educational expenditures. In the early and mid-1970s, a number of state supreme courts ruled that financing public educational expenditures entirely from local property taxes was inherently discriminatory, since wealthier communities could provide better education with less fiscal effort, that is, lower tax rates[16]. The states were required to design a more equitable statewide financial arrangement which would somehow provide transfers from the wealthier to the poorer communities. Feldstein reasoned that the courts' decisions imply a fiscal solution which sets the elasticity of educational output with respect to wealth equal to zero ($E_{Ed,W} = 0$). He suggested using a matching grant for this purpose, in which the matching rate applied to any one community is inversely proportional to its wealth. To achieve this goal, one needs reliable econometric estimates of the price and income (wealth) elasticities of educational expenditures independent of a new grant program. These estimates can then be used to design the required matching rates.

To see how this would work, suppose it is possible to estimate a constant elasticity demand-for-education equation across communities of the form:

$$Ed = CP^{\alpha}W^{\beta} \tag{59}$$

16 *Serrano v. Priest* in California was the landmark decision. Refer to *Serrano v. Priest*, L.A. 29820, Superior Court No. 938254.

where:

Ed = a measure of educational output per capita.

P = the price of a unit of educational output.

W = a measure of per capita community wealth.

α, β = the price and wealth elasticities.

C = a constant term embodying all other factors influencing the demand for education.

Rewriting, Equation (59) in log form:

$$\log Ed = C' + \alpha \log P + \beta \log W \qquad (60)$$

Next, define a matching aid formula that makes the net-of-aid price a function of wealth according to the constant elasticity form:

$$\mathrm{P} = \mathrm{W}^{\mathrm{k}} \qquad (61)$$

or

$$\log P = k \log W \qquad (62)$$

where:

k = the elasticity of the net price with respect to wealth.

Substituting Equation (62) into (60) yields:

$$\log Ed = C' + \alpha(k \log W) + \beta \log W = C' + (\alpha k + \beta) \log W \qquad (63)$$

With this matching program:

$$\frac{\partial \log Ed}{\partial \log W} = E_{Ed,W} = \alpha k + \beta \qquad (64)$$

Setting $E_{Ed,W} = 0$ implies:

$$k = -\beta / \alpha \qquad (65)$$

Thus, the required matching rate elasticity just equals the ratio of the wealth and price elasticities of education within the state, at least for a log-linear demand for education function. Feldstein (1975) estimated an education equation for a cross section of Massachusetts communities to demonstrate his technique. The required matching rate elasticity for Massachusetts turned out to be between .33 and .37.

It is worth repeating that matching grants for which the matching rate varies with respect to income or wealth have no role in the first-best theory of federalism and are at best only suggested by second-best considerations. Nonetheless, if the law requires neutralizing the effect of wealth on educational opportunity within states, then Feldstein's grant-in-aid formula provides a direct way of achieving this goal.

References

Atkinson, A.B. and J.E. Stiglitz (1980) *Lectures on Public Economics*, McGraw-Hill, UK.

Berglas, E. (1976) "Distribution of tastes and skills and the provision of local public goods", *Journal of Public Economics*, 6, pp.409-23.

Bradbury, K., H. Ladd, M. Perrault, A. Reschovsky, and J. Yinger (1984) "State Aid To Offset Fiscal Disparities Across Communities", *National Tax Journal*, June.

Buchanan, J. (1965) "An Economic Theory of Clubs", *Economica*, 32, pp.1-14.

Feldstein, M. (1975) "Wealth Neutrality and Local Choice in Public Education", *American Economic Review*, March.

Hotelling, H. (1929) "Stability in competition", *Economic Journal*, 39, pp.41-58.

LeGrand, J. (1975) "Fiscal Equity and Central Government Grants to Local Authorities", *Economic Journal*, 85(3) No.339, pp.531-47.

McGuire, M.C. (1974) "Group segregation and optimal jurisdictions", *Journal of Political Economy*, 82, pp.112-32.

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

Pauly, M.V. (1967) "Clubs, commonality, and the core", *Economica*, 34, pp.314-24.

Samuelson, P.A. (1954) "The Pure Theory of Public Expenditure", *Review of Economics and Statistics*, 36(4), pp.387-9.

Stiglitz, J.E. (1977) "The theory of local public goods" in *The Economics of Public Services*, M.S. Feldstein and R.P. Inman (eds.), Macmillan, London.

Stiglitz, J.E. (1979) "Local public goods and matching", unpublished, Oxford University.

Tiebout, C.M. (1956) "A Pure Theory of Local Expenditures", *Journal of Political Economy*, 64, pp.416-24.

Westhoff, F. (1977) "Existence of equilibria in economies with a local public good", *Journal of Economic Theory*, 14, pp.84-112.