

NBER WORKING PAPER SERIES

AN OPTIMAL TAXATION APPROACH
TO FISCAL FEDERALISM

Roger H. Gordon

Working Paper No. 1004

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

October 1982

The research reported here is part of the NBER's research program in Taxation. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

AN OPTIMAL TAXATION APPROACH TO FISCAL FEDERALISM*

Roger H. Gordon

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

In a Federal system of government, each unit of government decides independently how much of each type of public good to provide, and what types of taxes, and which tax rates, to use in funding the public goods. In this paper we explore what types of problems can arise from this decentralized form of decision-making. In particular, we describe systematically the types of externalities that one unit of government can create for nonresidents, through both its public goods decisions and its taxation decisions. The paper also explores briefly what the central government might do to lessen the costs of decentralized decision-making.

AN OPTIMAL TAXATION APPROACH TO FISCAL FEDERALISM*

Roger H. Gordon

Bell Laboratories
Murray Hill, New Jersey 07974

In a Federal system of government, each unit of government decides independently how much of each type of public good to provide, and what types of taxes, and which tax rates, to use in funding the public goods. Many papers have worried about whether any unit of government will act in the best interests of its citizens.¹ In this paper, we ask the broader question: assuming that each unit of government does in fact act in the best interests of its own citizens, will the collection of units of government together act in the best interests of all their citizens. Stated differently, what types of problems can arise from decentralized decision-making?

The main objective of this paper will be to describe systematically the types of externalities that a unit of government can create for nonresidents through both its public goods decisions and its taxation decisions. While a unit of government, acting solely in the interests of its own citizens, would ignore these externalities, overall social welfare would require that they be taken into account.

We will also explore briefly what the central government might do to lessen the costs of decentralized decision-making. We will explore, for example, the merits of centralizing provision of the public good, of restricting the types of taxes that local governments can use, of subsidizing certain types of local government activity, or of transferring tax revenue from the central government to local governments.

Previous papers have laid some of the ground work for this study. For example, Arnott and Grieson [1981] examined the optimal behavior of a local government, when nonresidents pay some of the taxes and use some of the public facilities. The analysis of local government behavior here will generalize their study to include such complications as residential mobility and endogenous prices. Starrett [1980] provides a nice treatment of externalities among units of government in a setting where each community imposes lump sum taxes on its residents.

The analysis here generalizes Starrett's model to incorporate use of distorting taxes by communities.

I. General characteristics of the model

The model has the following general characteristics. We assume that there are a finite set of nonoverlapping communities, which together comprise a self-sufficient economic region. We therefore examine just a two-tier Federal system — a central government and many local or regional governments.²

Individuals choose to live in that community which is most desirable, given their tastes. They may, however, work elsewhere, enjoy public services elsewhere, and buy goods or own property elsewhere, though mobility costs will, to varying degrees, implicitly limit the attractiveness of these activities. Similarly firms in each community can hire factors from other communities and sell their output in other communities, though transportation costs again limit the attractiveness of this. Because of transportation costs, prices of goods and factors will vary by community.

Each unit of government may impose taxes on each of the factors employed by firms in its community, and on each of the goods sold within (though not necessarily consumed within) the community. We thus have focused the analysis on what are commonly referred to as source based factor taxes and origin based commodity taxes. We will comment below on how the results change with destination based commodity taxes or residence based income taxes — taxes imposed on goods where consumed and on factors where owned.

Public services are produced by each unit of government. They are made available free of charge, without exclusion.

Communities are assumed to choose tax and expenditure policies so as to maximize a Benthamite social welfare function which includes the utility levels of its existing residents. We therefore assume away a range of public choice problems, not because we think they are

unimportant, but so as to isolate the problems created by decentralized decision-making. In choosing these policies, we assume that each community takes as given the policies of other communities.

II. Notation

Let us now develop the specific notation used in the model. Commodities will be indexed by j and communities by k . Thus p_{jk} represents the price received by a firm for the j th good in community k , while v_{jk} represents the price paid by the firm for the j th factor in community k . In general, prices vary by location. Similarly, q_{jk} represents the price paid by consumers for the j th good, while w_{jk} represents the price received for the j th factor, both in community k . The difference $s_{jk} = q_{jk} - p_{jk}$ represents the excise tax paid on a unit of good j , while the difference $t_{jk} = v_{jk} - w_{jk}$ is the tax on the j th factor, imposed by community k . The vectors of prices in community k are denoted by p_{*k} , v_{*k} , w_{*k} , q_{*k} , s_{*k} , and t_{*k} , while the extended vectors of all prices are denoted by p_{**} , v_{**} , w_{**} , q_{**} , s_{**} , and t_{**} .

The firms in community k collectively choose a vector of outputs y_{**}^k and a vector of inputs x_{*k} . Note that their inputs must be hired in community k , but their outputs can be sold anywhere. The firms' choices must satisfy the implicit production function $f^k(y_{**}^k, x_{*k}) = 0$.³ Transportation costs, for shipping output to other communities, are reflected in negative entries for some of the elements of y_{**}^k . The firm will choose y_{**}^k and x_{*k} so as to maximize profits $p_{**} \cdot y_{**}^k - v_{*k} \cdot x_{*k}$, subject to the production constraint. (Here, $p_{**} \cdot y_{**}^k$ denotes the inner product of the two vectors.) It follows that maximum profits π_k are implicitly a function of the prices faced by the firm: $\pi_k = \pi_k(p_{**}, v_{*k})$. For simplification, we also assume that the production functions have constant returns to scale, and that firms behave competitively, so that in equilibrium, $\pi_k = 0$. Let us denote aggregate output sold in community n by $Y_{*n} = \sum_k y_{*n}^k$.

From duality theory, we know that $\frac{\partial \pi_k}{\partial p_{jn}} = y_{jn}^k$, and also that $\frac{\partial \pi_k}{\partial v_{jk}} = -x_{jk}$.⁴ We will use these results in the analysis below.

We turn next to the modeling of individual behavior. We assume that there are a finite number of *types* of individuals. There are many individuals of each type, each with the same tastes, the same human capital, and the same ownership of other factors. Since individuals are mobile, all individuals of a given type will have the same utility. Let u^{ik} represent the number of individuals of type i who choose to locate in community k .

We will treat u^{ik} as a continuous variable, thereby assuming that an individual's location decision has only marginal impact on other aspects of the model. Eliminating the discrete choice aspect of individual locational decisions is the primary motivation underlying the assumption of many types of individuals, each type spread over many communities.

Each individual faces a vector of prices q_{**} for goods and a vector of factor prices w_{**} . In addition, each benefits from the availability of the extended vector of public service levels Q_{**} , where Q_{jk} is the level of the j th service in the k th community.

We assume in general that each type of individual has a unique type of human capital, i.e., one that is not a perfect substitute for the human capital of other types of individuals. (Perfect substitutability will be a special case.) The purpose of this assumption is to provide some mechanism to prevent all individuals of a given type from concentrating in the same one community. If they do, their wage rate there will be bid down, while it will rise in other communities, inducing them to spread out. Let h_i denote the human capital of type i , where h_i is a vector of length equal to the number of types of individuals with zero in all entries except the i th, and one in the i th entry.

Individuals of type i will also own a vector of nonhuman factors z_{**}^{ni} of various types in various communities. Let z_{**}^i denote the combined vector of human and nonhuman factors h_i and z_{**}^{ni} , where z_{**}^i will be of the same dimensions as the vector of firm inputs x_{**} .

In addition, individuals are assumed to be adversely affected by the degree of congestion c_k in their community. Congestion may not only reduce the quality of a given supply of public goods (e.g. roads or schools), but it may also reduce the quality of the environment (e.g. noise or pollution). We leave unspecified what factors contribute to congestion, though presume that the number of residents and the level of production in the community are included among them.

Individuals of type i in community k will maximize their utility subject to the prices they face, their endowments of factors, the level of congestion in the community, and the availability of public services. We can therefore denote their utility level indirectly, by

$$V^{ik} = V^i(q_{**}, w_{**}, z_{**}^i, c_k, Q_{**}, k).$$

Here, k appears explicitly since the attractiveness of purchasing or selling goods in any given community depends on the community in which the individual lives. (As with the firms, individuals must buy some of the transportation output in order to participate in the markets in other communities.)

From duality theory, we know that $\frac{\partial V^{ik}}{\partial q_{jn}} = -\alpha_{ik} y_{jn}^{ik}$, where α_{ik} is the marginal utility of income of the i th person in the k th community⁵ and where y_{jn}^{ik} is the demand by the i th person in the k th community for the j th good sold in the n th community. Similarly, $\frac{\partial V^{ik}}{\partial w_{jn}} = \alpha_{ik} x_{jn}^{ik}$, where x_{jn}^{ik} is the amount of the j th factor supplied in the n th community. Note that the individual need not supply all of the z_{**}^i to the market. For example, he can consume leisure.

The treatment of land as a factor merits comment. Since land in community k cannot be supplied as an input in community l , we assume that each production function recognizes that the required transportation costs here are indefinitely large. Also, identical nonhuman factor ownership for individuals of a given type living in different communities may seem implausible when land is considered as a factor. We assume here that housing is a service to be purchased, one of whose inputs is land in that community. Individuals living in a given community will

buy relatively more of the housing service in that community, but will still own just a representative share of the land in that community.

Let us next consider the activities of the local governments. Tax revenues in community k will equal $T_k \equiv s_{*k} \cdot Y_{*k} + t_{*k} \cdot x_{*k}$. These tax revenues will be used to buy a vector of factors b_{*k} , at revenue cost $R_k \equiv v_{*k} \cdot b_{*k}$. Budget balance implies that $R_k = T_k$. The production function for public services in community k will then be denoted by $g^k(Q_{*k}, b_{*k}) = 0$.

The modeling of the types of taxes communities can use is rather abstract, but includes as special cases most of the taxes that state or local governments use. For example, a sales tax is a uniform set of tax rates s_{*k} . A nonresidential property tax and a corporate income tax are both factor taxes on land and capital inputs to production. A residential property tax could be viewed as a tax on housing services sold within the community. A wage tax on employees would also be a factor tax, as would be a tax on natural resource inputs to production. Some taxes, such as an income tax on residents, are currently omitted, mainly for notational simplification.

In order to close the model, all that is needed in addition are resource constraints and market clearing constraints. In particular, we must assume that $\sum_{ik} u^{ik} x_{**}^{ik} = x_{**} + b_{**}$ in the factor markets, and that $\sum_{ik} y_{**}^{ik} = Y_{**}$ in the output market.

III. Characterization of optimal government behavior

Since the only taxes available to local governments are distorting, the model cannot in general lead to a Tiebout equilibrium with full efficiency.⁶ Therefore, each level of government faces a second-best problem of how to raise tax revenue with the least loss in welfare. In this section we will characterize the tax rates and public services levels that communities will choose.

In doing so, we assume that each level of government acts so as to maximize a Benthamite social welfare function.⁷ In particular, we assume that the welfare function of community k is⁸

$$W_k = \sum_i \omega_i u^{ik} V^{ik} ,$$

where individuals of type i are given welfare weight ω_i . Note that community k ignores the utility of nonresidents when making decisions. Since its decisions may affect the utility of nonresidents in a variety of ways, externalities will exist with decentralized decision-making. In order to characterize these externalities, we examine in addition the optimal behavior of local governments when they fully coordinate their behavior. In particular, we examine the optimal policy when the coordinated objective function is⁹

$$W = \sum_i \omega_i \sum_k u^{ik} V^{ik} .$$

Note that we have assumed the same welfare weights ω_i for each local government, so do not allow for inconsistent objectives. The motivation here is to isolate problems arising from decentralization of decision-making, per se.

In this context, we cannot assume that the initial income distribution is optimal, so that the social marginal utility of income for all individuals is the same. For any given type of individual i , utility levels are equalized due to mobility, and Benthamite social welfare function must give them equal weight. But marginal utilities of income may well vary across communities for these individuals. Without restrictions on mobility, this is an inevitable characteristic of the equilibrium.¹⁰ We therefore must examine distributional as well as efficiency considerations throughout the analysis.

1. Fully coordinated decision-making

In examining the costs of decentralized decision-making, we start by characterizing tax and spending behavior when all decisions are coordinated and then examine the decentralized equilibrium. Assume therefore that the coordinated government can choose the vectors of tax rates t_{**} and s_{**} , as well as public service inputs b_{**} and public service outputs Q_{**} , so as to

maximize social welfare. Its decision problem therefore is

$$\max_{t_{**}, s_{**}, b_{**}, Q_{**}} \sum_i \omega_i \sum_k u^{ik} v^{ik} + \mu [\sum_k (s_{**k} \cdot Y_{**k} + t_{**k} \cdot x_{**k} - v_{**k} \cdot b_{**k})] + \sum \gamma_k g_k \quad (A)$$

where μ is a Lagrangian multiplier on the government budget constraint, and γ_k is Lagrangian multiplier on the public goods production constraint in the k th community. Note that we have required only that the overall budget constraint be satisfied, and not that each community's budget constraint be satisfied separately.

When we carry out the differentiation, we derive the following results:

Proposition 1. Optimal coordinated government behavior can be characterized by the following set of first-order conditions:¹¹

a) For $s_{j\ell}$,

$$Y_{j\ell}(\mu - \bar{\theta}) + d\theta_{**} \cdot \frac{\partial I_{**}}{\partial s_{j\ell}} + \sum_k \frac{\partial C_k}{\partial s_{j\ell}} + \mu \sum_k \frac{\partial T_k}{\partial s_{j\ell}} - (\mu - \bar{\theta}) \sum_k \frac{\partial R_k}{\partial s_{j\ell}} = 0 \quad (1a)$$

b) For $t_{j\ell}$,

$$x_{j\ell}(\mu - \bar{\theta}) + d\theta_{**} \cdot \frac{\partial I_{**}}{\partial t_{j\ell}} + \sum_k \frac{\partial C_k}{\partial t_{j\ell}} + \mu \sum_k \frac{\partial T_k}{\partial t_{j\ell}} - (\mu - \bar{\theta}) \sum_k \frac{\partial R_k}{\partial t_{j\ell}} = 0 \quad (1b)$$

c) for $b_{j\ell}$,

$$(\gamma_\ell \frac{\partial g_\ell}{\partial b_{j\ell}} - \mu v_{j\ell}) + d\theta_{**} \cdot \frac{\partial I_{**}}{\partial b_{j\ell}} + \sum_k \frac{\partial C_k}{\partial b_{j\ell}} + \mu \sum_k \frac{\partial T_k}{\partial b_{j\ell}} - (\mu - \bar{\theta}) \sum_k \frac{\partial R_k}{\partial b_{j\ell}} = 0 \quad (1c)$$

d) For $Q_{j\ell}$,

$$\left(\sum_{ik} \omega_i u^{ik} \frac{\partial V^{ik}}{\partial Q_{jt}} + \gamma_j \frac{\partial g_j}{\partial Q_{jt}} \right) + d\theta_{**} \cdot \frac{\partial I_{**}}{\partial Q_{jt}} + \sum_k \frac{\partial C_k}{\partial Q_{jt}} + \mu \sum_k \frac{\partial T_k}{\partial Q_{jt}} - (\mu - \bar{\theta}) \sum_k \frac{\partial R_k}{\partial Q_{jt}} = 0 \quad (1d)$$

$$\text{where } \bar{\theta} = \frac{\sum_{ik} \omega_i u^{ik} \alpha_{ik}}{\sum_{ik} u^{ik}}$$

$$d\theta_{ik} = \omega_i \alpha_{ik} - \bar{\theta}$$

$$\frac{\partial I_{ik}}{\partial \sigma_{jt}} = u^{ik} (x_{**}^{ik} \cdot \frac{\partial w_{**}}{\partial \sigma_{jt}} - y_{**}^{ik} \cdot \frac{\partial q_{**}}{\partial \sigma_{jt}})$$

$$\frac{\partial C_k}{\partial \sigma_{jt}} = \sum_i \omega_i u^{ik} \frac{\partial V^{ik}}{\partial c_k} \frac{\partial c_k}{\partial \sigma_{jt}}$$

$$\frac{\partial T_k}{\partial \sigma_{jt}} = s_{*k} \cdot \frac{\partial Y_{*k}}{\partial \sigma_{jt}} + t_{*k} \cdot \frac{\partial x_{*k}}{\partial \sigma_{jt}}$$

$$\frac{\partial R_k}{\partial \sigma_{jt}} = b_{*k} \cdot \frac{\partial v_{*k}}{\partial \sigma_{jt}}$$

all for $\sigma_{jt} = s_{jt}, t_{jt}, b_{jt}, \text{ or } Q_{jt}$

Proof. The straightforward derivation is found in the Appendix.

Because of the generality of the model, the first-order conditions are rather messy, but the economic interpretation is still straightforward. Each first-order condition above consists of five terms. The first measures the direct impact of the policy change. For s_{jt} and t_{jt} , it measures the gain from transferring income from the representative individual to the government. Here, μ represents the welfare value of an extra dollar of government revenue, while $\bar{\theta}$ represents the welfare value of an extra dollar of income to the representative individual. If lump sum taxes existed, μ would equal $\bar{\theta}$. Since raising revenue creates efficiency losses, however, the value of an extra dollar of government revenue, μ , would normally exceed the value of a dollar to individuals, $\bar{\theta}$.

The first term in the first order condition for b_{jt} represents the value of the relaxed government production constraint relative to the social cost of purchasing the extra b_{jt} . The

first term in the equation for Q_{jt} represents the utility gain to individuals of the extra output relative to the social cost of the increased production.¹²

The second through the fifth terms collectively measure the efficiency and equity costs created in the process of transferring income to the government, or using the revenue to provide public services. Their size depends on the degree to which individual behavior, and equilibrium market prices, change in response to the government action. We see from equations (1a) and (1b) that the difference $(\mu - \bar{\theta})$ reflects the magnitude of these efficiency and equity costs created by raising government revenue.

The second term in each equation captures any distributional effects of the policy change. The expression $d\theta_{ik}$ measures the social value of a dollar to an individual of type i in community k , relative to the social value of a dollar to the "representative" individual. The expression $\frac{\partial I_{ik}}{\partial \sigma_{jt}}$ measures the income effect on individuals of type i in community k of the policy change. The inner product then measures whether the income transfers created by the policy change are weighted mainly towards the poor (positive value) or mainly towards the rich (negative value).¹³

The third term captures any impact of the policy change on the severity of congestion costs. Individuals may shift their residence, or production efforts may change location, in response to the policy change. A move towards less congested communities would be beneficial, and conversely. Individuals, in making their decisions, do not take into account these externalities.

The fourth term measures any indirect tax revenue effects of the policy change, as individuals change the amount or location of their market activity. As such, it captures any efficiency effects arising from this change in behavior. When prices reflect the social costs of commodities (or the marginal products of factors), changes in individual behavior have no efficiency consequences. But when taxes exist, one component of the price individuals face, the tax component, is not a true social cost. As a result, when an individual buys more of a taxed commodity (or supplies more of a taxed factor), there is a social gain since the utility

gain to the individual exceeds the social cost of the output (or the utility loss to the individual falls short of the marginal product) by an amount equal to the tax rate. In equations 1a and 1b, the most important element of this term is undoubtedly the response by individuals to the tax induced increase in price of the good $y_{j\ell}$ or the factor $x_{j\ell}$. This is the only part of the last four terms not assumed to equal zero when deriving the Ramsey rule for tax rates.¹⁴

The final term measures the social cost of raising any additional tax revenue required by changes in the costs of factor inputs to government production. The difference $(\mu - \bar{\theta})$ implicitly measures the social cost of transferring a dollar from individuals to the government.

2. Decentralized decision-making

We proceed next to provide a formal characterization of the fiscal decisions made by uncoordinated local governments. We assume that each local government acts independently, taking as given the fiscal decisions made by other local governments. Clearly, local governments do in part coordinate their behavior. The point here is not so much to provide a good positive description of local government behavior as to describe the types of problems which arise from decentralized decision-making. To what degree local governments have managed to alleviate these problems in practice through coordination is a separate question.

We assume then that the objective of each local government is to maximize the weighted sum of the utility of its own residents, subject to a balanced budget constraint, and subject to the public production function. That is, we assume that each local government behaves so as to:

$$\max_{s_{*f}, t_{*f}, b_{*f}, Q_{*f}} \sum_i \omega_i u^i [V^i + \mu_f [s_{*f} \cdot Y_{*f} + t_{*f} \cdot x_{*f} - v_{*f} \cdot b_{*f}] + \gamma_f \cdot g_f]$$

Note that any effects on other individuals or other units of government are ignored. Each local government carries out this maximization simultaneously, taking as given the actions of other local governments.

When we carry out the differentiation, we derive the following results:

Proposition 2. Decentralized government behavior can be characterized by the following set of first order conditions:

a) For s_{jt} ,

$$\begin{aligned} & (\mu_t Y_{jt} - \bar{\theta} u^{*t} \cdot y_{jt}^{*t}) + d\theta_{*t} \cdot \frac{\partial I_{*t}}{\partial s_{jt}} + \frac{\partial C_t}{\partial s_{jt}} \\ & + \mu_t \frac{\partial T_t}{\partial s_{jt}} - \mu_t \frac{\partial R_t}{\partial s_{jt}} + \bar{\theta} \sum_i u^{it} [x_{**}^{it}] \cdot \frac{\partial v_{**}}{\partial s_{jt}} - y_{**}^{it} \cdot \frac{\partial p_{**}}{\partial s_{jt}} = 0 \end{aligned} \quad (2a)$$

b) For t_{jt} ,

$$\begin{aligned} & (\mu_t x_{jt} - \bar{\theta} u^{*t} \cdot x_{jt}^{*t}) + d\theta_{*t} \cdot \frac{\partial I_{*t}}{\partial t_{jt}} + \frac{\partial C_t}{\partial t_{jt}} \\ & + \mu_t \frac{\partial T_t}{\partial t_{jt}} - \mu_t \frac{\partial R_t}{\partial t_{jt}} + \bar{\theta} \sum_i u^{it} [x_{**}^{it}] \cdot \frac{\partial v_{**}}{\partial t_{jt}} - y_{**}^{it} \cdot \frac{\partial p_{**}}{\partial t_{jt}} = 0 \end{aligned} \quad (2b)$$

c) For b_{jt} ,

$$\begin{aligned} & (\gamma_t \frac{\partial g_t}{\partial b_{jt}} - \mu_t v_{jt}) + d\theta_{*t} \cdot \frac{\partial I_{*t}}{\partial b_{jt}} + \frac{\partial C_t}{\partial b_{jt}} \\ & + \mu_t \frac{\partial T_t}{\partial b_{jt}} - \mu_t \frac{\partial R_t}{\partial b_{jt}} + \bar{\theta} \sum_i \frac{\partial I_{it}}{\partial b_{jt}} = 0 \end{aligned} \quad (2c)$$

d) For Q_{jt} ,

$$\begin{aligned} & (\sum_i \omega_i u^{it} \frac{dV^{it}}{dQ_{jt}} + \gamma_t \frac{\partial g_t}{\partial Q_{jt}}) + d\theta_{*t} \cdot \frac{\partial I_{*t}}{\partial Q_{jt}} \\ & + \frac{\partial C_t}{\partial Q_{jt}} + \mu_t \frac{\partial T_t}{\partial Q_{jt}} - \mu_t \frac{\partial R_t}{\partial Q_{jt}} + \bar{\theta} \sum_i \frac{\partial I_{it}}{\partial Q_{jt}} = 0 \end{aligned} \quad (2d)$$

Proof: The derivation is a simple modification of the derivation for Proposition 1.

The intuition underlying these first-order conditions is basically the same as for the fully coordinated case. Of the six terms in each equation, for all but the last, analogous terms appeared in the equations in Proposition 1. The key difference is that effects of fiscal decisions on residents of other communities and on the fiscal position of other governments are all

ignored here.

Let us examine the differences in the first-order conditions in Propositions 1 and 2 term by term. The first-term in each of the first-order conditions in Proposition 2 again measures the direct effects of the policy change. In equations 2a and 2b, however, while all the direct revenue effects of the tax change are taken into account, only the direct loss to *residents* from paying taxes is taken into account. Utility loss to nonresidents from paying taxes is ignored. To the extent then that nonresidents pay the taxes, as with sales taxes or wage taxes on nonresidents, these taxes will be used excessively when fiscal authority is decentralized. For taxes like a corporate income tax, this effect may be substantial.

A comparison of the first terms in equations 1d and 2d gives the obvious result that with decentralized decision making, the value of local public services to nonresidents is ignored. This type of benefit spillover has been discussed at length in the literature.

The second term continues to capture distributional effects of the policy change. With decentralized decision-making, however, only redistribution effects on the residents of the community are taken into account. The distributional pattern of income effects among nonresidents is ignored. For example, a sales tax increase may be paid essentially equally by rich and poor residents, but only poor nonresidents may have taken the time to travel in search of a cheaper price.

The third term captures the effect of any change in congestion in community q due to the policy change. Effects on congestion elsewhere are ignored. In the fully coordinated case, this term captured the difference in congestion costs when individuals or production move from one community to another. In the decentralized case, communities will tend to tax excessively congestion producing activities, since they experience any gains from decongestion, but ignore the associated loss from extra congestion elsewhere.

The fourth term captures the effect on tax revenues of any change in individual behavior. While in the fully coordinated case, effects on total tax revenues are taken into account, in the

decentralized case, effects only on the tax revenues of that community are taken into account. If cross-price elasticities are zero, then there are no effects on the tax revenues of other communities. However, in this context, where goods supplied in different communities are defined to be different goods, the assumption that cross price elasticities are zero is untenable. For example, the alternative extreme assumption, where a unit decrease in demand for a good in community k is associated with approximately a unit increase in demand for the same good in other communities, would frequently be quite plausible. In this case, with decentralization, tax rates would be set low, due to the mobility of demand. Since in aggregate, demand would be extremely inelastic, however, a uniform tax in all communities would be close to nondistorting. In general, elasticity of demand in one community may have little relation to the aggregate elasticity of demand, so that the set of tax rates arising in a decentralized setting need have little relation to those which would be chosen in a coordinated setting. Ability to relocate demand may be quite different from willingness to reduce demand when prices rise.

The interpretation of the fifth term is similar to that of the previous terms, but the externalities here are probably much less important. In a decentralized setting, the effect of the policy change on revenue requirements in that community only is taken into account. In a coordinated setting, the effect of the policy change on the revenue requirements of all communities would matter.

Finally, the sixth term in the equations for decentralized behavior did not appear explicitly previously. This term captures the income effects on residents of community l due to price changes caused by the policy change. This term is analogous to the terms of trade effect which appears in the optimal tariff literature. In the coordinated setting, the sum of these terms over all communities appeared in the course of the derivation. Due to competition in production, the value of inputs minus outputs in private production must remain zero so that in the

coordinated case this sum reduced to $\bar{\theta} \sum_k b_{*k} \cdot \frac{\partial v_{*k}}{\partial \sigma_{jl}}$. With decentralization, however,

communities would try to gain at the expense of other communities.

Propositions 1 and 2 together provide a formal characterization of the reasons why decentralized decision-making itself can lead to a less efficient (equitable) outcome. This arises because one community's decision affects in many ways the utility level of residents of other communities, yet these effects are ignored in the decision-making. The types of externalities that appeared in the equations resulting from a given community's decisions were:

- (1) Nonresidents may pay some of the taxes.
- (2) Nonresidents may receive some of the benefits from public services.
- (3) Congestion costs faced by nonresidents may change.
- (4) Tax revenues received in other communities may change due to the spillover of economic activity.
- (5) Resource costs for public services in other communities may change.
- (6) Output and factor price changes may favor residents over nonresidents.
- (7) Distributional effects among nonresidents would be ignored.

In certain special cases, externalities may tend to offset, in which case the costs of decentralized decision-making are minimized. For example, consider the decision-making of a central city surrounded by suburbs. The suburban residents benefit from using the city's public services, benefits ignored in setting public service levels. However, suburban residents also pay various taxes to the central city, such as sales taxes, wage taxes on nonresidents, etc. If the city were to consider a uniform expansion of both taxes and services, it may well be that the benefit spillover and the tax "spillin" essentially offset, leaving little net externality from these two terms.¹⁵

As another example, consider the externalities created by an individual decision to change residence as a result of a policy change. In his new community the individual pays extra taxes, but also adds to congestion; while in his old community he no longer pays (as much in) taxes, but he also no longer adds to congestion there. This is the setting of the Buchanan-Goetz

[1972] analysis. If the marginal costs and benefits of gaining or losing a resident just offset, then again there are no net externalities. That is, the congestion effects on other communities and the tax revenue effects on other communities may essentially offset, again leaving no net externality from these two terms.

This is also the setting of the Tiebout [1956] model, at least as formalized by McGuire [1974]. The above analysis makes clear how strong the assumptions must be to get a Tiebout type result. To get a Tiebout result, where the public good is supplied in a decentralized setting with no efficiency costs at all, there must be no net externalities and no net equity or efficiency costs arising from government behavior. That is, the sum of the seven sources of externalities listed above must be zero for each of the equations in Proposition 2, so that communities would in effect act in a coordinated fashion. In addition, the sum of the second, third, fourth, and fifth terms in each of the equations in Proposition 1, which together measure the equity and efficiency effects of government behavior, must equal zero.

How does the analysis change if we have destination-residence based taxes rather than source-origin based taxes. In particular, what if we tax factors where they are owned and goods where they are consumed (rather than purchased)? In this case, the tax is paid entirely by residents so that the first terms in equations 1a vs. 2a and 1b vs. 2b will look the same. All other externalities will remain, however.

IV. Analysis of Possible Federal Government Remedies

In principle, local governments could negotiate together to correct any of the externalities described above. However, given the difficulty of conducting negotiations among thousands of local government units, few of the above externalities are likely to have been corrected for in this manner. In this section, we therefore explore a number of policies which the Federal government might use in response to the inefficiencies resulting from decentralized decision-making.¹⁶

1. Centralized provision of public goods

One obvious policy is to have the Federal government provide all public goods, so that interactions among communities can be taken into account. However, the Federal government would normally be constrained to provide a uniform level of public services, and charge uniform tax rates, even though individual preferences differ. This may occur due to constitutional constraints or due to inadequate information about local preferences. One of the key advantages of decentralization is the resulting diversity of policies.

If we accept the constraint requiring uniform public service quality, and uniform tax rates, by each level of government within its jurisdiction, then it is possible to develop a theory for the optimal degree of decentralization. The optimal centralized uniform service level, and tax rates, would be inferior to the optimal diversified service levels described in Proposition 1, and may or may not be preferable to the diversified service levels, and tax rates, with decentralized decisions described in Proposition 2. A detailed comparison here would provide guidelines concerning whether to (or, when there are more than two levels of government, the degree to which to) decentralize the provision of any given public service.

2. Categorical grants

Local governments will to a degree choose the wrong level and composition of expenditures, since the size of the spillover to nonresidents will vary by type of expenditure. Categorical or matching grants provide one device by which the Federal government can induce local governments to shift their package of public services towards those which provide the most benefits to nonresidents. An example of a good with large spillovers where such grants have developed in the U.S. is interstate highways.

3. Tax deductibility of local government taxes

Local governments may also choose too small a level of public services in general. Not only do nonresidents benefit from extra public services, but also surrounding communities may

receive increased tax revenue as economic activity migrates in response to higher tax rates. On the other hand, nonresidents pay some of the taxes used to finance extra services, and also suffer from extra congestion as economic activity shifts into their communities.

If, on net, nonresidents tend to benefit (lose) from expanded public services, then there are grounds for an overall Federal subsidy (tax) to local public services. One way this is done in the U.S. is through allowing local public expenditures to be deductible under the Federal income tax. This particular policy is unattractive, however, in that expenditures are deductible only for the approximately 20% of tax payers who itemize (generally those with high incomes). Also, among those who itemize, the subsidy rate equals the individual's marginal tax rate, which varies substantially across individuals (being highest again for those with the highest incomes). A refundable tax credit, however, would not be subject to these problems.

4. Revenue sharing

Both previous policies change relative prices, and were designed to internalize the effects of spillovers. They have commonly been argued for on these grounds. (See, for example, Oates [1972]). The above analysis also suggests a rationalization for direct income transfers from the Federal government to local governments.

To make the point starkly, consider a tax on the return to capital, where capital is completely mobile, has by assumption the same productivity everywhere, but is in fixed supply. Since by assumption capital is perfectly mobile, it will all flow to the community with the lowest tax rate. Communities will compete for this tax base, and in equilibrium the tax rate on capital in each community will be driven to zero. However, since also by assumption capital is in fixed supply, a uniform tax on capital in all locations is completely nondistorting. Therefore, with centralized decision-making, the optimal tax rate on capital would be very high.

The central government could then impose a uniform tax rate on capital, and redistribute the revenues in something approximating a nondistorting way, allowing local governments either to expand expenditures or to reduce other distorting taxes. In effect, the central

government provides the mechanism by which local governments can coordinate their tax policies and take into account the effect each local government's action has on the revenues of other local governments.

In general, the Federal government will not be indifferent to how the revenue is distributed across local governments. For example, the size of the efficiency (equity) gain when a community lowers its taxes, or increases its expenditures, in response to a dollar received in revenue sharing may vary widely by community. The Federal government would then have the incentive to give relatively more in revenue sharing to communities where the efficiency (equity) gain would be largest.

To the extent that the amount of revenue a community gets back depends on its actions, however, complications arise due to these price distortions. For example, consider the situation where all the revenues collected from a community are returned to that community. Then, the community can completely offset the effects of the Federal action by lowering its own tax rate by an amount equal to the Federal tax rate. In principle, the decentralized equilibrium would remain unchanged by the Federal action. However, communities would normally be constrained to charge positive tax rates. If the Federal tax rate exceeds the decentralized tax rate, as we expect, then the Federal action would accomplish the desired end.

5. Federal regulation of permissible local tax bases

In the case of revenue sharing, we considered the situation where in the decentralized equilibrium a particular tax rate is relatively too low. The Federal government can then impose a higher tax rate. In situations where nonresidents pay an important share of a particular tax, however, the tax rate may be set too high. Examples of this situation may be severance fees, and perhaps state corporate income taxes. What remedies exist in this case?

The most obvious, and extreme, remedy is to forbid local governments from using such tax bases at all. In some cases, this may be an improvement, though a zero tax rate may be inefficiently low. An alternative might be to legislate a maximum permissible tax rate, set at

the desired value.

V. Conclusion

There may be many advantages to decentralizing government decision-making. Local governments, being "closer to the people," may better reflect individual preferences. The diversity of policies of local governments allows individuals to move to that community best reflecting their tastes. Competition among communities should lead to greater efficiency and innovation. However, this paper has shown the many ways in which decentralized decision-making can lead to inefficiencies, since a local government will ignore the effects of its decisions on the utility levels of nonresidents.

The paper shows that certain taxes may be used excessively, such as when nonresidents pay an important fraction of the tax. Other taxes may be used too little, since for example the increase in tax revenue in neighboring communities, as economic activity spills over to avoid a tax increase, would not be taken into account when communities act independently. These efficiency problems with decentralized decision-making imply the potential for efficiency gains through regulation by the Federal government of what tax bases local governments can use, and through transfer of Federal tax revenue to local governments.

APPENDIX

Proof of Proposition 1. Direct differentiation of equation A with respect to $s_{j\ell}$ gives:

$$\begin{aligned} \sum_i \omega_i \sum_k u^{ik} \left[\frac{\partial V^{ik}}{\partial q_{**}} \cdot \frac{\partial q_{**}}{\partial s_{j\ell}} + \frac{\partial V^{ik}}{\partial w_{**}} \cdot \frac{\partial w_{**}}{\partial s_{j\ell}} + \frac{\partial V^{ik}}{\partial c_k} \frac{\partial c_k}{\partial s_{j\ell}} \right] + \sum_i \omega_i \sum_k V^{ik} \frac{\partial u^{ik}}{\partial s_{j\ell}} \quad (\text{A.1}) \\ + \mu \left[Y_{j\ell} + \sum_k (s_{*k} \cdot \frac{\partial Y_{*k}}{\partial s_{j\ell}} + t_{*k} \frac{\partial x_{*k}}{\partial s_{j\ell}} - b_{*k} \cdot \frac{\partial v_{*k}}{\partial s_{j\ell}}) \right] = 0 \end{aligned}$$

Note that the second term equals zero, since V^{ik} has the same value for all k , and since the

$\sum_k u^{ik}$ is fixed. Note also that $\frac{\partial q_{ik}}{\partial s_{j\ell}} = \frac{\partial p_{ik}}{\partial s_{j\ell}}$ except for $i = j$ and $k = \ell$, when $\frac{\partial q_{j\ell}}{\partial s_{j\ell}} =$

$1 + \frac{\partial p_{j\ell}}{\partial s_{j\ell}}$. Also, $\frac{\partial w_{**}}{\partial s_{j\ell}} = \frac{\partial v_{**}}{\partial s_{j\ell}}$. Using duality theory, we can then reexpress the first term as

$$\sum_i \omega_i \sum_k u^{ik} \left[-\alpha_{ik} y_{j\ell}^{ik} - \alpha_{ik} y_{**}^{ik} \cdot \frac{\partial p_{**}}{\partial s_{j\ell}} + \alpha_{ik} x_{**}^{ik} \cdot \frac{\partial v_{**}}{\partial s_{j\ell}} + \frac{\partial V^{ik}}{\partial c_k} \frac{\partial c_k}{\partial s_{j\ell}} \right].$$

If we substitute for $\omega_i \alpha_{ik}$ the expression $\bar{\theta} + d\theta_{ik}$, the first term becomes the following:

$$\begin{aligned} \bar{\theta} \left[-\sum_{ik} u^{ik} y_{j\ell}^{ik} - \sum_{ik} u^{ik} y_{**}^{ik} \cdot \frac{\partial p_{**}}{\partial s_{j\ell}} + \sum_{ik} u^{ik} x_{**}^{ik} \cdot \frac{\partial v_{**}}{\partial s_{j\ell}} \right] \quad (\text{A.2}) \\ + \sum_{ik} d\theta_{ik} \left[-u^{ik} y_{**}^{ik} \cdot \frac{\partial q_{**}}{\partial s_{j\ell}} + u^{ik} x_{**}^{ik} \frac{\partial w_{**}}{\partial s_{j\ell}} \right] + \sum_i \omega_i \sum_k u^{ik} \frac{\partial V^{ik}}{\partial c_k} \frac{\partial c_k}{\partial s_{j\ell}} \end{aligned}$$

Since $\sum_{ik} u^{ik} y_{**}^{ik} = Y_{**}$ and $\sum_{ik} u^{ik} x_{**}^{ik} = x_{**} + b_{**}$, the first term simplifies to:

$$\begin{aligned} -\bar{\theta} Y_{j\ell} - \bar{\theta} \left[Y_{**} \cdot \frac{\partial p_{**}}{\partial s_{j\ell}} - x_{**} \cdot \frac{\partial v_{**}}{\partial s_{j\ell}} \right] \quad (\text{A.3}) \\ + d\theta_{**} \cdot \frac{\partial I_{**}}{\partial s_{j\ell}} + \sum_k \frac{\partial C_k}{\partial s_{j\ell}} + \bar{\theta} \sum_k \frac{\partial R_k}{\partial s_{j\ell}} \end{aligned}$$

That profits remain zero in equilibrium implies that the second term in equation (A.3) equals zero. Equation 1a in the text follows simply. Equations 1b, 1c, and 1d follow in a similar fashion.

REFERENCES

- Arnott, Richard and Ronald E. Grieson, "Optimal Fiscal Policy for a State or Local Government," *Journal of Urban Economics*, 9, 1981, pp. 23-48.
- Bradford, David and Wallace Oates, "Suburban Exploitation of Central Cities and Governmental Structure," in Hal Hochman and George Peterson, eds., *Redistribution Through Public Choice*, New York, Columbia University Press, 1974.
- Buchanan, James, *The Demand and Supply of Public Goods*, Chicago, Rand McNally, 1968.
- Buchanan, James and C. J. Goetz, "Efficiency Limits of Fiscal Mobility: An Assessment of the Tiebout Model," *Journal of Public Economics*, 1, 1972, pp. 25-44.
- Hamilton, Bruce W., "Zoning and Property Taxation in a System of Local Governments," *Urban Studies*, 12, 1975, pp. 205-211.
- McGuire, Martin C. "Group Segregation and Optimal Jurisdictions," *Journal of Political Economy*, 82, 1974, pp. 112-132.
- Mieszkowski, Peter, "The Property Tax: An Excise Tax or a Profits Tax?," *Journal of Public Economics*, 1, 1972, pp. 73-96.
- Mirrlees, James, "The Optimal Town," *Swedish Journal of Economics*, 74, 1972.
- Niskanen, William, "The Peculiar Economics of Bureaucracy," *American Economic Review*, 58, 1968, p. 293-305.
- Oates, Wallace E., *Fiscal Federalism*, New York, Harcourt Bruce Jovanovich, Inc., 1972.
- Samuelson, Paul, "The Pure Theory of Public Expenditure," *Review of Economics and Statistics*, 1954, pp. 387-9.
- Starrett, David A., "Measuring Externalities and Second Best Distortions in the Theory of Local Public Goods," *Econometrica*, 48, 1980, pp. 627-642.
- Stiglitz, Joseph E., "The Theory of Local Public Goods," in Martin S. Feldstein and Robert P. Inman, eds., *Economics of Public Services*, London, Macmillan, pp. 274-333.
- Tiebout, C., "A Pure Theory of Local Public Expenditure," *Journal of Political Economy*, 64, 1956, pp. 416-424.

FOOTNOTES

- * This paper was presented at the ISPE Conference on Taxation in Federal Systems in Canberra, Australia, on August 25, 1982. I would like to thank Fischer Black, Dieter Bös, Bhajan Grewel, Michelle White, and especially Charles E. McLure, Jr., for comments on an earlier draft. The views expressed in this paper are those of the author, and do not represent those of Bell Laboratories or of the Bell System.
1. See, for example, Buchanan [1968] and Niskanen [1968].
 2. Nothing of substance would be changed if the analysis were extended to three levels of government.
 3. For simplification, we ignore any contribution of public goods to production.
 4. To show this, note that $\pi_k = p_{**} \cdot y_{**}^k - v_{*k} \cdot x_{*k}$, evaluated at the optimal values of y_{**} and x_{*k} . Differentiating with respect to p_{jn} gives:

$$\frac{\partial \pi_k}{\partial p_{jn}} = y_{jn}^k + \sum_{i \neq j} p_{i\ell} \frac{\partial y_{i\ell}^k}{\partial p_{jn}} - \sum_i v_{ik} \frac{\partial x_{ik}}{\partial p_{jn}}$$

When y_{**} and x_{*k} are chosen so as to maximize profits subject to the production constraint, we have first-order conditions: $p_{i\ell} = -\lambda \frac{\partial f^k}{\partial y_{i\ell}^k}$ and $v_{ik} = \lambda \frac{\partial f^k}{\partial x_{ik}}$. Substituting and simplifying gives: $\frac{\partial \pi_k}{\partial p_{jn}} = y_{jn}^k - \lambda \frac{\partial f^k}{\partial p_{jn}}$. Since $f^k = 0$ always, the result follows.

5. While utility is the same across communities for the i th type, marginal utility need not be the same, a point emphasized by Mirrlees [1972].
6. If the supply of a factor has zero elasticity, as with land, then a tax on that factor is nondistorting. However, even in this setting, a Tiebout equilibrium may still not arise for many of the reasons described below. For example, if new residents add more (or less) to congestion costs than they do to tax revenues, then locational decisions will still create externalities. Hamilton [1975] has shown that if communities can impose quantity restrictions, e.g. zoning, then a Tiebout type equilibrium might be possible. We do not allow for such restrictions here.
7. These welfare functions could be reinterpreted as merely representing Pareto optimizing behavior by each government. The objective of a Pareto optimizing government would be to maximize the utility level of type ℓ individuals subject to the constraints that $V^{ik} \geq \bar{V}^{ik}$ for all other types ik , for some utility levels \bar{V}^{ik} . Put algebraically, the objective function, subject to constraints, would be

$$\max V^{\ell k} + \sum_{i \neq \ell} \omega_i (V^{ik} - \bar{V}^{ik})$$

where the ω_i represent Lagrangian multipliers. If the Lagrangian weights here are the same as the welfare weights in the text, optimal behavior will also be the same.

8. We assume that the utility of the residents at the time of the decision enters into the objective function, rather than the utility of those who would be present in response to any policy change.
9. Optimal behavior when coordination occurs is not necessarily intended here as a description of policy if all decisions were made by the central government. For example,

we will assume below that a central government would be required to provide uniform service levels, and charge uniform tax rates, in all jurisdictions.

10. For further discussion, see Mirrlees [1972].
11. While first-order conditions must be satisfied at the optimal outcome, they may also be satisfied elsewhere. Second-order conditions in this context need not necessarily be satisfied. See, for example, Stiglitz [1977]. We assume in the discussion that the desired equilibrium has, in fact, been reached.
12. Note that, ignoring all but the first terms, the sum of equations 1c and 1d gives the Samuelson [1954] optimal condition for pure public goods.
13. The tax incidence literature, which analyzes the degree to which labor, capital, or some other group, bears the burden of particular taxes, focuses on the nature of the equilibrium price changes which result from tax policy. For example, to what degree does a tax on capital income result in a rise in the before tax interest rate (due to a drop in the equilibrium capital stock), thereby lessening the loss to capital owners. In the analysis here, the direct effect of the tax payment appears in the first term, while the offsetting effect of price changes appears in the second term. We do not attempt to analyze the magnitude of these offsetting price changes. Presumably, the effect of marginal changes in any one community's tax policy on market prices (the context of these first order conditions) is small, though the effect of these changes in aggregate could be large. See Mieszkowski [1972] for further discussion.
14. Technically, in deriving the Ramsey rule, the Slutsky equation would be used to decompose the fourth term into price and income effects. All but the own price effects are then assumed to equal zero.
15. For further discussion, see Bradford and Oates [1974].
16. If these policies are pursued, yet local governments do in fact coordinate their policies, then there may be overcorrection for the externalities, with resulting inefficiencies.