Economica (2006) **73**, 725–748
doi:10.1111/j.1468-0335.2006.00506.x

Excludable and Non-excludable Public Inputs: Consequences for Economic Growth

By INGRID OTT† and STEPHEN J. TURNOVSKY‡

†*University of Lüneburg* ‡*University of Washington*

Final version received 21 June 2005.

Non-excludable and excludable public inputs are introduced into an endogenous growth model. We derive the equilibrium growth rate and design the optimal tax and user-cost structure, emphasizing the role of congestion and its consequences for the government's budget. The latter comprises fee and tax revenues that are used to finance the public inputs, although they may generate insufficient revenue to do so entirely. We extend the model to allow for monopoly pricing of the user fee by the government. Most of the analysis is conducted for general production functions consistent with endogenous growth, but the CES technology is also considered.

INTRODUCTION

A vast literature has evolved emphasizing the role of public investment as a determinant of economic growth. Among the earliest contributions, Arrow and Kurz (1970) is particularly significant, while the seminal work of Barro (1990) has been especially influential with respect to the contemporary endogenous growth literature. Stimulated by Aschauer's (1989) findings pertaining to the high rate of return on public capital, several authors have explored various aspects of the role of public capital in stimulating economic growth. For example, Baxter and King (1993) and Fisher and Turnovsky (1998) examine the interaction between public and private capital in a Ramsey-type growth model, while Futagami *et al*. (1993) perform a similar type of analysis in an endogenous growth context. Glomm and Ravikumar (1997) analyse the growth effects of productive public spending in an overlapping-generations setup, extending the framework to include majority voting, and alternative compositions of publicly provided goods and services; they also indicate private alternatives to public provision, but do not pursue this in detail. Cassou and Lansing (1998) focus on the relationship between private and public capital, showing how the evolution of the ratio between the two in the US economy since 1925 is broadly consistent with the predictions of a simple Ramsey growth model. They also show how the productivity slowdown in the US economy since the 1970s is more likely due to increasing tax rates than to non-optimal public investment policy.

Much of the literature, including Barro, treats the public input as a pure public good, freely available without restrictions or impediments to all agents in the economy. But the public goods literature identifies many different characteristics that most public goods in fact exhibit, notably the presence of 'rivalry' and/or 'excludability'; see e.g. Cornes and Sandler (1996). Thus, the treatment of a public input as a pure public good is extreme, as has been long acknowledged.[1]

In response to this, much of the recent literature analysing the impact of public expenditure and investment on economic growth allows for non-excludable public goods, which, because of congestion, are nevertheless subject to rivalry. Several alternative formulations of congestion have been adopted; see e.g. Barro and Sala-i-Martin (1992), Glomm and Ravikumar (1994), Turnovsky (1996, 2000), and Eicher and Turnovsky (2000). In particular, Eicher and Turnovsky emphasize the restrictions that must be imposed on the form of congestion function if an endogenous growth model is to sustain an equilibrium balanced growth path.

In addition to rivalry, a second key feature of many public goods is excludability. This means that individuals have access to the good if and only if they are willing to pay a 'user fee' for the service it provides. The costs of using the input may thus be unequivocally assigned to the users, something that is not possible for a pure public good, and potential users will be denied access to it unless they are willing to pay the necessary fee. Under this financing scheme, market provision of the public input is basically possible. Examples of public goods that are often excludable include highways, schools, universities, national parks and television, which may require fees or licences. In addition, publicly provided private goods, like water or electric power supply, for which governments levy user fees, also exist.

In contrast to the treatment of rivalry, the consequences of excludability of public goods and its financing by a user cost, as well as the choice between tax financing and user cost financing for economic growth, has received little attention, despite its practical importance, particularly in European countries. Ott (2001) focuses on the growth impact of an entirely excludable public production input subject to potential congestion. The optimal financing implications are derived for a government that provides the public input at competitive prices. The monopolistic provision of excludable public goods has been discussed by Brito and Oakland (1980), although not in the context of growth models.[2]

The objective of the present paper is to develop a growth model that includes both excludable and non-excludable public goods as productive inputs, both of which may be subject to some degree of congestion. What we have in mind is the following. A firm, as part of its production process, needs to ship its finished output to market. It has the choice of using a motorway, for which it pays a user fee, or a surface road that runs parallel to the motorway and is financed out of tax revenues. The two roads are clearly substitutes in the productive process, and the important question is the optimal provision of the two forms of public input and their pricing structure. Two main features distinguish our analysis from previous contributions: the introduction of partial excludability of the productive public input, and the possibility of monopolistic pricing by the government in an economy experiencing ongoing growth.

The provision of publicly provided infrastructure characterized by exclusion is quite widespread and is therefore a plausible assumption. Within the European Union, different systems for financing infrastructure exist: while some countries levy highway tolls only on trucks, other countries also charge private individuals for their use of the motorway. In addition, toll systems differ in their design from country to country: some tolls are time-based (e.g. the Euro-Vignette System in Belgium, Denmark, Luxembourg, the Netherlands and Sweden), and others are based on distance (see e.g. Germany, Italy or France). With the introduction of the user fee system in Germany, a transition from tax to fee financing for the provision and maintenance of the infrastructure is put in place. In addition, though still tentative, the introduction of highway user fees for private individuals is under discussion.[3]

While in most countries taxation still accounts for the largest part of government revenues, in many countries there has been a recent trend towards more user fees; see e.g. Wassmer and Fisher (2002) for the United States or European Commission (2001) for the EU. One important argument advanced in favour of fees is the revenue effect, meaning that they may increase total government revenues, thereby reducing the budget deficit and giving the government the financial flexibility to pursue other goals. This requires that charging user fees does not simply lead to a different structure of government revenues in favour of fees, but in fact generates additional net revenues. But even in cases where the overall revenues are not necessarily increased, efficiency considerations may still induce policy-makers to have a preference for one revenue mix over another.[4]

The model we employ is a straightforward extension of the Barro (1990) model, modified to include both a conventional non-excludable public input, financed out of tax revenues, and an excludable public input that requires a user fee. Both goods are rival, which we specify by introducing congestion. Our main results are presented as a series of propositions describing the interaction between these two forms of input, both in production and with respect to their financing.

Beginning with a centrally planned economy, we derive the first-best optimal shares of the two forms of public inputs as a benchmark. The first key objective is to characterize the structure of the optimal income tax and user fee that will replicate the first-best equilibrium. These are expressed in terms of: (i) the partial production elasticities of the two inputs, and (ii) their respective degrees of congestion. We show how the existence of congestion in either input raises the income tax and lowers the user fee. We then briefly turn to the case where the government provides non-optimal amounts of the public inputs, and characterize the tax-user fee structure that is necessary to correct for the two externalities that arise in that case: (i) the non-optimal provision of the public good, and (ii) the congestion effects.

We examine in detail the implications of tax versus user-fee financing for the government's budget. In particular, we find that the user fee will fully finance the excludable input if and only if there is no congestion. Whether or not the total revenue generated—taxes plus user fee—suffices to finance the government's overall budget depends critically upon the degrees of congestion, in both types of government input.

The fact that the government is the unique supplier of the public input presents it with the opportunity to price as a monopolist in the case of the excludable input. While this turns out to have no effect on optimal tax policy, it does have important consequences for the setting of the user fee and thus for the overall revenue. Indeed, we find that it can now fully finance the excludable input out of the user fee if the degree of monopoly power equals or exceeds the optimal tax rate.

The remainder of the paper is structured as follows. After setting out the underlying analytical structure in Section I, in Section II we derive the equilibrium in the centrally planned economy. Section III then derives the equilibrium in the decentralized economy, while Section IV provides a general characterization of the optimal tax and pricing policies. These have consequences for the government budget, which are spelled out in Section V. The extension to allow for monopoly pricing by the government is undertaken in Section VI. Up to this point our analysis is based on the most general production function, consistent with sustaining ongoing growth. All that this requires is that it be constant returns to scale in the three productive inputs, i.e. private capital and the two public inputs. Section VII briefly discusses the special case of the constant elasticity of substitution production function, thereby enabling us to focus explicitly on the role of factor substitutability. Section VIII concludes, while technical details are relegated to the Appendix.

I. THE ANALYTICAL FRAMEWORK

Production technology and public inputs

The economy is populated by n identical individuals who consume and produce a single good. Individual output is determined by privately owned capital, k, and the aggregate flow of public services. The individuals may be excluded from at least a part of these services. To capture the feature of excludability, the public input is split into an excludable part, E_S, and a non-excludable component, G_S. The individual agent's production function,

$$(1) \qquad y = F(k, E_S, G_S),$$

is homogeneous of degree one in the three inputs.[5] It is assumed that the productive services derived by the representative individual from a given amount of public expenditure depend upon the usage of his individual capital stock relative to aggregate usage. This describes the situation of relative congestion that is introduced via typical congestion functions; see e.g. Barro and Sala-i-Martin (1992), Eicher and Turnovsky (2000):

$$(2a) \qquad E_S = E\left(\frac{k}{K}\right)^{\varepsilon_E}, \quad 0 \leqslant \varepsilon_E \leqslant 1,$$

$$(2b) \qquad G_S = G\left(\frac{k}{K}\right)^{\varepsilon_G}, \quad 0 \leqslant \varepsilon_G \leqslant 1,$$

where $K = nk$ denotes the aggregate stock of private capital.[6] The exponents ε_E and ε_G parameterize the degree of congestion for either component of the public production input. The case $\varepsilon_E = \varepsilon_G = 0$ corresponds to non-rival pure public inputs that, independent of the size of the economy, are fully available to each individual. There is no congestion. The other limit, $\varepsilon_E = \varepsilon_G = 1$, reflects a situation of proportional (relative) congestion. Given an agent's individual capital stock, only if E and G increase at the same rate as the economy, as measured by the aggregate capital stock, do the levels of service provided to any individual remain constant. The public good is then like a private good, in that, since $K = nk$, each of the n individuals receives his proportionate share of the service; $E_S = E/n$.[7] The cases $0 < \varepsilon_E < 1$, $< 0 < \varepsilon_G < 1$ reflect situations of partial (relative) congestion, in the sense that, given the individual stock of capital, government spending can increase at slower rate than K and still provide a fixed level of services to the firm.[8]

Fiscal instruments and monopoly power

To finance the provision of the public input, the government needs to raise revenues. Most growth models with public inputs assume non-excludability, so that the only way to finance the provision of these goods is through taxes.[9] An important feature of the model developed here is that the government has at its disposal an additional fiscal instrument. Because of the possibility of exclusion, the government may levy user fees on the individual usage, E, which reflects the price each individual has to pay if he decides to employ E for production. As will be shown later, the optimal user fee will equal the marginal cost of provision.[10] But one also has to take into account of the fact that the government might behave as a monopolist in the provision of the excludable public good. Monopoly power is formalized via the degree of monopoly, denoted by ω, which reflects the negative reciprocal of the price elasticity of demand for E. Denoting the user fee by q and the distortionary tax on income by τ, we assume that the government balances its budget at each instant by levying a lump-sum tax, l, on each individual, in accordance with

$$nl = E + G - (n\tau y + nq(\omega)E) \equiv D, \tag{3}$$

where the right-hand side of (3) defines the 'primary budget deficit', D.[11] Writing it this way emphasizes the extent to which total government revenues from income taxation plus user fees deviates from total expenditures, thus enabling us to consider the revenue effect of the user fee.

Aggregate resource constraint

Output can be either consumed, or used for the provision of the public inputs, or accumulated as capital. Thus, the aggregate resource constraint is expressed by

$$n\dot{k} = ny - E - G - nc, \tag{4}$$

where c denotes consumption per capita.

Welfare

The agent's lifetime utility is represented by the intertemporal isoelastic utility function which depends only on consumption,

$$W \equiv \int_0^\infty \frac{c^{1-\sigma}}{1-\sigma} e^{-\rho t} dt, \quad \rho > 0, \ \sigma > 0, \tag{5}$$

where $1/\sigma$ denotes the intertemporal elasticity of substitution and ρ is the rate of time preference.

II. Central Planner

As a benchmark, we begin by considering a centrally planned economy in which the decision-maker can choose the resource allocation directly. To sustain an equilibrium of ongoing growth, both types of government expenditure must be tied to the scale of economy. This can be achieved most conveniently by assuming that the government sets its expenditures for E and G as fractions of aggregate output, $Y \equiv ny$, namely

(6a) $$E = eY, \; 0 < e < 1,$$

(6b) $$G = gY, \; 0 < g < 1.$$

An expansion in government expenditure is parameterized by increases in the expenditure shares, e and g. We analyse the case in which the government acts as a benevolent social planner, determining consumption, the rate of capital accumulation and both public inputs, to maximize the intertemporal utility function of the representative agent, (5), subject to the capital accumulation equation, (4). The social planner is aware of any possible congestion effects, thereby internalizing the link between individual and aggregate capital, $K = nk$. Using this relationship, the congestion functions (2) become $E_S = En^{-\varepsilon_E}$ and $G_S = Gn^{-\varepsilon_G}$, and, together with equations (2) and (6), the production function (1) can be rewritten as

(7) $$y = F(k, En^{-\varepsilon_E}, Gn^{-\varepsilon_G}) = F(k, eyn^{1-\varepsilon_E}, gyn^{1-\varepsilon_G}).$$

As a consequence of the homogeneity assumption, the equilibrium individual production function as perceived by the central planner turns out to be linear in capital and thus, for appropriate preferences, can sustain an equilibrium of ongoing growth.[12] The equilibrium production function is given by

(8) $$y = \phi(en^{1-\varepsilon_E}, gn^{1-\varepsilon_G})k, \; \phi_1 > 0, \; \phi_2 > 0,$$

where ϕ, which reflects both the marginal and average productivity of capital, is an increasing function of both public inputs.

We find it advantageous to conduct the social planner's optimization problem in two stages. We first determine the equilibrium in which the expenditure shares, e and g, are set arbitrarily; then in the second stage e and g are set optimally, along with individual consumption and capital accumulation. By first setting the expenditure shares arbitrarily, we are able to establish the relationship between government expenditure and the equilibrium growth rate—clearly an important relationship in its own right, and one that has been widely studied both theoretically and empirically. Moreover, knowing what happens when expenditure shares are set arbitrarily facilitates our understanding of optimal expenditure policy. From equations (12) and (14) below, for example, it becomes clear what the policy-maker needs to do in order to achieve the optimal growth rate or to maximize welfare (which coincide in this case), and which externalities the optimal policy is correcting. If we restricted our attention to the optimal policies alone, these insights would be lost. We shall refer to the equilibrium in which e and g are set arbitrarily as the 'second-best equilibrium' and the case where e and g are chosen optimally, as the 'first-best equilibrium'.

Arbitrarily set fractions of e and g (second-best optimum)

If the planner sets the expenditure parameters e and g arbitrarily, the optimization problem is simply to maximize welfare (5) subject to the resource constraint

$$(4') \qquad \dot{k} = (1 - e - g)y - c,$$

where y is given by equation (8). Optimizing over consumption, c, and capital, $\dot{k}$, leads to the standard optimality conditions

$$(9a) \qquad c^{-\sigma} = \lambda,$$

$$(9b) \qquad (1 - e - g)\frac{\partial y}{\partial k} = \rho - \frac{\dot{\lambda}}{\lambda},$$

where λ denotes the shadow price of capital. Equation (9a) equates marginal utility to the shadow value of an additional unit of capital, λ, while (9b) equates the social rate of return on capital to the rate of return on consumption. Combining (9a) and (9b), the second-best equilibrium growth rate is given by[13]

$$(10) \qquad \tilde{\varphi} = \frac{1}{\sigma}\left[(1 - e - g)\phi(en^{1-\varepsilon_E}, gn^{1-\varepsilon_G}) - \rho\right].$$

Differentiating (10) with respect to e and g, respectively, we obtain

$$(11a) \qquad \frac{\partial \tilde{\varphi}}{\partial e} = \frac{1}{\sigma e}\frac{y}{k\eta_k}[(1 - g)\eta_E - e(1 - \eta_G)],$$

$$(11b) \qquad \frac{\partial \tilde{\varphi}}{\partial g} = \frac{1}{\sigma g}\frac{y}{k\eta_k}[(1 - e)\eta_G - g(1 - \eta_E)],$$

where η_k, η_E, η_G are the elasticities of output as specified by (1) with respect to the three productive factors, k, E, G, respectively. Setting equations (11) to zero, we see that the growth-maximizing shares of the two types of public input are given by

$$(12a) \qquad e^* = \eta_E = \frac{\eta_e}{1 + \eta_e + \eta_g},$$

$$(12b) \qquad g^* = \eta_G = \frac{\eta_g}{1 + \eta_e + \eta_g},$$

where η_e, η_g are the elasticities of output with respect to the shares of the two forms of public input, e, g, respectively.[14] Thus, increasing either form of government expenditure will increase the growth rate until its share of output equals its corresponding productive elasticity. It is important to emphasize that, except for the Cobb–Douglas production function, these production elasticities are not constant but vary with the public inputs, E and G, as well as with other parameters.[15] This relationship will become apparent in our treatment of the CES production function in Section VII.

Optimally set fractions of e and g (first-best optimum)

It is straightforward to show that, when e and g are chosen optimally this leads to two further optimality conditions that can be conveniently summarized by

$$\text{(13a)}\qquad \begin{aligned}\operatorname{sgn}\left(\frac{\partial W}{\partial e}\right) &= \operatorname{sgn}((1-e-g)\eta_e - e) = \operatorname{sgn}\left(\frac{\eta_e}{1+\eta_e} - \frac{e}{1-g}\right)\\ &= \operatorname{sgn}\left(\frac{\partial\tilde{\varphi}}{\partial e}\right),\end{aligned}$$

$$\text{(13b)}\qquad \begin{aligned}\operatorname{sgn}\left(\frac{\partial W}{\partial g}\right) &= \operatorname{sgn}((1-e-g)\eta_g - g) = \operatorname{sgn}\left(\frac{\eta_g}{1+\eta_g} - \frac{g}{1-e}\right)\\ &= \operatorname{sgn}\left(\frac{\partial\tilde{\varphi}}{\partial g}\right),\end{aligned}$$

from which we infer that, for either form of public input, its qualitative impact on the welfare of the representative agent is identical to its qualitative effect on the growth rate. It immediately follows from (12)–(13) that the growth-maximizing expenditure shares given in (12) are also welfare-maximizing. The equilibrium optimal growth rate is thus given by

$$\text{(14)}\qquad \begin{aligned}\varphi^* &= \frac{1}{\sigma}\left[(1-e^*-g^*)\phi(e^*n^{1-\varepsilon_E}, g^*n^{1-\varepsilon_G}) - \rho\right]\\ &= \frac{1}{\sigma}\left[(1-\eta_E^*-\eta_G^*)\phi(\eta_E^*n^{1-\varepsilon_E}, \eta_G^*n^{1-\varepsilon_G}) - \rho\right],\end{aligned}$$

where e^*, g^* are the growth- and welfare-maximizing expenditure shares, as given in (12), and η_E^*, η_G^* are the corresponding production elasticities, evaluated at the optimum. Equations (12) and (14) thus comprise the first-best optimum. We therefore conclude that the well-known Barro (1990) proposition pertaining to the coincidence of growth- and welfare-maximizing government expenditures extends to both types of public input, and indeed extends beyond the Cobb–Douglas production function to the quite general specification adopted in (1).[17]

III. Equilibrium in the Decentralized Economy

We turn now to the representative agent in the decentralized economy. The individual's production function is given by equation (1). As noted, the individual has to pay an income tax, τ, a lump-sum tax, l, and the user fee, q, if he uses the excludable part of the government input, all of which he takes as parametrically given. For the present, we abstract from the monopolistic pricing of the excludable public input, delaying our discussion of this aspect until Section VI. Both public inputs are subject to relative congestion as represented by equation (2). In contrast to the social planner, the individual does not realize the negative external effect of capital accumulation. Thus, given the homogeneity, the production function as perceived by the individual is given by

$$\text{(15)}\qquad y = F\left(k, E\left(\frac{k}{K}\right)^{\varepsilon_E}, G\left(\frac{k}{K}\right)^{\varepsilon_G}\right) = \phi\left(\frac{E}{k}\left(\frac{k}{K}\right)^{\varepsilon_E}, \frac{G}{k}\left(\frac{k}{K}\right)^{\varepsilon_G}\right)k.$$

The individual's optimization problem is to choose the time paths for individual consumption, capital accumulation and his use of the excludable public input to maximize utility as given by equation (5) subject to the rate of capital accumulation,

$$\dot{k} = (1-\tau)y - c - qE - l, \tag{16}$$

and output y as given by (15). The new feature is the appearance of the user fee in the agent's budget constraint (16). The optimality conditions are

$$c^{-\sigma} = \lambda, \tag{17a}$$

$$(1-\tau)\frac{\partial y}{\partial k} = \rho - \frac{\dot{\lambda}}{\lambda}, \tag{17b}$$

$$(1-\tau)\frac{\partial y}{\partial E} = q. \tag{17c}$$

Equation (17a) coincides with (9a) for the social planner, while (17b) now equates the after-tax marginal product of private capital to the rate of return on consumption. The equilibrium marginal product of private capital, r, derived from (15) is given by[18]

$$r \equiv \frac{\partial y}{\partial k} = \phi(\cdot)(1-(1-\varepsilon_E)\eta_E - (1-\varepsilon_G)\eta_G). \tag{18}$$

Since individuals ignore the consequences of their own activities on the aggregate economy, the individually perceived marginal product of capital includes an externality. Agents overestimate the resulting marginal product if the public input is congested, that is if $\varepsilon_E > 0$ and/or $\varepsilon_G > 0$. Equation (17c) is the formal statement of exclusion. As individuals have to pay directly for the use of the excludable public input, E, they determine their optimal usage by equating its net marginal revenue product to its marginal cost, q (the user fee).

We assume that the government sets the user fee to ensure that the demand for the excludable public good, chosen by the private sector, equals the supply, set by the government. Using the relationship $E = eny$, the market equilibrium in the decentralized economy is thus formally represented by the following two equations:

$$\varphi = \frac{1}{\sigma}[(1-\tau)\phi(\cdot)(1-(1-\varepsilon_E)\eta_E - (1-\varepsilon_G)\eta_G) - \rho], \tag{19a}$$

$$(1-\tau)\eta_E = qne. \tag{19b}$$

Given the expenditure shares e, g, and tax rate, τ, these two equations jointly determine the equilibrium growth rate, φ, and the market-clearing user fee, q. Rewriting (19b) in the form

$$\frac{(1-\tau)\eta_E}{q} = ne = E \tag{19b$'$}$$

highlights the equilibrating role the user fee plays in equating the private demand for the excludable public input (the left-hand side) to the government-determined supply (the right-hand side).

To illustrate the budgetary consequences of a shift from a distortionary income tax to a user fee, we focus on the Cobb–Douglas production function, when η_E is constant. In this case (19b) requires $\mathrm{d}q = -(\eta_E/ne)\,\mathrm{d}\tau > 0$, which from (19a) implies an unambiguous reduction in the growth rate. This has consequences for both the government's primary deficit, given by the right-hand of (3), and the present value of its deficit, as the economy evolves along its equilibrium growth path. Substituting for (6a), (6b) and (8) into (3), and using (19b), the short-run primary deficit and its present value (the intertemporal primary deficit) are, respectively,

$$\text{(20a)} \qquad D = n\phi(\cdot)k[e + g - \tau - (1-\tau)\eta_E],$$

$$\text{(20b)} \qquad \begin{aligned} V \equiv \int_0^\infty D\mathrm{e}^{-r(1-\tau)t}\mathrm{d}t &= \frac{n\phi(\cdot)k_0[e + g - \tau - (1-\tau)\eta_E]}{r(1-\tau) - \varphi} \\ &= \frac{n\phi(\cdot)k_0[e + g - \tau - (1-\tau)\eta_E]}{(1 - 1/\sigma)r(1-\tau) + \rho/\sigma}. \end{aligned}$$

With the productive elasticity being less than 1 (see (12a)), a reduction in the income tax raises the short-run deficit unambiguously, while the effect on the intertemporal deficit satisfies

$$\operatorname{sgn}\left(\frac{\partial V}{\partial \tau}\right) = \operatorname{sgn}(\rho(\eta_E - 1) - r(1 - e - g)(\sigma - 1)).$$

Since $1 > e + g$, for a switch to a user tax to reduce the intertemporal deficit, it is necessary for the intertemporal elasticity of substitution to exceed unity ($\sigma < 1$), a condition that contradicts the bulk of the empirical evidence.[19] Thus, we conclude that, at least for the Cobb–Douglas production function, a switch towards a user fee is unlikely to augment government revenue. Additional revenue stimulated by a higher growth rate over time will be dominated by the reduction of the current revenue flow.

IV. General Characterization of Optimal Tax and User Cost

From a welfare point of view, optimal fiscal policy should allow for the provision of the optimal quantities of E and G and internalize, if necessary, any negative external effect of non-optimal capital accumulation that may otherwise prevail in the decentralized economy. This involves three objectives: (i) attaining the optimal growth rate; (ii) maintaining market equilibrium for the excludable public input; and (iii) ensuring government solvency. We will show that (i) and (ii) can be achieved by the appropriate choice of income tax and user fee. Given these choices and the revenues they generate, equation (3) determines the necessary choice of lump-sum tax to maintain solvency. We begin with the analysis for optimally set expenditure shares and then turn to the case where the government sets e and g arbitrarily.

Government sets e and g optimally

In this case the expenditure shares are determined in accordance with equations (12), that is $e^* = \eta_E$ and $g^* = \eta_g$. Thus, the market clearing condition, (19b), simplifies to

$$(19'') \qquad 1 - \tau = qn.$$

Equating the equilibrium growth rate (19a) to the first-best optimal growth rate (15), and using (20) yields the first-best optimal income tax rate and user fee as functions of the optimally set government expenditure shares (or, equivalently, the corresponding partial production elasticities, where * identifies the optimum) and the congestion parameters

$$(21a) \qquad \tau^* = \frac{\varepsilon_E e^* + \varepsilon_G g^*}{1 - (1 - \varepsilon_E)e^* - (1 - \varepsilon_G)g^*} = \frac{\varepsilon_E \eta_E^* + \varepsilon_G \eta_G^*}{1 - (1 - \varepsilon_E)\eta_E^* - (1 - \varepsilon_G)\eta_G^*},$$

$$(21b) \qquad q^* = \frac{1}{n}\frac{1 - e^* - g^*}{1 - (1 - \varepsilon_E)e^* - (1 - \varepsilon_G)g^*} = \frac{1}{n}\frac{1 - \eta_E^* - \eta_G^*}{1 - (1 - \varepsilon_E)\eta_E^* - (1 - \varepsilon_G)\eta_G^*}.$$

Note again that, except in the case of the Cobb–Douglas production function, $e^* = \eta_E^*$, $g^* = \eta_G^*$ vary with the quantities of the public inputs used, as well as with the congestion parameters.[20] The presence of congestion associated with either input causes individuals to overestimate the social marginal product of capital, generating an incentive to over-accumulate private capital. The resulting growth rate in the decentralized economy becomes sub-optimally high. Hence a positive tax on income is required in order to reduce the incentive to accumulate capital, and thereby correct for this externality. At the same time, the income tax reduces the after-tax marginal product of the excludable public input, E. As the individual demand for E requires equating the (after-tax) marginal product of E to its marginal cost, an increase in τ reduces the marginal product, thereby reducing individual demand. In order to ensure that demand for the excludable good is maintained equal to the optimally set supply, the government decreases the user fee as consequence of an increase in τ. We may summarize this with our first proposition.

Proposition 1. Assume that the government sets the expenditure shares of the excludable and non-excludable public goods optimally. The optimal income tax and the optimal user fees are functions of the partial production elasticities of the inputs, together with their respective congestions. The existence of congestion in either good raises the income tax and reduces the user fee.

The result that congestion favours an income tax is consistent with Barro and Sala-i-Martin's (1992) conclusion, although in their model they interpret the income tax as an approximation to a user fee and note its superiority over lump-sum taxation; in our case the comparison is between the income tax and an explicit user fee.

The mechanism described above, according to which the income tax internalizes the congestion, has one counterintuitive implication: namely, that congestion in the *non-excludable* public good reduces the equilibrium user fee for the *excludable* good. Intuitively, one might have expected that congestion in the non-excludable good would raise the demand for the excludable input, thereby raising its user fee. On the other hand, the fact that congestion in the excludable good, by reducing the marginal product of that input, reduces the user fee is quite intuitive.

Government sets g and e arbitrarily

This case brings out the relationship between the financing instruments on the one hand, and the deviations of the actual expenditure shares from their respective optima on the other. We now consider the second-best growth rate of the centrally planned economy, as determined in equation (10), as a reference. In order for the growth rate in the decentralized economy, (19a), to replicate this second best optimum, we require

$$(22) \qquad 1 - e - g = (1 - \tau)(1 - (1 - \varepsilon_E)\eta_E - (1 - \varepsilon_G)\eta_G).$$

Moreover, given the government's choice of its public inputs, q must be set in accordance with (19b) so as to ensure goods market clearance for the excludable public input. Solving (22) and (19b), the corresponding second-best optimal income tax rate and user fee are now given by

$$(23a) \qquad \hat{\tau} = \frac{e - (1 - \varepsilon_E)\eta_E + g - (1 - \varepsilon_G)\eta_G}{1 - (1 - \varepsilon_E)\eta_E - (1 - \varepsilon_G)\eta_G},$$

$$(23b) \qquad \hat{q} = \left(\frac{1}{ne}\right) \frac{\eta_E(1 - e - g)}{1 - (1 - \varepsilon_E)\eta_E - (1 - \varepsilon_G)\eta_G}.$$

In this case the second-best optimal tax and user fee depend upon actual expenditure shares as well as the production elasticities and congestion parameters. There are now two externalities that need to be corrected: (i) the degree of congestion, and (ii) the deviations in actual expenditures, e and g, from their respective optima. Because the productive elasticities are functions of the actual and optimal expenditure shares, comparison of $\hat{\tau}$, $\hat{q}$ with the first-best optimal values τ^*, q^* is not in general practical, although it becomes feasible in the case of the CES production function, discussed in Section VII below.

One comparison of some interest is that

$$(24a) \qquad \frac{\partial \hat{\tau}}{\partial \eta_I} < 0, \ \frac{\partial \hat{q}}{\partial \eta_I} > 0, \ I = E, G,$$

$$(24b) \qquad \frac{\partial \tau^*}{\partial \eta_I} \geqslant 0, \ \frac{\partial q^*}{\partial \eta_I} \leqslant 0, \ I = E, G.$$

Holding the fraction of government expenditures fixed, an increase in the productivity of either public input (as measured by the productive elasticities, η_E, η_G) raises output and the tax base. This permits the second-best tax rate to be reduced, while the higher productivity, yielding enhanced productive benefits, allows a higher user fee to be charged. But since the higher productive elasticity induces more usage of either input, thereby reducing its marginal productivity, this in general raises the first-best tax rate and reduces the optimal usage fee. The exception is the polar case, where both inputs are pure public goods, in which case (21) implies $\tau^* = 0, q^* = 1/n$, independent of the productive elasticities.

V. Optimally Set Expenditure Shares: Budgetary Implications

It is evident that the degrees of congestion associated with the two types of public input have important consequences for their mode of financing and therefore for government budget balance. To focus on this important issue, we assume that the government sets both expenditure ratios at their respective optima, namely $e^* = \eta_E$ and $g^* = \eta_G$, in accordance with (12). To sustain this equilibrium, the revenue-generating fiscal instruments must satisfy (21). The issue we want to address is the extent to which each type of public good can be individually financed entirely from its designated source of revenue—the non-excludable input from tax revenues, the excludable input from the user fee—as well as the extent to which aggregate expenditure on the two public goods can be financed out of both revenue sources, taken together.

Given n agents, total revenues earned from these two sources equal $n\tau y + nqE$, and we are concerned with the extent to which this is compatible with total public expenditure $G + E$. Recalling the expenditure rules, (6a), (6b), and with expenditure shares set optimally, the government can balance its budget using these two instruments alone (i.e. without lump-sum tax financing) if and only if

$$\tau + nqe^* = e^* + g^*. \tag{25}$$

As we shall see, the extent to which this is possible depends crucially upon the degrees of congestion and the optimal fiscal policy they induce.

To address this issue, it is convenient to rewrite (21a) and (21b) in the form

$$\tau^* = g^* + \frac{\varepsilon_E e^* - g^*[(1-\varepsilon_G)(1-g^*) - (1-\varepsilon_E)e^*]}{1-(1-\varepsilon_E)e^* - (1-\varepsilon_G)g^*}, \tag{26a}$$

$$nq^*e^* = e^* - \frac{(\varepsilon_E e^* + \varepsilon_G g^*)e^*}{1-(1-\varepsilon_E)e^* - (1-\varepsilon_G)g^*}, \tag{26b}$$

and thus

$$\tau^* + q^*ne^* = g^* + e^* + \frac{(1-g^*-e^*)[\varepsilon_E e^* - (1-\varepsilon_G)g^*]}{1-(1-\varepsilon_E)e^* - (1-\varepsilon_G)g^*}. \tag{26c}$$

These expressions highlight how the relationship between revenues and expenditures depends upon the degree of congestion. From these equations we can derive the following

Proposition 2. (i) The revenue generated by the user fee suffices to finance the excludable public input if and only if the optimal tax rate is zero. This occurs if and only if neither public input is subject to congestion.

(ii) The total revenue generated suffices to finance total public expenditure if and only if the optimal ratios of the two inputs satisfies $e^*/g^* = ((1-\varepsilon_G)/\varepsilon_E)$. In the case where congestion is uniform across the two inputs, $\varepsilon_E = \varepsilon_G = \varepsilon$, the government budget (25) will balance if and only if $\varepsilon = g^*/(e^* + g^*)$, i.e. if and only if the degree of congestion equals the fraction of non-excludable in total public expenditure.

Further insight is obtained by discussing special cases, and the following will be considered.

(i) $\varepsilon_E = \varepsilon_G = 0$. In case of an absence of congestion in either public input, as noted, the optimal fiscal policy (21a) and (21b) reduces to

$$\tau^* = 0, \tag{27a}$$

$$q^* = \frac{1}{n}. \tag{27b}$$

With a zero income tax rate, the after-tax marginal product of E is not distorted and the optimal user fee equals the marginal cost of providing E. Substituting these optimal tax rates in (25), we see that, while the expenditure of the excludable good is self-financing (consistent with Proposition 3), the expenditure on the non-excludable good is not. The provision of these optimally supplied public goods is sustainable only as long as the government has at its disposal (positive) lump-sum taxation, which may be employed to finance the non-excludable good.

(ii) $\varepsilon_E = \varepsilon_G = 1$. Suppose that both public inputs are proportionally congested. In this case optimal fiscal policy, (21a) and (21b), becomes

$$\tau^* = e^* + g^* > g^*, \tag{28a}$$

$$q^* = \frac{1}{n}(1 - e^* - g^*) < \frac{1}{n}. \tag{28b}$$

The striking aspect of this result is that tax revenues alone suffice to finance the entire production of the optimally provided public inputs. Thus, although we introduce excludability, it is the *income tax* that internalizes the external effect for both parts of the public input, reducing the incentive that would otherwise exist to over-accumulate capital. In addition, the government receives positive user fee revenues (which would, however, be insufficient to finance the entire amount of E). These revenues are the consequence of the individual demand for E as described in (19b). The positive income tax rate reduces the after-tax marginal product of E. If e is set optimally, the market-clearing condition of E requires a user fee that is below the optimal level, but still positive. Thus, since total revenues generated exceed the expenditure required, the excess revenues should be rebated via a growth-neutral fiscal instrument, such as a (negative) lump-sum tax.

(iii) $\varepsilon_E = 0$, $\varepsilon_G = 1$. We now assume that the excludable part is not congested, whereas the non-excludable part of the public production input is proportionally congested. The optimal tax rate and user fee satisfy

$$e^* + g^* > \tau^* = \frac{g^*}{1 - e^*} > g^* > 0, \tag{29a}$$

$$q^* = \left(\frac{1}{n}\right)\frac{1 - e^* - g^*}{1 - e^*} < \frac{1}{n}, \tag{29b}$$

which upon substitution can be seen to satisfy the balanced budget condition (25). While the revenue generated by the user fee is insufficient to finance the excludable part of the public input, E, the income tax revenue exceeds the amount necessary to finance the non-excludable component, G. Indeed, the excess tax revenues generated equal precisely the amount necessary to make up the shortfall to finance fully the excludable public input. Thus, with tax revenues in part subsidizing the excludable input, the government can balance its budget without the need to impose a lump-sum tax.

(iv) $\varepsilon_E = 1$, $\varepsilon_G = 0$. Finally, we assume that the excludable part of the public production input is proportionally congested, while the non-excludable part is a pure public good. The optimal tax rate and fee then are given by

$$\text{(30a)} \qquad \tau^* = \frac{e^*}{1-g^*} = g^* + \frac{e^* - g^*(1-g^*)}{1-g^*} \gtreqless g^*,$$

$$\text{(30b)} \qquad q^* = \left(\frac{1}{n}\right)\frac{1-e^*-g^*}{1-g^*} < \frac{1}{n},$$

which implies that

$$\text{(31)} \qquad \tau^* + nq^*e^* = e^* + g^* + \frac{e^*(1-e^*) - g^*(1-g^*)}{1-g^*}.$$

Again, the revenues from the user fee are insufficient to finance the optimal amount of the excludable good. But now, excess tax revenues may or may not arise. Several cases need to be distinguished. First, if $e^* = g^*(1-g^*)$, then the tax revenues are sufficient to provide exactly the efficient amount of the non-excludable good, G. If $e^* < g^*(1-g^*)$, tax revenues are insufficient to finance even the non-excludable part of the public input. In both these cases there is a budgetary shortfall. Only if $e^* > g^*(1-g^*)$ does the government generate sufficient tax revenues to finance the non-excludable good. In this case there will still be a budgetary shortfall as long as $g^*(1-g^*) > e^*(1-e^*)$. The total budget will be exactly balanced without an additional instrument if and only if $e^* = g^*$. If $(1-g^*) > e^* > g^*$ excess total revenues are generated, leaving resources available that can be redistributed back to the agents via a lump-sum rebate.

These results highlight how the equilibrium user fee declines with the degree of congestion in the excludable input. This is a consequence of the assumed exogeneity of congestion, and because the individual demand for E requires the marginal revenue and marginal cost of E to be equalized. As the two fiscal instruments, τ and q are linked together (see equation (19b)), it is the income tax rate that internalizes the external effects of capital utilization. Thus, the optimal income tax rate is positive if either public good is subject to congestion, reducing the after-tax marginal product of E. Therefore, q^* must be reduced below marginal cost for the market-clearing condition (19b) to be met.

We may summarize these special cases as follows.

Proposition 3. (i) If $\varepsilon_E = \varepsilon_G = 0$, the expenditure of the excludable input is financed entirely by the optimal user fee, whereas the non-excludable part must be financed via a growth neutral instrument.

(ii) If $\varepsilon_E = \varepsilon_G = 1$, the optimal income tax and user fee yield excess revenues which can be rebated in a growth-neutral manner.

(iii) If $\varepsilon_E = 0$, $\varepsilon_G = 1$, the government budget (25) is balanced. The excess tax revenue exactly covers the shortfall generated by the fee revenue.

(iv) If $\varepsilon_E = 1$, $\varepsilon_G = 0$, whether or not the income tax revenues suffice to finance the revenue shortfall associated with the user fee depends upon the relative sizes of the optimal expenditure shares.

VI. Monopoly Pricing

Thus far, we have assumed that the government provides the excludable part of infrastructure by entering the market for final output and selling at a competitive market-clearing price to the private sector. But as it is the sole supplier of the public production inputs, it is reasonable to analyse the consequences of its acting as a monopolist. As Brito and Oakland (1980) note, abundant examples of monopoly provision exist (e.g. motorways, airports, parks, museums), so this is clearly an important case to consider.

Assume now that the government recognizes that it faces a downward-sloping aggregate demand function for the excludable good, $q = q(E)$, $q' < 0$. The marginal revenue from providing E monopolistically is $q.(1-\omega)$, where ω is the degree of monopoly power defined by $\omega = -1/\delta_{E,q}$, where $\delta_{E,q}$ denotes the price elasticity of demand for E. Analogous to (19b), market clearance in the provision of E now requires

$$(32) \qquad (1-\tau)\eta_E = nq(1-\omega)e.$$

This relation, together with the growth rate given in (19a), describes the market equilibrium under monopolistic behaviour.

Again, we analyse the optimal fiscal policy in the sense of replicating the first-best optimal growth rate in (10) for the optimally set expenditure shares, e^* and g^*. It is immediately evident that the optimal income tax rate is unaffected by monopoly pricing and thus coincides with that given in (21a). In contrast, the optimal user fee is directly influenced by monopolistic behaviour. It is derived analogously to (21b) and is given by

$$(21\mathrm{b}') \qquad \begin{aligned} q^* &= \left(\frac{1}{(1-\omega)n}\right)\frac{(1-e^*-g^*)}{[1-(1-\varepsilon_E)e^*-(1-\varepsilon_G)g^*]} \\ &= \left(\frac{1}{(1-\omega)n}\right)\frac{1-\eta_E^*-\eta_G^*}{1-(1-\varepsilon_E)\eta_E^*-(1-\varepsilon_G)\eta_G^*}. \end{aligned}$$

The optimal user fee increases with an increase in the degree of monopoly and exceeds q^*, as determined in (21b). Analogous to (26), we may express the optimal fiscal policy in the form

$$(26\mathrm{a}') \qquad \tau^* = g^* + \frac{\varepsilon_E e^* - g^*[(1-\varepsilon_G)(1-g^*)-(1-\varepsilon_E)e^*]}{1-(1-\varepsilon_E)e^*-(1-\varepsilon_G)g^*},$$

$$(26\mathrm{b}') \qquad nq^*e^* = e^* + \frac{\omega(1-e^*-g^*)-(1-\omega)(\varepsilon_E e^*+\varepsilon_G g^*)}{(1-\omega)[1-(1-\varepsilon_E)e^*-(1-\varepsilon_G)g^*]}e^*,$$

and thus

$$(26\mathrm{c}') \qquad \tau^* + q^*ne^* = g^* + e^* + \frac{(1-g^*-e^*)[(1-\omega)[\varepsilon_E e^*-(1-\varepsilon_G)g^*]+\omega e^*]}{(1-\omega)[1-(1-\varepsilon_E)e^*-(1-\varepsilon_G)g^*]}.$$

Although it does not affect the optimal tax, the presence of monopoly power still plays an important role in the overall structure of optimal fiscal policy. Most significantly, we see that the user fee generates more revenue under monopolistic pricing, thus reducing the potential amount of lump-sum tax financing required to balance the budget. Moreover, it can now generate more revenue than is required to fully finance the excludable public input. Recalling (21a), (26b′) implies, $q^* \gtreqless \frac{1}{n}$ according to whether $\omega \gtreqless \tau^*$. We can further identify from (26c′) a critical degree of monopoly power, $\omega^* < 1$, such that if ω exceeds ω^* total revenues will always cover total expenditure costs.[21]

We may summarize the impact of monopoly pricing with a further proposition.

Proposition 4. (i) It is possible for certain degrees of monopoly to realize excess revenues out of the user fee, something that is not possible if the government provides the excludable part of public input under competitive pricing.

(ii) The user fee can fully finance the provision of the excludable input if $\omega = \tau^*$. If the degree of monopoly exceeds (falls short of) the optimal income tax, the financing contribution of the fee exceeds (falls short of) the financing requirement for the excludable input.

(iii) There is a critical degree of monopoly power $\omega^* < 1$ such that, if $\omega > \omega^*$, total revenues from the income tax and user fee will always exceed total expenditure costs.

VII. The CES Production Function

We now specify the production technology to be a constant elasticity of substitution (CES) production function that is homogeneous of degree one in the three inputs. Specializing the production function in this way not only facilitates the study of optimal fiscal policy, but is also convenient for analysing the consequences of different degrees of substitution between the inputs for optimal fiscal policy. Thus, the production function (1) becomes

$$(1') \qquad y = \left[\alpha k^{-\xi} + \beta G_S^{-\xi} + \gamma E_S^{-\xi}\right]^{-\frac{1}{\xi}} \quad 0 < \alpha < 1, 0 < \beta < 1, 0 < \gamma < 1, \alpha + \beta + \gamma = 1,$$

where $\theta \equiv 1/(1+\xi)$, $0 < \theta < \infty$, denotes the elasticity of substitution between the three inputs. Utilizing the congestion function (2) and the expenditure shares as given by (6), the equilibrium production function, (7), can be expressed in the linear 'AK form':

$$(7') \qquad y = \alpha^{-1/\xi}\left[1 - \beta g^{-\xi} n^{-\xi(1-\varepsilon_G)} - \gamma e^{-\xi} n^{-\xi(1-\varepsilon_E)}\right]^{1/\xi} k.$$

Centrally planned economy

As in Section II, we begin by summarizing the equilibrium growth rate in the centrally planned economy. For arbitrarily set expenditure shares, the equilibrium (second-best) growth rate is

$$(10') \qquad \tilde{\varphi} = \frac{1}{\sigma}\left[(1-e-g)\alpha^{-1/\xi}\left(1 - \beta g^{-\xi} n^{-\xi(1-\varepsilon_G)} - \gamma e^{-\xi} n^{-\xi(1-\varepsilon_E)}\right)^{1/\xi} - \rho\right].$$

Following the procedure employed for the general production function in Section II, we can verify that the growth-maximizing and welfare-maximizing expenditure shares coincide, being given by

$$(13') \qquad g^* = \beta^{\frac{1}{1+\xi}} n^{-\frac{\xi(1-\varepsilon_G)}{1+\xi}}; \quad e^* = \gamma^{\frac{1}{1+\xi}} n^{-\frac{\xi(1-\varepsilon_E)}{1+\xi}},$$

respectively. In addition, we can compute the production elasticities directly from (1′), together with the congestion functions (2), to obtain

$$(33) \qquad \eta_G = \beta g^{-\xi} n^{(\varepsilon_G - 1)\xi}; \quad \eta_E = \gamma e^{-\xi} n^{(\varepsilon_E - 1)\xi}.$$

Combining (13′) and (33) yields

$$(34) \qquad \eta_G = (g^*)^{1+\xi} g^{-\xi}; \quad \eta_E = (e^*)^{1+\xi} e^{-\xi}.$$

These expressions bring out the point made earlier that in general the production elasticity depends upon the usage of the productive input, as well as the degree of congestion. The exception is the Cobb–Douglas production function, $\xi = 0$, when $\eta_G = \beta, \eta_E = \gamma$. Note further from (34) that, when g and e are set optimally, this equation implies $\eta_G = g^*$, $\eta_E = e^*$, consistent with (12).

Dividing the two expressions in (13′) implies

$$(35) \qquad \left(\frac{g}{e}\right)^* = \left(\frac{\beta}{\gamma}\right)^{\frac{1}{1+\xi}} n^{\frac{\xi(\varepsilon_G - \varepsilon_E)}{1+\xi}},$$

from which we see that the optimal ratio of non-excludable to excludable public inputs depends upon (i) their productivity, (ii) the elasticity of substitution and (iii) their differential degrees of congestion. We may note the following three important cases.

(i) If $\xi \to \infty$ and thus $\theta = 0$ (Leontief production function), then

$$\left(\frac{g}{e}\right)^* \to n^{(\varepsilon_G - \varepsilon_E)},$$

so that the ratio of their optimal usage depends only upon their differential congestion.

(ii) If $\xi = 0$ and thus $\theta = 1$ (Cobb–Douglas production function), then

$$\left(\frac{g}{e}\right)^* = \frac{\beta}{\gamma},$$

so that the ratio of their optimal usage depends only upon their relative productivity and is independent of the degree of congestion.

(iii) If $\xi = -1$ and thus $\theta \to \infty$ (perfect substitutes),

$$\left(\frac{g}{e}\right)^* = \begin{cases} 0 & \text{if } \beta n^{\xi\varepsilon_G} < \gamma n^{\xi\varepsilon_E} \\ 1 & \text{if } \beta n^{\xi\varepsilon_G} = \gamma n^{\xi\varepsilon_E} \\ \infty & \text{if } \beta n^{\xi\varepsilon_G} > \gamma n^{\xi\varepsilon_E} \end{cases}.$$

Thus, in the case where the two public inputs are perfect substitutes, the entire public input should take the form of the one having the higher 'congestion-adjusted' productivity.

Decentralized economy

Analogous to (15), the representative agent perceives the production function in the form:

$$
\begin{aligned}
y &= \left[\alpha k^{-\xi} + \beta\left(G\left(\frac{k}{K}\right)^{\varepsilon_G}\right)^{-\xi} + \gamma\left(E\left(\frac{k}{K}\right)^{\varepsilon_E}\right)^{-\xi}\right]^{-\frac{1}{\xi}} \\
&= \left[\alpha + \beta\left(\frac{G}{k}\right)^{-\xi}\left(\frac{k}{K}\right)^{-\xi\varepsilon_G} + \gamma\left(\frac{E}{k}\right)^{-\xi}\left(\frac{k}{K}\right)^{-\xi\varepsilon_E}\right]^{-\frac{1}{\xi}} k.
\end{aligned} \tag{36}
$$

Taking the derivative of (36) with respect to k, in equilibrium the agent's perceived private marginal physical product of capital is

$$\frac{\partial y}{\partial k} = \frac{y}{k}(1 - (1-\varepsilon_E)\eta_E - (1-\varepsilon_G)\eta_G),$$

where $y/k \equiv (\eta_k/\alpha)^{1/\xi} \equiv \phi(\cdot)$. Omitting for simplicity the monopolistic pricing effect, the market equilibrium in the decentralized economy is again given by (19), the only difference being in the specification of the function $y/k \equiv \phi(\cdot)$.

The second-best optimal fiscal policy is again specified by (23). But in light of relationships (34), we can express it in the following intuitive way:

$$\hat{\tau} = \frac{g\left[1-(1-\varepsilon_G)(g^*/g)^{1+\xi}\right] + e\left[1-(1-\varepsilon_E)(e^*/e)^{1+\xi}\right]}{1-(1-\varepsilon_G)g(g^*/g)^{1+\xi} - (1-\varepsilon_E)e(e^*/e)^{1+\xi}}, \tag{37a}$$

$$\hat{q} = \left(\frac{1}{n}\right)\frac{(e^*/e)^{1+\xi}(1-e-g)}{1-(1-\varepsilon_G)g(g^*/g)^{1+\xi} - (1-\varepsilon_E)e(e^*/e)^{1+\xi}}. \tag{37b}$$

When written in this way it is quite explicit how the optimal fiscal policy is correcting for two distortions: (i) congestion and (ii) the deviations of the actual expenditure shares from their respective optima. From (37a) and (37b), we can derive the following.[22]

Proposition 5. (i) If both g and e are set at their respective optima, g^*, e^*, then $\hat{\tau} = \tau^*, \hat{q} = q^*$ as given in (21).

(ii) If g is set optimally, then $\hat{\tau} \gtrless \tau^*$ and $\hat{q} \lesseqgtr q^*$ according to whether $e \lesseqgtr e^*$.

(iii) If e is set optimally, then $\hat{\tau} \gtrless \tau^*$ and $\hat{q} \lesseqgtr q^*$ according to whether $g \lesseqgtr g^*$.

VIII. Conclusions

Many public goods are characterized by two key attributes: rivalry and excludability. While the role of rivalry has been widely considered in the growth literature, excludability has not. In this paper we have introduced both non-excludable and excludable public inputs into a simple endogenous growth model. Our focus has been on deriving the equilibrium growth rate and designing the optimal tax and user-cost tax structure. Our results emphasize the role of congestion in determining this optimal structure, and the consequences this has in turn for the government's budget. The latter consists of fee and tax revenues that are used to finance the entire public production input and that may or may not suffice to satisfy the financing requirements for the entire input. If no congestion arises, a user fee set at marginal cost yields the optimal amount of the excludable public input, while the non-excludable input must be financed via a growth neutral tax. If either form of the public input is congested, it is optimal to levy a positive income tax to internalize the external effect. At the same time, owing to the interdependence between the optimal fees and income taxes, the optimal user fee is reduced, thereby decreasing the corresponding fee revenues. Then the financing contribution of the fees is not sufficient to provide the optimal amount of the excludable input. This result changes if the government passes on user fees at marginal costs but makes use of monopoly pricing that might be accomplished if the government is the unique supplier. It is then possible for certain degrees of monopoly to realize excess revenues out of the user fee that might replace (non-distortionary) taxes in order to finance the entire infrastructure.

We end with two caveats and suggestions for further research on this important topic. First, by introducing government inputs as flows into production, the equilibrium we derive always places the economy on its balanced growth path. This has the analytical advantage of simplifying the characterization of the optimal tax and pricing structure. But much of the recent literature analysing the role of publicly provided productive inputs recognizes that they should be treated more appropriately as stocks than flows, thereby introducing public as well as private capital. This observation was made early on by Arrow and Kurz (1970) and is also recognized in some of the more contemporary endogenous growth models; see e.g. Futagami *et al.* (1993) and Turnovsky (1997). The effect of this is to introduce transitional equilibrium dynamics, suggesting that the optimal financing policies will involve time-varying tax rates and user fees, as the decentralized economy seeks to track the first-best optimal path.[23]

Another limitation of the analysis is that it does not fully capture the linkage between congestion and the user fee. It takes the degree of congestion as given and determines the corresponding equilibrium user fee. One of the motivations for imposing a user fee is to reduce congestion, in which case the equilibrium level of congestion would become endogenously determined along with the user fee. Extending the model in this direction would be an important step.

APPENDIX

Relationship between general production function and the intensive form

As the production function

$$\text{(A1)} \qquad y = F(k, E_S, G_S)$$

is assumed to be homogeneous of degree one in its three arguments k, E_S, G_S, Euler's theorem implies that

$$\text{(A2)} \qquad y = \frac{\partial F}{\partial k}k + \frac{\partial F}{\partial E_S}E_S + \frac{\partial F}{\partial G_S}G_S \equiv F_1 k + F_2 E_S + F_3 G_S.$$

Substituting from the relationships

$$E_S = En^{-\varepsilon_E} = en^{1-\varepsilon_E}y, \quad G_S = Gn^{-\varepsilon_G} = gn^{1-\varepsilon_N}y,$$

we can rewrite (A2) as

(A3) $$\left(1 - F_2 en^{1-\varepsilon_E} - F_3 gn^{1-\varepsilon_G}\right)y = F_1 k.$$

Now take the total differential of (A1) to obtain

$$\mathrm{d}y = F_1 \mathrm{d}k + F_2 \mathrm{d}E_S + F_3 \mathrm{d}G_S,$$

which, holding n, e, g constant, implies

(A4) $$\left(1 - F_2 en^{1-\varepsilon_E} - F_3 gn^{1-\varepsilon_G}\right)\mathrm{d}y = F_1 \mathrm{d}k.$$

Equations (A3) and (A4) imply that $\mathrm{d}y/y = \mathrm{d}k/k$, so that any production function having the above homogeneity properties can be written in the 'AK form':

(A5) $$y = \phi(en^{1-\varepsilon_N}, gn^{1-\varepsilon_G})k,$$

as represented by (8) in the text.

Relationships between elasticities

The following relationships between $\eta_e, \eta_g, \eta_E, \eta_G$ hold. First, rewriting (A2), we have

(A6) $$\frac{\partial F}{\partial k}\frac{k}{y} + \frac{\partial F}{\partial E_S}\frac{E_S}{y} + \frac{\partial F}{\partial G_S}\frac{G_S}{y} = 1,$$

which in elasticity form can be written as

(A7) $$\eta_k + \eta_{E_S} + \eta_{G_S} = 1.$$

Using the fact that $E_S = En^{1-\varepsilon_E}$ and $G_S = Gn^{1-\varepsilon_G}$,

$$\eta_{E_S} \equiv \frac{\partial F}{\partial E_S}\frac{E_S}{y} = \frac{\partial F}{\partial E}\frac{E_S}{n^{1-\varepsilon_S}y} = \frac{\partial F}{\partial E}\frac{E}{y} \equiv \eta_E,$$

and similarly $\eta_{G_S} = \eta_G$, so that (A7) can be written in the equivalent form as

(A8) $$\eta_k + \eta_E + \eta_G = 1.$$

To derive the relationships between the elasticities in the aggregate quantities and in the shares, we rewrite equation (A1) as

(A9) $$y = F(k,\ en^{1-\varepsilon_E}y,\ gn^{1-\varepsilon_G}y).$$

Taking derivatives of this with respect to e, g, respectively, we obtain

(A10a) $$\frac{\partial y}{\partial e} = \frac{F_2 n^{1-\varepsilon_E}y}{1 - F_2 n^{1-\varepsilon_E}e - F_3 n^{1-\varepsilon_G}g} > 0,$$

(A10b) $$\frac{\partial y}{\partial g} = \frac{F_3 n^{1-\varepsilon_G}y}{1 - F_2 n^{1-\varepsilon_E}e - F_3 n^{1-\varepsilon_G}g} > 0.$$

Using the above fact that

$$\eta_E \equiv \eta_{E_S} \equiv \frac{\partial F}{\partial E_S}\frac{E_S}{y} = F_2 en^{1-\varepsilon_E},$$

and analogously for η_G, (A10a) and (A10b) imply that

(A11a) $$\eta_e = \frac{\eta_E}{1 - \eta_E - \eta_G} = \frac{\eta_E}{\eta_k},$$

$$\text{(A11b)}\quad \eta_g = \frac{\eta_G}{1-\eta_E-\eta_G} = \frac{\eta_G}{\eta_k}.$$

Derivation of marginal physical product of capital in a decentralized economy

Differentiating (15) with respect to k yields

$$\text{(A12)}\qquad \frac{\partial y}{\partial k} = F_1 + \varepsilon_E E F_2 \frac{k^{\varepsilon_E - 1}}{K^{\varepsilon_E}} + \varepsilon_G G F_3 \frac{k^{\varepsilon_G - 1}}{K^{\varepsilon_G}}.$$

Imposing the equilibrium condition $K = nk$, and using the fact that $\varepsilon_{E_S} = \varepsilon_E$ and $\varepsilon_{G_S} = \varepsilon_G$, this simplifies to

$$\text{(A13)}\qquad \begin{aligned} \frac{\partial y}{\partial k} &= F_1 + \varepsilon_E F_2 \frac{E}{k} + \varepsilon_G F_3 \frac{G}{k} \\ &= \frac{y}{k}\left[\frac{F_1 k}{y} + \varepsilon_E F_2 \frac{E}{y} + \varepsilon_G F_3 \frac{G}{y}\right]. \end{aligned}$$

Using the above definitions of elasticities, (A8), and the relationship $y = \phi(\cdot)k$ in (8), (A13) immediately yields the expression in the text, namely

$$\text{(A14)}\qquad \frac{\partial y}{\partial k} = \phi(1 - (1-\varepsilon_E)\eta_E - (1-\varepsilon_G)\eta_G).$$

ACKNOWLEDGMENTS

This paper was begun when Ingrid Ott visited the University of Washington. Stephen Turnovsky's research was supported in part by the Castor Endowment at the University of Washington. The constructive suggestions of an anonymous referee are gratefully acknowledged.

NOTES

1. For example, Thompson (1974) argues that even national defence, often regarded as the prototypical pure public good, is subject to a form of congestion.
2. Other authors who discuss the provision of excludable public goods include Burns and Walsh (1981), Brennan and Walsh (1985) and Fraser (1996), although in all cases from a very different perspective from that we shall be adopting.
3. Another important example of infrastructure that is excludable, but nevertheless still predominantly provided by governments, is airports. The structure of airport ownership of in Europe is quite diverse. While the British Airport Authority privatized its entire airport system, on the European continent at most 50% of any airport is owned by private agents. The excludability arises from the need to purchase landing rights to use the airport.
4. For a discussion of these issues see e.g. Dewees (2002). Another argument for user fees (not addressed here) pertains to a potential incentive effect, whereby the user fee may reduce congestion arising from a suboptimally high usage of the public input.
5. We assume that labour is supplied inelastically.
6. Eicher and Turnovsky (2000) discuss at some length alternative specifications of congestion, some of which have their genesis in the urban economics literature; see e.g. Edwards (1990). They draw the distinction between 'relative' congestion, as specified in equation (2), and 'absolute' congestion, whereby (2a), for example, would take the form $Es = EK^{-\chi}$, say. As Eicher and Turnovsky note, since, unlike relative congestion, absolute congestion is in general inconsistent with endogenous growth, we adopt the specification given in (2). In this specification congestion operates through the aggregate use of capital, so that, as long as agents own capital—and, being identical, we assume that they all do—increasing the number of people in the economy will generate congestion and reduce the services provided by the public input.
7. As a terminological point, Barro and Sala-i-Martin (1992) describe the public service in this case as being excludable.

8. Although we do not discuss this case, we should not necessarily rule out congestion parameters in excess of unity. This describes a situation in which congestion is so great that the public good must grow faster than the economy in order for the level of services provided to remain constant. This case is unlikely at the aggregate level, but may well be plausible for local public goods (see Edwards 1990).
9. See e.g. Barro (1990), Rebelo (1991) and Futagami *et al.* (1993). Some analyses also allow for government borrowing, which typically is equivalent to lump-sum tax financing; see e.g. Turnovsky (1996), Ireland (1994) and Bruce and Turnovsky (1999).
10. Note that all the fiscal instruments that we consider are constant, so that problems of time-inconsistent policies do not arise. A discussion of optimal public investment with and without government commitment is provided by Azzimanti-Renzo *et al.* (2003).
11. We define the term 'primary budget deficit' to be government expenditures less tax revenues from current activity. Note that D may be positive or negative, depending upon the combination of fiscal instrument chosen. This terminology may differ from other usages which may include tax revenues on government interest payments.
12. The relationship between the basic production function, (1) and the 'AK form' (8) is discussed in the Appendix.
13. The derivation of (10) is straightforward. Taking the time derivative of (9a), combining with (9b), and recalling (8), immediately yields the equilibrium growth rate of consumption. Assuming the balanced growth path along which consumption and capital grow at the same rate, (4′) yields the consumption–capital ratio consistent with this assumption.
14. The relationships between the two sets of elasticities are found in the Appendix, (A11a) and (A11b). These relationships have been utilized in deriving the second pair of equalities in (12a) and (12b).
15. This means that solving explicitly for the optimal government expenditure shares may involve solving a highly nonlinear pair of equations that may, or may not, yield closed-form solutions.
16. We should emphasize that precisely the same first-best equilibrium as defined by (12) and (14) would obtain if, instead of the two-stage optimization procedure we have adopted, we were to choose the optimal *levels* of government expenditure directly by maximizing intertemporal utility, (5), subject to the resource constraint, (4), and the production function, (1).
17. However, its robustness should not be overstated. Turnovsky (2000) discusses a number of important circumstances in which it ceases to hold. These include: (i) the introduction of risky technology, (ii) the government input as a stock rather than as a flow and (iii) adjustment costs associated with investment.
18. See the Appendix for the derivation of this equation.
19. The empirical estimates of the intertemporal elasticity of substitution are quite far-ranging consensus estimates probably lying in the range 0.3–0.4. However, we should note that estimates in excess of unity have been obtained; see e.g. Vissing-Jørgensen and Attanasio (2003), in which case we cannot rule out the possibility that the switch to the user tax may reduce the intertemporal deficit.
20. We shall illustrate this aspect in the context of the CES production function in Section VII below.
21. $\omega^* = ((1-\varepsilon_G)g^* - \varepsilon_E e^*)/((1-\varepsilon_G)g^* + (1-\varepsilon_E)e^*) < 1$, which varies with congestion and the optimal expenditure ratios.
22. The results of Proposition 5 continue to hold if, instead of (1′), output is determined by a two-level production of the form $y = Ak^{\alpha}X^{1-\alpha}$, $X = [\beta G_S^{-\xi} + (1-\beta)E_S^{-\xi}]^{-1/\zeta}$.
23. For example, Turnovsky (1997) shows how the optimal tax rate is time-varying in the case where public capital is non-excludable.

REFERENCES

Arrow, K. J. and Kurz, M. (1970). *Public Investment, the Rate of Return, and Optimal Fiscal Policy*. Baltimore: Johns Hopkins University Press.

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of Monetary Economics*, **23**, 177–200.

Azzimanti-Renzo, M., Sarte, P. and Soares, J. (2003). *Optimal Public Investment and Time Consistency*. Federal Reserve Bank of Richmond Working Paper 03–10.

Barro, R. J. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy*, **98**, 103–25.

——— and Sala-i-Martin, X. (1992). Public finance in models of economic growth. *Review of Economic Studies*, **59**, 645–61.

BAXTER, M. and KING, R. G. (1993). Fiscal policy in general equilibrium. *American Economic Review*, **83**, 315–34.
BRENNAN, H. G. and WALSH, C. (1985). Private markets in (excludable) public goods: a re-examination. *Quarterly Journal of Economics*, **99**, 811–19.
BRITO, D. L. and OAKLAND, W. H. (1980). On the monopolistic provision of excludable public goods. *American Economic Review*, **70**, 691–704.
BRUCE, N. and TURNOVSKY, S. J. (1999). Budget balance, welfare, and the growth rate: 'dynamic scoring' of the long-run government budget. *Journal of Money, Credit, and Banking*, **31**, 162–86.
BURNS, M. E. and WALSH, C. (1981). Market provision of price-excludable public goods: a general analysis. *Journal of Political Economy*, **89**, 166–91.
CASSOU, S. P. and LANSING, K. (1998). Optimal fiscal policy, public capital and the productivity slowdown. *Journal of Economic Dynamics and Control*, **22**, 911–35.
CORNES, R. and SANDLER, T. (1996). *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge: Cambridge University Press.
DEWEES, D. N. (2002). Pricing municipal services: the economics of user fees. *Canadian Tax Journal*, **50**, 586–99.
EDWARDS, J. H. Y. (1990). Congestion function specification and the 'publicness' of local public goods. *Journal of Urban Economics*, **27**, 80–96.
EICHER, T. and TURNOVSKY, S. J. (2000). Scale, congestion and growth. *Economica*, **67**, 325–46.
EUROPEAN COMMISSION (2001). *European Transport Policy for 2010: Time to Decide*, White Paper. Luxembourg: European Communities.
FISHER, W. H. and TURNOVSKY, S. J. (1998). Public investment, congestion, and private capital accumulation. *Economic Journal*, **108**, 399–413.
FRASER, C. D. (1996). On the provision of excludable public goods. *Journal of Public Economics*, **60**, 111–30.
FUTAGAMI, K., MORITA, Y. and SHIBATA, A. (1993). Dynamic analysis of an endogenous growth model with public capital. *Scandinavian Journal of Economics*, **95**, 607–25.
GLOMM, G. and RAVIKUMAR, B. (1994). Public investment in infrastructure in a simple growth model. *Journal of Economic Dynamics and Control*, **18**, 1173–87.
GLOMM, G. and RAVIKUMAR, B. (1997). Productive government expenditures and long-run growth. *Journal of Economic Dynamics and Control*, **21**, 183–204.
IRELAND, P. N. (1994). Supply-side economics and endogenous growth. *Journal of Monetary Economics*, **33**, 559–71.
OTT, I. (2001). Congestion, excludability, and endogenous growth. *ifo-Studien*, **4**, 453–67.
REBELO, S. (1991). Long-run policy analysis and long-run growth. *Journal of Political Economy*, **99**, 500–21.
THOMPSON, E. A. (1974). Taxation and national defense. *Journal of Political Economy*, **82**, 755–82.
TURNOVSKY, S. J. (1996). Optimal tax, debt, and expenditure policies in a growing economy. *Journal of Public Economics*, **60**, 21–44.
——— (1997). Fiscal policy in a growing economy with public capital. *Macroeconomic Dynamics*, **1**, 615–39.
——— (2000). *Methods of Macroeconomic Dynamics*, 2nd edn. Cambridge, Mass.: MIT Press.
VISSING-JØRGENSEN, A. and ATTANASIO, O. P. (2003). Stock-market participation, intertemporal substitution, and risk-aversion. *American Economic Review, Papers and Proceedings*, **93**, 383–91.
WASSMER, R. W. and FISHER, R. C. (2002). Interstate variation in the use of fees to fund K-12 public education. *Economics of Education Revenue*, **21**, 87–100.