Optimal Taxation and Public Production II: Tax Rules

Peter A. Diamond, James A. Mirrlees

The American Economic Review, Volume 61, Issue 3 (Jun., 1971), 261-278.

Your use of the JSTOR database indicates your acceptance of JSTOR's Terms and Conditions of Use. A copy of JSTOR's Terms and Conditions of Use is available at http://www.jstor.ac.uk/about/terms.html, by contacting JSTOR at jstor@midas.ac.uk, or by calling JSTOR at 0161 275 7919 or (FAX) 0161 275 6040. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The American Economic Review is published by American Economic Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.ac.uk/journals/aea.html.

The American Economic Review
©1971 American Economic Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor@midas.ac.uk.

©1999 JSTOR

http://www.jstor.ac.uk/
Mon Jan 4 13:16:55 GMT 1999

Optimal Taxation and Public Production II: Tax Rules

By PETER A. DIAMOND AND JAMES A. MIRRLEES*

In Part I of this paper which appeared in the March 1971 issue of this *Review*, we set out the problem of using taxation and government production to maximize a social welfare function. We derived the first-order conditions, and considered the argument for efficiency in aggregate production. Here in Part II we consider the structure of optimal taxes in more detail. Part I contained five sections, and Part II begins at Section VI. In the sixth and seventh sections we consider commodity taxation in one- and many-consumer economies. In the eighth section we consider other kinds of taxes; and in the ninth, public consumption. In the tenth section we consider a rigorous treatment of the problem, giving a sufficient condition for the validity of the first-order conditions. To begin, we shall restate the notation and basic problem.

Notation

p	producer prices
q	consumer prices
t	taxes ($t=q-p$)
$x^h(q)$	net demand by consumer h (incomes are assumed to equal zero) $h=1, 2, \ldots, H$
$u^h(x^h)$	utility function of consumer h
$v^h(q)$	indirect utility function of consumer h $v^h(q)=u^h(x^h(q))$
$X(q)$	aggregate net demand $X(q)=\sum_h x^h(q)$
$U(x^1, \ldots, x^H)$	social welfare function
$V(q)$	indirect social welfare function $V(q)=U(x^1(q), \ldots, x^H(q))$
$W(u^1, \ldots, u^H)$	special case of an individualistic social welfare function, assumed for some of the analysis below.

With this notation before us again, we can restate the welfare maximization problem as that of selecting q to

$$\text{(33)} \qquad \begin{aligned} &\text{Maximize } V(q) \\ &\text{subject to } G(X(q)) \leqq 0 \end{aligned}$$

where G represents the aggregate production constraint. This problem gave rise to the first-order conditions ((19) and (22)) which were equivalently stated as

$$\text{(34)} \qquad \begin{aligned} \frac{\partial V}{\partial q_k} &= \lambda \sum p_i \frac{\partial X_i}{\partial q_k} \\ &= -\lambda \frac{\partial}{\partial t_k} \left(\sum t_i X_i \right) \\ &\qquad (k = 1, 2, \ldots, n) \end{aligned}$$

Equations (34) were derived only for $k=2, \ldots, n$. But we can see that they hold also for $k=1$; for, on multiplying by q_k and adding, we have

$$\sum_{k=1}^{n} \left[\frac{\partial V}{\partial q_k} - \lambda \sum_i p_i \frac{\partial X_i}{\partial q_k} \right] q_k = 0$$

by the homogeneity of degree 0 of V and the X_i. Equation (34) states that the impact of a price rise on social welfare is proportional to the cost of meeting the change

* Massachusetts Institute of Technology and Nuffield College, respectively. The remainder of the matching footnote in Part I is appropriate here too.

in demand induced by the price rise. Alternatively the impact of a tax increase on social welfare is proportional to the induced change in tax revenue (all calculated at fixed producer prices).

VI. Optimal Tax Structure—One-Consumer Economy

For one consumer and an individualistic welfare function (so that V coincides with v, the indirect utility function of the only consumer in the economy), we can express directly the derivative of social welfare with respect to q_k ($v_k = -\alpha x_k$ where α is the marginal utility of income—see equation (5) of Part I). For this case we can then explore the structure of taxation in more detail. The formulation of the first-order conditions using compensated demand derivatives is due to Paul Samuelson (1951). We begin by stating the familiar Slutsky equation:

$$\frac{\partial x_i}{\partial q_k} = s_{ik} - x_k \frac{\partial x_i}{\partial I} \tag{35}$$

where s_{ik} is the derivative of the compensated demand curve for i with respect to q_k, and $\partial x_i/\partial I$ is the derivative of the uncompensated demand with respect to income (evaluated at $I=0$ in our case). We shall make use of the well-known result that $s_{ik} = s_{ki}$.

Substituting into the first-order conditions (34) we have:

$$\begin{aligned} -\alpha x_k &= -\lambda \frac{\partial}{\partial t_k}\left(\sum t_i x_i\right) \\ &= -\lambda\left(x_k + \sum t_i \frac{\partial x_i}{\partial t_k}\right) \\ &= -\lambda x_k - \lambda \sum t_i s_{ik} \\ &\quad + \lambda x_k \sum t_i \frac{\partial x_i}{\partial I} \\ &\qquad k = 1, 2, \ldots, n \end{aligned} \tag{36}$$

Rearranging terms, we can write this in the form:

$$\frac{\sum_i t_i s_{ik}}{x_k} = \frac{\alpha + \lambda - \lambda \sum t_i \dfrac{\partial x_i}{\partial I}}{\lambda} \tag{37}$$

The point to be noticed is that the right-hand side of this equation is independent of k. Call it $-\theta$. Finally, using the symmetry of the Slutsky matrix, we write the first-order conditions as:

$$\frac{\sum_i s_{ki} t_i}{x_k} = -\theta \tag{38}$$

Multiplying by $t_k x_k$ and summing, we obtain

$$\theta \sum_k t_k x_k = -\sum_{k,i} t_k s_{ki} t_i \geq 0, \tag{39}$$

by the negative semi-definiteness of the Slutsky matrix. Thus θ has the same sign as net government revenue.

The left-hand side of (38) is the percentage change in the demand for good k that would result from the tax change if producer prices were constant, the consumer were compensated so as to stay on the same indifference curve, and the derivatives of the compensated demand curves were constant at the same level as at the optimum point:

$$\begin{aligned} \Delta x_k &= \sum_i \int_0^{t_i} \frac{\partial x_k}{\partial t_i} dt_i = \sum_i \int_0^{t_i} s_{ki} dt_i \\ &= \sum_i s_{ki} \int_0^{t_i} dt_i = \sum_i s_{ki} t_i \end{aligned} \tag{40}$$

In fact, it is not possible for all these derivatives to be constant. But if the optimal taxes are small, it is approximately true that the optimal tax structure implies an equal percentage change in compensated demand at constant producer prices.

We can also calculate the actual changes in demand arising from the tax structure (assuming price derivatives of demand and production prices are constant) by resubstituting from the Slutsky equation (35). Then, upon substitution, we have:

$$\sum_i \frac{\partial x_k}{\partial q_i} t_i + \frac{\partial x_k}{\partial I} \sum t_i x_i = -\theta x_k;$$

or

$$\frac{\sum_i \frac{\partial x_k}{\partial q_i} t_i}{x_k} = -\theta - x_k^{-1} \frac{\partial x_k}{\partial I} \sum t_i x_i \tag{41}$$

The actual changes in demand (again assuming constant derivatives) induced by the tax structure differ from proportionality with a larger than average percentage fall in demand for goods with a large income derivative.

Three-Good Economy

In the case of a three-good economy, we can obtain an expression for the relative ad valorem tax rates of the two taxed goods. This argument is similar to that of W. J. Corlett and D. C. Hague, who discussed the direction of movement away from proportional taxation that would increase utility. In the three-good case, with good one untaxed, the first-order conditions (38) become

$$\begin{aligned} s_{22}t_2 + s_{23}t_3 &= -\theta x_2 \\ s_{32}t_2 + s_{33}t_3 &= -\theta x_3 \end{aligned} \tag{42}$$

Solving these equations we have

$$t_2 = \theta \frac{s_{23}x_3 - s_{33}x_2}{s_{22}s_{33} - s_{23}^2}, \quad t_3 = \theta \frac{s_{32}x_2 - s_{22}x_3}{s_{22}s_{33} - s_{23}^2} \tag{43}$$

Notice that the denominator here is positive, by the properties of the Slutsky matrix. We convert these into elasticity expressions, defining the elasticity of compensated demand by

$$\sigma_{ij} = \frac{q_j s_{ij}}{x_i} \tag{44}$$

Equation (43) can then be written

$$\frac{t_2}{q_2} = \theta'(\sigma_{23} - \sigma_{33}), \quad \frac{t_3}{q_3} = \theta'(\sigma_{32} - \sigma_{22}), \tag{45}$$

where

$$\theta' = \frac{\theta x_2 x_3}{q_2 q_3 (s_{22}s_{33} - s_{23}^2)}$$

We now substitute for σ_{23} and σ_{33}, using the adding-up properties of compensated elasticities,

$$\begin{aligned} \sigma_{23} &= -\sigma_{22} - \sigma_{21}, \\ \sigma_{32} &= -\sigma_{33} - \sigma_{31} \end{aligned} \tag{46}$$

This gives us

$$\begin{aligned} \frac{t_2}{q_2} &= \theta'(\sigma_{21} + \sigma_{22} + \sigma_{33}), \\ \frac{t_3}{q_3} &= \theta'(\sigma_{31} + \sigma_{22} + \sigma_{33}) \end{aligned} \tag{47}$$

The interesting case to consider is where labor ($x_1 < 0$) is the untaxed good, while goods 2 and 3 are consumer goods ($x_2 > 0$, $x_3 > 0$). Then θ' has the same sign as net government revenue. For definiteness, suppose that government revenue is positive so that $\theta' > 0$. Equation (47) shows that

$$\frac{t_2}{q_2} \gtreqless \frac{t_3}{q_3} \text{ according as } \sigma_{21} \lesseqgtr \sigma_{31} \tag{48}$$

The tax rate is proportionally greater for the good with the smaller cross-elasticity of compensated demand with the price of labor. (It is possible that one commodity is subsidized, but it has to be the one with the greater cross-elasticity.)

Examples

The implications of the above model are very diverse, depending upon the nature of the demand functions. A simple example will show how the theory can be used. If we define ordinary demand elasticities by the usual formula

$$\epsilon_{ik} = q_k x_i^{-1} \frac{\partial x_i}{\partial q_k}, \tag{49}$$

we can rewrite the optimal taxation formula in the form

$$v_k = q_k^{-1} \lambda \sum p_i x_i \epsilon_{ik} \tag{50}$$

When the welfare function is individualistic, equation (5) applies, so that equation (50) may be written as

$$(51) \qquad -\alpha q_k x_k = \lambda \sum p_i x_i \epsilon_{ik}$$

or

$$q_k p_k^{-1} = -\frac{\lambda}{\alpha} \sum_i \frac{p_i x_i}{p_k x_k} \epsilon_{ik}$$

If we have a good whose price does not affect other demands (implying a unitary own price elasticity), equation (51) simplifies to yield the optimal tax of that good:

$$(52) \qquad \begin{array}{l} \text{If } \epsilon_{ik} = 0 \ (i \neq k) \quad \text{and} \quad \epsilon_{kk} = -1, \\ \text{then } q_k p_k^{-1} = \lambda \alpha^{-1} \end{array}$$

where $q_k p_k^{-1}$ equals one plus the percentage tax rate. Recalling that α is the marginal utility of income while λ reflects the change in welfare from allowing a government deficit financed from some outside source, their ratio gives a marginal cost (in terms of the numeraire good) of raising revenue. Thus the optimal tax rate on such a good gives the cost to society of raising the marginal dollar of tax.

An example of a utility function exhibiting such demand curves is the Cobb-Douglas, where only labor is supplied. As an example consider:

$$(53) \qquad u(x) = b_1 \log (x_1 + \omega_1) + \sum_{i=2}^{n} b_i \log x_i$$

If we choose labor as the untaxed numeraire, all other goods satisfy (52) and we see that the optimal tax structure is a proportional tax structure.

It is easy to exhibit examples where the optimal tax structure is not proportional. Consider the example:

$$(54) \qquad \begin{array}{c} u(x) = \sum b_i \log (x_i + \omega_i), \\ \sum b_i = 1, \ \omega_i \neq 0 \end{array}$$

The demands arising from these preferences are:

$$(55) \qquad x_i = q_i^{-1} b_i \sum q_j \omega_j - \omega_i$$

Therefore the demand elasticities are:

$$(56) \qquad \begin{array}{ll} \epsilon_{ik} = b_i \omega_k x_i^{-1} \dfrac{q_k}{q_i} & (k \neq i) \\ \epsilon_{kk} = -b_k x_k^{-1} \sum\limits_{j \neq k} \omega_j \dfrac{q_j}{q_k} & \end{array}$$

Substituting in the formula for the optimal taxes,

$$(57) \qquad \begin{aligned} -\alpha q_k x_k &= \\ &\lambda \left[\sum_{j \neq k} b_j \frac{p_j}{q_j} \omega_k q_k - b_k \frac{p_k}{q_k} \sum_{j \neq k} \omega_j q_j \right] \\ &= \lambda \sum_j \left[b_j \omega_k \frac{p_j q_k}{q_j} - b_k \omega_j \frac{p_k q_j}{q_k} \right] \end{aligned}$$

Since the assumption $\sum b_j = 1$ allows us to write the demand functions (55) in the form:

$$(58) \qquad q_k x_k = \sum_j [b_k \omega_j q_j - b_j \omega_k q_k],$$

we can deduce from (57) and (58) that

$$(59) \qquad \begin{aligned} \sum_j \Bigg[b_j \omega_k q_k \left(\frac{p_j}{q_j} - \frac{\alpha}{\lambda} \right) \\ - b_k \omega_j q_j \left(\frac{p_k}{q_k} - \frac{\alpha}{\lambda} \right) \Bigg] = 0 \end{aligned}$$

These equations allow us to calculate p for any given q, and in that way give the optimal taxation rules. In general, taxes will not be proportional. As one example of this, consider the following three-good case.

Sample Calculation

Let us combine the above two examples by considering a three-good economy (one-consumer good and two types of labor) with preferences as in (54). This example will be used to show that limited tax possibilities (represented by the same proportional tax on goods 2 and 3) intro-

duces the desirability of aggregate production inefficiency.

Example e. Assume that preferences satisfy

$$u = \log x_1 + \log (x_2 + 1) + \log (x_3 + 2) \tag{60a}$$
$$x_1 > 0, \quad x_2 > 1, \quad x_3 > -2;$$

while private production possibilities are

$$y_1 + y_2 + y_3 \leqq 0, \tag{60b}$$
$$y_1 \geqq 0, \quad y_2 \leqq 0, \quad y_3 \leqq 0;$$

and the government constraint is

$$1.02z_1 + z_2 \leqq 0 \tag{60c}$$
$$z_1 \geqq 0, \quad z_2 \leqq 0, \quad z_3 \leqq -0.1$$

Thus the government needs good 3 for public use and can produce good 1 from good 2, but only less efficiently than the private sector can.

Since we know that production efficiency is desired, we have

$$q_1 = p_1 = p_2 = 1, \qquad z_1 = z_2 = 0$$

From the first-order conditions (59) and market clearance given the demands (58), we obtain two equations to determine q_2 and q_3:

$$q_2(q_3^{-1} - 1) = 2q_3(q_2^{-1} - 1)$$
$$(q_2 + 2q_3)(q_2^{-1} + q_3^{-1} + 1) = 8.7$$

These have a unique positive solution

$$q_2 = 0.94494, \quad q_3 = 0.90008$$

which give

$$x_1 = 0.9150, \quad x_2 = -0.0316, \quad x_3 = -0.9834$$
$$u = -0.1045$$

If we now require the same tax rate on goods 2 and 3 and at the same time impose production efficiency, then $q_2 = q_3 = q$, and the tax rate is determined by the market clearance equation. We obtain

$$3q + 6 = 8.7; \qquad \text{i.e., } q = 0.9$$

Then demands are

$$x_1 = 0.9, \quad x_2 = 0, \quad x_3 = -1$$

and

$$u = -0.1054$$

Notice that the economy is still on the production frontier even though both input prices are lower in this case. If we introduce inefficiency with $p_2 > 1$, so that $y_2 = 0$ and $x_2 = z_2$, we can increase utility. Market clearance now requires

$$(q_2 + 2q_3)((1.02)^{-1}q_2^{-1} + q_3^{-1} + 1) = 8.7$$

At prices $q_2 = .92$, $q_3 = .90008$ for example, we have, $x_1 = 0.9067$, $x_2 = -0.0144$, $x_3 = -0.9926$, and $u = -0.1051$.

VII. Optimal Tax Structure—Many-Consumer Economy

As we noted in Section III of Part I, the equations for optimal taxation with a single consumer which do not reflect the particular form of V are also valid for many consumers. To pursue the analysis further, we must find an expression for V_k, the derivative of social welfare with respect to the kth consumer price.

With an individualistic welfare function, we have

$$V(q) = W(v^1(q), v^2(q), \ldots, v^H(q)) \tag{61}$$

Differentiating with respect to q_k, we obtain

$$V_k = \sum_h \frac{\partial W}{\partial u^h} v_k^h = -\sum_h \frac{\partial W}{\partial u^h} \alpha^h x_k^h \tag{62}$$

The term α^h is the marginal utility of income of consumer h. Therefore

$$\beta^h = \frac{\partial W}{\partial u^h} \alpha^h \tag{63}$$

is the increase in social welfare from a unit increase in the income of consumer h. We have

$$-V_k = \sum_h \beta^h x_k^h, \tag{64}$$

or the derivative of welfare with respect to a price equals the "welfare-weighted" net consumer demand for commodity k. The necessary condition for optimal taxation makes V_k proportional to the marginal contribution to tax revenue from raising the tax on good k.

$$\sum_h \beta^h x_k^h = \lambda \frac{\partial T}{\partial t_k}, \tag{65}$$

where $T = \sum t_i X_i$ is total tax revenue, and the derivative is evaluated at constant producer prices (i.e., on the basis of consumer excess demand functions alone). We also have the alternative formula

$$\sum_h \beta^h x_k^h = -\lambda \sum_i p_i \frac{\partial X_i}{\partial q_k} \tag{66}$$

Example f. Before turning to interpretations of the optimal tax formulae like those above, let us consider an example.

We will assume that each consumer has a Cobb-Douglas utility function,

$$u^h = b_1^h \log (x_1^h + \omega^h) + \sum_2^n b_i^h \log x_i^h, \qquad \sum_1^n b_i^h = 1 \tag{67}$$

Choosing good 1 as numeraire, we saw in Section VI that with a one-consumer economy, taxation would be proportional. This will not, in general, be true in a many-consumer economy where each consumer has this utility function. The individual demand curves arising from this utility function are:

$$\begin{aligned} x_i^h &= q_i^{-1} b_i^h q_1 \omega^h, \qquad i = 2, 3, \ldots, n \\ x_1^h &= -(1 - b_1^h)\omega^h \end{aligned} \tag{68}$$

Notice that $\partial x_i^h/\partial q_k = 0 (k \neq i \neq 1)$ and $\partial x_i^h/\partial q_i = -x_i^h/q_i$ $(i \neq 1)$.

Assuming an individualistic welfare function, the first-order conditions (66) are in this case

$$\sum_h \beta^h x_k^h = \lambda p_k q_k^{-1} \sum_h x_k^h \qquad (k = 2, \ldots, n) \tag{69}$$

This implies the following formula:

$$\frac{q_k}{p_k} = \lambda \frac{\sum_h x_k^h}{\sum_h \beta^h x_k^h} = \lambda \frac{\sum_h b_k^h \omega^h}{\sum_h \beta^h b_k^h \omega^h} \qquad (k = 2, \ldots, n) \tag{70}$$

To complete the determination of the optimal taxes, we must find the relationship between λ, p_1, and q_1. This is obtained from the Walras identity. The value of net consumer demand in producer prices is equal to minus the profit in production. (Alternatively, we could determine λ so that the government budget is balanced.) That is

$$-p_1 \sum_h (1 - b_1^h)\omega^h + \sum_{i=2}^n \sum_h p_i q_i^{-1} b_i^h q_1 \omega^h = \gamma, \tag{71}$$

where γ is the maximized profit of production net of government needs $(= \sum_{i=1}^n p_i z_i)$. Substituting from (70) and rearranging, we obtain

$$\begin{aligned} \frac{q_1}{p_1} &= \lambda \frac{\sum_h (1 - b_1^h)\omega^h + \gamma p_1^{-1}}{\sum_{i=2}^n \sum_h \beta^h b_i^h \omega^h} \\ &= \lambda \frac{\sum (1 - b_1^h)\omega^h + \gamma p_1^{-1}}{\sum \beta^h (1 - b_1^h)\omega^h} \end{aligned} \tag{72}$$

The number γp^{-1} is determined by the technology and the government expenditure decision, and therefore depends on p (unless $\gamma = 0$).

Equations (70) and (72) determine the optimal tax rates. If the social marginal utilities, β^h, are independent of taxation, the optimal tax rates can be read off at

once. This is true if W has the special form $\sum_h v^h$; for in that case $\beta^h = 1/\omega^h$. It should be noticed that, although each household's social marginal utility of income is unaffected by taxation, it is desirable to have taxation in general. If households with relatively low social marginal utility of income predominate among the purchasers of a commodity, that commodity should be relatively highly taxed. Although such taxation does nothing to bring social marginal utilities of income closer together, it does increase total welfare.

In general, taxation does affect social marginal utilities of income. The β^h depend on the tax rates, and equations (70) do not, therefore, give explicit formulae for the optimum taxes. In the case $W = -\mu^{-1} \sum_h e^{-\mu v^h}$, $\mu > 0$, so that there is a stronger bias toward equality than in the additive case, it can be verified quite easily that the optimum taxes have to satisfy

$$(73) \quad \frac{q_k}{p_k} \sum_h b_k^h (\omega^h)^{-\mu} \prod_{i=2}^{n} (b_i^h)^{-\mu b_i^h} q_i^{\mu b_i^h} = \lambda \sum_h b_k^h \omega^h \qquad (k = 2, 3, \ldots, n)$$

In this case, marginal utilities of income are brought closer together.[1] It is not immediately obvious from the equations (10) that the q are determined given the p. However, it can be shown that, in the present example, the first-order conditions must have a unique solution.[2] In fact, the relations (70) (along with (72)) would, if followed by government, certainly lead to maximum welfare if production were perfectly competitive, since any state of the economy satisfying these conditions maximizes welfare, and the maximum is unique for the welfare function considered. Unfortunately this convenient property is not general.

From equation (70) we can identify two cases where optimal taxation is proportional. If the social marginal utility of income is the same for everyone ($\beta^h = \beta$, for all h), then equation (70) reduces to $q_k p_k^{-1} = \lambda/\beta$. In this case there is no welfare gain to be achieved by redistributing income, and so no need to tax differently (on average) the expenditures of different individuals. Thus the optimal tax formula has the same form as in the one-consumer case. When the β^h do differ, taxes are greater on commodities purchased more heavily by individuals with a low social marginal utility of income. If, for example, the welfare function treats all individuals symmetrically and if there is diminishing social marginal utility with income, then there is greater taxation on goods purchased more heavily by the rich.

The second case leading to proportional taxation occurs when demand vectors are proportional for all individuals, $x^h = \rho^h x$, and thus $b_k^h = b_k$ for all h. With all individuals demanding goods in the same proportions, it is impossible to redistribute income by commodity taxation implying that the tax structure again assumes the form it has in a one-consumer economy.

Optimal Tax Formulae

The description in Section VI of some possible interpretations of the optimal tax formula carries over to the many-consumer case. Thus, as was true there con-

[1] If $\mu < 0$, utilities and marginal utilities are moved further apart.

[2] It is easily verified that $v^h = \delta_h + \sum_i b_i \log (q_1/q_i)$, where the δ_h are constants. Consequently

$$V(q) = -\mu^{-1} \sum_h e^{-\mu\delta_h} \prod_i (q_1/q_i)^{-\mu b_i^h}$$

which is a concave function of $(q_1/q_2, q_1/q_2, \ldots, q_1/q_n)$. Also, aggregate demand is

$$X_i(q) = \sum_h b_i^h \omega^h \cdot (q_1/q_i), \quad X_1(q) = -\sum_h (1 - \alpha_i^h)\omega^h$$

If the production set is convex, the set of $(q_1/q_2, \ldots, q_1/q_n)$ for which $(X_1, X_2, \ldots, X_n)$ is feasible is also convex. Thus the optimum q is obtained by maximizing a concave function of $(q_1/q_2, \ldots, q_1/q_n)$ over a convex set, and is therefore uniquely defined by the first-order conditions.

sumer price elasticities but not producer price elasticities enter the equations, and at the optimum the social marginal utility of a price change is proportional to the marginal change in tax revenue from raising that tax, calculated at constant producer prices. Analysis of the change in demand can also be carried out, but is naturally more complicated. Assuming an individualistic welfare function, the first-order conditions can be written[3]

$$\sum \beta^h x_k^h = \lambda \sum_h \sum_i t_i \frac{\partial x_i^h}{\partial q_k} + \lambda \sum_h x_k^h \tag{74}$$

From the Slutsky equation, we know that

$$\begin{aligned} \frac{\partial x_i}{\partial q_k} &= s_{ik} - x_k \frac{\partial x_i}{\partial I} = s_{ki} - x_k \frac{\partial x_i}{\partial I} \\ &= \frac{\partial x_k}{\partial q_i} - x_k \frac{\partial x_i}{\partial I} + x_i \frac{\partial x_k}{\partial I} \end{aligned} \tag{75}$$

Substituting from (75) in (74) we can write the optimal tax formula as equation (76). Rearranging terms we can write equation (76) as (77). With constant producer prices, equation (77) gives the change in demand as a result of taxation for a good with constant price-derivatives of the demand function (or for small taxes). Considering two such goods, we see that the percentage decrease in demand is greater for the good the demand for which is concentrated among:

(1) individuals with low social marginal utility of income,
(2) individuals with small decreases in taxes paid with a decrease in income,
(3) individuals for whom the product of the income derivative of demand for good k and taxes paid are large.

[3] We neglect the possibility of a free good when the first-order condition would be an inequality.

VIII. Other Taxes

Thus far we have examined the combined use of public production and commodity taxation as control variables. It is natural to reexamine the analysis when additional tax variables are included in those controlled by the government. In particular, in the next subsection we will briefly consider income taxation; but first, let us examine a general class of taxes such that the consumer budget constraint depends on consumer prices and on tax variables. We shall replace the budget constraint $\sum q_i x_i = 0$ by the more general constraint $\phi(x, q, \zeta) = 0$, where ζ represents a shift parameter to reflect the choice among different systems of additional taxation (for example, the degree of progression in the income tax). Let us note that this formulation continues to assume that all taxes are levied on consumers and that there are no profits in the economy.

The key assumption to permit an extension of the analysis above is an independence of the two constraints on the planner. We need to assume that the choice of tax variables does not affect the production

$$\sum_h \beta^h x_k^h = \lambda \sum_h \sum_i t_i \frac{\partial x_k^h}{\partial q_i} + \lambda \sum_h \sum_i t_i \left(x_i^h \frac{\partial x_k^h}{\partial I} - x_k^h \frac{\partial x_i^h}{\partial I} \right) + \lambda \sum_h x_k^h \tag{76}$$

$$\frac{\sum_h \sum_i t_i \frac{\partial x_i^h}{\partial q_i}}{\sum_h x_k^h} = \frac{1}{\lambda} \frac{\sum_h \beta^h x_k^h}{\sum_h x_k^h} - 1 + \frac{\sum_h \left(\sum_i t_i \frac{\partial x_i^h}{\partial I} \right) x_k^h}{\sum_h x_k^h} - \frac{\sum_h \left(\sum_i t_i x_i^h \right) \frac{\partial x_k^h}{\partial I}}{\sum_h x_k^h} \tag{77}$$

possibilities, and further that the choice of a production point does not affect the set of possible demand configurations. In particular, this formulation implies that producer prices do not affect consumer budget constraints. Thus the income tax, to fit this formulation, needs to be levied on the wages that consumers receive, not on the cost of wages to the firm. Similarly it is assumed that there are no sales tax deductions from the income tax base.

We know already that in such a case, optimal production is efficient. We may therefore concentrate upon the case in which all production is controlled by the government, and the production constraint is that $x_1 = g(x_2, x_3, \ldots, x_n)$. We have to choose $q_2, q_3, \ldots, q_n, \zeta$ to

$$\text{(78)} \quad \text{maximize } V(q, \zeta) \text{ subject to } X_1(q, \zeta) = g(X_2(q, \zeta), \ldots, X_n(q, \zeta))$$

As before we introduce a Lagrange multiplier λ. Differentiation with respect to q_k yields the familiar

$$\text{(79)} \qquad V_k = \lambda \sum_i p_i \frac{\partial X_i}{\partial q_k},$$

where the producer price p_i is $\partial g/\partial x_i$ $(i = 2, 3, \ldots, n)$, and $p_1 = 1$. Differentiation with respect to the new tax variable provides the similar equation

$$\text{(80)} \qquad \frac{\partial V}{\partial \zeta} = \lambda \sum_i p_i \frac{\partial X_i}{\partial \zeta}$$

We have an alternative form for (79), namely,

$$\text{(81)} \qquad V_k = -\lambda \frac{\partial T}{\partial t_k}$$

In exactly the same way, we obtain from (80) a formula involving the effect of the new tax on total tax revenue,

$$\text{(82)} \qquad V_\zeta = -\lambda \frac{\partial T}{\partial \zeta}$$

Income Taxation

Nothing that we have said suggests that commodity taxation is superior to income taxation. The analysis has only considered the best use of commodity taxation. It is natural to go on to ask how one employs both commodity taxation and income taxation. The formulation of income taxation raises a problem. If the planners are free to select any income tax structure and if there are a finite number of tax payers, the tax structure can be selected so that the marginal tax rate is zero for each taxpayer at his equilibrium income (although this does not necessarily bring the economy to the full welfare maximum). This eliminates much of our problem, but like lump sum taxation, seems to be beyond the policy tools available in a large economy. The natural formulation of this problem is for a continuum of tax payers, since then no man can have a tax schedule tailor-made for him. (This approach is taken by Mirrlees.) However, we shall here take the alternative route by assuming a limited set of alternatives for the income tax structure.

If only commodity taxation is possible, the tax paid by a household that purchases a vector x^h is

$$\text{(83)} \qquad T^h = \sum_i t_i x_i^h$$

To add income taxation to the tax structure, we can select a subset of commodities, L, e.g., labor services, and tax the value of transactions on this subset, so that

$$I^h = \sum_{i \text{ in } L} q_i x_i^h$$

where I is "taxable income." Then

$$\text{(84)} \qquad T^h = \sum_i t_i x_i^h + \tau(I^h, \zeta),$$

where τ is a fixed continuously differentiable function depending on a parameter ζ, and is the same for all consumers. With a

tax on services (x_i negative) we would expect τ to be decreasing in its tax base, with a derivative between zero and minus one. In terms of the notation employed above, we can define the budget constraint $\phi(x^h, q, \zeta)$ by

$$\begin{aligned}\phi(x^h, q, \zeta) &= \sum p_i x_i^h + T^h \\ &= \sum q_i x_i^h + \tau\left(\sum_{i \text{ in } L} q_i x_i^h, \zeta\right)\end{aligned} \tag{85}$$

Here we can regard q and ζ as the policy variables. Thus the consumer's budget constraint can be expressed in a form depending on consumer prices and independent of producer prices.

The first-order conditions for optimal income taxation are just the conditions (79) and (80), interpreted for this special case. The social marginal utility of a tax variable change is proportional to the marginal change in tax revenue calculated at constant producer prices. In the case of an individualistic welfare function, we can give more explicit formulae for the welfare derivatives, V_k and V_ζ:

$$V_k = \sum_h \beta^h x_k^h \left(1 + \delta_k \frac{\partial \tau^h}{\partial I}\right) \tag{86}$$

$$V_\zeta = \sum_h \beta^h \frac{\partial \tau^h}{\partial \zeta}, \tag{87}$$

where $\delta_k = 1$ if k is in L, 0 if k is not in L; and $\tau^h = \tau(I^h, \zeta)$.

These equations are derived from the first-order conditions for maximizing u^h subject to $\phi = 0$, noticing that, for example, the budget constraint implies that

$$\sum_k \frac{\partial \phi}{\partial x_k} \frac{\partial x_k}{\partial \zeta} + \frac{\partial \phi}{\partial \zeta} = 0$$

Combining (82) and (87), we obtain

$$\sum \beta^h \frac{\partial \tau^h}{\partial \zeta} = \lambda \frac{\partial T}{\partial \zeta} \tag{88}$$

Thus, at the optimum, for any two different kinds of change in the income tax structure, the social-marginal-utility weighted changes in taxation (consumer behavior held constant) are proportional to the changes in total tax revenue (both income and commodity tax revenue, calculated at fixed producer prices, with consumer behavior responding to the price change).

IX. Public Consumption

From the start, we have considered the government production decision as constrained by $G(z) \leqq 0$. The presence of a fixed bundle of public consumption was therefore included in the model (and would show itself by $G(0)$ being positive). This is unsatisfactory and was assumed to keep as uncluttered as possible a naturally complicated problem. We can now consider a choice among vectors of public consumption which affect social welfare directly. (We shall assume that the government controls all production, thus ignoring public expenditures which affect private production rather than consumer utility.) Let us denote by e the vector of public consumption expenditures. (Items of public consumption which are difficult to measure can be described by the inputs into their production.) The presence of public consumption alters our problem in three ways. First, public consumption represents public production (or purchases) which are not supplied to the market. Thus market clearance becomes $X = z - e$.

Second, the presence of public consumption affects private net demand, which must now be written $X(q, e)$. Third, the level of public consumption directly affects the social welfare function (by affecting individual utility in the case of an individualistic welfare function).

We can restate the basic maximization problem as

$$(89)\quad \underset{q,e}{\text{Maximize}}\ V(q, e)$$
$$\text{subject to } G(X(q, e) + e) \leqq 0$$

The presence of e in the problem will not affect the equations obtained by differentiating a Lagrangian expression with respect to q. Thus the presence of alternative bundles of public consumption does not alter the rules for the optimal tax structure. Nor would we expect it to affect the conditions which imply production efficiency at the optimum. We can therefore replace the inequality in (89) with an equality and differentiate the Lagrangian expression with respect to e_k:

$$(90)\quad \frac{\partial V}{\partial e_k} - \lambda\left[\sum G_i \frac{\partial X_i}{\partial e_k} + G_k\right] = 0$$

Since

$$(91)\quad \sum G_i \frac{\partial X_i}{\partial e_k} = \sum p_i \frac{\partial X_i}{\partial e_k} = \sum (q_i - t_i) \frac{\partial X_i}{\partial e_k} = \frac{\partial}{\partial e_k}\left(\sum q_i X_i - \sum t_i X_i\right) = -\frac{\partial}{\partial e_k}\left(\sum t_i X_i\right),$$

we can write (90) as

$$(92)\quad \frac{\partial V}{\partial e_k} = -\lambda \frac{\partial}{\partial e_k}\left(\sum t_i X_i\right) + \lambda G_k$$

Equations (92) show how the optimal level of public consumption depends on:

(i) the direct contribution of public consumption to welfare (measured by $\partial V/\partial e_k$);

(ii) the effect of public consumption on tax revenue (measured by $\partial \sum t_i X_i/\partial e_k$); and

(iii) the direct cost of public consumption (G_k).

There are three differences between this theory and that of public goods in the presence of lump sum taxation (as developed, for example, by Samuelson (1954)). Because social marginal utilities of income are not equated, the expression $\partial V/\partial e_k$ cannot be reduced to a sum of marginal rates of substitution, but depends on the weights given to the different beneficiaries of public consumption:

$$(93)\quad \frac{\partial V}{\partial e_k} = \sum_h \frac{\partial W}{\partial u^h} \frac{\partial u^h}{\partial e_k}$$

Second, the cost associated with the raising of government revenue implies that the impact of public consumption on revenue is a relevant part of the first-order conditions. Third, for the same reason, the cost of public consumption is measured in terms of the cost to the government of raising revenue to finance the expenditures (in terms of the one-consumer equation, λ may not be equal to α, the marginal utility of income).

The first-order conditions for the provision of public goods can be expressed in another way, showing the relationships between the marginal cost and "willingness to pay." Write r_k^h for the marginal rate of substitution between public good k and income for the hth household. Then $\partial u^h/\partial e_k = \alpha^h r_k^h$, where α^h is the hth household's marginal utility of income. The social marginal utility of the hth household's income, β^h, is $(\partial W/\partial u^h)\alpha^h$. Consequently, from (93)

$$(94)\quad \frac{\partial V}{\partial e_k} = \sum_h \beta^h r_k^h$$

Then, from (92)

$$(95)\quad G_k = \sum_h \left[\frac{\beta^h}{\lambda} r_k^h + \frac{\partial}{\partial e_k} \sum_i t_i x_i^h\right]$$

Thus the marginal cost of producing the public good should be equated to a sum, over all households, of the price which the household is just willing to pay for a

marginal increment in the level of provision, weighted by the marginal "social worth" of the household's income, and adjusted for the effect of the level of provision on net tax payments by the household.[4]

In the discussion of public consumption thus far it has been assumed that there were no possible fees associated with the provision of public goods. This would be appropriate for national defense or preventive medicine, but not for goods where licenses can be required from users. The optimal level of license fees will not, in general, be zero. Indeed we may be able to associate with any good more complicated pricing mechanisms than the single fixed price considered above. In particular, there are the familiar examples of two-part tariffs (a license fee for use of a facility plus a per unit charge on the amount of use), and prices depending on quantity of sales. Formally these can be treated in a fashion similar to the income taxes considered above; the set of goods over which the tax is defined is now a consumption good rather than labor. With a two-part tariff, this would imply a tax function which was not continuous at the origin.

Presumably the introduction of more general pricing and taxing schemes gives an opportunity for increasing social welfare, just as the progressive income tax gives such an opportunity. In practice, the ignored costs of tax administration may severely limit the number of complicated pricing schemes which can increase welfare. We would expect the analysis done above to be basically unchanged by the addition of these possibilities, although a two-part tariff will cause aggregate demand to have discontinuities. In practice we would expect these discontinuities to be small relative to aggregate demand, and formally, they could be eliminated by the device of a continuum of consumers.

[4] Another case can be treated in a similar manner: that of limited government production of a good, which is also being produced privately, when government production is given away rather than being sold. Since the government production rule given above does not reduce to the first-order condition in producer prices, we would not find aggregate production efficiency for the sum of these two sources of production.

X. The Optimal Taxation Theorem

In the earlier discussion, we employed calculus techniques to obtain the first-order conditions for the optimal tax structure. However, the valid use of Lagrange multipliers is subject to certain restrictions, which in the present case have no very obvious economic significance. This section provides a rigorous analysis of conditions under which the tax formulae (34) are indeed necessary conditions for optimality, and in particular provides economically meaningful assumptions that ensure their validity. The reader should be warned that the discussion is highly technical.

One might hope to provide a rigorous analysis by using the well-known Kuhn-Tucker theorem for differentiable (not necessarily concave) functions. This theorem requires a certain "constraint qualification" to be satisfied. Let us apply it and see how far we get. We wish to

$$\text{Maximize } V(q)$$
$$\text{subject to } g(X(q)) \leqq 0 \quad \text{and} \quad q \geqq 0,$$

where g is a (vector) production constraint such that $g(X) \leqq 0$ if, and only if, X is in G. Given that V, X, and g are differentiable, and that the Kuhn-Tucker constraint qualification is satisfied, we have the first-order conditions

$$(96) \quad V'(q^*) = \frac{\partial V}{\partial q} \leqq p \cdot \frac{\partial X}{\partial q} = p \cdot X'(q^*),$$

where $p = \lambda \cdot g'(X(q^*))$ for a vector of Lagrange multipliers λ, and is therefore a support or tangent hyperplane to G at $X(q^*)$. Since V and X are homogeneous

of degree zero, $[V'(q^*)-p\cdot X'(q^*)]\cdot q^*=0$: consequently $\partial V/\partial q_i=p\cdot(\partial X/\partial q_i)$ for i such that $q_i^*>0$.

To express the first-order conditions in this form, we naturally expect to assume that V and X are continuously differentiable: to that extent, the differentiability assumptions are innocuous. The assumption that the production set can be described by a finite number of continuously differentiable inequality constraints that satisfy the constraint qualification is less satisfactory. The constraint qualification is an assumption about the functions g: one can violate it by changing the functions g without changing the actual constraint set, G. Some such assumption is required to avoid not unreasonable counter-examples, as we shall see below. But it is not at all obvious how one would check whether a particular example that failed to satisfy the constraint qualification could be put right by describing G by a better behaved set of inequalities. We should like to use a constraint qualification that depends on the properties of the set G (and X) rather than the particular functions g; and we should like the assumption to be more amenable to economic interpretation. The theorem we prove below contains such an assumption, for the case where G is convex and has an interior.

Before stating the theorem let us consider an example in which the first-order conditions are not satisfied at the optimum.

Example g. Consider the one-consumer economy. In the case shown in Figure 10, the offer curve is tangent to the production frontier at the optimum production point. As q varies, the vector $X(q)$ traces out the offer curve. Thus, holding q_2 constant, the vector $\partial X(q)/\partial q_1$ is tangent to the offer curve at $X(q^*)$. Therefore if p is the vector of producer prices, which is tangent to the production frontier at $X(q^*)$, $p\cdot\partial X(q^*)/\partial q_1=0$. The same is true for the derivatives with respect to q_2. But there is no reason why $V'(q^*)$ should be zero: therefore the above first-order conditions may not be satisfied at the optimum.

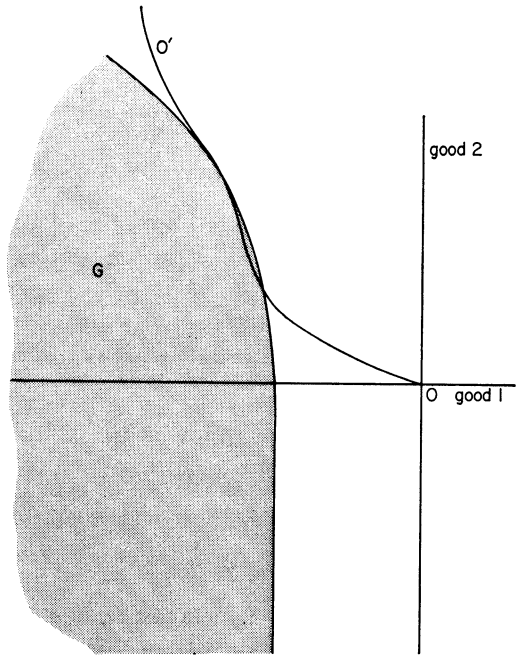

FIGURE 10

We shall make an assumption ruling out tangency between the frontier of the production set and the offer curve:

> For any p, q ($q\geqq 0$, $p\neq 0$) such that $X(q)$ is in G and $p\cdot X(q)\geqq p\cdot x$ for all x in G, $p\cdot X'(q)\geqq 0$.

The qualification takes this particular form because we also have the constraint $q\geqq 0$. Let us note that for $q>0$ the condition $p\cdot X(q)\geqq 0$ is equivalent to $p\cdot X'(q)\neq 0$, because X is homogeneous of degree zero. The qualification asserts that *for any possible competitive equilibrium (under commodity taxation) there is a consumer price change which will decrease the value of equilibrium demand, measured in producer*

prices. By the aggregate consumer budget constraint, $q \cdot X = (p+t) \cdot X = 0$. Therefore the assumption says that at any possible equilibrium point on the production frontier, it is possible to increase tax revenue. Thus the first-order conditions may not be applicable if the optimal point represents a local tax revenue maximum. Returning to example g, we see that $p \cdot X' = 0$ at the optimum, or equivalently $\partial(t \cdot X)/\partial t = 0$, although the derivatives of V are not necessarily zero there.

We now state and prove the theorem.[5]

THEOREM 5: *Assume an optimum, (X^*, q^*) exists; that $V(q)$ and $X(q)$ are continuously differentiable; and that G is convex and has a nonempty interior. Assume furthermore that there is no pair of price vectors (p, q) for which*

$$X(q) \text{ maximizes } p \cdot x \text{ for } x \text{ in } G, \quad p \neq 0, \text{ and } \quad p \cdot X'(q) \geqq 0 \tag{97}$$

Then there exists p^* such that

$$X^* \text{ maximizes } p^* \cdot x \text{ for } x \text{ in } G, \text{ and } \quad V'(q^*) \leqq p^* \cdot X'(q^*)$$

PROOF:

Let $P = \{p \mid p \cdot X^* \geqq p \cdot x, \text{ all } x \text{ in } G\}$. P is the cone of normals to G at X^*, including the zero vector. It is a nonempty, closed, convex cone.

[5] It should be noticed that when the constrained optimum is (locally) an unconstrained maximum, the producer prices satisfying the theorem are zero. This happens if optimal production is in the interior of the production set and may happen if it is on the frontier. The theorem can be weakened in a complicated manner by replacing the nontangency qualification by two conditions. One is an analog of the Kuhn-Tucker Constraint Qualification providing for the existence of an arc in the attainable set. The other use of nontangency occurs when V' is in $\overline{B}$ but not in B. If it is assumed that when there is tangency, the cone of normals is polyhedral, B will be closed. The Kuhn-Tucker theorem is then a special case of the weakened version of theorem 5 when G is the nonnegative orthant. The Kuhn-Tucker theorem is very much easier to prove, however.

We write V' for $V'(q^*)$ and X' for $X'(q^*)$. Consider the set

$$B = \{v \mid v \leqq p \cdot X', \text{ some } p \text{ in } P\}$$

We have to show that V' is in B. We do this by showing first, that if V' is in $\overline{B}$, the closure of B, in fact V' is in B; and then that V' must be in $\overline{B}$.

If V' is in $\overline{B}$, there exist sequences $\{v_n\}$ and $\{p_n\}$, p_n in P, such that

$$v_n \leqq p_n \cdot X', \qquad v_n \to V' \quad (n \to \infty) \tag{98}$$

Either $\{p_n\}$ is bounded or it is not. If not, we can find a subsequence on which

$$\|p_n\| \to \infty, \qquad \frac{p_n}{\|p_n\|} \to \bar{p} \neq 0$$

Then, dividing (98) by $\|p_n\|$ and letting $n \to \infty$ on the subsequence, we obtain $\bar{p} \cdot X' \geqq 0$ while $\bar{p}$, $\neq 0$, is in P. This possibility is excluded by assumption (97). Therefore $\{p_n\}$ is bounded, and has a limit point p, in P. Equation (98) implies that $V' \leqq p \cdot X'$. The conclusion of the theorem is thus established on the assumption that V' is in $\overline{B}$.

Suppose, on the contrary, that V' is not in $\overline{B}$. We shall derive a contradiction by a sequence of lemmas.

LEMMA 5.1:

$\overline{B}$ is pointed. That is, v and $-v$ both belong to $\overline{B}$ only if $v = 0$.

PROOF:

If v, $-v$ is in $\overline{B}$, we have sequences such that

$$v_n^1 \leqq p_n^1 \cdot X', \qquad v_n^2 \leqq p_n^2 \cdot X', \tag{99}$$

$$v_n^1 \to v, \qquad v_n^2 \to -v \tag{100}$$

If $v \neq 0$, it cannot be the case that p_n^1 and p_n^2 both tend to zero. Suppose, for example, p_n^1 does not, and take a subsequence on which

$$\|p_n^1\| \to \pi_1 \leqq \infty,$$

$$p_n^1/\|p_n^1\| \to p^1, \neq 0$$

If $p_n^1+p_n^2 \to 0$, $p_n^2/\|p_n^1\| \to -p^1$, and therefore $-p^1$ is in P. This is impossible, since, G having a nonempty interior, P is pointed. (If p, $-p$ are in P, $p\cdot x$ is constant for x in G, but a hyperplane has no interior.) We can therefore take a subsequence on which

$$\|p_n^1 + p_n^2\| \to \pi, \qquad 0 < \pi \leqq \infty,$$

$$\frac{p_n^1 + p_n^2}{\|p_n^1 +_n^2 p\|} \to p, \neq 0, \in P$$

From (99) (adding and dividing by $\|p_n^1+p_n^2\|$) and (100), we now have

$$(101) \qquad p\cdot X' \geqq \operatorname{Lim} \frac{v_n^1 + v_n^2}{\|p_n^1 + p_n^2\|} = 0$$

This contradicts (97), since p is in P and $p\neq 0$, and thereby establishes the lemma.

LEMMA 5.2: *If C is a pointed, closed, convex cone, there exists a vector p such that for all non-zero z in C, $p\cdot z<0$.*

PROOF:

By the duality theorem for convex cones $C^{++}=C$, where C^+ is the dual cone, $\{p \mid p\cdot z \leqq 0,\ z \text{ is in } C\}$. Clearly, if C^+ is pointed, C has a nonempty interior: for if interior C is empty, $p\cdot z=0$ for some nonzero p and all z in C, and then p and $-p$ both belong to C^+. Under the assumptions of the theorem, C is closed and pointed. Therefore C^{++} is pointed, and C^+ has an interior point p.

$$p\cdot z < 0 \quad (\text{all nonzero } z \text{ in } C)$$

Otherwise, if $p\cdot z=0$, we can easily find a sequence $\{p_n\}$ on which $p_n \to p$ and $p_n\cdot z>0$, so that p_n is not in C^+.

LEMMA 5.3: *If V' is not in $\overline{B}$, there exists r such that*

$$(102) \qquad V'\cdot r > 0$$

$$(103) \qquad v\cdot r < 0 \qquad (v \in B)$$

PROOF:

The closed convex cone $\overline{B}+\{\lambda V' \mid \lambda \leqq 0\}$ is pointed. Thus there exists an r such that

$$v\cdot r + \lambda V'\cdot r < 0$$
$$(v \in \overline{B},\ \lambda \leqq 0,\ v, \lambda \text{ not both zero})$$

Putting $v=0$ and $\lambda=-1$ we obtain (102); putting $\lambda=0$ we obtain (103).

LEMMA 5.4: *Let r be a vector satisfying (102) and (103). For some $\delta>0$,*

$$(104) \quad X(q^* + \theta r) \in G \quad (0 \leqq \theta \leqq \delta)$$

PROOF:

Assume not. Then for some sequence $\{\theta_n\}$, $\theta_n>0$, $\theta_n \to 0$,

$$X(q^* + \theta_n r) \notin G$$

Since G is convex, this implies that

$$X(q^*) + \frac{\lambda}{\theta_n}[X(q^* + \theta_n r) - X(q^*)] \notin G$$

for $\lambda \geqq \theta_n$. Letting $n\to\infty$, we deduce, for any $\lambda>0$, that

$$X(q^*) + \lambda X'\cdot r$$
$$= \operatorname*{Lim}_{n\to\infty}\left[X(q^*) + \lambda\frac{X(q^* + \theta_n r) - X(q^*)}{\theta_n}\right]$$

is not in the interior of G. It follows that the half-line $\{X(q^*)+\lambda X'\cdot r \mid \lambda>0\}$ can be separated from the interior of G by a hyperplane with normal $p\neq 0$:

$$p\cdot X(q^*) + \lambda p\cdot X'\cdot r \geqq p\cdot x$$
$$(\lambda > 0,\ x \in \operatorname{Int} G)$$

Letting $\lambda\to 0$ we have $p\in P$. Letting $x\to X^*$ we have

$$p\cdot X'\cdot r \geqq 0,$$

which contradicts (103) since $p\cdot X'$ is in B. The lemma is proved.

Since q^* is optimal, (104) implies that

$$V(q^* + \theta r) \leqq V(q^*) \qquad (0 \leqq \theta \leqq \delta)$$

Therefore,

$$V' \cdot r = \lim_{\theta \to 0} \frac{1}{\theta} [V(q^* + \theta r) - V(q^*)] \leqq 0$$

This, however, contradicts (102). The hypothesis of Lemma 5.3, that $V' \in \overline{B}$, is therefore false. The proof of the theorem is thus complete.

In reaching our results that the first-order conditions for optimum taxes (96) hold in general, we have assumed that the production set, G, is convex. But one common argument for government control of production is nonconvexity of the production set. This is not a question we are primarily concerned with in this paper. However, some extensions of the theorem do hold. As an example, assume the frontier of G is differentiable at X^*, so that p can be uniquely defined as the normal at X^* and that G is not thin in the neighborhood of X^*—i.e., there exists a ball with center on the normal through X^*, contained in G and containing X^*. Applying the theorem to this ball we get the validity of the first-order conditions (96) using the producer prices defined by the normal.

As in general welfare economics, two uniqueness problems may arise when considering the application of the first-order conditions to achieve an optimum. In the first place, there may be more than one pair of price vectors, (p, q), that satisfy the first-order conditions and allow markets to be cleared. This is similar to the problem that arises when we attempt to define optimum production and distribution by first-order conditions in the presence of a non-convex production set. It is noteworthy that, if lump sum transfers are excluded as a feasible policy, this problem may arise even when the production set is convex. There is no reason why the demand functions should have any of the nice convexity properties which ensure that first-order conditions imply global maximization. Only in particular cases, such as that discussed in footnote 2 above (where rigorous argument is possible without appeal to theorem 5), will the first-order conditions lead to a unique solution.

The second problem is that the tax policies one might like to employ may not uniquely determine the behavior of the system. The lump sum redistribution of wealth required in standard welfare economics does not carry with it any guarantee that the desired competitive equilibrium is the unique one consistent with the optimal wealth distribution (although if the wrong equilibrium is achieved, this should be easily noticed). Similarly, in the present case, if we employ taxes rather than consumer prices as the government control variables, the equilibrium of the economy may not be unique.[6] But if consumer prices are used as the control variables—and why not?—the demand functions give us a unique equilibrium position, so long as preferences are strictly convex.

XI. Concluding Remarks

Welfare economics has usually been concerned with characterizing the best of attainable worlds, accepting only the basic technological constraints. As economists have been aware, the omitted constraints on communication, calculation, and administration of an economy (not to mention political constraints) limit the direct applicability of the implications of this theory to policy problems, although great insight into these problems has certainly been acquired. We have not at-

[6] For a discussion of multiple equilibria in a related problem, see E. Foster and H. Sonnenschein.

tempted to come directly to grips with the problem of incorporating these complications into economic theory. Instead, we have explored the implications of viewing these constraints as limits on the set of policy tools that can be applied. There are many sets of policy tools which might be examined in this way. Specifically, we have assumed that the policy tools available to the government include commodity taxation (and subsidization) to any extent. For these tools we have derived the rules for optimal tax policy and have shown the desirability of aggregate production efficiency, in the presence of optimal taxation. We have also considered expansion of the set of policy tools in such a way that we continue to have the condition that production decisions do not change the class of possible budget constraints. For example, this condition is still preserved when one includes poll taxes, progressive income taxation, regional differences in taxation, taxation on transactions between consumers, and most kinds of rationing. This type of expansion of the set of policy tools does not alter the desirability of production efficiency, nor does it alter the conditions for the optimal commodity tax structure, although in general the tax rates themselves will change. We have, unfortunately, ignored the cost of administering taxes. Presumably optimization by means of sets of policy tools that do not, because the cost of administration, include the full scope of commodity taxation, will not lead to the same conclusions.

Let us briefly consider the type of policy implications that are raised by our analysis. In the context of a planned economy our analysis implies the desirability of using a single price vector in all production decisions, although these prices will, in general, differ from the prices at which commodities are sold to consumers.

As an application of this analysis to a mixed economy, let us briefly examine the discussion of a proper criterion for public investment decisions. As has been widely noted, there are considerable differences in western economies between the intertemporal marginal rates of transformation and substitution. This has been the basis of analyses leading to investment criteria which would imply aggregate production inefficiency because they employ an interest rate for determining the margins of public production which differs from the private marginal rate of transformation. One argument used against these criteria is that the government, recognizing the divergence between rates of transformation and substitution, should use its power to achieve the full Pareto optimum, bringing these rates into equality. When this is done, the single interest rate then existing will be the appropriate rate to use in public investment decisions. We begin by presuming that the government does not have the power to achieve any Pareto optimum that it chooses. Then from the maximization of a social welfare function, we argued that the government will, in general, prefer one of the non-Pareto optima to the Pareto optima, if any, that can be achieved. At the constrained optimum, which is the social welfare function maximizing position of the economy for the available policy tools, we saw that the economy will still be characterized by a divergence between marginal rates of substitution and transformation, not just intertemporally, but also elsewhere, e.g., in the choice between leisure and goods. However, we concluded that in this situation we desired aggregate production efficiency. This implies the use of interest rates for public investment decisions which equate public and private marginal rates of transformation.

We have obtained the first-order conditions for public production, but we have not considered the correct method of evaluating indivisible investments. This

is one problem that deserves examination. In examining the optimal tax structure, we have briefly considered the tax rates implied by particular utility functions. This analysis should be extended to more general and more interesting sets of consumers. Further, we have not examined in any detail the uniqueness and stability of equilibrium, that is, the question whether there are means of achieving in practice an equilibrium which is close to the optimum.

Finally, we would like to emphasize the assumptions which seem to us most seriously to limit the applications of this theory.[7] We have assumed no costs of tax administration and no tax evasion. And we have assumed constant-returns-to-scale and price-taking, profit-maximizing behavior in private production. Pure profits (or losses) associated with the violation of these assumptions imply that private production decisions directly influence social welfare by affecting household incomes. In such a case, it would presumably be desirable to add a profits tax to the set of policy instruments. Nevertheless, aggregate production efficiency would no longer be desirable in general; although it may be possible to get close to the optimum with efficient production if pure profits are small. We hope, nevertheless, that the methods and results of this paper have shown that economic analysis need not depend on the simplifying, but unrealistic, assumption that the perfect capital levy has taken place.[8]

[7] These assumptions are viewed in the context of equilibrium theory. There is no need here to go into the limitations inherent in current equilibrium theory.

[8] A recent paper by Clarence Morrison also deals with marginal cost pricing as a special case of optimal pricing.

REFERENCES

W. J. Corlett and D. C. Hague, "Complementarity and the Excess Burden of Taxation," *Rev. Econ. Stud.*, 1953, *21*, No. 1, 21–30.

E. Foster and H. Sonnenschein, "Price Distortion and Economic Welfare," *Econometrica*, Mar. 1970, *38*, 281–97.

H. Kuhn and A. Tucker, "Nonlinear Programming," in J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley 1951.

J. A. Mirrlees, "An Exploration in the Theory of Optimum Income Taxation," *Rev. Econ. Stud.*, Apr. 1971, *38*, forthcoming.

C. C. Morrison, "Marginal Cost Pricing and the Theory of Second Best," *Western Econ. J.*, June 1969, *7*, 145–52.

P. A. Samuelson, "Memorandum for U.S. Treasury, 1951," unpublished.

———, "The Pure Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.