Journal of Public Economics 6 (1976) 55–75. © North-Holland Publishing Company

THE DESIGN OF TAX STRUCTURE: DIRECT VERSUS INDIRECT TAXATION*

A.B. ATKINSON

University of Essex, Wivenhoe Park, Colchester, England

J.E. STIGLITZ

Stanford University, Stanford, CA 94305, U.S.A.

Revised version received February 1976

1. Introduction

The recent literature on optimal taxation may be seen as attempting to clarify the structure of the arguments advanced to support changes in the tax system, tracing the implications of taxes and quantifying (analytically) the trade-offs between the various objectives of tax policy. This literature has examined the optimal structure for particular types of taxation taken in isolation, such as the optimal rates of excise tax and the optimal income tax schedule. Our purpose, on the other hand, is to provide a broader framework and to consider the interaction between different kinds of taxation. To illustrate this, we reexamine the age-old question of direct versus indirect taxation and the relationship of these taxes to the goals of efficiency, vertical equity and horizontal equity.

After describing in section 2 the general framework of the analysis, and arguing that any treatment of the choice of tax structures must be centrally concerned with distributional considerations, we begin in section 3 with the extension of the classic Ramsey formula for optimal excise taxation to include vertical equity objectives. This was considered by Diamond and Mirrlees (1971), but the results

*This is a revised and condensed version of the paper given at the ISPE meeting under the title 'Alternative approaches to the distribution of income.' This in turn was based on part II of 'The structure of indirect taxation,' Cowles Foundation, 1970 [part I appeared as Atkinson and Stiglitz (1972)] and on the draft of chapter 15 of *Lectures on Public Economics*, University of Essex, 1971. Parts of the paper have been presented by the first author at seminars at the universities of Essex, Harvard and Namur, and by the second author at Chicago, National Bureau of Economic Research–West and Stanford, and they are grateful to participants in these seminars for their helpful comments. This work was supported in part by National Science Foundation Grant SOC74-22182 at the Institute for Mathematical Studies in the Social Sciences at Stanford University, and in part by the Guggenheim and Ford Foundations.

given here are in a rather different form.[1] The rest of the paper is concerned with the case where the government can employ both income and excise taxes. In section 4 it is shown that the existence of an optimal linear income tax may lead to quite different results. Section 5 introduces the possibility of a general non-linear income tax, and argues that under a relatively wide class of conditions – separability between leisure and consumption – the optimal tax system can rely solely on income taxation. This brings out clearly the importance of considering simultaneously the whole range of tax instruments open to the government. Finally, section 6 examines the relationship between vertical and horizontal equity, and the implications of differences in tastes.

2. The basic framework for taxation

The general problem of taxation of individuals may be posed as follows. There are a large number of people in any economy who differ with respect to a number of characteristics, in particular their endowments and tastes. On the basis of certain ethical premises, it is decided that individuals with different characteristics should pay varying amounts of tax. If we could observe these characteristics costlessly and perfectly, that would be the end of the analysis: we would simply impose a lump sum tax on individuals, with the amount differing according to their characteristics. The theory of optimal taxation would then be concerned simply with deriving, on the basis of the specified ethical premises, what the functional relationship between characteristics and taxes 'ought to be.'[2]

It is the difficulties associated with observing characteristics which make the theory of taxation an interesting and difficult problem. The theory may be seen as being concerned with the choice of certain easily observable characteristics which are related systematically to the unobservable characteristics in which we are really interested. It is thus part of what has come to be called the 'theory of screening.' The use of these surrogate characteristics gives rise to a number of problems similar to those discussed in the screening literature [see, for example, Spence (1973) and Stiglitz (1975)].

(1) Many of the characteristics which may be used for screening are, at least to some extent, under the control of the individual, and basing a tax on these is inevitably distortionary.

(2) Almost all characteristics which may be used for screening are imperfect; that is, the surrogate characteristics employed to determine tax liability are not perfectly correlated with the characteristics with which we are really concerned.

[1] This section includes the distributional results referred to in our earlier paper [Atkinson and Stiglitz (1972)].

[2] Another potentially important function of the tax system – to provide signals concerning the demand for public goods – is not discussed here.

(3) There are costs (e.g. of administration) associated with even nondistortionary screening systems.

This general view of taxation shows that the analysis of tax systems must be inherently concerned with individual differences. As a consequence, the treatment of, say, optimal excise taxation in a world where individuals are assumed to be identical is at best of limited relevance. In what follows we assume that people differ with respect to their abilities (earning power) and their tastes, although for the main part of the paper (sections 3–5) we concentrate on differences in ability. For simplicity, we assume that this can be measured by a single parameter, n, so that an individual of ability n_1 can do in $1/n_1$ hours what an individual of ability n_2 can perform in $1/n_2$ hours. We assume, however, that ability is not observable directly. What one can identify depends on the nature of the employment relationship. The following are three of the most important possibilities, where L is the number of labor hours worked, e is the level of effort, and income is given by $Y = neL$.

(i) Income is observable, but effort and labor time are not. This makes sense for unincorporated businesses, although not necessarily for employees.

(ii) Wages per hour ($w \equiv ne$) are observable, but not labor hours and hence not income. This applies where individuals may have several jobs and it may be difficult to keep track of them. It should be noted that where effort is unobservable, one cannot infer ability, even when one can observe the wage rate.

(iii) Both wages and hours are observable, but since effort is not, ability cannot be inferred.

Case (i) corresponds to that where income taxation is employed (Y is the surrogate characteristic), case (ii) to that where there is a wage tax (w is the surrogate characteristic), and in case (iii) there is a choice of screening devices.

In addition to income and wages, other economic variables on which taxation might be based are purchases by different individuals of different commodities. In a world where income and wages are unobservable, but purchases of certain luxuries are observable, the latter may provide the best screening device. Whether such purchases remain good screening devices when income and wages *are* observable is one of the questions to which we address ourselves in this paper. Still other economic variables that may be useful as screening devices are the sources of income: e.g. the government could distinguish between salaried and wage workers, between earned and unearned income, or, within unearned income, between dividends and capital gains. For the purposes of this paper, however, we consider only labor income and do not distinguish among types of jobs, except in terms of the wages they pay. There are certain other distinctions, such as the sex, age, and marital status of the worker, which are relatively costless to observe. An argument can be made for differentiation on this basis [see Boskin (1973)], but again, for present purposes, we ignore these distinctions.

Thus, if x_i are the individual's purchases of commodity i, we can describe a general tax system as a relationship between potentially observable characteristics, x_i, Y and w, and his tax payments:

$$T = T(x, Y, w).$$

In practice, almost all tax systems possess a high degree of separability, and indeed are often linear in some or all of the arguments. There are good reasons why this is so. Not only are there greater costs of calculating tax liabilities when nonseparable and nonlinear tax systems are employed, but also there are significantly higher costs of record-keeping and enforcement (with linear commodity taxes, for example, no record of the number of units purchased by a given person need be kept). Thus although separability and linearity have great analytical advantages, and will be assumed in much of what follows, there are also strong economic grounds for making these assumptions.

Within this framework, we can consider the following taxes.

Excise tax: $T = \sum_i t_i x_i = t \cdot x$,

where $t \cdot x$ denotes the inner product of the two vectors. In the simplest case the tax rates t_i are constant, but in certain situations (e.g. housing subsidies) the tax may be nonproportional. Taxes may also be income-related, $t_i(x_i, Y)$, or wage-related $t_i(x_i, w)$, the latter applying, for example, to job-related subsidies.

Income tax: $T = T(Y)$.

In certain cases the tax base may depend on the consumption of commodities (e.g. medical care), so that $T = T(Y, x_i)$; it may be constrained to be linear (constant marginal tax rate) or allowed to vary freely.

Wage tax: $T = \tau(w)L$.

Again the tax schedule may be constrained to be linear. (The problem of the optimal wage structure in a socialist country may be viewed as determining the function τ.)

Thus the theory of optimal taxation must be concerned with the choice of tax base as well as the structure of taxes imposed. A full analysis would, of course, begin with the general function $T(x, Y, w)$ and examine its properties. The difficulty with such a completely general approach is that it does not appear – at least at this juncture – to lead to any simple or clear prescriptions. In this paper we attempt a less ambitious task and focus primarily on the relationship between excise and income taxation. This piecemeal approach has obvious limitations, but we hope that it is sufficient to demonstrate the importance of a unified treatment of the choice of tax base and the optimal design of tax rates. As a preliminary to this, we review in the next section the main results regarding excise taxes viewed in isolation; then, in sections 4–6, we examine the interaction with income tax.

3. Excise taxes and distribution

The optimal structure of indirect taxation, and particularly whether there should be differential rates of tax, is an old issue which has recently be re-examined in a series of papers. Much of this literature has ignored differences in endowments and has concentrated on efficiency aspects. At the same time, it has been recognized that the policy prescriptions would need to be modified when distributional considerations were introduced. This aspect of the problem was first discussed by Diamond and Mirrlees (1971); their treatment was, however, somewhat different from that given below.

We assume that there are N individuals, denoted by a superscript h. Each individual has a well-behaved utility function defined over the n commodities and labor,[3]

$$U^h = U^h(x, L). \tag{1}$$

The individual maximizes utility subject to the budget constraint

$$q \cdot x = w^h L^h, \tag{2}$$

where q is the price of the commodity to the consumer, and w^h is his after-tax wage. The solution leads to individual demand and labor supply functions. Substituting these back into the utility function gives the indirect utility function $V^h(q, w^h)$. There is no loss of generality (with the assumptions made below) in letting labor be the numeraire and in assuming it to be untaxed (a proportional tax on labor income is simply equivalent to a uniform commodity tax). This will be done throughout the analysis. Finally, we denote by X_i the total demand for good i summed over all individuals $(\Sigma_h x_i^h)$.

At this stage it is assumed that the only taxes open to the government are proportional excise taxes at the rate t_i on commodity i, and that no lump-sum taxes or subsidies are allowed.[4] For simplicity, we take producer prices as fixed and normalize them at unity, so that $q_i = 1 + t_i$. We assume that the govern-

[3]The labor variable may be treated more generally as a vector, including elements such as hours, effort, etc.

[4]Such a restriction makes sense in the context of the general approach taken in this paper only if 'individuals' are not directly observable as individuals: e.g., with a lump-sum subsidy, they could collect twice under different 'names' or with a lump-sum tax they disappear into the bush.

ment wishes to raise a given amount of revenue,

$$R \equiv \sum_h t \cdot x^h \geqq \bar{R}, \tag{3}$$

and that subject to this constraint it aims to maximize a social welfare function of the Bergson form $G(U^1, \ldots, U^N)$, where G is increasing in all its arguments. Forming the Lagrangian

$$\mathscr{L} = G(V^h) + \lambda\left[\sum_h t \cdot x^h - \bar{R}\right], \tag{4}$$

straightforward manipulation yields the result that the first-order conditions imply[5]

$$\frac{\sum_h \left[\sum_k t_k (S_{ik}^h)\right]}{X_i} = -\left[1 - \sum_h b^h \left(\frac{x_i^h}{X_i}\right)\right], \qquad i = 1, \ldots, n, \tag{5}$$

where

$$S_{ik}^h = \left(\frac{\partial x_i^h}{\partial p_k}\right)_{\bar{U}},$$

the compensated price derivative;

$$b^h = \frac{\beta^h}{\lambda} + \frac{\partial R}{\partial I^h},$$

the *net* social marginal utility of income for household h, using government income as numeraire;

$$\beta^h = \frac{\partial G}{\partial V^h} \frac{\partial V^h}{\partial I^h},$$

the gross social marginal utility of income (consumption) accruing to household h; and

$$\frac{\partial R}{\partial I^h} = \sum_k t_k \frac{\partial x_k^h}{\partial I^h},$$

the marginal tax paid by household h on receiving an extra dollar of income.

[5]This by making use of the fact that $\partial V/\partial q_i = -x_i^h \alpha^h$, where α^h is the private marginal utility of income of individual h, and of the Slutsky equation

$$\frac{\partial x_k}{\partial q_i} = S_{ki} - x_i \frac{\partial x_k}{\partial I},$$

where $\partial x_k/\partial I$ is the derivative with respect to income (evaluated at $I = 0$ in this case) and S_{ki} is the compensated price term.

In interpreting b^h, note that there are two effects of transferring a dollar to the hth household: the direct effect, which is just β^h/λ measured in government revenue, plus an indirect effect – the effect of the transfer on government income. It may also be noted that the mean ($\bar{b}$) is the net value of giving an equal lump-sum payment to everyone. Thus, if uniform lump-sum payments or taxes were allowed, the government would set them at a level such that $\bar{b} = 1$. The implications of this are explored in the next section.

The left-hand side of (5) has the usual interpretation of the proportional reduction of the consumption of the ith commodity along the compensated demand schedules. We can immediately see that this is no longer necessarily the same for all commodities. Sufficient conditions for it to be independent of i are *either* that b^h be the same for all h *or* that x_i^h/X_i be the same for all commodities (there are no goods which are consumed disproportionately by rich or poor). In general, where these are not satisfied, the compensated reduction in demand with the optimal tax structure is smaller:[6] (1) the more the good is consumed by individuals with a high net social marginal utility of income, (2) the more the good is consumed by households with a high marginal propensity to consume taxed goods.

Eq. (5) can be rewritten in two ways which will prove useful in the subsequent discussion:

$$\sum_h \sum_k t_k S_{ik}^h = -X_i(1-\bar{b}r_i), \qquad i = 1, \ldots, n, \tag{5'}$$

where

$$r_i = \sum_h \left(\frac{x_i^h}{X_i}\right)\left(\frac{b^h}{\bar{b}}\right); \tag{6}$$

and

$$\sum_h \sum_k t_k S_{ik}^h = -X_i[(1-\bar{b})-\bar{b}\phi_i], \qquad i = 1, \ldots, n, \tag{5''}$$

where $\phi_i \equiv r_i - 1$ is the normalized covariance between the consumption of the ith commodity and the net social marginal utility of income [a result derived independently by Diamond (1975)]. In the first of these formulae, r_i is a generalization of the 'distributional characteristic' of Feldstein (1972a) and (1972b). It shows that if $\bar{b}$ is large, i.e. if there would be large gains from a uniform lump-sum payment, then distributional considerations are to be weighted more heavily.

[6]Diamond and Mirrlees (1971) derived the analogous expression for the uncompensated changes. Since the uncompensated reductions in demand with the optimal tax structure are not the same even without distributional considerations, to make comparisons with the Ramsey results more direct, we have employed compensated derivatives. In the uncompensated form, Diamond and Mirrlees have identified a third factor determining the percentage reduction in demand: it will be greater the more the demand for the commodity is concentrated among individuals for whom the product of the income derivative of demand for that good and total taxes paid is large.

The extension of the Ramsey formula given above is relatively general. In particular, it allows individuals to differ with respect to both tastes and endowments; other taxes (e.g. a lump-sum tax) may be imposed; and not all commodities need be taxed. (As in the earlier Ramsey analysis, the result does, however, depend on there being either constant returns to scale in production or 100 per cent profits taxes – see Stiglitz and Dasgupta (1971).) However, to obtain detailed results on the optimal tax structure, we need to make more specific assumptions about the nature of differences between individuals and the form of the utility function. Here, and until section 6, we assume that everyone has the same tastes, that effort is not a variable, and that individuals differ solely with respect to their ability (wage rate). For ease of analysis, we assume a continuum of individuals, and replace the summation signs in the previously derived formulae by integrals. We let F represent the distribution function of abilities, where we normalize such that $F(\infty) = 1$. The special case of the utility function we consider for purposes of illustration is that where all individuals have independent compensated demand schedules. Eqs. (5′) and (5″) then give

$$\frac{t_i}{1+t_i} = \frac{1-\bar{b}r_i}{\bar{\varepsilon}_i} = \frac{(1-\bar{b})-\bar{b}\phi_i}{\bar{\varepsilon}_i}, \tag{7}$$

where $\bar{\varepsilon}_i$ is the weighted average compensated price elasticity, the weights being the consumption of the different individuals.[7]

In the case where everyone is identical, (7) reduces to the familiar formula that taxes should be inversely proportional to demand elasticities. Eq. (7) provides a simple adjustment to this formula for distributional considerations. The value of r_i depends now solely on the social marginal valuation of income received by different households and on the proportion of total consumption which goes to them. In particular, it depends on the degree of aversion to inequality. If β is constant, i.e. society is indifferent with regard to the distribution, then the optimal tax formula is the familiar one. But if the social marginal valuation of income falls with w, this tends to increase the tax rate on goods which are primarily consumed by those at the top of the scale.[8]

[7]The first-order conditions need careful interpretation since they may not lead to a unique solution. Where the price elasticity varies with q_i there may be multiple solutions, and the optimal tax structure may involve taxing at different rates two goods with identical demand curves.

[8]That is, letting r_i be a function of ρ, some measure of inequality aversion with $\rho = 0$ corresponding to no inequality aversion, then $r_i(0) = 1$, for all i, and

$$r_i(\rho)-r_i(0) = \frac{\Sigma\,(x_i^h-\bar{x}_i)\,(b^h-\bar{b})}{\bar{b}X_i} \gtrless 0 \quad \text{as} \quad \frac{\partial x^h}{\partial b^h} \gtrless 0\,,$$

i.e. households which consume more of x_i (relative to mean consumption $\bar{x}_i$) have a higher or lower valued net marginal social utility of income. (For the meaning of inequality aversion, see Atkinson (1970) and Diamond–Stiglitz (1974).) Because of our normalization, $\bar{x}_i = X_i$.

A formula similar to (7) was given by Feldstein (1972a,b), but he did not bring out the inherent conflict between equity and efficiency considerations. With an additively separable utility function and constant marginal utility of leisure, demands depend on the ratio of commodity price to wage. This means that a commodity with a low elasticity of demand appears from an efficiency standpoint to be a good candidate for taxation, but that since the consumption of such a commodity rises only slowly with w, this points to low tax rates for equity reasons. Which of these factors will predominate depends on the form of the social welfare function and on the shape of the distribution of abilities.

One especially simple case to examine is that where the government maximizes the sum of utilities – the classical utilitarian case – and where the compensated demand curves have constant elasticity. In table 1, we present the value of ϕ_i and the associated form of eq. (7) for the Pareto and lognormal distributions. For the Pareto distribution, it follows that where the government would like to make a uniform lump-sum transfer to everyone ($\bar{b} > 1$), the tax

Table 1
Values of distributional characteristics: Pareto and lognormal distributions.

(a) Pareto distribution: $f = \delta \tilde{w}^{\delta} w^{-(1+\delta)}$ (where it is required that $\delta > \varepsilon_i$):

$$\phi_i = \frac{-\varepsilon_i}{\delta(1+\delta-\varepsilon_i)}, \qquad \frac{t_i}{1+t_i} = -\frac{(\bar{b}-1)}{\varepsilon_i} + \frac{\bar{b}}{\delta(1+\delta-\varepsilon_i)}.$$

(b) Lognormal distribution (where $(e^{\sigma^2}-1)^{1/2}$ is the coefficient of variation):

$$\phi_i = e^{-\varepsilon_i \sigma^2} - 1, \qquad \frac{t_i}{1+t_i} = -\frac{(\bar{b}-1)}{\varepsilon_i} + \frac{\bar{b}(1-e^{-\varepsilon_i \sigma^2})}{\varepsilon_i}.$$

rate rises with the elasticity of demand; this is therefore a sufficient condition for equity to outweigh efficiency considerations and for goods with a high price elasticity to be taxed more heavily. It may also be noted that the magnitude of the distributional term falls with δ, or as the distribution of abilities becomes less unequal [for the same mean, see Chipman (1974)]. For the lognormal distribution, if $\bar{b} > 1$ and σ is small, then again the distributional considerations dominate; but if σ is not small, then as the elasticity of demand increases, the tax rate may at first increase (for low elasticities, distributional considerations are more important) and then decrease (for high elasticities, efficiency dominates).[9]

4. Excise taxes with an optimal linear income tax

Thus far we have considered indirect taxation in isolation from the rest of the tax system, and in particular we have not examined how the possibility of

[9]This may be seen by expanding the term exp $(-\varepsilon_i\sigma^2)$ and first considering terms of order σ^2, and then of order σ^4 (it is assumed that $\sigma^2\varepsilon_i < 1$).

employing *direct* taxes affects the optimal structure of *indirect* taxation. How does the existence of a progressive income tax affect the balance between equity and efficiency considerations in determining the optimal rates of excise taxation?

A first step towards considering the interaction between direct and indirect taxation may be taken by a relatively straightforward modification of the analysis of the previous section. The simplest progressive income tax is that where there is an exemption level and a proportional rate of tax both above and below this level (the tax below the exemption level being a negative income tax, so that the taxpayer receives a supplement from the revenue). Such a linear income tax schedule can readily be incorporated into the model we have been discussing, since wages are the only source of income and a uniform tax on all commodities is equivalent to a proportional tax on wages. The only difference therefore is in the exemption level, which can be introduced by supposing that the government provides a lump-sum payment identical in amount (E) to all individuals (if E is negative, it is a lump-sum tax). We assume an additive, symmetric, social welfare function and write the Lagrangian

$$\mathscr{L} = \int_0^\infty [G\{V(t, E)\} + \lambda\{t\cdot x - E - \bar{R}\}]\mathrm{d}F. \tag{8}$$

The indirect utility function now depends on E, where $\partial V/\partial E = \alpha$, the marginal utility of income. The first-order conditions give:

$$\frac{\partial \mathscr{L}}{\partial t_i} = \int_0^\infty \left[(\lambda - G'\alpha)x_i + \lambda \sum_k t_k \frac{\partial x_k}{\partial t_i}\right]\mathrm{d}F = 0, \qquad i = 1, \ldots, n, \tag{9}$$

$$-\frac{\partial \mathscr{L}}{\partial E} = \int_0^\infty \left[(\lambda - G'\alpha) - \lambda \sum_k t_k \frac{\partial x_k}{\partial I}\right]\mathrm{d}F = 0. \tag{10}$$

Since $\beta = G'\alpha$, (10) is equivalent to $\bar{b} = 1$, as the previous section indicated. Thus, with an optimal linear income tax, the percentage reduction of consumption along the compensated demand schedule is simply *equal* to the normalized covariance between consumption of the commodity and the net marginal social utility of income (eq. (5″) with $\bar{b} = 1$).

If β were constant, that is, if society were indifferent regarding the distribution, then $\{t_i = 0, \text{ all } i\}$ would provide a solution to the first-order conditions, and if there were a positive revenue requirement, it would all be raised by a poll tax ($E < 0$). This is a quite intuitive result, since we should expect that efficiency considerations taken on their own would dictate using solely a lump-sum tax. Where the government is concerned with the distribution of income, i.e. β is a decreasing function of w, then indirect taxes would in general be employed. The question, however, is whether they would be employed with differential rates, since, as we have seen, a uniform indirect tax is equivalent – in this model – to a proportional income tax.

The point at issue may be illustrated by one very special example. Suppose that the utility function is quadratic (an example used by Ramsey), that the cross-terms are zero, and that the marginal utility of leisure is constant:

$$U = \sum_i \left(a_i x_i - \frac{c_i}{2} x_i^2 \right) - vL \,. \tag{11}$$

In the absence of the income tax, it may be shown that the optimal tax rates vary according to $a_i(1-\bar{b})$, and would in general differ across commodities. However, the introduction of an income tax with the exemption level E means that $\bar{b} = 1$, and that the optimal tax structure is uniform. It follows that no indirect taxation need be employed, and that the optimum may be achieved simply through a linear income tax. (Another example is the linear expenditure system.)

Where the utility function is more general, but the compensated demands are still independent, we can see from eq. (7) that $t_i/(1+t_i) = -\phi_i/\bar{\varepsilon}_i$. We may note two features of this result. Firstly, it implies that there is no case for subsidizing normal goods; an increase in the lump-sum subsidy is always superior. Secondly, the tax rates depend on the level of revenue to be raised only through the dependence on the covariance of x_i with net marginal social utility of income. With a constant marginal utility of leisure and $G' = 1$, ϕ_i is independent of the level of revenue to be raised – any increase in $\bar{R}$ is met by a reduction in E. Hence for sufficiently large $\bar{R}$, the tax system is regressive.

From table 1 we can derive the optimal tax rates in the constant elasticity case. For the Pareto distribution, the tax is higher on goods with a higher price elasticity (which is also the elasticity with respect to w). With $\delta = 3.0$, the tax rates vary from 9.5 percent with $\varepsilon = 0.5$ to 16.7 percent with $\varepsilon = 2.0$. In the case of the lognormal, it is quite possible for the tax rate to fall with ε: for example, if $(\sigma^2\varepsilon)$ is sufficiently less than 1 for third and higher powers to be neglected, then the tax rate may be approximated by $\sigma^2 - \varepsilon\sigma^4/2$, which gives the following results (where all individuals work).

	ε		
σ^2	0.5	1.0	2.0
0.16	15%	15%	13%
0.24	23%	21%	18%

The fact that the tax structure may be regressive (i.e. the rates fall with ε) may appear to conflict with the intuitive notion discussed above that efficiency considerations would point to the use of a poll tax and that it is concern for the distribution which leads to the use of commodity taxes. However, when distributional objectives are relevant, indirect taxes play two roles. Firstly, by taxing luxuries at a higher rate they may increase the progressitivity of the tax system; secondly, they provide an alternative source of revenue, allowing the regressive poll tax to be reduced or converted into a lump-sum payment. In the latter case, the revenue would be raised in the distortion-minimizing way, and the final tax structure would balance the two sets of considerations.

Going back to the general formulation (5″), we can see that

$$\int_0^\infty \left(\sum_k S_{ik}t_k\right)\mathrm{d}F = \int_0^\infty (x_i - X_i)\,(b-\bar{b})\mathrm{d}F\,, \tag{12}$$

so that the reduction of consumption along the compensated demand curve is simply equal to the covariance between the consumption of that good and the net marginal social utility of income. For small variance, the tax structure may be approximated by taking a Taylor series expansion of the RHS of (12),

$$x_i\phi_i \approx \frac{\mathrm{d}x_i}{\mathrm{d}w}\frac{\partial b}{\partial w}\sigma_w^2\,,$$

where σ_w^2 is the variance of wages (abilities). Thus, the percentage reduction (along the compensated demand curve) in consumption is exactly proportional to the uncompensated derivative of the commodity with respect to the wage. If there is constant marginal utility of leisure and separable demand functions, we obtain

$$x_i\phi_i \approx q_i\left(\frac{\partial x_i}{\partial q_i}\right)_U \frac{\partial b}{\partial w}\sigma_w^2\,,$$

so

$$\frac{t_i}{1+t_i} \approx \frac{\partial b}{\partial w}\sigma_w^2\,,$$

independent of i: i.e. to the first order of approximation, there should be uniform taxation.

Expanding ϕ_i further shows that to the second order of approximation, differences in tax rates depend on the concavity or convexity of the demand functions $(\partial^2 x_i/\partial q_i^2)$ and the third moment of the ability distribution, parameters for which we are unlikely to obtain robust estimates.

The examples given above show that the results described in the previous section may need significant modification where the government is able to employ income taxation, even where this is restricted to a simple linear schedule. In the next section we examine the relationship between direct and indirect taxation where the income tax schedule may be freely varied.

5. Excise taxes and optimal income taxation

We assume that the income tax schedule is differentiable,[10] but apart from that may be of any form. We also allow for the possibility that the tax rate on commodities may be a function of the level of consumption.[11] The individual with wage w faces a budget constraint

$$\sum_i (x_i + t_i(x_i)) = wL - T(wL), \tag{13}$$

and the first-order conditions for utility maximization are[12]

$$U_i = \frac{(1+t_i')(-U_L)}{w(1-T')}, \qquad i = 1, \ldots, n. \tag{14}$$

The government maximizes the social welfare function subject to

$$\int_0^\infty \left[\sum_i t_i(x_i) + T(wL)\right] \mathrm{d}F = \bar{R},$$

or

$$\int_0^\infty \left[wL - \sum_i x_i - \bar{R}\right] \mathrm{d}F = 0. \tag{15}$$

This problem may be treated in a number of different ways. In the heuristic argument which follows, we take $x_2, \ldots, x_n$ and L as the control variables, treating U as a state variable, and making use of the fact that x_1 depends on $U, x_2, \ldots, x_n$ and L. Moreover,

$$\frac{\mathrm{d}U}{\mathrm{d}w} = \frac{-U_L L}{w} \equiv -U_L \theta(w, L). \tag{16}$$

The Hamiltonian may then be written

$$H = \left[G(U) + \lambda(wL - \sum_i x_i - \bar{R})\right] f - \mu \theta U_L, \tag{17}$$

[10]See Mirrlees (1971). In general this need not be the case. For an analysis of such non-differentiabilities within the context of this class of 'screening' problems, see Stiglitz (1974a).

[11]Actually we could have considered a general tax function of the form $T(x, L, w)$. In fact, for this particular problem, the results for the more restrictive, but practically more important, tax structure involving separability assumed here are identical to those in which the separability is dropped. This may be seen most easily by observing that nowhere in the analysis is the separability restriction on the tax function actually used.

[12]For an interior solution; we do not consider the case where labor supply is zero, although the analysis could easily be modified.

where f is the density function. Maximizing H with respect to x_i, we obtain as necessary conditions

$$-\lambda\left[\left(\frac{\partial x_1}{\partial x_i}\right)_U + 1\right] - \frac{\mu}{f}[U_{L1}\left(\frac{\partial x_1}{\partial x_i}\right)_U + U_{Li}]\theta = 0\,. \tag{18}$$

From (14) it is immediate that

$$\left[\frac{\partial x_1}{\partial x_i}\right]_U = -\frac{U_i}{U_1} = -\frac{(1+t_i')}{(1+t_1')}\,. \tag{19}$$

Thus we can rewrite (18) as

$$\lambda\left[\frac{1+t_i'}{1+t_1'} - 1\right] = \frac{\mu\theta U_i}{f}\,\frac{\mathrm{d}\log\left(\frac{U_i}{U_1}\right)}{\mathrm{d}L}\,. \tag{20}$$

Without loss of generality, we set $t_1' = 0$. Hence

$$\frac{t_i'}{1+t_i'} = \frac{\mu\theta\alpha}{\lambda f}\,\frac{\mathrm{d}\log\left(\frac{U_i}{U_1}\right)}{\mathrm{d}L}\,. \tag{21}$$

Tax rates are simply proportional to the rate at which the marginal rate of substitution between commodity i and commodity 1 changes with a change in the consumption of leisure.

From this analysis we obtain at once an interesting result. If the utility function is weakly separable between labor and all consumption goods (taken together), then no commodity taxation need be employed ($t_i = 0$). It is immediate that we could have allowed U to depend on n as well, as long as we maintain our separability hypothesis: $U = U(V(x_1, \ldots, x_n), L, n)$. With the greater flexibility provided by the nonlinear income tax schedule, the result found for special cases in the previous section now holds for much more general utility functions. The assumption of separability between consumption and labor may well be regarded as a reasonable first approximation for our purpose; and even if it is in fact empirically rejected, it is a useful benchmark case.[13] From the results given above, it follows that goods which are complementary (in the Edgeworth, not the more usual Hicksian, sense) with leisure ($U_{iL} < 0$) will face lower tax rates, whereas substitutes face higher tax rates. Finally, it is interesting to note that relative tax rates are independent of the social welfare function, so that they may be viewed as conditions for constrained Pareto optimality.[14]

[13]Where a subset of commodities is separable from labor, then the commodities in this group should all be taxed at the same rate.

[14]We are indebted to J.A. Mirrlees for pointing this out in his discussion of the paper at the Paris conference.

There are three interesting applications of the results given above which should be mentioned briefly [see also Atkinson (1974)]. First, if the goods are interpreted as consumption at different dates, then the analysis shows that the conventional presumption in favor of consumption rather than income taxation may be interpreted as assuming separability between leisure and consumption. Perhaps a more reasonable structure of preferences in this context is

$$U = U_1(c_1, L) + U_2(c_2),$$

in which case whether there should be an interest income tax or subsidy depends on the complementarity or substitutability (in the Edgeworth sense) between the first-period consumption and labor. The second application is to the question of the differential treatment of safe and risky assets: x_i is then treated as purchases of the ith security. Our theorem then says that where the individual maximizes $V(L)+EU(Y)$, there should be no differential treatment of risky assets [Stiglitz (1970) and Atkinson and Stiglitz (1972)].

The third application is to the use of quotas of specific allocations for distributing certain goods. Some economists [e.g. Tobin (1970)] have argued that there exist certain inelastically supplied commodities (medical care, at least in the short run) where quotas might be desirable. Such quota systems can be viewed as an extreme nonlinear commodity tax-subsidy scheme: below the quota the price is zero, above the quota, infinite. Viewed this way, the question of the desirability of quotas is equivalent simply to the question of whether it is optimal to have such an extreme form of progression for some particular commodity. The import of our theorem is that, provided the separability assumption is satisfied, not only should no quota be employed for such commodities, but not even a tax should be imposed. The result does not depend on the supply elasticities for the commodities in question.[15]

The basic intuition behind the argument that quotas might be desirable for inelastic commodities was that, if commodities are elastically supplied, then individuals should be allowed to trade off consumption of one good against the other: an individual's increased consumption of vanilla ice cream cones does not deprive someone else of his consumption of vanilla ice cream cones. When commodities are inelastically supplied, then there is no production inefficiency introduced by quotas. But prices serve as signals not only for the production of goods but also for the allocation of goods among individuals (the conventional exchange model). So long as tastes differ, the use of quotas will result in exchange inefficiency.

But, it might be argued, if we had a separable utility function, a first-best

[15] In our proof, we assume an elastic supply of all commodities, but it is easy to establish that, provided profits (rents) are fully taxed, the results are true for any production technology (including the limiting case of a perfect inelastically supplied commodity).

solution would entail allocating the same amount of the given good to everyone (if they had the same utility function) and hence we could achieve a first-best allocation of this particular good, with no loss of production efficiency. Such an argument, though plausible at first sight, fails to recognize the second-best nature of the problem we are considering: satisfying one of the first-best conditions (equating marginal utilities of consuming this particular good) does not necessarily represent an improvement when the other conditions are not satisfied.

A more plausible argument is that if we are able to discriminate among those with higher incomes by charging them a higher price (e.g. by having price an increasing function of quantity consumed) we would improve welfare, since such a differential price imposes a higher cost on those with lower marginal utilities of income. But there is a cost in deadweight loss from such differential pricing, and the import of our theorem is that in the central case examined, the cost outweighs the gains.[16]

6. Differences in tastes and horizontal equity

The existence of differences in tastes among individuals of the same ability raises issues in the design of the tax structure which we have not yet taken into account. In the conventional treatment, the principle of horizontal equity – that people who are in all relevant senses identical ought to be treated identically – plays an important role. In this section, we discuss, necessarily briefly, the nature of this principle as well as its implications for the design of tax policy. We first point out that the principle of horizontal equity may be in direct conflict with the utilitarian maximum even when tastes are identical; next we examine the case where tastes differ and show that the principle does not imply, as some have suggested, uniform taxation; finally, we consider more generally the status of horizontal equity as an objective of government policy.

The literature on optimal taxation has typically assumed that the redistributive goals of the government may be represented by maximizing a Bergsonian social welfare function, such as $G(U)$ defined above, and has not discussed the relationship between this and the concept of horizontal equity. Some earlier authors have taken the view that there is no conflict: 'the requirements of horizontal and vertical equity are but different sides of the same coin' [Musgrave (1959, p. 160)]. However, this need not be so. It is quite possible that the maximization of a Bergsonian social welfare function may indicate that individuals with identical tastes and endowments should be taxed at different rates (if this

[16]Spence (1975) and Weitzman (1974) have discussed this issue in a partial equilibrium context. The fact that their results differ from those given here is attributable to the fact that the presence of the optimal income tax has important implications for the role to be played by other distributive mechanisms, as we have emphasized throughout this paper.

is feasible), thus violating conventional notions of horizontal equity [see Atkinson and Stiglitz (1976)].[17]

The point is that if the feasible set of allocations is not convex (as it may be when only indirect taxes are employed), optimality may entail treating otherwise identical individuals differently.[18] An even stronger conflict has been noted by Stiglitz (1974b), where horizontal equity may conflict with the principle of Pareto optimality. Even before we introduce taste differences, therefore, there is a possible conflict between horizontal equity and the maximization of a social welfare function of the type usually assumed.

If we now introduce differences in tastes, the immediate consequence is that we must confront the interpersonal comparability question, which we have ignored thus far. When individuals have the same indifference curves, it is natural simply to use the same cardinal number of the indifference curves for different individuals. But when tastes differ, this is no longer so. Even if everyone had the same homothetic indifference maps, we must still decide which indifference curve for individual 1 corresponds to a given curve for individual 2.

The point is that the utilitarian system evaluates taxes in terms of the individual's ability to derive utility from goods and leisure, and in this respect may be contrasted with the alternative criterion of 'ability to pay,' that is, of basing taxation on opportunity sets. When the only differences are those in the ability to produce, then a utilitarian ethic leads to redistribution from those with 'better' opportunity sets to those with 'poorer'. There is no conflict between it and the ability-to-pay approach. But this may arise as soon as tastes differ. Suppose individual 1 has a higher productivity, so that his budget constraint lies outside

[17] Consider the simplest possible case of labor and a single consumption good (C), with two identical individuals. We assume that lump-sum taxes (poll taxes) are not admissible. The utilitarian problem may be formulated as

$$\max V(q_1)+V(q_2),$$

subject to

$$\tau_1 C_1+\tau_2 C_2 = \bar{R},$$

with first-order conditions

$$V_{q_i}(q_i) = -\lambda\left(C_i+\tau_i\frac{\partial C_i}{\partial q_i}\right),$$

where λ is the Lagrange multiplier associated with the constraint. It is obvious that

$$q_1 = q_2 = q^* = 1+\tau^*,$$

where

$$2\tau^* C(q^*) = \bar{R},$$

satisfies the first-order conditions. But

$$V_{qq}+\lambda\left(\tau_i\frac{\partial^2 C_i}{\partial q_i{}^2}+\frac{2\partial C_i}{\partial q_i}\right)$$

may well be positive at $q_i = q^*$, which would mean that this represents a local minimum.

[18] Analogous results in different contexts have been noted by Stiglitz (1974b) and Mirrlees (1972).

that of individual 2. The ability-to-pay criterion would indicate that individual 1 paid more tax, but there are obviously numberings of their indifference curves which lead to the opposite result with the utilitarian objective.

In order to contrast these two approaches, let us suppose that tastes may be represented by a single parameter, γ, so that the indirect utility function may be written as $V(q, w, \gamma)$. The utilitarian principle recognizes such taste differences as a legitimate basis for discrimination, and the government maximizes $G[V(q, w, \gamma)]$. On the other hand, if we introduce the concept of horizontal equity and interpret this as meaning that differences in tastes are not 'relevant' characteristics on which discrimination ought to be based, then this has two implications. Firstly, it introduces a cardinalisation $V(1, w, \gamma) = \tilde{V}(1, w)$, so that only endowments, w, and consumer prices (normalized at unity before tax) are relevant. Secondly, it constrains the government in levying taxes ($q \neq 1$) to maintain

$$V(q, w, \gamma) = \tilde{V}(q, w). \tag{22}$$

Suppose that the government were to adopt this version of horizontal equity; what would be the implications for the optimal tax structure? It is popularly believed that it would require uniform taxation. If two individuals are identical in all respects except that one likes chocolate ice cream and the other likes vanilla, a system which taxes chocolate ice cream at a higher rate is felt to be horizontally inequitable.[19] This is not however necessarily correct, as may be seen from the following example:

$$U = \sum_i (A_i(\gamma))^{(1/\varepsilon_i)} \frac{x_i^{1-(1/\varepsilon_i)}}{1-(1/\varepsilon_i)} - vL.$$

(It should be noted that we are assuming that there are no differences between people in the marginal utility of leisure, and that ε_i is independent of γ.) Let us further assume that A_i is independent of γ, for $i = 3, \ldots, n$, and that $A_1 = \gamma$. The requirement of normalization is then that $A_2(\gamma)$ is such that $V(1, w, \gamma) = V(1, w)$: i.e. that all those with the same w have the same pre-tax utility. Using this, it can be shown that the horizontal equity condition (22) requires that[20]

$$q_1^{1-\varepsilon_1} = q_2^{1-\varepsilon_2}. \tag{23}$$

[19]Pigou (1947) gives a nice example: 'When England and Ireland were united under the same taxing authority, it was strongly argued that, owing to the divergent tastes of Englishmen and Irishmen, it was improper to subject them to the same tax formulae in respect of beer and whiskey.' The tax on spirits, more generally consumed in Ireland, was more than two-thirds of the price, whereas the tax rate on beer was only about one-sixth of the price.

[20]It may be noted that (for $\varepsilon_i \neq 1$).

$$V(q, w, \gamma) = \sum_i \frac{A_i(\gamma)}{(\varepsilon_i - 1)} \left[\frac{(vq_i)}{w}\right]^{1-\varepsilon_i}.$$

The condition for horizontal equity is not, therefore, uniform taxation; only if the price elasticity is the same – as it may well be in the chocolate/vanilla ice cream case – would uniform tax rates be horizontally equitable. This may be related to the argument made by Pigou (1947, p. 77):

> Suppose that there are two persons of equal income and general economic status, that in the aggregate of their tastes they are similar, in the sense that they would get equal satisfactions from equal incomes if they were permitted to spend them as they chose, but that one likes and purchases commodity *A* and not commodity *B*, the other commodity *B* and not commodity *A*. Suppose, further, that taxes are imposed upon commodities *A* and *B* in such ways that both these persons pay the same amount of tax. It will not necessarily follow that they suffer equal real burdens. If the demand of one for his commodity is more elastic than the demand of the other for his, the former will suffer the larger hurt.

The model just described is a very simple one, but it brings out clearly the conflict between horizontal equity and the maximization of a social welfare function of the Bergson type. For example, where $G' = 1$ (the classical utilitarian case), the latter leads to the first-order condition,

$$1-\frac{1}{q_i}=\frac{1-\bar{b}r_i}{\varepsilon_i},$$

as before. This is not in general consistent with the requirement of horizontal equity, eq. (23).

This raises the important issue of the status of the horizontal equity principle. It is often suggested that horizontal equity is in some sense prior to vertical equity: 'it is sometimes said that the horizontal aspect is more basic and less controversial' [Musgrave and Musgrave (1973, p. 199)]. Most authors, including Musgrave and Musgrave, go on to argue that neither is more basic than the other; however, this ignores the conflict which we have seen to arise between the two principles, at least in the form presented here. Faced with this potential conflict, it might seem more reasonable to view the social welfare function as lexicographic. For certain classes of goods, probably those marked by considerable diversity of tastes, the horizontal equity requirement is imposed, and the government then maximizes a Bergsonian social welfare function subject to this constraint. As Pigou (1947, p. 51) put it, 'the ideal of least sacrifice has to be pursued subject to a handicap.' The optimal structure of taxation, and the choice between direct and indirect taxes, will depend on how wide is the range of goods covered by constraints such as (23).

7. Concluding comments

In this paper, we have attempted to present a framework within which we can evaluate the appropriateness of different tax bases and to apply this framework to the classical question of the use of direct versus indirect taxation.

The general framework employed may be summarized as follows. The necessity for any form of taxation other than a uniform lump-sum tax arises from the fact that individuals have differing characteristics (endowments or tastes). If we could observe all relevant characteristics costlessly and perfectly, we should be able to achieve a first-best solution. However, in practice we have to make use of surrogate characteristics, which are related systematically to the characteristics on which we would like to differentiate individuals, but which are not perfectly correlated and which are, to some extent, under the control of the individual. Certain ethical principles, notably those which fall under the rubric of horizontal equity, limit further the set of surrogates which may be used. Having established an admissible class of characteristics, the problem then becomes one of determining which are to be employed (the choice of tax base) and the structure of the tax schedule.

The application of this framework to the direct/indirect tax problem led to the following results. Firstly, if the government had no distributional objectives and was concerned solely with efficiency, it may employ only direct taxation and this would take the form of a poll tax. This is a very straightforward prescription, but it has the implication, which runs counter to much popular belief, that the use of indirect taxation stems from a pursuit of distributional objectives. The extent to which indirect taxes are employed to this purpose – that is, purchases of different commodities are used as a screening device – depends on the form of consumer preferences and on the restrictions (if any) on the type of income taxation employed. If a general income tax function may be chosen by the government, we have shown that, where the utility function is separable between labor and all commodities, no indirect taxes need be employed. In this case, the use of consumption of particular commodities as a screening device offers no benefit. Finally, we have seen that horizontal equity considerations may impose constraints on the structure of taxes which may be levied.

Throughout the paper, we have stressed the importance of the interactions between different taxes, and the fact that a piecemeal approach may be misleading. In section 4, for example, it was shown that in the quadratic case considered by Ramsey (plus constant marginal utility of leisure and independence) the introduction of an optimal linear income tax meant that indirect taxation was no longer necessary. The Ramsey-style results would, therefore, only be relevant where there were constraints on the use of income taxation. Such interactions are equally a warning that the results given in this paper should be treated with considerable caution. For this and other reasons, such as the failure to incorporate the costs of administration,[21] the theory may be more useful in illuminating the structure of the argument than in providing definite answers to policy issues.

[21]See Heller and Shell (1974) for an attempt to introduce administration costs into the analysis of optimal taxation.

References

Atkinson, A.B., 1970, On the measurement of inequality, Journal of Economic Theory 2, 244–263.

Atkinson, A.B., 1974, Housing allowances, income maintenance and income taxation, International Economic Association conference, Turin.

Atkinson, A.B., and J.E. Stiglitz, 1972, The structure of indirect taxation and economic efficiency, Journal of Public Economics 1, 97–119.

Atkinson, A.B. and J.E. Stiglitz, 1976, Lectures on Public Economics, forthcoming.

Boskin, M.J., 1973, Optimal tax treatment of the family, Memorandum 143 (Center for Research in Economic Growth, Stanford University, Stanford, CA).

Chipman, J.S., 1974, The welfare ranking of Pareto distributions, Journal of Economic Theory 9, 275–282.

Diamond, P.A., 1975, A many-person Ramsey tax rule, Journal of Public Economics 4, 335–342.

Diamond, P.A. and J.A. Mirrlees, 1971, Optimal taxation and public production, American Economic Review 61, 8–27 and 261–278.

Diamond, P.A. and J.E. Stiglitz, 1974, Increases in risk and in risk aversion, Journal of Economic Theory 9, 337–360.

Feldstein, M.S., 1972a, Distributional equity and the optimal structure of public prices, American Economic Review 62, 32–36.

Feldstein, M.S., 1972b, Equity and efficiency in public pricing, Quarterly Journal of Economics 86, 175–187.

Heller, W.P. and K. Shell, 1974, On optimal taxation with costly administration, American Economic Review 74, papers and proceedings, 338–345.

Mirrlees, J.A., 1971, An exploration in the theory of optimum income taxation, Review of Economic Studies 38, 175–208.

Mirrlees, J.A., 1972, Population policy and the taxation of family size, Journal of Public Economics 1, 169–198.

Musgrave, R.A., 1959, The theory of public finance (McGraw-Hill, New York).

Musgrave, R.A. and P.B. Musgrave, 1973, Public finance in theory and practice (McGraw-Hill, New York).

Pigou, A.C., 1947, A study in public finance (Macmillan, London).

Spence, M., 1973, Job market signalling, Quarterly Journal of Economics 87, 355–379.

Spence, M., 1975, Nonlinear prices and welfare, Technical Report 158, (IMSSS, Stanford University, Stanford, CA).

Stiglitz, J.E., 1970, Taxation, risk-taking and the allocation of investment in a competitive economy, in: M. Jensen, ed., Studies in the theory of capital markets (forthcoming).

Stiglitz, J.E., 1974a, Monopoly and imperfect information (Oxford and Stanford University) mimeo.

Stiglitz, J.E., 1974b, The efficiency wage hypothesis, surplus labor and the distribution of income in LDCs, Technical Report 152 (IMSSS, Stanford University, Stanford, CA).

Stiglitz, J.E., 1975, The theory of 'screening', education and the distribution of income, American Economic Review 65, 283–300.

Stiglitz, J.E. and P.S. Dasgupta, 1971, Differential taxation, public goods, and economic efficiency, Review of Economic Studies 38, 151–174.

Tobin, J., 1970, On limiting the domain of inequality, Journal of Law and Economics 13, 263–278.

Weitzman, M.L., 1974, Is one price system or rationing more effective in meeting true needs for a deficit economy? (MIT, Cambridge, MA) mimeo.