LECTURE

TWELVE

THE STRUCTURE OF INDIRECT TAXATION

12-1 INTRODUCTION

In most countries excise taxes are levied on commodities at different rates. This is certainly true in the United States, where the rates of taxation vary widely. In European countries there has been a move towards uniformity of tax rates with the introduction of a value added tax (VAT). In the United Kingdom, VAT replaced a purchase tax which had rates varying from 50 per cent on items such as jewellery and cameras to $12\frac{1}{2}$ per cent on clothing, footwear and furniture. Even with VAT, however, differential rates have been maintained in most countries, and typically quite a wide range of goods have a zero rate. The rationale for these systems of indirect taxation, and for the changes made, needs however to be examined. Are there good reasons for taxing goods at different rates? Is the move in European countries towards a more uniform structure of indirect taxation desirable on efficiency or distributional grounds?

According to conventional wisdom, there is a definite preference for a uniform rate structure, and this view appears to influence government policy-making. The British Government, when announcing the introduction of a value added tax, claimed that:

> a more broadly-based structure..., by discriminating less between different types of goods and services, would reduce the distortion of consumer choice.... Selective taxation gives rise to distortion of trade and of personal consumption patterns, and can lead to the inefficient allocation of resources. [HMSO, 1971, p. 3]

This case is based on efficiency considerations; i.e., that a differentiated structure has greater distortionary effects. A second, and quite different, argument for a uniform system of taxation is that of equity between consumers: "a general sales tax or added-value tax on all expenditure at a

single rate... would be fair as everyone would pay the same tax on all their expenditure" (Wheatcroft, 1969, p. 26). Similarly, "non-uniformity results in discrimination against those people having particular preference for the more heavily taxed goods" (Due, 1963, p. 285).

In assessing these arguments for uniform taxation, it is helpful to discuss the efficiency and equity aspects separately, since the considerations involved are different. For this reason we focus in the first part (Sections 12-1–12-4) of this lecture on a model where all individuals are identical, and are assumed to be treated identically.[1] No redistributional issues therefore arise, and we concentrate on the efficiency question as to whether, from the allocational standpoint, a uniform tax is preferable to a differentiated structure. This question is first discussed in the context of the partial equilibrium framework used in most textbooks and then extended to a general equilibrium treatment in Sections 12-2–12-4. In Section 12-5, distributional considerations are introduced and the balance between equity and efficiency considered.

Partial Equilibrium Analysis

In contrast to the view of the British Government, the standard textbook analysis of the structure of indirect taxation suggests that uniform rates are not in fact necessarily desirable from an efficiency standpoint. In this section, we show how this can be demonstrated by a simple partial equilibrium analysis,[2] where there are no cross-price effects and relevant income derivatives are zero.

Let us assume that the supply of good k is perfectly elastic at price p_k, so that the equilibrium in the absence of taxation is at point E in Fig. 12-1. The effect of an *ad valorem* tax at rate t_k is to raise the consumer price from p_k to $p_k(1+t_k)$. The after-tax equilibrium is at point B. In this partial equilibrium framework the distortion caused by the tax is often measured by the loss of consumer surplus over and above the revenue raised, the "excess burden". If we take the area $ABECD$ as a measure of the loss of consumer surplus, the excess burden is represented by the shaded area BCE. Let us denote the consumer price by q_k and write the demand curve, following Marshall, as $q_k(X_k)$. The excess burden caused by the tax on good k may then be seen from Fig. 12-1 to equal:

$$B_k \equiv \int_{X_k^t}^{X_k^0} q_k dX_k - p_k(X_k^0 - X_k^t) \qquad (12\text{-}1)$$

Area $BEFGC$ – area $CEFG$

[1] The reason why this is an assumption and not an implication is explained below in the section on horizontal equity.

[2] See, for example, Hicks (1947, Ch. X). For a more formal argument, which draws attention to the limitation of her analysis, see Bishop (1968).

where X_k^0 denotes the equilibrium quantity before the tax is introduced, X_k^t that after the tax is introduced. From this it follows that

$$\frac{\partial B_k}{\partial t_k} = -q_k \frac{\partial X_k^t}{\partial t_k} + p_k \frac{\partial X_k^t}{\partial t_k} = -p_k t_k \frac{\partial X_k^t}{\partial t_k} \tag{12-2}$$

where the term in q_k arises from differentiating the lower limit of integration and the second step follows from the fact that $q_k = p_k(1+t_k)$. The excess burden is therefore zero for infinitesimal taxes (i.e., evaluating at $t_k = 0$). As noted by Samuelson (1964a), all consumer surplus terms are of second order.

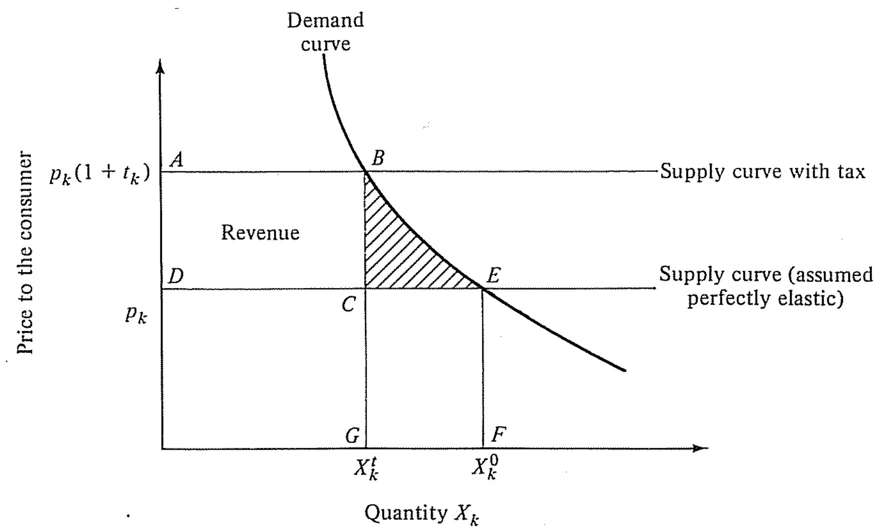


Figure 12-1 Excess burden from tax on good k.

Suppose now that the government chooses the tax rates on different goods $(t_1, \ldots, t_n)$ in such a way as to raise a specified revenue with the minimum total excess burden. The revenue condition is properly seen in terms of the government's purchasing a fixed amount of real commodities (government spending), but with fixed producer prices we can treat it as a financial constraint:

$$R \equiv \sum_{k=1}^{n} t_k p_k X_k^t = R_0 \tag{12-3}$$

where R_0 is the required level. This constrained maximization problem may be formulated in terms of the Lagrangean:

$$\mathscr{L} = -\sum_{k=1}^{n} B_k + \lambda(R - R_0) \tag{12-4}$$

The first-order conditions for the choice of t_k are therefore

$$\frac{\partial B_k}{\partial t_k} = \lambda \frac{\partial R}{\partial t_k} = \lambda p_k X_k^t + \lambda p_k t_k \frac{\partial X_k^t}{\partial t_k} \quad \text{for all } k \tag{12-5}$$

Combining this with Eq. (12-2), we obtain

$$\frac{-t_k}{X_k^t}\frac{\partial X_k^t}{\partial t_k} = \frac{\lambda}{1+\lambda} \tag{12-6}$$

or[3]

$$\frac{t_k}{1+t_k} = \frac{\theta}{\varepsilon_k^d} \tag{12-7}$$

where θ is equal to $\lambda/(1+\lambda)$ and ε_k^d is the elasticity of demand for good k.

A solution satisfying these first-order conditions (the precise status of these conditions is discussed in Section 12-2) involves therefore the tax rate on good k being in inverse proportion to the price elasticity of demand. In the extreme case of a good demanded completely inelastically (or a factor supplied by households completely inelastically), the excess burden is zero and all revenue, or as much as feasible, should be raised by taxing this commodity. Apart from this, the optimal tax structure can be uniform only where all goods have the same elasticity of demand. In general, "the best way of raising a given revenue ... is by a system of taxes, under which the rates become progressively higher as we pass from uses of very elastic demand or supply to uses where demand or supply are progressively less elastic" (Pigou, 1947, p. 105) (although we have not discussed the case where *supply* is less than perfectly elastic—see Lecture 15).[4]

This finding, although typically reported in public finance texts, is often regarded with considerable scepticism. Musgrave relegates it to a footnote

[3] This step may be seen if we rewrite the left-hand side as

$$\left(\frac{t_k}{1+t_k}\right)\left[\frac{p_k(1+t_k)}{X_k^t}\right]\left[-\frac{\partial X_k^t}{\partial p_k(1+t_k)}\right]$$

which equals $t_k/(1+t_k)$ times the elasticity of demand. The reader may like to compare this condition with the choice of $(1-\alpha)$, which is equivalent to a tax rate, in the capitalist's revenue-maximizing problem of Lecture 4 (cf. Eq. (4-29)), and to the price-setting condition for a monopolist, where the excess of price over marginal cost may be seen as equivalent to a tax.

[4] A further feature of the optimal tax structure may be noted in the case where the demand curve is linear: $X = a - bq$. Then

$$X_k^t - X_k^0 = -bp_k t_k$$

From (12-6), this gives

$$X_k^t - X_k^0 = -\theta X_k^t = \frac{-\theta}{1+\theta} X_k^0$$

i.e., the proportionate reduction in demand is the same for all commodities. As shown in the next section, this carries over in a weaker form to more general models.

and comments that "the theorem is arrived at within the framework of the old welfare economics of inter personal utility comparison. It belongs in the welfare view of the ability-to-pay approach and does not fit the context of the present argument" (Musgrave, 1959, p. 149n). However, Musgrave's own analysis is a special case of that described above. As pointed out by Bishop (1968, p. 212n), Musgrave's conclusion that "a general *ad valorem* tax is preferable to a system of selective excises imposed at differential rates" (Musgrave, 1959, p. 148) assumes a fixed supply of labour. The argument of the previous paragraph indicates that in this case all revenue should be raised by taxing labour; and, ignoring saving, this is equivalent to a uniform excise tax. Other writers have expressed reservations about the strength of the assumptions. Prest, for example, dismisses the results with the comment that "such restrictive assumptions have to be made in order to derive a solution, that they would appear to have little practical significance" (Prest, 1975, p. 53). However, he offers nothing in their place.

The assumptions underlying the partial equilibrium framework are indeed restrictive, requiring in effect that there be no income effects and that cross-price elasticities be zero. In the remainder of this Lecture, we adopt a general equilibrium approach, beginning with the classic paper by F. P. Ramsey, "A Contribution to the Theory of Taxation", published in the *Economic Journal* in 1927. This article provided the foundation for Pigou's discussion of the question in his textbook, and for the partial equilibrium treatment we have described.

12-2 THE RAMSEY TAX PROBLEM

The opening paragraph of Ramsey's article is worth quoting in full, since it sets out clearly the problem that had been posed to him by Pigou, and the framework within which he set about answering it:

> The problem I propose to tackle is this: a given revenue is to be raised by proportionate taxes on some or all uses of income, the taxes on different uses being possibly at different rates; how should these rates be adjusted in order that the decrement of utility may be a minimum? I propose to neglect altogether questions of distribution and considerations arising from the differences in the marginal utility of money to different people; and I shall deal only with a purely competitive system with no foreign trade. Further I shall suppose that, in Professor Pigou's terminology, private and social net products are always equal or have been made so by State interference not included in the taxation we are considering. I thus exclude the case discussed in Marshall's *Principles* in which a bounty on increasing-return commodities is advisable. Nevertheless we shall find that the obvious solution that there should be no differentiation is entirely erroneous. [Ramsey, 1927, p. 47]

The treatment here differs from that of Ramsey in certain respects, and we relax the assumptions about distribution (Section 12-5) and allow for externalities (Lecture 14). However, in this section we keep the specification as simple as possible.

The Model

Since the initial aim is to focus on efficiency considerations, it is assumed that all consumers are identical, and face identical tax rates, and that the objective of the government is to maximize the welfare of a "representative" individual. On the production side, it is assumed that there are fixed producer prices for all goods and a fixed wage rate w for labour. Labour is to be the only factor supplied by households, and they have no other source of income. Since the producer prices are fixed, we can without loss of generality set them at unity, so that the consumer price of good k is given by $q_k = 1+t_k$.

The structure of the problem is that the government is maximizing subject to the demand and supply functions of individuals, which are themselves based on solving a constrained maximization problem. The representative consumer supplies L units of labour (where L is measured as a fraction of the working day) and consumes X_i of good i ($i = 1,\dots,n$). He is assumed to maximize $U(\mathbf{X}, L)$ subject to the budget constraint

$$\sum_{i=1}^{n} q_i X_i = wL \tag{12-8}$$

It may be noted that there is assumed to be no tax on wage income, but if there is no other source of income for the consumer this involves no loss of generality. Suppose that a tax of τ is imposed on wage income. The consumer's budget constraint becomes

$$\sum_i q_i X_i = w(1-\tau)L \tag{12-8$'$}$$

(the summation runs from 1 to n unless otherwise indicated). As far as the consumer is concerned, this is equivalent to a situation where there is no wage tax and q_i is increased to $q/(1-\tau)$; i.e., for the tax rate to become

$$t_i' = \frac{1+t_i}{1-\tau} - 1 = \frac{\tau+t_i}{1-\tau} \tag{12-9}$$

The government revenue in the latter case is

$$\sum_i t_i' X_i = \sum_i \left(\frac{\tau+t_i}{1-\tau}\right) X_i \tag{12-10}$$

which may be compared with that in the case of the wage tax:

$$\sum_i t_i X_i + \tau wL = \sum_i t_i X_i + \frac{\tau}{(1-\tau)} \sum_i (1+t_i) X_i = \sum_i \left(\frac{\tau+t_i}{1-\tau}\right) X_i \tag{12-11}$$

where we have substituted for wL from (12-8′). The government revenue is also unaffected. A tax on wage income is therefore equivalent in this model to a uniform tax on all goods. This depends on the fact that there is no other source of income (such as profit income) and that we cannot tax the consumer's labour endowment (i.e., leisure).[5]

The government aims then to maximize individual welfare subject to the revenue constraint and the individual conditions for utility maximization. Following Diamond and Mirrlees (1971), the problem may conveniently be treated in terms of the indirect utility function $V(\mathbf{q}, w)$. Forming the Lagrangean

$$\mathscr{L} = V(\mathbf{q}, w) + \lambda\left(\sum_i t_i X_i - R_0\right) \qquad \text{(12-12)}$$

gives first-order conditions for the tax rate t_k

$$\frac{\partial V}{\partial q_k} = -\lambda\left(X_k + \sum_i t_i \frac{\partial X_i}{\partial q_k}\right) \quad \text{for } k = 1, \ldots, n \qquad \text{(12-13)}$$

Writing α for the marginal utility of income to the consumer, and using the properties of the indirect utility function ($\partial V/\partial q_k = -\alpha X_k$),

$$\sum_i t_i \frac{\partial X_i}{\partial q_k} = -\frac{(\lambda - \alpha)}{\lambda} X_k \quad \text{for } k = 1, \ldots, n \qquad \text{(12-14)}$$

These equations may be transformed using the Slutsky relationship

$$\frac{\partial X_i}{\partial q_k} = S_{ik} - X_k \frac{\partial X_i}{\partial M} \quad \text{for all } i, k \qquad \text{(12-15)}$$

where S_{ik} is the derivative of the compensated demand curve and $\partial X_i/\partial M$ denotes the income effect (evaluated at $M = 0$). Substituting, we obtain

$$\sum_i t_i S_{ik} = -\left(1 - \sum_i t_i \frac{\partial X_i}{\partial M} - \frac{\alpha}{\lambda}\right) X_k \quad \text{for } k = 1, \ldots, n \qquad \text{(12-16)}$$

Using the symmetry of the Slutsky terms ($S_{ik} = S_{ki}$), and introducing θ for the coefficient of X_k in (12-16),[6]

$$\sum_i t_i S_{ki} = -\theta X_k \quad \text{for } k = 1, \ldots, n \qquad \text{(12-17)}$$

[5] What is required is that the demand functions be homogeneous of degree zero in consumer prices. The demand functions do not have this property if the consumer receives lump-sum transfers (pays lump-sum taxes) in nominal units, or where there are profits from the production sector. See Dixit and Munk (1977).

[6] From this expression, one can see the relation with the partial equilibrium formula (12-6). If there are no income effects ($\partial X_i/\partial M = 0$), and if $S_{ik} = 0$ for $i \neq k$, then Eq. (12-17) becomes $t_k(-S_{kk}) = \theta X_k$.

Discussion of the Results

The formulation (12-17) is due to Samuelson (1951), who gave the following interpretation: the left-hand side is the change in the demand for good k that would result if the consumer were compensated to stay on the same indifference curve and the derivatives of the compensated demand curves were constant. In fact, it is not possible for the latter condition to be satisfied for all commodities, but for small taxes it is approximately true that the optimal tax structure involves an equal proportionate movement along the compensated demand curve for all goods (since θ is independent of k).[7] The importance in this formula of the *compensated* derivatives accords with intuition: the income effect would arise with any form of taxation, and the distortion stems from the substitution effect. We may note that multiplying (12-17) by t_k and summing gives

$$\sum_k \sum_i t_k S_{ki} t_i = -\theta R_0 \qquad \text{(12-18)}$$

The left-hand side can be shown to be negative (using the negative semi-definiteness of the Slutsky matrix), so that θ has the same sign as government revenue.

A further interpretation may be given for θ by examining the effect of allowing the government to levy a lump-sum tax T. The Lagrangean then becomes

$$\mathscr{L} = V(\mathbf{q}, T) + \lambda\left(T + \sum_i t_i X_i - R_0\right) \qquad \text{(12-19)}$$

From this we can see that, using the definition of θ,

$$\frac{\partial \mathscr{L}}{\partial T} = -\alpha + \lambda\left(1 - \sum_i t_i \frac{\partial X_i}{\partial M}\right) = \theta\lambda \qquad \text{(12-20)}$$

(since $\partial V/\partial T = -\alpha$ and $\partial X_i/\partial T = -\partial X_i/\partial M$). Now suppose that the government were allowed to make a small increase dT in the lump-sum tax (moving away from the optimum described above), where the commodity taxes are adjusted so that the revenue constraint continues to hold. Then $d\mathscr{L}/dT = \partial\mathscr{L}/\partial T$ (since $\partial\mathscr{L}/\partial t_k = 0$ for all k from the first-order conditions) and, since the revenue constraint continues to hold, $d\mathscr{L} = dV = \lambda\theta dT$. Welfare rises and θ measures the benefit *expressed in terms of revenue* from being able to switch from the (optimal) indirect tax system to lump-sum taxation.

At this point we may note the consequences of relaxing the assumption of fixed producer prices. Suppose that production takes place under constant returns to scale (as in Diamond and Mirrlees, 1971). The

[7] It has not been demonstrated that (12-17) holds for labour, but this can readily be shown.

government revenue constraint is replaced by a production constraint:

$$wL = F(X_1, \ldots, X_n) + R_0 \tag{12-21}$$

where w is fixed (labour is again the *numeraire*) and the right-hand side gives the labour requirements of the private sector, $F(X)$, and government revenue, R_0, expressed in terms of labour. The first-order conditions become

$$\frac{\partial V}{\partial q_k} = -\lambda\left(w\frac{\partial L}{\partial q_k} - \sum_i F_i \frac{\partial X_i}{\partial q_k}\right) \tag{12-22}$$

Since there are constant returns to scale, there is no pure profit income, so that differentiating the consumer's budget constraint yields

$$\frac{w\partial L}{\partial q_k} = X_k + \sum_i q_i \frac{\partial X_i}{\partial q_k} \tag{12-23}$$

Profit maximization in the private sector implies $F_i = p_i$, where p_i are producer prices,[8] so it follows that

$$\frac{\partial V}{\partial q_k} = -\lambda\left(X_k + \sum_i t_i \frac{\partial X_i}{\partial t_k}\right)$$

(noting that $q_i - p_i = t_i$). We are therefore back with condition (12-13). The *form* of the first-order conditions is therefore unaffected in the case of constant returns to scale (non-constant returns are discussed in Lecture 15); on the other hand, the producer prices in general vary with changes in the tax rates.

The analysis so far has been based on the first-order conditions, and we should note that their necessity has been asserted, not demonstrated. This point is discussed in detail by Diamond and Mirrlees (1971, Section X), who set out a constraint qualification such that the use of Lagrange multipliers is indeed valid, and the formulae given earlier are necessary for optimality. They also provide a valuable discussion of the question of uniqueness. There are two problems. First, the specification of the tax rates may not uniquely determine the behaviour of the system. Second, there may be more than one solution to the first-order conditions:

> 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. [Diamond and Mirrlees, 1971, p. 276]

It is possible to construct examples where more than one value of t_k satisfies the first-order conditions.[9] Moreover, it is quite possible that there are two goods, j and k, with identical demand conditions, leading to identical first-order conditions, where the global optimum involves an asymmetric solution ($t_j \neq t_k$). For this reason, one needs to be careful in drawing conclusions about uniformity of taxes from the first-order conditions. The mere fact that the conditions for t_j and t_k are identical in form does not imply that they should be set equal. (There are also important implications for horizontal equity, taken up later.)

[8] The private sector maximizes $\Sigma p_i X_i - F(X_1, \ldots, X_n)$.

[9] For further discussion, see Harris (1975), Atkinson and Stiglitz (1976), and Atkinson (1977b).

An Example

In order to illustrate the results, let us take the case where there are two goods and labour, and a non-negative revenue requirement. The conditions (12-17) then become:

$$\begin{aligned} t_1 S_{11} + t_2 S_{12} &= -\theta X_1 \\ t_1 S_{21} + t_2 S_{22} &= -\theta X_2 \end{aligned} \tag{12-24}$$

Solving,

$$\begin{aligned} t_1 &= (\theta/S)(S_{12}X_2 - S_{22}X_1) \\ t_2 &= (\theta/S)(S_{21}X_1 - S_{11}X_2) \end{aligned}$$

where $S = S_{11}S_{22} - S_{12}^2$ and is positive by the properties of the Slutsky matrix. Defining the elasticities of compensated demand, and setting $p_i = 1$,

$$\varepsilon_{ij} = q_j S_{ij}/X_i \tag{12-25}$$

we have

$$\frac{t_1}{1+t_1} = \frac{t_2}{1+t_2}\left(\frac{\varepsilon_{12} - \varepsilon_{22}}{\varepsilon_{21} - \varepsilon_{11}}\right) \tag{12-26}$$

Let us introduce the notation that good 0 is leisure (i.e., minus leisure). From the properties of the Slutsky terms, we know that:[10]

$$\sum_{j=0}^{2} q_j S_{ij} = 0 \tag{12-27}$$

(where $q_0 = w$). So that:

$$\begin{aligned} \varepsilon_{10} + \varepsilon_{11} + \varepsilon_{12} &= 0 \\ \varepsilon_{20} + \varepsilon_{21} + \varepsilon_{22} &= 0 \end{aligned} \tag{12-28}$$

Hence, substituting in (12-26),

$$\frac{t_1}{1+t_1} = \frac{t_2}{1+t_2}\left[\frac{-(\varepsilon_{11} + \varepsilon_{22}) - \varepsilon_{10}}{-(\varepsilon_{11} + \varepsilon_{22}) - \varepsilon_{20}}\right] \tag{12-29}$$

[10] This may be seen from the fact that $S_{ij} = E_{ij}$, where E is the expenditure function, and E_i is homogeneous of degree zero in prices.

It follows that $\varepsilon_{10} > \varepsilon_{20}$ implies $t_1 < t_2$ (ε_{ii} being negative). At the optimum, the good with the larger cross-elasticity of compensated demand with the price of labour (leisure) has the smaller tax rate.[11] This is the basis of the result reached by Corlett and Hague (1953), Meade (1955, p. 30), Harberger (1964) and others that we should tax more heavily goods that are complementary with leisure.

Exercise 12-1 Derive the optimal tax structure where the utility functions have the Cobb–Douglas form

$$U = \sum_{i=1}^{n} a_i \log X_i + A \log(1-L) \tag{12-30a}$$

where $A + \Sigma_i a_i = 1$ and $n = 2$.

Exercise 12-2 For the utility function

$$U = \sum_{i=1}^{n} A_i \frac{X_i^{1-1/\varepsilon_i}}{1-1/\varepsilon_i} - vL \tag{12-30b}$$

show that the income terms $(\partial X_i/\partial M)$ and cross-price terms are zero. Derive the optimal tax structure where ε_i are (positive) constants.

12-3 APPLICATION OF THE RAMSEY RESULTS

The general formulation given in the previous section provides important insights into the nature of the solution, but does not yield much in the way of concrete results. Equation (12-17) does not, for example, suggest which goods should be taxed more heavily, and the two-good example cannot readily be extended. In order to obtain more definite results, Ramsey himself made a number of special assumptions on the demand side equivalent to the partial equilibrium analysis described in Section 12-1. From this it might appear that we have to choose between definite results based on highly restrictive assumptions and more general models yielding only limited conclusions. However, it is possible by adopting an alternative approach to derive results midway in generality, and these are discussed in this section, together with some of the numerical applications. We retain for the present the assumption of identical individuals.

Alternative Formulation

The analysis in the previous section used the "dual" price variables as controls open to the government and exploited the properties of the indirect

[11] The elasticities are typically functions of the prices, and hence the tax rates, and there may be multiple solutions to (12-24).

utility function. (In the next section, we show how this relates to the expenditure function.) For many purposes, the dual approach provides a neat and compact treatment, and it has been widely adopted. On the other hand, in some cases the "primal" approach, using the quantities as controls, may aid understanding. In this section, we show how formulating the model in this way leads to an alternative form of the optimal tax conditions. We are in fact returning to Ramsey's original way of setting up the problem, since he worked with the direct utility function.

Let us therefore take as control variables for the government the quantities $X_1, \ldots, X_n$ and L, with the tax rates being obtained as functions of the control variables from the conditions for individual utility maximization. With this "primal" approach, we have to ensure that the consumer budget constraint is satisfied (see Atkinson and Stiglitz, 1972). For this purpose, we make use of the individual utility maximization conditions

$$U_i = \alpha q_i \quad i = 1, \ldots, n$$
$$-U_L = \alpha w \tag{12-31}$$

From these, the condition that the individual be on his offer curve may be written (substituting in the budget constraint and eliminating α),

$$\sum_i U_i X_i + U_L L = 0 \tag{12-32}$$

The Lagrangean then becomes[12]

$$\mathscr{L} = U(\mathbf{X}, L) + \lambda\left(wL - \sum_i X_i - R_0\right) + \mu\left(\sum_i U_i X_i + U_L L\right) \tag{12-33}$$

and the first-order conditions

$$U_k = \lambda - \mu U_k\left(1 + \sum_i \frac{U_{ik} X_i}{U_k} + \frac{U_{Lk} L}{U_k}\right) \quad \text{for } k = 1, \ldots, n \tag{12-34}$$

Let us now define

$$H^k \equiv -\left(\sum_i \frac{U_{ik} X_i}{U_k} + \frac{U_{Lk} L}{U_k}\right) \quad \text{for } k = 1, \ldots, n \tag{12-35}$$

and substitute for $U_k = \alpha(1+t_k)$. This yields

$$(1+t_k)[1-\mu(H^k-1)] = \lambda/\alpha \tag{12-36}$$

There is in addition the condition with respect to L

$$U_L = -\lambda w - \mu U_L\left(1 + \sum_i \frac{U_{iL} X_i}{U_L} + \frac{U_{LL} L}{U_L}\right) \tag{12-34'}$$

[12] In the revenue constraint we have used the fact that

$$\sum_i t_i X_i = \sum_i (q_i - 1) X_i = wL - \sum_i X_i$$

If we define the corresponding expression

$$H^L \equiv -\left(\sum_i \frac{U_{iL}X_i}{U_L} + \frac{U_{LL}L}{U_L}\right) \tag{12-35'}$$

and substitute $U_L = -\alpha w$, we obtain

$$\mu(1-H^L) = \frac{\lambda-\alpha}{\alpha} \tag{12-36'}$$

Eliminating μ between (12-36) and (12-36') gives[13]

$$\frac{t_k}{1+t_k} = \frac{\lambda-\alpha}{\lambda}\left(\frac{H^k-H^L}{1-H^L}\right) \tag{12-37}$$

While this equation does not in general provide an explicit formula for the optimal tax rate (since the terms H^k depend on the tax rates), it does allow us to draw a number of conclusions about the optimal structure. First, the partial equilibrium results can be seen as polar cases of this formula. Suppose on the one hand that $(-H^L)$ tends to infinity, which corresponds to a completely inelastic supply of labour $(-U_{LL} \to \infty)$; then the limit of (12-37) is a uniform tax on all goods at rate $(\lambda-\alpha)/\alpha$. Since we have seen that a uniform rate of tax on all goods is equivalent to a tax on labour alone, this corresponds to the conventional prescription that a factor in completely inelastic supply should bear all the tax. On the other hand, if H^L tends to zero, we have the case of a completely elastic supply of labour (constant marginal utility of income). If in addition we assume that $U_{ij} = 0$ for $i \neq j$ we have the conditions required for the validity of partial equilibrium analysis (no income effects and independent demands). Since[14]

$$U_{kk}\frac{\partial X_k}{\partial q_k} = \alpha \quad \text{implies} \quad H^k = \frac{1}{\varepsilon_k^d}$$

the optimal tax

$$\frac{t_k}{1+t_k} = \frac{\lambda-\alpha}{\lambda}H^k = \frac{\lambda-\alpha}{\lambda}\frac{1}{\varepsilon_k^d} \tag{12-38}$$

as obtained in Section 12-1. This shows that the formula (12-37) may be seen as a "weighted average" of two polar tax systems: the uniform tax and taxes proportional to H^k. Where between these two extremes the optimal tax system depends on H^L.

Secondly, the formulation (12-37) suggests one case where the results may be particularly simple—that where the utility function is directly additive. This implies that there exists some monotonic transformation of the utility function such that $U_{ij} = 0$ for $i \neq j$. Since H^k is invariant with respect to such transformations,[15] this means that

$$H^k = \frac{-U_{kk}X_k}{U_k}$$

But by differentiating the first-order conditions for utility maximization, we can see that this is inversely proportional to the income elasticity of demand for k:

$$U_{kk}\frac{\partial X_k}{\partial M} = q_k\frac{\partial \alpha}{\partial M} = U_k\frac{1}{\alpha}\frac{\partial \alpha}{\partial M}$$

or

$$H^k\frac{1}{X_k}\frac{\partial X_k}{\partial M} = \frac{-1}{\alpha}\frac{\partial \alpha}{\partial M} \tag{12-39}$$

We have therefore the interesting result that *when the utility function is directly additive, the optimal tax rate depends inversely on the income elasticity of demand.* Necessities should be taxed more heavily than luxuries. This has important implications for the conflict between equity and efficiency, which are discussed further below. Direct additivity is a restrictive assumption; it is however considerably less restrictive than the assumptions required for partial equilibrium analysis to be valid (for $H^L \neq 0$, direct additivity does not imply zero cross-price effects). Moreover, direct additivity is assumed in many demand studies, e.g., the linear expenditure system discussed below.

Finally, the primal approach adopted in this section has been used by Deaton (1979) to discuss the conditions under which the optimal structure is uniform. He shows that the optimal tax conditions are identical for all goods if there is implicit separability between leisure and goods; i.e., where the expenditure function can be written $e[w, f(\mathbf{q}, U), U]$. Combined with weak separability between goods and leisure, this implies unitary expenditure elasticities (Sandmo, 1974a).[16] In considering these results, the earlier qualification concerning non-uniqueness of the first-order conditions should be borne in mind: the fact that the right-hand sides of (12-37) may be equal for two goods does not necessarily imply uniformity.

[13] Equation (12-37) can also be obtained from the results of the previous section by inverting Eq. (12-17)—see Atkinson and Stiglitz (1972). For an alternative approach using the Antonelli matrix, see Deaton (1979).

[14] Differentiating $U_k = \alpha q_k$ where α is by assumption constant, and dividing by α.

[15] Suppose U is replaced by $G(U)$; then $G_i = G'U_i$, $G_{ij} = G'U_{ij} + G''U_iU_j$. This means that

$$H^k = \sum_i\left(\frac{-G_{ik}X_i}{G_k}\right) = \sum_i\left(\frac{-U_{ik}X_i}{U_k}\right) - \frac{G''}{G'}\sum_i U_iX_i$$

but the second term disappears (using the budget constraint) establishing that H^k is invariant.

[16] Sandmo shows that it implies equal compensated elasticities with respect to the wage. See also Sadka (1977). The earlier statement in Atkinson and Stiglitz (1972, p. 105) was unclear, although it was not intended to carry the interpretation placed on it by Sadka.

Example of Linear Expenditure System

One function widely used in the empirical study of demand is the Stone–Geary function which generates the linear expenditure system:

$$U(\mathbf{X}, L) = \left\{ \delta^{1-\rho} \left[\prod_{i=1}^{n} \left(\frac{X_i - \gamma_i}{a_i} \right)^{a_i} \right]^{\rho} + (1-\delta)^{1-\rho}(L_0 - L)^{\rho} \right\}^{1/\rho} \quad (12\text{-}40)$$

where $\Sigma a_i = 1$. The parameters γ_i correspond to "committed" consumption, the a_i are share parameters, and ρ measures the ease of substitution between goods and leisure ($\sigma \equiv 1/(1-\rho)$ is the elasticity of substitution). From the first-order conditions for utility maximization and the budget constraint, we obtain the demand functions and the labour supply function:

$$q_k X_k = q_k \gamma_k + a_k \left(wL_0 - \sum_i q_i \gamma_i \right) Z \quad k = 1, \ldots, n \quad (12\text{-}41a)$$

$$wL = ZwL_0 + (1-Z) \sum_i q_i \gamma_i \quad (12\text{-}41b)$$

where

$$Z \equiv 1 - \frac{(1-\delta)^{1-\rho}(L_0 - L)^{\rho}}{U^{\rho}} \quad (12\text{-}41c)$$

(i.e., Z is a measure of the contribution of "goods" to total utility). The expenditure on good k consists of the committed expenditure ($q_k \gamma_k$) plus a fraction, $a_k Z$, of the remaining income, where the latter is defined as "full income", wL_0.

In order to apply the optimal tax formula (12-37), we need to calculate H^k and H^L. It is left as an exercise to the reader to show that

$$H^L = -(1-\rho)\frac{L}{L_0 - L}$$

$$H^k = \frac{X_k}{X_k - \gamma_k} - \rho\xi$$

where

$$\xi \equiv \sum_i \frac{a_i X_i}{X_i - \gamma_i} = 1 + \frac{1}{Z} \frac{\Sigma q_i \gamma_i}{wL_0 - \Sigma q_i \gamma_i}$$

(the last step substitutes for U_i and uses the fact that $\Sigma U_i X_i = \alpha w L$). Substituting in the condition (12-37) for optimal tax rates,

$$\frac{t_k}{1+t_k} = \frac{\lambda - \alpha}{\lambda} \frac{X_k/(X_k - \gamma_k) + (1-\rho)[L/(L_0 - L)] - \rho\xi}{1 + (1-\rho)[L/(L_0 - L)]} \quad (12\text{-}42)$$

In applying this formula so as to arrive at some illustrative calculations of the optimal tax rates, there are two difficulties. First, we require estimates of both commodity demand functions and the labour supply function.

Simultaneous estimation of both is relatively rare, and where it has been undertaken, for example by Abbott and Ashenfelter (1976), it has been on the basis of prior assumptions about the value of ρ (they set $\rho = 0$). In view of this, the calculations below are based on assumed values of ρ. Second, Eq. (12-42) is not an explicit formula for t_k, since X_k, Z and L depend on t_k. These problems are taken up in turn.

Let us first consider the influence of ρ. A useful benchmark is the case of unitary elasticity of substitution ($\sigma = 1$, $\rho = 0$). The relative tax rates on goods k and j are then

$$\frac{t_k/(1+t_k)}{t_j/(1+t_j)} = \frac{X_k/(X_k - \gamma_k) + L/(L_0 - L)}{X_j/(X_j - \gamma_j) + L/(L_0 - L)} \quad (12\text{-}43)$$

In order to get some feel for this, let us suppose that γ varies from 0 (luxury) to $\frac{2}{3}X$ (necessity) and that $L = \frac{1}{2}L_0$. This implies a range of taxes such that the rate on the necessity is double that on the luxury. For elasticities of substitution less than unity ($\rho < 0$), the weight on the term $L/(L_0 - L)$ tends to rise, and the effect of this is increased by the third term in the numerator of the right-hand side of (12-42). Where $\rho > 0$, we would expect the behaviour to depend more critically on its value. Moreover, it is important to note the sensitivity to the precise specification of the substitutability between goods and leisure. If in (12-40) the product Π were replaced by its logarithm, then the term $-\rho\xi$ would not appear on the right-hand side of (12-42). This in turn would imply that, as $\rho \to 1$ (i.e. $\sigma \to \infty$), which is the limiting case of a perfectly elastic labour supply, the relative tax rates would depend on the ratio of $X_k/(X_k - \gamma_k)$. This is the case considered by Atkinson and Stiglitz (1972), where we solve the resulting equations for t_k, using data for the United Kingdom.[17]

The "back of the envelope" calculations above give some feeling for the considerations likely to be important. An alternative approach, leading to precise computations, is that based on algorithms similar to those discussed in Lecture 6. Harris and MacKinnon (1979), for example, make use of demand parameters a_i and γ_i estimated for Canada and postulated values of L_0 and δ. Producer prices are fixed, labour is taken as the untaxed good, and the government has a fixed revenue requirement in labour units. The solution to the first-order conditions is then calculated for different values of σ. The results show, for example, a range of 19.6 per cent (transport) to 21.2 per cent (food) when $\sigma = 0.3$, widening to 18.9 to 26.5 per cent when $\sigma = 1.1$, and 13.9 to 57.5 per cent when $\sigma = 3.0$. For a larger revenue requirement, the range is even wider.

Where analytical conclusions are difficult to obtain, numerical results are undoubtedly valuable. At the same time, it would be a mistake to read

[17] Thus, although the results are a limiting case in the sense that $H^L = 0$, the form chosen for the utility derived from *goods* can be varied to give greater divergencies from uniform taxes.

too much into them, not least because the linear expenditure system is a rather restrictive specification. The main point of calculations of optimal tax rates is to throw light on the role of different considerations and what appear to be the sensitive features. In particular, it has allowed us to illustrative the role of labour supply, and the fact that the results may depend crucially on the specification. It does not however follow that there would be a substantial welfare loss if the function were incorrectly specified (and hence a sub-optimal tax policy chosen). For this, and other reasons, we need to explore the position away from the optimum, and this is the subject of the next section.

12-4 PARTIAL WELFARE IMPROVEMENTS AND TAX REFORM

The literature on optimal taxation has been criticized for directing too much attention at characterizing the optimum and not considering the process by which it can be attained. Feldstein, for example, has distinguished between tax *design*, or "tax laws being written *de novo* on 'a clean sheet of paper'" (quoting Woodrow Wilson), and tax *reform*, which takes "as its starting point the existing tax system and the fact that actual changes are slow and piecemeal" (Feldstein, 1976d, p. 77). We now consider therefore whether we can identify changes in tax rates that represent a partial welfare improvement in that, although falling short of the optimum, they represent a step in the right direction.

Partial Welfare Improvements

As is now well known from the literature on second-best, this is a difficult area. Reforms that may appear to move in the correct direction can turn out on closer inspection to reduce welfare. Intuition can be very misleading. None the less, the optimum tax results discussed in the previous sections provide some insights. For this purpose, we go back to the dual formulation of Section 12-2. In that case we were in effect evaluating possible changes in policy in terms of their effect on the indirect utility function. If we denote a possible variation in tax rates by the vector **dt**, then, by the properties of the indirect utility function,

$$dV = -\alpha \mathbf{X}' \cdot \mathbf{dt} \tag{12-44}$$

The effect of this variation on the revenue is

$$dR = \mathbf{t}' \cdot \mathbf{dX} + \mathbf{X}' \cdot \mathbf{dt} \tag{12-45}$$

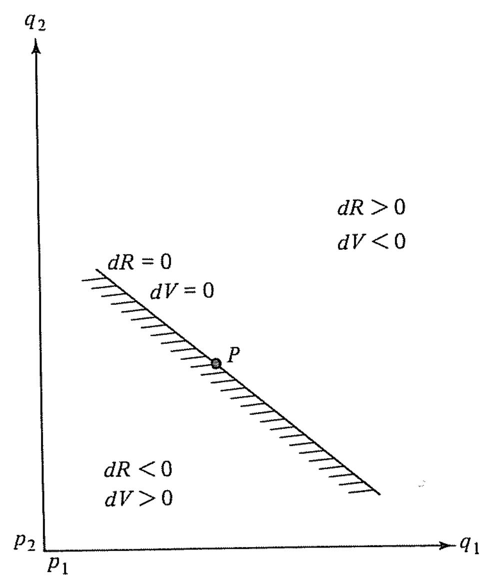
The solution involves identifying conditions under which variations satisfying (12-45) could not achieve an increase in welfare. In geometric terms, the condition for optimality is that the half-space of welfare-improving changes ($dV > 0$) be disjoint from the closed half-space of changes that satisfy the revenue constraint ($dR \geqslant 0$). This interpretation, due to Dixit (1975), is illustrated in Fig. 12-2a. This also brings out why a move towards lump-sum taxes from the distortionary tax optimum raises welfare (in effect $dR < 0$)—as shown by Atkinson and Stern (1974).

What happens however if we are not at the optimum? Is it possible to reach straightforward conclusions about directions for welfare improvement? We consider first a shift from distortionary taxation. Suppose that the government is able to raise revenue by other means and as a result R can be reduced ($dR \leqslant 0$). Does it follow that $dV > 0$? The answer is not necessarily affirmative. This may be seen geometrically in the two-tax case from Fig. 12-2b. The condition $dR \leqslant 0$ defines a closed half-space of *local* revenue-reducing changes, defined by $dR = 0$ in the diagram, with revenue being greater above this line. Correspondingly, we can define the open half-space of local strictly welfare-improving changes ($dV > 0$), with welfare higher below the line $dV = 0$. In the case shown in Fig. 12-2b, it is clearly possible to move from the point P to a new set of taxes where revenue is lower but welfare is lower. This is illustrated by PA. Even though it may appear intuitive that a switch from distortionary to lump-sum taxation raises welfare, this is not everywhere the case. (A formal treatment, using the properties of the expenditure function and allowing for more general assumptions about production is provided by Dixit, 1975. See also Dixit and Munk, 1977.)

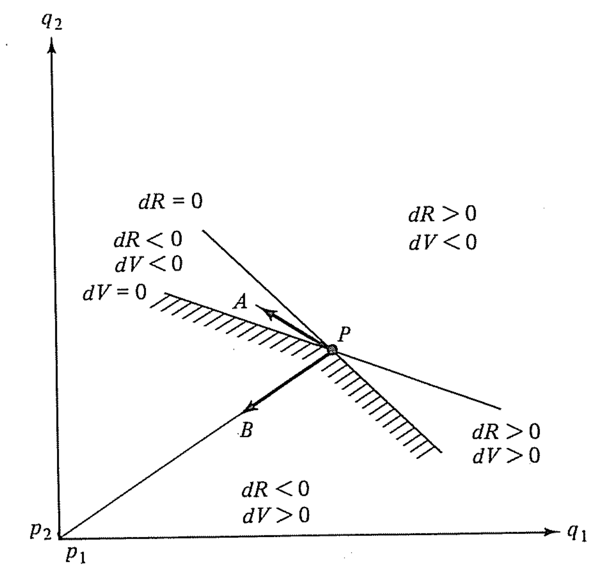
The negative result just described is illustrative of those in the second-best literature which led to a general pessimism. As Dixit notes, "some particular rules that were at one time thought to be intuitively plausible by some economists turned out to be wrong, and this failure received a great deal of publicity" (Dixit, 1975, p. 122). On the other hand, this pessimism does not seem warranted. Even though it is not true, as we have seen, that any move to lump-sum taxation is necessarily welfare-improving, there are many directions in which tax changes may be welfare-improving—the hatched area in Fig. 12-2b. The issue is one of characterizing the directions of feasible welfare-improving change.

In order to illustrate the possibility of constructive rather than negative second-best results, we may note that it can be shown that under rather general conditions a proportionate reduction in the distortion raises welfare (Foster and Sonnenschein, 1970, and Bruno, 1972). To see this intuitively, suppose that all taxes are reduced proportionately with a compensating adjustment in lump-sum taxation, T, to maintain overall revenue. If the proportionate reduction is $db(>0)$, then

$$\mathbf{dt} = -(\mathbf{q}-\mathbf{p})db = -\mathbf{t}db \tag{12-46}$$



(a)



(b)

Figure 12-2 Directions of welfare improvement: (a) existing taxes optimal; (b) existing taxes non-optimal.

The change in lump-sum tax is

$$dT = -\mathbf{t}'\,\mathbf{dX} - \mathbf{X}'\,\mathbf{dt} \tag{12-47}$$

The change in welfare is

$$\begin{aligned} dV &= -\alpha\mathbf{X}'\,\mathbf{dt} - \alpha dT \\ &= \alpha\mathbf{t}'\,\mathbf{dX} \end{aligned} \tag{12-48}$$

(using (12-47)). Now the change in demands consists of a substitution term and an income term. The substitution component is given by

$$\alpha\mathbf{t}'\,S\,\mathbf{dt} = -\alpha\mathbf{t}'\,S\mathbf{t}\,db \tag{12-49}$$

where S is the Slutsky matrix. From the negative semi-definiteness of S, it follows that this component is non-negative. If the income effect were such that the individual became worse off, which could only happen if T rises, this implies that the excise revenue collected from him goes up as his level of welfare falls. If we rule out this apparently perverse case, then the individual must be better off. (For a precise statement of the condition, see Dixit, 1975, p. 107.) Geometrically, the effect is that a move in the direction PB raises welfare (the origin being drawn at $t_1 = t_2 = 0$).

Another example of a "constructive" second-best result is that given by Corlett and Hague (1953). In the context of a simple model, with two consumption goods and labour, the latter being the untaxed *numeraire*, they show that, subject to one qualification, beginning with an initial situation of uniform taxes, welfare can be increased by raising the tax on the good "more complementary" with leisure, while lowering the other tax so that revenue is unchanged. From the earlier condition for an optimum (page 375), this represents a move "towards" the optimum. The qualification is to rule out what they call the "crazy" case, where an increase in the tax rate on one good lowers total revenue. The extension to n goods is left as an exercise.

Exercise 12-3 In a model with n goods and (untaxed) labour, derive the conditions under which a small revenue-neutral departure from uniform taxation increases welfare if all commodities whose prices are lowered are better substitutes for the *numeraire* than all those whose prices are raised. Illustrate geometrically for the case $n = 2$. (See Dixit, 1975, Theorem 6.)

The relation between optimal taxation and tax reform can be considered further. For a number of reasons policy-makers may be unwilling, or unable, to make large changes in the tax structure. The reasons include the fact that our knowledge of the relevant production and demand parameters is typically limited to the neighbourhood of the current position, and even here there may be considerable uncertainty. (A factor working in the opposite direction is that there are fixed costs to tax reform—which would point to infrequent changes.) In view of this, a number of writers have characterized the problem as one of choosing from neighbouring equilibria—or of designing the optimal tax change subject to a constraint on their overall magnitude. Thus Diewert (1978) imposes in effect the constraint

$$\sum_i (\Delta t_i)^2 \leqslant 1 \tag{12-50a}$$

and Dixit (1979) considers

$$\sum_i |\Delta t_i| \leqslant 1 \tag{12-50b}$$

This raises the question of the *process* of tax reform, where there is a clear parallel with the literature on planning algorithms (e.g. Heal, 1973). At each point we need to ask whether there is a feasible, welfare-improving step which can be made; and we need to ask whether the sequence of "tax reforms" converges and, if so, what are the characteristics of the limiting solution. These issues have been discussed by, among others, Guesnerie (1977) and Fogelman, Quinzii and Guesnerie (1978), in the context of a many-consumer model. Among the general features of their results are the difficulties posed by the basic non-convexity of the set of equilibria (already discussed in Section 12-2) and the demonstration that inefficiency in the production sector may be necessary *temporarily* in the process of tax reform. If the process of tax reform is subject to a constraint of the kind described, and is required to be welfare-improving, then the condition of production efficiency that characterizes the full optimum (under certain conditions—see Lecture 15) may not apply on the route to the optimum.

12-5 OPTIMAL TAXATION IN A MANY-PERSON ECONOMY

To this point we have assumed that all individuals are identical, and, as will be argued in Lecture 14, the Ramsey analysis is of limited policy relevance in this context. The extension to a many-consumer economy by Diamond and Mirrlees (1971), developed in Diamond (1975b), Mirrlees (1975) and Atkinson and Stiglitz (1976), is therefore of considerable importance.

Taxation and Redistribution

We now assume that there are H households, denoted by a superscript h, so that the indirect utility function of the hth household is V^h. The objectives of the government are assumed to be represented by maximizing the social

welfare function

$$\psi(V^1, V^2, V^3, \ldots, V^H)$$

where ψ is increasing in all arguments. The government's maximization problem may be formulated in terms of the Lagrangean:

$$\mathscr{L} = \psi(\mathbf{V}(\mathbf{q})) + \lambda\left[\sum_{i=1}^{n} t_i\left(\sum_{h=1}^{H} X_i^h\right) - R_0\right] \tag{12-51}$$

This gives first-order conditions

$$\sum_h \frac{\partial \psi}{\partial V^h}\alpha^h X_k^h = \lambda\left[H\bar{X}_k + \sum_i t_i\left(\sum_h \frac{\partial X_i^h}{\partial t_k}\right)\right] \tag{12-52}$$

where $\bar{X}_k = \sum_h X_k^h/H$. Denoting $(\partial\psi/\partial V^h)\alpha^h$ by β^h (the social marginal utility of income accruing to household h), and using the Slutsky relationship, this may be written as

$$\sum_i t_i\left(\sum_h S_{ik}^h\right) = -\left[H\bar{X}_k - \sum_h \frac{\beta^h}{\lambda} X_k^h - \sum_i t_i\left(\sum_h X_k^h \frac{\partial X_i^h}{\partial M^h}\right)\right] \tag{12-53}$$

In order to help interpret this, let us define

$$b^h = \frac{\beta^h}{\lambda} + \sum_i t_i \frac{\partial X_i^h}{\partial M^h} \tag{12-54a}$$

$$\bar{b} = \sum_h \frac{b^h}{H} \tag{12-54b}$$

b^h is the net social marginal valuation of income, measured in terms of government revenue. It is *net* in the sense of measuring the benefit of transferring \$1 to household h *allowing* for the marginal tax paid on receiving this extra \$1. Equation (12-53) may then be written:

$$\frac{\sum_i t_i \sum_h (S_{ik}^h/H)}{\bar{X}_k} = -\left[1 - \sum_h \frac{b^h}{H}\left(\frac{X_k^h}{\bar{X}_k}\right)\right] \quad k = 1, \ldots, n \tag{12-55}$$

The left-hand side has the same interpretation as before: it is a proportional reduction in the consumption of the kth commodity along the compensated demand schedule. In contrast, the right-hand side is no longer necessarily the same for all commodities. It is independent of k if b^h is the same for all h or if $X_k^h/\bar{X}_k$ is the same for all commodities (there are no goods that are consumed disproportionately by rich or poor). In general, the compensated reduction in demand with the optimal tax structure is smaller:[18]

1. the more the good is consumed by individuals with a high social marginal valuation of income;
2. the more the good is consumed by households with a high marginal propensity to consume taxed goods.

Equation (12-55) can be rewritten in two ways which prove useful in the subsequent discussion:

$$\frac{\sum_i t_i \sum_h S^h_{ik}}{H} = -\bar{X}_k(1-\bar{b}r_k) \quad \text{for } k = 1,\ldots,n \tag{12-56}$$

where

$$r_k = \frac{\sum_h \left(\frac{X^h_k}{\bar{X}_k}\right)\left(\frac{b^h}{\bar{b}}\right)}{H} \tag{12-57}$$

and

$$\frac{\sum_i t_i \sum_h S^h_{ik}}{H} = -\bar{X}_k[(1-\bar{b})-\bar{b}\phi_k] \quad \text{for } k = 1,\ldots,n \tag{12-58}$$

where $\phi_k = r_k - 1$ is the normalized covariance between the consumption of the kth commodity and the net social marginal valuation of income. In the first of these formulae, r_k is a generalization of the distributional characteristic of Feldstein (1972a, 1972b). It shows that, if the average value of the net social marginal valuation, $\bar{b}$, is large, that is there would be large gains from a uniform lump-sum payment, then distributional considerations are to be weighted more heavily.

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 depends on there being either constant returns to scale in production or 100 per cent profits taxes—see Lecture 15.) 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. In order to facilitate this, we assume now that everyone has the same tastes, and that individuals differ solely with respect to their wage rate, w.

[18] Diamond and Mirrlees (1971) derive 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 the comparison with the Ramsey results more direct, we have employed compensated derivatives.

The easiest utility function to consider is the Cobb–Douglas given in Eq. (12-30a). In this case, however, with identical tastes, the optimal tax system is uniform, since $X^h_k/\bar{X}_k$ is the same for all k. Where individuals consume goods in the same proportions, it is not possible to use indirect taxes to redistribute income—they impose the same percentage burden on everyone. In view of this, we consider the more interesting situation of non-unitary expenditure elasticities. This is more complicated, and in order to simplify the analysis we assume that all individuals have identical utility functions:

$$U = \sum_i G_i(X_i) - L \tag{12-59}$$

(N.B. Exercise 12-2 is the constant elasticity version of this form), so that the first-order conditions give

$$\alpha = 1/w^h \quad \text{and} \quad G_i' = (q_i/w^h) \tag{12-60}$$

It follows that there are no income effects on the demand for goods and that the demand schedules are independent. The model is therefore equivalent to the partial equilibrium analysis of Section 12-1. While highly restrictive, it does allow us to examine the consequences of incorporating redistributional goals.

With this special assumption, the condition for optimality becomes

$$\frac{t_k}{1+t_k} = \frac{1-\bar{b}-\bar{b}\phi_k}{\bar{\varepsilon}_k} \quad \text{for all } k = 1,\ldots,n \tag{12-61}$$

where $\bar{\varepsilon}_k$ is the elasticity of the aggregate demand. In the situation where everyone is identical, this reduces to the familiar formula that the taxes should be inversely proportional to demand elasticities. Equation (12-61) provides a simple adjustment to this formula for distributional considerations. The term ϕ_k depends on the social marginal valuation of income received by different households and on the proportion of total consumption that goes to them. In particular, it depends on the degree of aversion to inequality. If β is constant, that is, if 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 that are primarily consumed by those at the top of the scale.

A formula similar to (12-61) was given by Feldstein (1972a, 1972b), but he did not bring out the inherent conflict between equity and efficiency considerations. With this utility function, the demands depend on the ratio of the commodity price to the wage (see (12-60)). 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 the wage, this points to low tax rates for equity reasons. Which of these factors predominates depends on the government's objectives and on the shape of the distribution of abilities. This is illustrated by the simple example where the government maximizes the sum of utilities, the demand curves have constant elasticity, and wage rates are distributed (continuously) according to the Pareto distribution; i.e.,

$$b^h = \alpha^h/\lambda, \quad X_k^h = A_k(w^h/q_k)^{\varepsilon_k} \tag{12-62}$$

and the density function of wage rates is

$$dF = \mu \mathbf{w}^{\mu} w^{-(1+\mu)} dw \quad \text{for } w \geqslant \mathbf{w} \tag{12-63}$$

On the assumption that $\mu > \varepsilon_k$ all k, it may be calculated that

$$r_k = \int_{\mathbf{w}}^{\infty} \left(\frac{X_k^h}{\bar{X}_k}\right)\left(\frac{b^h}{\bar{b}}\right) dF = \frac{\int_{\mathbf{w}}^{\infty} w^{\varepsilon_k - 1} dF}{\left(\int_{\mathbf{w}}^{\infty} w^{\varepsilon_k} dF\right)\left(\int_{\mathbf{w}}^{\infty} w^{-1} dF\right)} \tag{12-64}$$

Hence

$$\frac{t_k}{1+t_k} = -\frac{\bar{b}-1}{\varepsilon_k} + \frac{\bar{b}}{\mu(1+\mu-\varepsilon_k)} \tag{12-65}$$

It follows that, where the government would like to make a uniform lump-sum transfer to everyone ($\bar{b} > 1$), the tax 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 μ, or as the distribution of abilities becomes less unequal (for the same mean, see Chipman, 1974).

Exercise 12-4 Examine the optimal structure of taxation where wage rates are distributed lognormally and the demand functions are of the form given in Exercise 12-2. (See Feldstein, 1972a, and Atkinson and Stiglitz, 1976).

The special case considered above is not of course intended to be realistic, its purpose being to illustrate some of the factors at work. In any actual application, more general demand functions need to be employed, coupled with realistic assumptions about the distribution of endowments and a range of assumptions about the form of the social welfare function. For examples of such empirical calculations, see Deaton (1977), Heady and Mitra (1977), and Harris and MacKinnon (1979).

Horizontal Equity

As noted in Lecture 11, much of the literature on optimal taxation has assumed that the redistributive goals of the government may be represented by maximizing a social welfare function, such as $\psi(\mathbf{V})$ defined above, and has not discussed the relationship between this and the concept of horizontal equity. In the kind of second-best problem we are considering, horizontal inequity is not ruled out by the maximization of a social welfare function. This is a further example of the problems caused by the non-convexities referred to in Section 12-2, and, as shown in general terms in Lecture 11, it may be possible to raise social welfare by taxing identical individuals at different rates. This is discussed with particular reference to indirect taxes in Atkinson and Stiglitz (1976).

It is for this reason that the specification of the problem in terms of each individual facing identical tax rates is an assumption, not an implication of welfare maximization. On the other hand, it may not be an unreasonable assumption. As argued in Lecture 11, the most appealing interpretation of horizontal equity may be that it imposes certain prior constraints on the instruments the government can employ. The constraint that all individuals face the same rates of indirect tax may well appear reasonable in this context, and thus provide a justification for the assumption made (implicitly) in much of the literature.

The introduction of differences in tastes makes the problem even more severe. This is because the social welfare function approach evaluates taxes in terms of the individual's ability to derive utility from goods and leisure, in contrast to the criterion of "ability to pay", which bases taxation on opportunity sets. When the only differences are those in ability to produce, then maximizing ψ 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 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 social welfare function ψ.

In order to illustrate the relationship between these objectives, let us suppose, as in Lecture 11, that tastes may be represented by a single parameter γ, so that the indirect utility function may be written as $V(\mathbf{q}, w, \gamma)$. The social welfare function approach recognizes such taste differences as a legitimate basis for discrimination, and the government maximizes $\psi(\mathbf{V}(\mathbf{q}, w, \gamma))$. 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 for discrimination, then this has two implications. First, it introduces a cardinalization, $V(\mathbf{p}, w, \gamma) = \bar{V}(\mathbf{p}, w)$, so that only endowments (w) and consumer prices (normalized at before-tax levels) are relevant. Second, it constrains the government in levying taxes ($\mathbf{q} \neq \mathbf{p}$) to maintain

$$V(\mathbf{q}, w, \gamma) = \bar{V}(\mathbf{q}, w) \tag{12-66}$$

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 that taxes chocolate ice cream at a higher rate is felt to be horizontally inequitable.[19] This is not however necessarily correct. In Atkinson and Stiglitz (1976), we give an example, based on the independent compensated demands–no income effect case considered earlier, that shows that horizontal equity does not imply uniform taxation where the elasticity of demand differs between the goods in question. The horizontal equity condition (12-66) implies in fact

$$q_1^{1-\varepsilon_1} = q_2^{1-\varepsilon_2} \tag{12-67}$$

where ε_i denotes the (constant) elasticity of demand for good i, and the taste differences are multiplicative (affecting X_1 and X_2 only). The condition for horizontal equity is not necessarily, therefore, uniform taxation; only if the price elasticity is the same—as of course it may be in the chocolate–vanilla ice cream case—would uniform tax rates be horizontally equitable.

Finally, we may note that this example also brings out the conflict between horizontal equity and maximization of a social welfare function. The condition (12-66) is not in general consistent with the maximization of $\Psi(V)$. On the other hand, on the interpretation of horizontal equity as a constraint on instruments—what we have identified as the "means", rather than "ends", approach—there is no necessary conflict. The horizontal equity criterion is logically prior, imposing constraints on the choice of tax policy. On this basis, discrimination against chocolate ice cream lovers would be ruled out *a priori*.

12-6 CONCLUDING COMMENTS

One of the main functions of the second-best literature has been to show that certain common preconceptions about the desirable policy changes are not necessarily correct, and that intuitive arguments based on first-best

[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.

considerations may be misleading. This is well illustrated by the arguments for a uniform structure of indirect taxation. As we have seen, the efficiency argument is far from convincing; nor does horizontal equity necessarily imply a general sales tax.

These counter-examples to conventional wisdom have led to a degree of pessimism about second-best tax policy. This is however unwarranted in the sense that, starting from an arbitrary initial tax structure, there is likely to be a large number of tax reforms that potentially raise welfare. Moreover, it is possible to obtain some insight into the role of different factors, particularly efficiency and equity considerations. At the same time, the characterization of an optimal tax structure, and of the process by which it can be attained, requires detailed investigation of the appropriate model. There are not typically simple rules with wide applicability.

READING

The classic articles are Ramsey (1927) and Samuelson (1951). A basic reference is Diamond and Mirrlees (1971), which stimulated much of the recent interest. A clear introduction to the literature is provided by Sandmo (1976b). The reader may also like to consult, on the material of Section 12-3, Atkinson and Stiglitz (1972); on Section 12-4, Dixit (1975); and on 12-5, Atkinson and Stiglitz (1976).