2

The Theory of Optimal Commodity and Income Taxation: An Introduction

Nicholas Stern

What types of goods should be taxed? How progressive should the income tax be? What should be the balance between the taxation of commodities and the taxation of income? These questions are obviously central to public finance and have occupied many of the leading economists of the last two centuries from Smith, Mill, Dupuit, Edgeworth, and Wicksell to Pigou and Ramsey. The period since 1970, however, has seen a tremendous growth in the formal analysis of these problems, and the main purpose of this chapter is to give the reader an introduction to this newer literature. Much of it has been technical, but I shall try here to offer a broad understanding of the methods of approach, the type of arguments used, and the main conclusions reached. An introduction is provided to some of the theories that later chapters will construct and apply to problems of developing countries.

The models to be examined require modifications and extensions before they can be applied to developing countries. Nevertheless, they do provide a number of fairly general and robust lessons, and I shall pay special attention to their identification. The three purposes of the chapter are, therefore, to provide an introduction to the literature on optimal taxation, bringing out the main elements of the approach; to identify general principles; and to establish a point of departure for many of the theories examined in this book.

The next section contains a description of the approach and scope of the modern theories. We shall indicate briefly some historical antecedents, bring out the important elements of the analysis, and underline some of the assumptions that require relaxation before the theories can be applied directly to developing countries.

This chapter is based on my paper "Optimum Taxation and Tax Policy" (Stern, 1984), which was written during a visit to the Fiscal Affairs Department of the International Monetary Fund in the summer of 1983. I am very grateful to the IMF and to Vito Tanzi, director of the department. I received helpful comments from E. Ahmad, A. B. Atkinson, A. S. Deaton, G. Frewer, D. M. G. Newbery, V. Tanzi, and participants at a seminar in the IMF. Special thanks are due to J. A. Mirrlees. All errors are mine.

In the third section I shall set out some of the main results of the theory of optimal taxation. I begin with commodity taxation and the Ramsey rule for the one-consumer economy and then examine its extension to an economy with many consumers. Optimal income taxation, following the approach of Mirrlees (1971), will be presented. I shall examine the appropriate combination of income and commodity taxation, often discussed under the heading of the balance between direct and indirect taxation. The fourth section discusses the specification and use of a social welfare function, an element that is central to the analysis previously described.

Some applications of the theory to discussions of tax policy are presented in the fifth section. It is shown that the simple principles embodied in this section can be used to discriminate among arguments in discussions of public policy. This is, of course, one of the main purposes of theoretical inquiry in economics—to establish which of the many possible intuitive and informal arguments are well founded. Tax reform (that is, the welfare analysis of a small movement from a given initial position) is briefly introduced in the sixth section, and its close relation to the theory of optimality is shown.

The Scope of Modern Theories

Much of the discussion in the nineteenth century was concerned with the enunciation of general principles to guide tax policy. One example was the argument between those who espoused the benefit principle (which states that those who benefit should pay) and those who argued that taxation should be based on ability to pay. This last concept was itself discussed extensively in terms of whether equal absolute or proportional or marginal sacrifice was appropriate where sacrifice was related to utility of (say) income. The argument included a discussion of whether the base should be income, expenditure, or wealth. For an analysis of this discussion and some of the classic statements, the reader may consult Musgrave and Peacock (1967).

The analysis of the questions of public finance in terms of a collection of principles continues and characterizes much of the literature to the present (see, for example, Musgrave and Musgrave, 1980). The modern theories of public economics have much in common with the traditional approach in that they take up two of the important themes of efficiency and equity. A central feature of the modern approach is that efficiency and equity are defined and combined in terms of the criterion of classical welfare economics, a Bergson-Samuelson social welfare function. They are firmly individualistic in that the behavior of consumers is modeled as utility maximization, and the welfare criterion counts as an improvement any change that makes one individual better off without making someone else worse off. The use and role of the social welfare function are discussed further below.

The unifying features of this approach provide substantial clarity and analytic power. There are, however, many interesting ethical economic issues that it leaves out. The approach is essentially "consequentialist," for example, in that policies are evaluated in terms of their consequences. We might argue that, in taxation as with other things, certain principles should be observed irrespective of their consequences. A case in point might be the kind of information the state should be allowed to use. Second, the consequences are evaluated solely in terms of changes in utility of the society's members. Again, this approach might exclude aspects of the consequences of a particular policy—concerning, for example, the rights that it grants individuals. For further discussions of some of these issues, see Sen and Williams (1982). Many of the questions that we are discussing here, however, concern whether a given rate of tax should be increased or decreased, and in this context the difficulties just raised may not be of overwhelming importance. They should not be dismissed, however, and may have considerable relevance for some aspects of social policy, for example, the question of which instruments of policy are admissible.

The analysis of taxation in the modern theory proceeds by first describing the effects of taxation and then applying criteria (usually a social welfare function) to evaluate those effects. This view splits the subject into a logically prior positive side and a subsequent normative side on which value judgments are introduced. In this chapter I shall concentrate on the normative, but it should be recognized that a large part of modern public economics is concerned with the positive: for example, more than half of a major textbook on the subject (Atkinson and Stiglitz, 1980) is devoted to the analysis of the consequences of taxes, and the analysis of normative issues does not begin until page 331. Examples of the positive issues are (1) the analysis of the consequences of income or wealth taxation for risk taking, (2) how different forms of company taxation will affect investment and the distribution of profits, (3) the effects of national debt and taxation on saving and growth, (4) how incomes of different households or groups are affected by a tax change (that is, by the incidence of a change), and so on. The application of careful and formal microeconomic theory to such questions has been a major feature of modern public economics. Questions of incidence will figure prominently in many chapters of the book.

Clearly, if it is difficult simply to calculate the consequences of policies, then the choice of the optimum among all policies may be intractable. After all, we are then searching through a set of options each of which presents analytical difficulties. The normative part of public economics has thus in the main been concerned with models rather simpler than those used for the analysis of positive questions only. For further discussion of the positive models, see Atkinson and Stiglitz (1980), and see Shoven (1983) for a discussion of applied general equilibrium models.

The modern theory of public economics takes as its point of departure the two basic theorems of welfare economics. The first of these states that a competitive equilibrium is Pareto efficient. The second states that a prescribed Pareto-efficient allocation can be achieved as a competitive equilibrium if prices are set appropriately and lump-sum incomes are allocated so that each

individual can buy the consumption bundle given in the allocation at the prices that will prevail. The important assumptions for the first theorem are the existence of a complete set of markets and the absence of externalities. For the second theorem we require in addition, for private producers, decreasing or constant returns to scale; for consumers, diminishing marginal rates of substitution; and for the government, the ability to arrange lump-sum transfers and taxes. The prescribed Pareto-efficient allocation is often called "the first best," and we say that the assumptions and policy tools of the second theorem allow us to achieve the first best. With the failure of the assumptions or more limited policy tools, we have a problem of the "second best." Occasionally "first best" and "Pareto efficient" are used interchangeably, but it seems preferable to reserve "first best" for the desired Pareto-efficient allocation (that is, the one selected among all those possible) rather than for *any* Pareto-efficient point. Obviously, some Pareto-efficient points may involve very unattractive distributions of welfare.

It is common to regard these results as requiring such restrictive assumptions as to be devoid of practical interest, yet it is remarkable that the first of the theorems is an essential part of the argument of those who argue in favor of the virtues of the market mechanism, and the second provides a valuable framework for public economics in that much of the subject is concerned with the investigation of what the government may do, particularly through taxation, when the assumptions required for the second theorem fail to apply. In this chapter we examine the main results of the part of the investigation that concentrates on the inability to achieve a desirable set of lump-sum transfers.

The tax policies that may appropriately be used to deal with externalities have been extensively discussed in the literature (for a classic statement, see Pigou, 1962). The theory of public sector pricing is close to that of commodity taxation in that the difference between price and marginal cost is analogous to a tax (see, for example, Boiteux, 1956), and our discussion of commodity taxation thus essentially includes the important topic of public sector pricing. Valuable and interesting work in public economics has been done on the problem of measuring marginal cost (see, for example, Drèze, 1964).

Recall that a lump-sum tax on an individual is a payment that the individual cannot alter by any action. Thus a tax on cigarettes is not lump sum because an individual can pay less by smoking less; similarly, a wealth tax is not lump sum because one can accumulate less. Clearly we would want to relate our lump-sum transfers and taxes to individual circumstances, yet at the same time the collection of information for those taxes—for example, data on earning power or wealth—will be such as to prevent them from being lump sum. The individual will discover what is being measured and will usually be able to adjust that dimension if it seems desirable to do so. Note that lump-sum taxes are not, in general, impossible. Differential taxation by sex or height is lump sum if it is assumed that no direct action would be taken to change these characteristics and that there would be no emigration. The difficult problem is the achievement of desirable lump-sum taxes.

This conclusion has implications of two sorts. First—a fairly robust general notion—there is an argument in favor of taxing things that individuals or firms cannot easily vary in response to taxation. An important example would be pure rent or monopoly profits, where these can be identified, as we shall shortly see more formally with regard to a case in commodity taxation. Second, we need a theory that addresses the problem of taxation in a world where lump-sum taxes are limited. This consideration leads us directly to the theory of optimal taxation.

In conclusion, it is interesting to note that much of the argument concerning public sector pricing and taxation that we have just been discussing was set forth by Wicksell in a remarkable article in 1896 (see Musgrave and Peacock, 1967). He notes the importance of marginal-cost pricing in the public sector and the importance of financing losses and other government activities by lump-sum taxation, for example, on land. This approach is linked directly to a Pareto improvement through his notion of unanimity.

The Standard Assumptions

As we have indicated, the theories of optimal taxation may be seen as an examination of the principles of taxation when we rule out lump-sum taxation. In order to focus on this question we retain the other assumptions of the standard competitive model that is used in classical welfare economics. Thus we assume that production takes place in firms in competitive conditions, with profits distributed to consumers; that there are no externalities; and that the price-taking consumers maximize utility. Concerning the tax tools to be examined, we usually make the further assumptions that, in the case of income taxation, income can be observed perfectly and, for commodity taxation, that all goods can be taxed. These assumptions are made for (good) reasons of tractability and to allow us to isolate the impact of the absence of lump-sum taxes.

Some of the other assumptions concerning the model and tax tools are relaxed in the next chapter and in other chapters of the book. Thus in chapter 3 we examine the possibility that distortions in the economy give rise to a divergence between shadow prices and producer prices, and we also consider the effects of taxes in noncompetitive markets. In part 4 the limitations on tax tools play an important role in much of the analysis, as does the explicit recognition that households are producing units as well as consuming units. In this chapter, however, we shall keep everything very simple, concentrating on the effects of taxes on consumers and on revenue. We shall see, however, that the extension of the standard theories of externalities and public goods to second-best economies is relatively straightforward. Production is pushed into the background, using the assumptions of competitive markets and fixed producer prices; in a sense we are assuming that the production side can be perfectly controlled, and we concentrate on the consumption side, where we

suppose information is weaker, thereby (for example) precluding the full set of lump-sum transfers. Production plays a prominent role in the next chapter.

Optimal Taxation

In this section we examine optimal commodity taxation, optimal income taxation, and the optimal combination, in that order. These topics might be seen as indirect taxation, direct taxation, and the old question of indirect versus direct taxation, which are often discussed rather vaguely and are complicated by different definitions of "direct" and "indirect." Under the optimal taxation approach, we specify the tools that are available in a precise and formal way and then analyze how they should be set. Before examining the analysis in detail, we may usefully consider the balance between direct and indirect taxation in general terms. We shall define a direct tax as one that is personal in that the rate(s) depend on the individual or household (varying, for example, according to income, wealth, age, family structure, and so on).

In standard welfare economics, lump-sum direct taxes may be used to achieve the first-best distribution of income for any set of value judgments, and indirect taxes enter the picture only on efficiency grounds, in particular as a way of dealing with externalities. In the one-consumer economy, furthermore, a poll tax will be the best way of raising revenue. We shall assume (reasonably) that lump-sum taxes are not possible and that there are many consumers. We then examine appropriate taxes on commodities and incomes, and we shall find that both efficiency and equity arguments will be involved in the selection of each type of tax. Furthermore, except in special cases, we would want to use both sorts of tax if both are available.

The question of direct taxation versus indirect does, however, remain interesting, because we can still concern ourselves with the adequacy of uniform commodity taxation. If commodity taxation is at a uniform rate, then the same effects can be achieved by a proportional tax on incomes; thus the indirect taxes could, in principle, be replaced by direct taxes. Whether this is the best way to collect them is, of course, another question. Uniformity is a substantive issue quite apart from the balance beween direct and indirect taxes, because it greatly simplifies administration. In the following discussion uniformity will thus be an important theme.

Commodity Taxation

Economists examining optimal taxation where lump-sum taxes are impossible have concentrated on commodity taxation and income taxation. Analysis of the former problem began with Ramsey (1927). Important papers by Boiteux (1956) and Samuelson (1951) were written shortly after the Second World War, but the subject expanded rapidly in the 1970s, following the Diamond-Mirrlees papers of 1971. The subject of optimal income taxation was created by Mirrlees (1971), discussed below. This has been by far the most influential paper in the area because it both specified the problem precisely and analyzed it in considerable depth; see also Fair (1971).

The Ramsey problem is to raise a given revenue from a consumer through the taxation of the commodities consumed in such a way as to minimize the loss in utility that arises from taxation. Ramsey considered the case of one consumer (or equivalently identical consumers who are treated identically), and so we have a simple efficiency problem in that distributional considerations are ignored (a point to which we shall return). Notice that the one-consumer case is a somewhat artificial vehicle; we could and should raise all the required revenue by a poll tax and have zero commodity taxes. It is best seen as an example with which to develop some intuition and some ways of interpreting the many-consumer case.

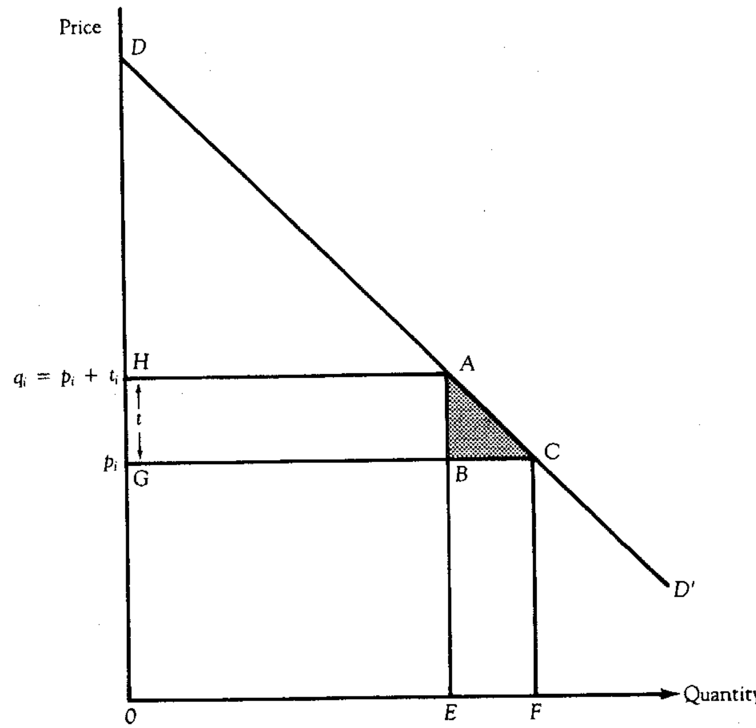
In interpreting the results from the Ramsey problem, and for further reference below, we will find it useful to have in front of us a brief description of the partial-equilibrium approach to the question. These two brief analyses will be used to demonstrate the methods and to develop some intuition that we shall call upon in later arguments. They are, however, obviously very simple and unsatisfactory in a number of ways.

The partial-equilibrium assumption here is that the demand for a good or commodity does not depend on the price of other goods, so that we can draw the familiar demand curve *DD'* (see figure 2-1).[1] We assume producer prices **p** are fixed, so that the effect of a tax vector **t** is to increase prices **q** faced by consumers from **p** to (**p** + **t**). The so-called deadweight loss from the taxation of the *i*th good is measured by the shaded triangle *ABC* in the figure. The motivation for this definition of deadweight loss is as follows: the state of affairs associated with a given tax, and thus consumer prices and demand, is evaluated by the sum of benefits to the consumers (measured by consumer surplus), to the government (measured by tax revenue), and to producers (measured by profits). Note that the sum is unweighted, so that one dollar to each group is regarded as equally valuable.

Profits here are assumed to be zero (producer prices are fixed, so that competition would drive profits to zero), and we therefore consider only consumer surplus plus government revenue. In the absence of taxation, government revenue is zero, and consumer surplus is the area below the demand curve and above the line *GC*. With taxation, government revenue is given by the rectangle *ABGH*, and consumer surplus is the area below the demand curve and above *AH*. The net loss, or deadweight loss, is thus the triangle *ABC*.

Where taxes are zero, then obviously government revenue is zero. When the tax is *GD*, so that demand is zero, revenue is also zero. Hence government revenue *ABGH* has a maximum for some level of tax $\hat{t}$ between zero and *GD*. It will thus never be optimal to have a tax rate above $\hat{t}$, because lowering the rate to $\hat{t}$ will increase both revenue and welfare. This argument has been standard in public finance since Dupuit first made it in 1844 (see Dupuit in Arrow and

Figure 2-1. *Deadweight Loss in Partial Equilibrium*



Scitovsky, 1969) and has recently been reemphasized in discussions of the Laffer curve. It also appears in chapters 14 and 16.

We examine the minimization of the sum across goods of triangles *ABC* (that is, total deadweight loss), subject to the constraint that the sum across goods of the rectangles *ABGH* (total tax revenue) is not less than a given figure. It is straightforward to show that, following this procedure, the tax as a proportion of the consumer price of each good should be inversely related to the elasticity of demand. Formally, $t_i/q_i = \mu/\varepsilon_i$, where μ is constant across goods and q_i, t_i, and ε_i are, respectively, consumer price, tax, and own-price elasticity of demand (as a positive number) for the *i*th good.

Following the work of Harberger (1954), who applied this approach to measure the deadweight losses caused by monopoly (the distance of price above marginal cost playing an analogous role to the tax), the empirical literature has included a number of calculations of such triangles. The more modern approach is to use explicit utility functions and "equivalent variations," thus avoiding the unattractive assumption that the demand for a good does not depend on the prices of other goods (see, for example, Hausman, 1981*a*; King, 1983*a*; and Rosen, 1978).

We now give a brief mathematical formulation of the central result in optimal commodity taxation, the single-person Ramsey rule. This rule dispenses with the partial-equilibrium assumption concerning demands and uses explicit utility functions. To keep things simple, we retain the assumption that producer prices are fixed, so that an increase in taxes implies an equal increase in consumer prices. Goods may be either bought or sold by consumers. Sales are treated as negative purchases. It is convenient to treat the sale of labor differently from the sale of other goods and to identify it separately as l in the utility function and the budget constraint. If w is the wage faced by the consumer, who has lump-sum income M, $\mathbf{x}$ is the vector of quantities transacted, and $\mathbf{q}\cdot\mathbf{x}$ denotes $\Sigma_i q_i x_i$, then the individual problem may be written:

$$\text{(2-1)} \qquad \underset{\mathbf{x},\, l}{\text{Maximize}} \quad u(\mathbf{x}, l) \qquad \text{subject to} \quad \mathbf{q}\cdot\mathbf{x} - w\,l = M.$$

Note that if the prices of all goods and labor are raised by taxation in the same proportion, so that $q_i = (1 + \tau)p_i$ and $w = (1 + \tau)w_p$, where w_p is the wage faced by producers (there is a wage subsidy), then we effectively have a lump-sum tax. The reason is that the proportional change in prices is simply equivalent to a reduction of M to $M/(1 + \tau)$, as may be seen by inspection of the budget constraint in problem 2-1. The revenue is $\tau M/(1 + \tau)$. In the one-consumer economy with lump-sum incomes, this form of taxation would be optimal, provided that the revenue requirement R does not exceed M. In this case, the optimal uniform tax rate τ is given by

$$\text{(2-2)} \qquad \frac{\tau}{1+\tau} = \frac{R}{M}.$$

Where there are no lump-sum incomes, proportional taxes (including the wage subsidy) raise no revenue. If the revenue requirement does exceed M, then distortionary taxes (that is, those not equivalent to lump-sum taxes) will be necessary.

If there are no lump-sum incomes ($M = 0$), then we may choose one good to be untaxed without loss of generality. It is convenient to make that good labor. For a formal discussion of the numeraire, see the appendix to chapter 3. When $M = 0$, the budget constraint is simply

$$\text{(2-3)} \qquad \mathbf{q}\cdot\mathbf{x} = w\,l.$$

Then, for the consumer, a tax rate $\hat{\tau}$ on wage income—reducing the posttax wage to $w(1 - \hat{\tau})$—is equivalent to raising prices to $q/(1 - \hat{\tau})$. We shall assume in what follows in this subsection that there are no lump-sum incomes and that labor is untaxed. Notice that, if we consider leisure L as the argument of the utility function and T as an endowment of time, then equation 2-3 becomes

$$\text{(2-3a)} \qquad \mathbf{q}\cdot\mathbf{x} + wL = wT.$$

In this sense the individual has an endowment of time that finances the purchase of goods and leisure. A lump-sum tax levied from this endowment would be first best, but we have assumed that this choice is impossible.

We consider, then, just one consumer whose individual demands $\mathbf{x}(\mathbf{q}, w)$ are a function of consumer prices only. The maximum utility an individual can achieve when facing prices $\mathbf{q}$ is written $V(\mathbf{q}, w)$: this is the indirect utility function. The problem then becomes to choose $\mathbf{t}$, or, equivalently, $\mathbf{q}$, to maximize $V(\mathbf{q}, w)$ (and thus to minimize utility loss) subject to the constraint that the tax revenue $\Sigma_k t_k x_k$ meets the requirement $\bar{R}$. $\bar{R}$ is the value at $\mathbf{p}$ of the bundle of goods and factors required by the government. We need not concern ourselves with the precise form of the bundle required, because the government can transform its revenue at prices $\mathbf{p}$ into whichever goods it desires.

Formally, then, we have the problem

$$\text{(2-4)} \qquad \underset{\mathbf{q}}{\text{Maximize}} \quad V(\mathbf{q}, w) \qquad \text{subject to} \quad R(\mathbf{t}) = \Sigma t_k x_k \geqq \bar{R}.$$

Taking a Lagrange multiplier for the constraint, λ, the first-order conditions for maximization are

$$\text{(2-5)} \qquad \frac{\partial V}{\partial t_i} + \lambda \frac{\partial R}{\partial t_i} = 0.$$

The discussion of optimal commodity taxation has concentrated very heavily on the interpretation and analysis of first-order conditions such as equation 2-5. The satisfaction of these conditions, however, does not guarantee that we have an optimum. Before we can have such a guarantee, generally speaking, the problem must be concave in the sense of maximizing a concave function over a convex set. The revenue constraint, however, is the product of taxes and demands, and furthermore the concavity properties of demands depend on third derivatives of the utility function. Thus the programming problem will not, in general, be concave. We shall in this chapter (and for much of the book) ignore the problem of establishing global optimality, and we do not usually even check for local optimality (through the second-order conditions). In doing so we follow the literature. It seems that little that is both general and interesting concerning the sufficiency of first-order conditions for global optimality has been established or perhaps is to be expected. The analysis of the first-order conditions can, however, yield valuable insights, and as necessary conditions for optimality, they deserve attention. We now investigate them in more detail.

Remembering that producer prices are fixed so that differentiation with respect to t_i and q_i are equivalent, we have

$$\text{(2-6)} \qquad \frac{\partial V}{\partial q_i} + \lambda\left(x_i + \sum_k t_k \frac{\partial x_k}{\partial q_i}\right) = 0.$$

Using $\partial V/\partial q_i = -\alpha x_i$ where α is the marginal utility of income and the standard decomposition of $\partial x_k/\partial q_i$ into a (symmetric) substitution effect and an income effect ($= s_{ik} - x_i\,\partial x_k/\partial M$), we have the Ramsey rule

$$\frac{\sum_k t_k s_{ik}}{x_i} = -\theta \tag{2-7}$$

where s_{ik} is the utility-compensated change in demand for the ith good when the kth price changes and where θ is a positive number independent of i. It is easy to show that θ is $-\alpha + \lambda(1 - \sum_k t_k\,\partial x_k/\partial M)$, and we explain in the appendix to this chapter that θ may be interpreted in terms of the benefits from a switch to lump-sum finance.

An intuitive interpretation of equation 2-7 is as follows. We can view $\sum_k t_k s_{ik}$ as the (compensated) change in demand for the ith good resulting from the imposition of the vector of small taxes $\mathbf{t}$. The typical term in the sum is

$$t_k\,\partial x_i/\partial t_k\Big|_{\text{constant utility}}$$

which is the change in the compensated demand for good i as a result of the increase in consumer price t_k if t_k is small. Summing across k gives the change arising from the vector of taxes. Strictly speaking, of course, the size of the taxes t_k is determined within the problem, and we are not really justified in assuming that taxes t_k are small. With this qualification, however, according to the Ramsey rule, the proportional reduction in compensated demand that results when the set of taxes is imposed should be the same for all goods.

This result is an important one and provides the main insight into tax rules that arise from the theory of optimal commodity taxation. It should be emphasized that proportional changes in *quantity* are equal in this rule. Thus, crudely speaking, quantities that are relatively insensitive to price will be taxed relatively more. It will be important later in our argument that this notion is, in general, very different from the proposition that taxation should be uniform, that is, that all proportional price changes should be equal. The result provides a generalization of the rule that taxes should be inversely related to elasticities of demand, which is familiar from the less rigorous partial-equilibrium treatment that we have just seen. The Ramsey rule offers an example of the general principle that efficient taxation is directed toward goods that cannot be varied by consumers. Note, however, that we need to exercise considerable care with substitutes and complements, a question that is suppressed by the partial-equilibrium approach. If we do take such care, we come directly to the emphasis on the (compensated) quantity reductions.

Given that labor was assumed to be untaxed and that there is an endowment of time, the Ramsey rule can be interpreted in terms of the complementarity with leisure and substitutability for leisure of the taxed consumer goods; a notable early example was the work of Corlett and Hague (1953). Goods that tend, relatively speaking, to complement leisure should bear the higher tax rate. Thus we can show (Deaton, 1981) that, if leisure is quasi-separable from

all goods, then the Ramsey rule gives uniform taxation of goods. Intuitively, quasi-separability means that all goods complement leisure equally. Formally, goods i and j are quasi-separable from leisure if the marginal rate of substitution between i and j is independent of leisure at constant utility (where compensation for a change in l occurs via a proportional change in the vector $(\mathbf{x}, l)$. Note that the issues of complementarity, substitutability, and separability with respect to leisure arise because there is an untaxed endowment of time, and in a sense we are trying to levy a tax on this lump-sum income. The conclusions would be expressed in terms of another good if there were a corresponding endowment of that good. The concept of complementarity is, of course, simply another way of considering elasticities. With quasi-separability, furthermore, preferences are such that uniform taxation rather than anything else brings about the equal proportional reductions that we are seeking.

As we have noted, in a sense the one-consumer economy is an awkward vehicle for the development of the argument. The reason is that lump-sum taxation (which we know, in general, is first best) becomes simply a poll tax, which, it might be argued, would be feasible. Alternatively, as we saw above when we dealt with fixed lump-sum incomes, the same result may be achieved equivalently through proportional taxation of *all* goods (including subsidies on factor supplies). The real case of interest is, of course, the many-consumer economy, and here the poll tax is, in general, not *by itself* the best way to raise revenue, and indirect taxation will be required. Indeed, in the many-consumer case the optimal poll tax will often be negative, that is, it will be a poll subsidy. Our discussion of the Ramsey rule should therefore be viewed as a development of the intuition for application in the more general case.

The Ramsey rule would seem to be rather inegalitarian in that it appears to direct commodity taxation toward "necessities," which we usually consider fairly insensitive to price. Still, by formulating the problem in terms of one consumer, we explicitly ignore distributional questions. The result can, however, be generalized to many consumers in a fairly straightforward way. We simply replace $V(\mathbf{q}, w)$ in problem 2-4 with the social welfare function $W(u^1, u^2, \ldots, u^H)$, where u^h is the utility function of the hth individual, which we again consider as a function of consumer prices $\mathbf{q}$ and the wage w^h. The total demand $\mathbf{X}(\mathbf{q}, w)$ is $\sum_h \mathbf{x}^h(\mathbf{q}, w^h)$ where $\mathbf{x}^h(\mathbf{q}, w^h)$ is the demand function for individual h. The rule, then, is no longer that the proportional reduction in compensated demand should be the *same* for all goods or commodities; our modification shows how it should vary across goods. The proportional reduction in quantity for a good should now be higher where the share of the rich in its total consumption is higher. Strictly, we are using "the rich" here to designate those people whose social marginal valuation of income is low. Following an argument similar to that used in the derivation of the Ramsey rule equation 2-7, we can show

$$\frac{\sum_h \sum_k t_k s_{ik}^h}{X_i} = -(1 - \bar{b}\,r_i), \tag{2-8}$$

where s_{ik}^h is the Slutsky term for household h; $\bar{b}$ is the average across households of b^h, the *net* social marginal valuation of income of household h; and r_i is the normalized covariance between the consumption of the ith commodity and the net social marginal valuation of income plus one (and is defined formally in equation 2-20 below.)[2] By *net* we mean the value of an extra dollar to individual h as perceived by the government plus any extra indirect tax revenue arising from the expenditure of the dollar (formally $b^h = \beta^h/\lambda + t \cdot \partial x^h/\partial M^h$ where β^h is the social marginal utility of income, M^h is lump-sum income, and λ is the Lagrange multiplier on the revenue constraint). The number r_i is a generalization of the distributional characteristic of a good introduced by Feldstein (1972) and indicates the (relative) extent to which a good is consumed by individuals with a high net social marginal valuation of income—the interpretation as the distributional characteristic is explained when the equation for r_i is presented below (equation 2-20).

Thus the proportional reduction of compensated demand denoted by the left-hand side of equation 2-8 embodies the efficiency arguments for taxing necessities introduced in the Ramsey rule, together with the distributional judgment as associated with the r_i on the right-hand side, which indicates taxation of luxuries. The implication of equation 2-8 for tax rates will depend on the way in which these two effects combine. Much will depend on the structure of preferences and the type of income tax tools available, as we shall see.

Public Goods and Externalities

Where lump-sum taxes are possible, a standard first-order condition for the optimality of the level of public good supply is that the sum across consumers of the marginal willingness to pay should be equal to the marginal cost. A crucial feature is that the sum across consumers is unweighted—because lump-sum taxes have been set optimally, the social marginal utility of income (β^h) is the same for all households. Where lump-sum taxes are not possible, the standard condition is modified in two ways. First, we must weight the willingness-to-pay by the β^h. Second, we must take into account the effect of an extra unit of the public goods on tax revenue. An improvement in the extent or quality of public broadcasting, for example, may produce an increase in the purchase of taxed radios. The first-order condition for the kth public good, level e_k, when revenue is raised by commodity taxes is, therefore,

$$\sum_h \beta^h \rho_k^h = \lambda p_k - \sum_i t_i \frac{\partial X_i}{\partial e_k} \tag{2-9}$$

where ρ_k^h is the marginal willingness to pay on the part of the hth household, p_k is the marginal cost at producer prices, and $\partial X_i/\partial e_k$ is the effect on total demand for private good i of a marginal increase in k (see Diamond and Mirrlees, 1971, for a formal derivation).

A similar analysis holds for externalities, as we should expect, inasmuch as

public goods are a special case. The standard Pigovian argument for the case where lump-sum taxes are possible is that, for any given private good, each household should face a household-specific tax equal to the marginal diseconomy inflicted on others, measured by their marginal willingness to pay to avoid an increase. Where lump-sum taxes are not possible, the appropriate marginal tax is given by the sum of the marginal willingness to pay weighted by β^h plus any marginal loss to tax revenue associated with demand shifts that follow from an increase in consumption of the private good.

Considered as a whole, the argument underlines two important general principles associated with policy analysis in economies without lump-sum taxes: (1) marginal costs or benefits have to be weighted by social marginal utilities of income and (2) the effect of policy changes on tax revenues through demand shifts must be treated explicitly.

Income Taxation

Both Adam Smith and John Stuart Mill argued that taxation should be linked to ability to pay, with the former stating, "Subjects should contribute in proportion to their respective abilities," and the latter arguing, "Whatever sacrifices the [government] require . . . should be made to bear as nearly as possible with the same pressure upon all."[3] The form that pressure should take was extensively discussed, and conclusions often reflected a notion of cardinal utility, linking income to some utility level. At various points it was suggested that the sacrifice of utility should be equal for everyone or that an equal proportion of utility should be sacrificed. Given a utility function (assumed to be the same for everyone) and one of these principles (say, equal absolute sacrifice), we can calculate a corresponding tax function. If income is Y and the tax payable is $T(Y)$, then, given some total revenue requirement, we can calculate T, assuming that Y is independent of the tax schedule, for each level of Y from

$$U(Y) - U(Y - T) = \text{constant}, \tag{2-10}$$

the condition for equal absolute sacrifice. For calculations in this framework, see Stern (1977). We can show, for example, that if $U'(Y) = Y^{-\varepsilon}$, then taxation is progressive (in that the marginal exceeds the average rate) for $\varepsilon > 1$. The logarithmic or Bernoulli form corresponding to $\varepsilon = 1$ gives proportional taxation.

These criteria are adduced, however, without any reference to guiding principles. From this point of view, the notion of "equal marginal sacrifice" set forth by Edgeworth has greater clarity in that it derives from the utilitarian objective of the sum of utilities. If we assume again that pretax income is independent of the tax schedule, and further that everyone has the same strictly concave utility function, then equal marginal utility implies equal posttax incomes. The marginal tax rate is thus 100 percent, casting the question of incentives in a very stark light. This problem of incentives was recognized very early in the discussion: see, for example, McCulloch (1845/1975), part 1, chapter 4: "Graduation is not an evil to be paltered with. Adopt and you will effectively paralyze industry. . . . The savages described by Montesquieu, who, to get at the fruit cut down the tree, are about as good financiers as the advocates of this sort of taxes" (p. 146).

The incentive and distribution aspects of the income tax have long been recognized. Perhaps surprisingly, a model that examined the distribution and the size of the cake simultaneously did not appear until 1971 (Mirrlees, 1971). The Mirrlees paper essentially created the subject of optimal income taxation. As we shall see, Mirrlees kept his model as simple as it could possibly be, given the issue at hand, but the problem is nonetheless not easy, and the analysis poses considerable technical difficulties. The reason is that the policy tool is the whole income tax *function*. Thus for each income we must specify the tax payment, and the optimization occurs in a space of all admissible functions. This situation contrasts with that manifested in the problems usually examined in standard microtheory (for example, consumer or producer behavior), where only a finite number of variables—for example, consumption of each type of good—is considered.

We can simplify the income tax problem considerably by confining our attention to a linear tax schedule that combines a lump-sum benefit or tax with a constant marginal tax rate. After the Mirrlees article was published, a number of papers examined the simpler problem (see Atkinson and Stiglitz, 1980, lecture 13, for references). In the discussion of numerical results, I shall concentrate on the linear case but shall begin by setting forth the Mirrlees nonlinear problem, explaining why it takes the form that it does. The main results for the nonlinear problem are then summarized, with numerical calculations for the linear case presented in conclusion to illustrate the sensitivity to the important parameters and in order to compare the computed tax rates with levels we see in practice. Some work has recently been done on an intermediate case with a finite number of individuals—we might interpret them as representative of certain groups—where the optimal tax schedule can be taken as piecewise linear (see Guesnerie and Seade, 1982; and Stern, 1982).

Given that the nonlinear problem will pose difficulties, it is sensible for us to begin by keeping the structure as simple as we can while retaining the question. Considered from this point of view, the model concerned with distribution and incentives must have two features: individuals should not be identical, and there must be an input over which individuals exercise choice. If individuals were identical, then the optimum would be given by a poll tax with zero marginal taxation (this is the standard result of welfare economics); if there were no incentive problem, as we have seen in our discussion of Edgeworth above, the marginal rate would be 100 percent. The Mirrlees model has individuals who differ only in their pretax wage or productivity, and incentives have only one aspect, labor supply. Thus in the model, labor is supplied by individuals, each of whom has an identical utility function in order to maximize the utility of consumption and leisure, in view of the pretax wage and the income tax schedule. The government chooses the income tax schedule so as to maximize a Bergson-Samuelson social welfare function, subject to the raising of some given amount of revenue.

All individuals have the same utility function $u(c, l)$, which depends on consumption c and labor supply l. Individuals differ in their wage rates w, and the distribution of w is described by the density function $f(w)$. We speak of an individual as being of type w. The government knows the distribution of w but cannot identify the w associated with a particular individual. If it *could* do so, then the optimum would be of the standard first-best type, with the lump-sum tax a function of w.

The problem is to choose a function $g()$ that relates posttax to pretax income in order to maximize

$$\int \phi(u)\, f(w)dw \tag{2-11}$$

subject to

$$\int [wl - g(wl)]\, f(w)dw = R \tag{2-12}$$

where (c, l) is chosen by the individual to maximize $u(c, l)$ subject to

$$c = g(wl) \tag{2-13}$$

and where $\phi(u)$ is a monotonic transform of utility (we introduce ϕ to provide a way of discussing different attitudes to inequality). The government revenue requirement is R, which is seen as fixed for the problem 2-11–2-13. At a later stage we shall ask how the solution varies with different values of R. The maximand 2-11 is a Bergson-Samuelson social welfare function of additive form; we add $\phi(u)$ across individuals. Equation 2-12 represents the revenue requirement; wl is pretax income, so that $wl - g(wl)$ is the tax payment by individual type w, and this is integrated or added across individuals. The constraint 2-13 represents the second-best nature of the problem, that is, it says individuals make their own choice subject to the budget constraint set by their wage and the government tax function. We can express this idea by saying that no individual would prefer the pretax income of some other individual (pretax income is essentially effort, which individuals select for themselves).

Before proceeding to results, we should note some particular features of the model. First, as specified, the model is static, and there is no saving. This provision keeps the structure as simple as possible. From a broader perspective, we might regard l as representing lifetime labor supply and c as lifetime consumption, but the treatment of l as a vector would take us too far afield at this point. We shall shortly consider a vector of different consumption goods, however. Second, equation 2-12 may be replaced in a general equilibrium framework by a production constraint, namely that total production, a function of total effective labor $\int (lw)\, f(w)\, dw$, must equal total consumption $\int cf(w)\, dw$ plus R. Here we assume that w measures productivity so that wl is effective tasks performed by person type w in hours l. It is then straightforward to show that this general equilibrium model is equivalent to the problem 2-11–2-13. Notice that relative wages and effective hours or tasks per clock hour are exogenous, so that the tasks performed by different individuals are perfect substitutes.

Third, note that the constraint 2-13 gives the model its special structure in that it embodies the incentive constraint. Without this constraint we could go to the first best using lump-sum taxation. It is interesting in this context that the first-best optimum would have utility *decreasing* in ability w if consumption and leisure are normal goods (see Mirrlees, 1979). Intuitively, high lump-sum taxes on people who are able would lead to work concentrated on the most productive people (note that there is no difference between individuals on the consumption side). In the income tax model we assume explicitly that the government *cannot* distinguish between types of individuals and measures only an individual's income (not hours of work or wage). Thus with the constraint embodied in equation 2-13, we must have utility nondecreasing in w—an individual of higher w always has the option of consuming the same amount as an individual of low w while doing less work.

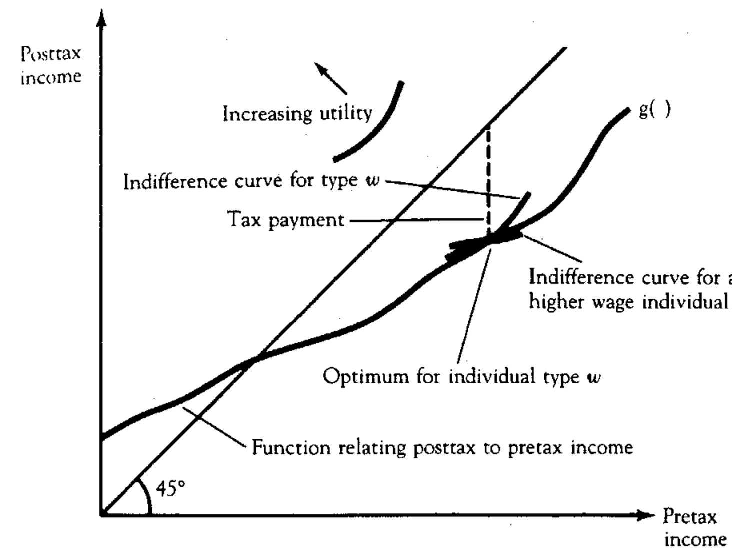
Fourth, we cannot guarantee that at the optimum $l(w) > 0$ for all individuals. Thus it may be optimal for some group of individuals with the lowest productivity to do no work.

We turn now to some results in the Mirrlees model of nonlinear taxation. The number of general results (in the sense that they are independent of functional forms) that are available are rather few. Moreover, these results themselves may not hold if we modify the model, for example to include complementarities between different types of labor (see, for example, Stern, 1982). The important ones in the Mirrlees model are (1) the marginal tax rate should be between zero and one; (2) the marginal tax rate for the person with the highest earnings should be zero; and (3) if the person with the lowest w is working at the optimum, then the marginal tax rate he or she faces should be zero.

I will not offer formal proofs of these propositions here but will give some intuitive arguments (see Mirrlees, 1971, and Seade, 1977, for the formal treatment). Let us consider first whether the marginal tax rate should ever exceed one. The implication would be that the reward for the marginal hour was negative. Hence in the model, no one would choose to work where the marginal tax rate exceeded one, and so we could replace any portion of the $g()$ function that slopes downward with a horizontal section without changing behavior (see figure 2-2), and we can confine our attention to tax schedules with marginal rates that do not exceed one.

We have illustrated the tax function and consumer choice in figure 2-2. For an individual with fixed w, we can draw indifference curves in the pretax-posttax income space, because the former represents work and the latter consumption. Through any point, the indifference curve for a person with higher w will, we suppose, be less steep than that for a person with lower w: at the given consumption level, the person with higher w is doing less work and

Figure 2-2. *The Budget Constraint and Choice of Hours of Work*



Source: Stern (1984), pp. 339–78.

will thus need less extra consumption to compensate him or her for doing the (lower) amount of extra work required for the extra dollar. In general, then, a person with higher w will locate to the right of (will earn more money than) the person with lower w, because at the optimum for the person with lower w [tangency with $g()$], the indifference curve for the person with higher w will intersect g from above (coming from the left).

The tax payment is given by the vertical distance from $g()$ to the 45 degree line. Note that a movement of an individual parallel to the 45 degree line keeps revenue constant. It is possible to use this feature to show that the marginal tax rate cannot fall below zero. If it were below zero at some income, then $g()$ would be steeper than 45 degrees, as would the indifference curve of any individual choosing that income. In this case, $g()$ would be steeper than 45 degrees and, intuitively, an equal revenue shift of person w in the southwest direction would take that person to a higher indifference curve.

An intuitive argument for the second result would be the following. Suppose, with some given income tax schedule, that the person with highest income earns \$Y pretax and the marginal tax rate is positive. Consider the option of lowering the marginal tax rate to zero for all incomes above \$Y. The top person may now decide to work more (the reward for the marginal hour has gone up) and, if so, will be better off. The government has lost no revenue, because the tax payment on the income \$Y has stayed constant. The utility of the top person has increased, that of others is no lower, government revenue is no lower, and we have therefore found a Pareto-improving change that meets the constraints. Accordingly, the given income tax schedule could not have been optimal, and the schedule that is optimal must have the property that the marginal tax rate at the top is zero. If people near the top elect to work more in response to the change, then they are both better off and pay more tax, so that the argument is reinforced.

We cannot deduce that, where there is no highest income and the distribution of skills includes individuals at or above any positive skill levels, the optimal tax rate tends to zero. There are cases (see Atkinson and Stiglitz, 1980, lecture 13; and Mirrlees, 1971) in which it does not (involving the Pareto distribution). We should remember, too, that the argument assumes that there are no externalities, so that making the top individual better off upsets no one. Furthermore, the "top" may be a very high level of income. Zero may be a poor approximation even within most of the top percentile. Nevertheless, the result is rather striking.

I shall not give the argument for the third result concerning the zero marginal rate at the bottom in any detail. It proceeds along the following lines. Suppose that on a given schedule the marginal rate at the bottom is greater than zero. Consider a change in the lower end of the tax schedule that has the sole effect of inducing the bottom person to do a little more work, thus moving a small amount along the schedule. To the first order in utility, that person is no worse off, because his indifference curve was tangential to the schedule. There is a first-order increase in tax revenue, however, because the marginal rate is positive. Hence, the given schedule is not optimal (for formal discussion of this result and the previous one, see Seade, 1977).

Thus, the general results in this particular model tell us that the marginal rate should be zero at the top and bottom. This finding contrasts strongly with the workings of many tax-cum-social-security systems, and we shall briefly return to this issue later.

Mirrlees (1971) presented a number of numerical calculations of the optimum nonlinear income tax, using the Cobb-Douglas utility function for consumption and leisure and wage distributions based on data for the United Kingdom. From these examples he concluded:

1. The optimal tax structure is approximately linear, that is, a constant marginal tax rate, with an exemption level below which negative tax supplements are payable.
2. The marginal tax rates are rather low ("I must confess that I had expected the rigorous analysis of income taxation in the utilitarian manner to provide arguments for high tax rates. It has not done so" [Mirrlees, 1971, p. 207]).
3. "The income tax is a much less effective tool for reducing inequalities than has often been thought" (Mirrlees, 1971, p. 208).

Ten years ago (Stern, 1976) I investigated a wider class of utility functions and in addition examined sensitivity with respect to the social welfare function and the level of government revenue, confining attention to linear taxation. I used the constant elasticity of substitution utility function

$$u(c, l) = [\alpha (1 - l)^{-\mu} + (1 - \alpha)c^{-\mu}]^{-1/\mu} \quad (2\text{-}14)$$

with welfare criterion

$$\frac{1}{(1 - \varepsilon)} \int_0^\infty [u(c, l)]^{1-\varepsilon} f(w)\, dw. \quad (2\text{-}15)$$

The tax function in the model is linear, so that the individual budget constraint is

$$c = (1 - t)wl + g \quad (2\text{-}16)$$

where t is the marginal tax rate and g the lump-sum grant (the same for everyone). The government budget constraint is

$$t \int wlf(w)dw = g + R \quad (2\text{-}17)$$

where, as before, R is an exogenous revenue requirement, and where the number of individuals is normalized to 1 so that g is the total payment on lump-sum grants.

The utility function of equation 2-14 has an elasticity of substitution between consumption and leisure of

$$\sigma = \frac{1}{1 + \mu}. \quad (2\text{-}18)$$

We may use empirical estimates of labor-supply functions to estimate σ, and these suggest a number near 0.4 on the basis of estimates for married males in the United States (Stern, 1976, p. 136). Where the elasticity is less than one, the labor-supply function (for positive g) is forward sloping for low wages and backward sloping for higher wages. Notice that the concept of labor supply in the models is much broader than the simple measure of hours used in the estimation of short-run supply functions. The Mirrlees labor-supply function corresponds to the limit as σ tends to 1 (μ tends to zero), and $\sigma = 0$ gives right angle indifference curves (zero substitution effect). We can show generally that with $\sigma = 0$ the optimal marginal rate is 100 percent. Note that this is zero *compensated* elasticity of labor supply and *not* inelastic labor supply. A selection of the results appears in table 2-1.

We may regard ε as analogous to the elasticity of the social marginal utility of income that is often used in analyses of measures of inequality using the Atkinson index (see Atkinson, 1970), because the utility function is homogeneous degree 1 in consumption and leisure (doubling each would double utility) and is thus itself analogous to income. The specification of ε then completes the statement of distributional value judgments. Note, how-

Table 2-1. *Calculations of Optimal Linear Marginal Tax Rates*

σ	$\varepsilon = 0$ t	$\varepsilon = 0$ g	$\varepsilon = 2$ t	$\varepsilon = 2$ g	$\varepsilon = 3$ t	$\varepsilon = 3$ g	$\varepsilon = \infty$ t	$\varepsilon = \infty$ g
$R = 0$ *(purely redistributive tax)*								
0.2	36.2	0.096	62.7	0.161	67.0	0.171	92.6	0.212
0.4	22.3	0.057	47.7	0.116	52.7	0.126	83.9	0.167
0.6	17.0	0.042	38.9	0.090	43.8	0.099	75.6	0.135
0.8	14.1	0.034	33.1	0.073	37.6	0.081	68.2	0.111
1.0	12.7	0.029	29.1	0.062	33.4	0.068	62.1	0.094
$R = 0.05$ *(equivalent to about 20 percent of GDP)*								
0.2	40.6	0.063	68.1	0.135	72.0	0.144	93.8	0.182
0.4	25.4	0.019	54.0	0.089	58.8	0.099	86.7	0.139
0.6	18.9	0.000	45.0	0.061	50.1	0.071	79.8	0.107
0.8	19.7	0.000	38.9	0.042	43.8	0.051	73.6	0.082
1.0	20.6	0.000	34.7	0.029	39.5	0.037	68.5	0.064
$R = 0.10$ *(equivalent to about 45 percent of GDP)*								
0.2	45.6	0.034	73.3	0.110	76.7	0.119	95.0+	—
0.4	35.1	0.000	60.5	0.065	65.1	0.076	89.3	0.112
0.6	36.6	0.000	52.0	0.036	57.1	0.047	83.9	0.081
0.8	38.6	0.000	46.0	0.016	51.3	0.026	79.2	0.057
1.0	40.9	0.000	41.7	0.002	47.0	0.011	75.6	0.039

Notes: In each pair of columns, the first entry is the marginal tax rate (percent) and the second is the lump-sum grant. $\varepsilon = 0$ corresponds (roughly) to an absence of aversion to inequality in incomes and $\varepsilon = \infty$ to the Rawlsian maxi-min. In Stern (1976), $(1 - \varepsilon)$ is used in place of ε. A central estimate of the elasticity of substitution σ might be 0.4. Total output in these models is about 0.25 (it is endogenous), so that $R = 0.05$ corresponds to government spending (excluding transfer payments) of about 20 percent of gross national product. Where the optimal tax rate is above 95 percent, the precise level and g were not calculated (a dash is shown for g). The level of the uniform grant g was not presented in the Stern (1976) tables. It satisfies the government budget constraint 2-17, that is, $tY = g + R$, where Y is national income per capita. Hence Y can be calculated from the values of t, g, and R presented.

Source: Based on Stern (1976), table 3.

ever, that the social marginal utility of income is independent of g for $\varepsilon = 0$ but does depend on w, so some redistribution is still desirable even in this case. Values of ε between 1 and 2 are quite commonly used. Dalton (1922/1967, pp. 68–69) argued that Bernoulli's law (or utility logarithmic in income and marginal utility, decreasing as the inverse of income), $\varepsilon = 1$, "gives a rather slow rate of diminution of marginal utility," and he saw $\varepsilon = 2$ "as best combining simplicity and plausibility" (although he was working in the context of equal absolute sacrifice). Whether these views of ε helped him when he subsequently became chancellor of the exchequer is a matter for speculation.

National product in the model is endogenous but is mostly near 0.25. Hence a revenue requirement of R of 0.05 corresponds to about 20 percent of GNP. The case $\varepsilon = 2$, $R = 0.05$, and $\sigma = 0.4$ gives a marginal tax rate of 54 percent. The expenditure of the 54 percent of GNP consists of 34 percent for transfer payments and 20 percent for expenditure on goods and services. These results

are not wildly out of line with tax rates (taking direct and indirect together) from a number of developed countries. Hence if we consider a wider class of cases than those used by Mirrlees, the computed tax rates may be rather higher.

Generally, the tax rates increase with the aversion to inequality ε and with the revenue requirement R but decrease with σ, the elasticity of substitution. The value of the (uniform) lump-sum grant moves in the same direction as tax rates but much more sharply. Thus, for example, in the case $\varepsilon = 2$, $R = 0.05$, the tax rate is halved as we move from σ of 0.2 to σ of 1.0, but the lump-sum grant is divided by four. For low values ($\varepsilon = 0$) of inequality aversion, the grant becomes very small (for example, zero to three significant figures for $R = 0.05$, $\sigma = 0.8$). The lump-sum grant is the money income of the poorest person (with zero wage). Thus it is never optimal to have negative grants; the poorest individuals could not then survive.

The Combination of Income and Commodity Taxes

The question of the appropriate combination of income and commodity taxation provides fertile ground for confusion. Many economists have claimed that the allocation effects of indirect taxes are inferior to those of direct taxes. The contention in its simple form is mistaken, because there is an excess burden or deadweight loss associated with the divergence between consumer and producer prices for labor and thus with the income tax, just as with other goods. A second example concerns the claim that we often hear, that a switch from income tax to indirect taxes such as a value-added tax would increase work effort. At the simple level, this statement is clearly false: an increase in prices (from the VAT), together with an increase in earnings (from the reduction in income tax), would leave the incentive to work unchanged. Perhaps the argument is intended to be subtler, depending on intertemporal allocations and expectations—on progressivity or on the existence of lump-sum incomes, for example—but it is usually presented in naive forms, such as "taxing spending rather than earning induces work."

As it happens, we can show that, under certain conditions, it would be desirable to tax income rather than goods, but those conditions are very special. The argument depends critically on particular features of the model and involves some difficulty. Furthermore, it is not easy to determine how the obvious divergence of the world from these special conditions should influence our views on the balance between direct taxation and indirect taxation. Thus, at least formally, the subject involves difficulty in analysis, and interpretation is not straightforward. We must beware of simple arguments or contentions such as the ones described.

I shall not present the details of the theorems on the optimal combination of income and commodity taxes but shall instead highlight the importance of the assumptions and give an intuitive feel for the results. There are essentially two theorems: the first deals with the case in which there is a linear income tax, and the second, the case in which there is a nonlinear income tax.

Note that if individuals are identical, then the basic theorem of welfare economics tells us that the first best can be reached with a lump-sum tax to raise the required government revenue and zero marginal taxation of income and goods. With different individuals, then, some combination of income and commodity taxes will be necessary, and each of these is distortionary in that marginal rates of substitution between labor and goods or among goods in consumption will not be equal to marginal rates of transformation in production. Notice that some distortionary taxation will always be optimal in second-best problems, because a marginal imposition of taxes from a position of zero taxation involves zero deadweight loss and will be desirable if it improves distribution.

For the first theorem we assume that a linear income tax is available in the form of a lump-sum grant or tax (the same for everyone) and a constant marginal rate on labor income. As we saw previously, a constant marginal rate on labor is, in this context, equivalent to a proportional tax rate on all goods (and a proportional adjustment to the lump-sum grant/tax), because we assume that there are no sources of income other than the lump-sum grant/tax and wages.

The first-order conditions for the optimal indirect taxes are given, as before, by equation 2-8. The condition for the optimality of the lump-sum grant is that $\bar{b} = 1$, that is, the grant is adjusted to the point where the benefit in terms of social welfare of the marginal dollar (the average of the social marginal utilities of income) is equal to the cost to the government (one dollar). Substituting this condition in equation 2-8 gives us

$$\sum_h \frac{\sum_k t_k s_{ik}^h}{X_i} = -(1 - r_i) \tag{2-19}$$

where

$$r_i = \sum_h \frac{x_i^h}{X_i} \cdot \frac{b^h}{\bar{b}}. \tag{2-20}$$

Recall that r_i is 1 plus the normalized covariance between consumption of the ith commodity by the hth household and the net social marginal utility of income b^h. From equation 2-20 we think of r_i as the distributional characteristic of good i: it is the weighted average of the x_i^h with weights $b^h/\bar{b}H$), divided by the average (X_i/H), and it measures the extent to which the ith good is consumed by people with a high net social marginal utility of income. If the government is indifferent to distributional considerations in that it sees b^h as equal for all households, then r_i will be equal to 1, and the right-hand side of equation 2-19 will be zero. Indirect taxes are zero and all revenue is raised through the lump-sum grant (tax), as in the case of identical individuals. Thus in this sense indirect taxes are desirable *because* distributional considerations arise: they may be seen as financing reductions in the poll tax (which bears most heavily on the worst off)—although it must be remembered that this consideration applies where the poll tax is feasible, and we are thinking of a combination of poll tax and indirect taxes.

We have seen that indirect taxes appear because we are interested in distribution, but this observation does not tell us what form the indirect taxes should take. The taxation of goods consumed by the rich provides some progressivity, but indirect taxes also have the effect of raising revenue to increase the progressive lump-sum grant (or to reduce the regressive tax), and the taxation of necessities may be an efficient way to raise this revenue (as in the Ramsey case). The way in which these two considerations balance each other depends quite critically on the form of the differences among the population and on the structure of demand functions, as we see from the first of the theorems below.

If we have an optimal linear income tax, individuals have identical preferences but differ in the wage rate, and the direct utility function has the Stone-Geary form, which gives rise to the linear expenditure system

$$U(\mathbf{x}, l) = \sum_{i=1}^{n} B_i \log (x_i - x_i^0) + B_0 \log (l_0 - l) \tag{2-21}$$

then the optimal indirect taxes are uniform, that is, the proportion of tax in consumer price (t_i/q_i) is the same for all goods. The result follows from equations 2-19 and 2-20, using $\bar{b} = 1$ and substituting for the specific form of the Slutsky terms derived from equation 2-21. The result was established by Atkinson (1977), and the derivation is provided in an appendix to this chapter.

Deaton (1979 and 1981) shows that it applies in a class of cases somewhat wider than the linear expenditure system. The important conditions are (1) that the Engel curves are linear and identical (that is, for each good everyone has the same constant marginal propensity to consume and the same minimum "requirement" $\mathbf{x}^0$); and (2) weak separability (see equation 2-22) between leisure and goods.[4] The extension allows the B_i and x_i^0 to depend on price and dispenses with the LES labor-supply formulation. Deaton also shows (1979) that, if a subgroup of goods satisfies these two conditions, then taxes should be uniform for the subgroup. These results have recently been extended by Deaton and Stern (1985), who show that differences in intercepts of the Engel curves across households do not disturb the uniformity result provided that they depend only on household characteristics and that there is an optimal system of family grants that depend on household structure.

The second theorem states that, if we have an optimal nonlinear income tax, individuals differ only in the wage rate, and the direct utility function has goods weakly separable from labor in the sense that utility can be written

$$u(\mathbf{x}, l) = u[\xi(\mathbf{x}), l] \tag{2-22}$$

where ξ is a scalar function. Optimal indirect taxes are then uniform. Weak separability involves the marginal rate of substitution between goods being independent of labor/leisure. The validity of the separability assumption is not easily judged (similar considerations arise in our discussion of complementarity in the Ramsey problem). Introspection may be as useful as or more useful than econometrics here, because the properties are very hard to determine statistically (and see Deaton's chapter 4 below).

An intuitive argument, due to Mirrlees, would be the following. At the optimum there will be a distribution across households of the pair (ξ, l) where ξ is the subutility. This optimum must have the feature that every ξ is supplied through the goods vectors $\mathbf{x}$ to households at minimum cost, because the separability assumption implies that it is only the level of ξ that affects the incentive problem and not its makeup in terms of goods. By the usual marginal arguments the minimum cost way of supplying the given ξ levels to the different households requires marginal rates of substitution in consumption to equal marginal rates of transformation in production and hence consumer prices proportional to producer prices. That is, we require uniform taxes. We could, presumably, recast this argument as a proof, although the standard discussion uses the calculus of variations (see, for example, Atkinson and Stiglitz, 1980, chap. 14). Intuitively, differences arise only in labor, which itself separates out from the utility function. Indirect taxation, then, cannot improve upon a flexible tax instrument that concentrates on labor income. Note that the more sophisticated income tax in the second theorem allows us to make a weaker assumption regarding preferences.

We shall consider the importance and interpretation of these two theorems shortly, but first we must recognize an important point. The taxes that emerge from optimal tax models depend critically on the combination of three sets of assumptions: (1) the form of differences between households; (2) the range of tax tools assumed to be available; and (3) the structure of preferences. These assumptions are made before specific parameter values, social welfare judgments, and revenue requirements are entered in the model, and the results will also be sensitive to these subsequent selections. The importance of examining the availability of tax tools makes it clear that we must analyze taxation and expenditure together, not separately, because a number of expenditure policies (particularly subsidized rations but also public goods) take the form of a lump-sum grant, and some of these can be differentiated across individuals.

The Social Welfare Function

In the approach just described, judgments concerning the appropriate relationship between efficiency and equity are embodied in the social welfare function. In this section we examine ways of viewing the social welfare function and how examples can be selected.

There are several possible (related) interpretations of the social welfare function. First, we can regard it as representing the views of a commentator with given values who is trying to suggest ways of improving a given tax or policy system. Second, we can see the social welfare function as forming part of

a dialogue about policy. Thus it would be possible to imagine a discussion in which an economist calculated that a proposed tax system could be optimal only if the government had certain values. This is the so-called inverse optimum problem examined, for example, by Ahmad and Stern (1984). The government might then respond that either the taxes or the implied values are not what it has in mind. By moving back and forth between values and outcomes, the economist could help the government form a view of what its values are. Third, an analysis may be viewed in terms of interest groups. The social welfare function could be specified so as to reflect mainly or only the welfare of certain groups of individuals. The analysis could then help understand the policies that are proposed or might be proposed by these groups. Fourth, we can see it as arising from a notion of justice in which individuals try to judge what is fair or just from behind a "veil of ignorance," disassociating themselves from knowledge of who they are or will be (see Harsanyi, 1955, or Rawls, 1971). Fifth, the model could be viewed as a positive description of government behavior. There are doubtless further interpretations.

The authors of the various chapters in this book are not committed to a single view and might emphasize different interpretations. I find the first two views of particular interest, and I shall examine them in more detail below in the course of analyzing ways in which the social welfare function might be specified. It should already be clear, however, that the approach cannot sensibly be criticized on the grounds that it presupposes a benevolent and monolithic government. A little careful thought makes plain several different and useful interpretations of the approach. It therefore seems rather odd that this last criticism should surface from time to time.

How, then, might we go about constructing a social welfare function? First, it should be clear that we would not, in general, want to settle on a single social welfare function, because we should expect discussion of values to reveal differences, even after careful analysis has rejected some views as unattractive or inconsistent. Nevertheless, we do want to be able to indicate interesting ranges within which the values could reasonably lie and to grasp the meaning of different specifications.

The approach that I find particularly helpful involves the use of examples and the inverse optimum problem. The simplest course is to think of the following question. Suppose that we can transfer income between individuals without problems of incentives but that some income is lost on the way (perhaps it melts). Consider two individuals A and B. Then let us suppose that we would be prepared to lose as much as a proportion δ of a marginal transfer from A to B and still regard the change as an improvement (but that if the loss were more than δ it would be seen as a deterioration). Then

$$\frac{\beta_A}{\beta_B} = (1 - \delta) \tag{2-23}$$

where β_A and β_B are the welfare weights. In this way we could in principle construct a system of welfare weights for all households. If we take a very simple world where the weights depend only on income M (we can imagine the discussion taking place for a given set of prices), then we can think of a functional form describing the relation between β and M. If, for example, relative β depends only on relative incomes, then we can write

$$\beta = kM^{-\varepsilon}. \tag{2-24}$$

Then in the above example,

$$\left(\frac{M_B}{M_A}\right)^{\varepsilon} = 1 - \delta. \tag{2-25}$$

If, therefore, we were prepared to lose as much as half the unit in a transfer from A to B, if A had twice the income of B, then ε would be equal to 1. Similarly, if δ were three-quarters, then ε would be 2.

The dependence of the relative welfare weights only on relative income cannot be accepted without question. In dealing with cases of extreme poverty, for example, we might consider writing $\beta = k(M - \hat{M})^{-\varepsilon}$ where $\hat{M}$ is some minimum subsistence level, so that $\beta \to \infty$ as $M \to \hat{M}$. This procedure again has its drawbacks; in any sample one is likely to find incomes below $\hat{M}$. "Minimum subsistence" does not really mean what it says if many people live with less. The specification of forms such as equation 2-24 requires careful thought and discussion.

The purpose of thinking through our values in a very simple case is that we may find it easy to comment directly on the policy. We can then infer the values in a straightforward way and use them in more complicated contexts. Moreover, experience with the use of optimal tax theory can help us with the problem of inferring values from decided policies. Thus, for example, we showed in the previous section how the optimal marginal tax rate could be related to distributional judgments and the elasticity of substitution between consumption and leisure. If we can estimate or specify the latter and can form some idea of the appropriate marginal income tax rate (and of the uniform lump-sum transfer), then we can infer our distributional values. Similarly, Ahmad and Stern (1984) showed how it is possible to work back from a system of commodity taxation to the underlying values. We therefore now have experience with models that can be of some assistance in a dialogue with policymakers on their preferred ways of specifying the social welfare function.

We should not, however, confine our attention to the second of our interpretations of the social welfare function, which involves government directly. The individual forming a personal judgment of policy will want to examine his (or her) values and should also find such discourse helpful. Thus the models should be considered useful for intelligent political and economic argument as well as in policymaking. Finally, we should note that the social welfare function is only the final step in the consequentialist approach. Most of the work in that approach is associated with calculating the incidence of taxes—that is, how they affect the different households of the economy.

Hence disagreements concerning the specification or use of the social welfare function should not be an excuse for avoiding the major task.

Applications to General Arguments

One of the main purposes of economic theory is to distinguish between correct and incorrect arguments and to help establish reliable rather than unreliable intuition. In this section I shall try to distill from the theory some lessons of this type. I shall begin by setting forth three general principles that emerge from the analysis and shall then turn to the question of uniformity of indirect taxation, commenting briefly on the income tax. The principles will be stated in summary form before their foundation and interpretation are discussed.

1. Tax revenue is raised most efficiently by taxing goods or factors with inelastic demand or supply. Note that this abstracts from distributional questions and that inelasticity refers to compensated demands and supplies. Care should be taken with the pattern of complements and substitutes in that we are looking at the impact on quantities of the whole set of taxes.
2. Taxation concerned with distribution and with externalities or market failures should as much as possible go to the root of the problem. Thus for distribution we should look for the sources of inequality (for example, land endowments or earned incomes) and should concentrate taxation there. In the case of externalities, we should attempt to tax or to subsidize directly the good or activity that produces the externality. Note, however, that it will often be impossible to deal completely with an issue directly, and this limitation will have very important consequences for other policies.
3. We must recognize that it will be impossible to deal perfectly with questions of distribution and market failure directly. The former, for example, require strictly a full set of lump-sum taxes. Thus the target-instrument approach may be treacherous in a second-best world. In this context, a range of policy tools will be required, and we shall need to ask how any particular policy affects all of our objectives—including distribution. The optimal policy for any one tax will often be very sensitive to assumptions concerning the availability and levels of other taxes.

The principles will be discussed in turn and will be related to the preceding analysis. The investigation started with the basic theorems of welfare economics, which clearly establish that the first-best way of raising revenue is a set of lump-sum taxes. The tax payment itself is then completely inelastic in that the behavior of individuals cannot affect the payment. This finding suggests the first principle. The discussion of the Ramsey problem concerning indirect taxes led in the same direction but cautioned us that the compensated demands were relevant. Any system of taxation will have an income effect; we distinguish among them by the "excess burden," or distortions in compensated demands.

The pattern of substitutes and complements will in general be of considerable importance. Away from the optimum, for example, a small increase in indirect taxation may yield a great deal of revenue at little cost if it leads to a sharp switch in demand to goods that are heavily taxed. We must beware of notions of increasing marginal distortion. In a second-best world, for example, we *cannot* assume, even in the one-consumer economy, that a reduction in indirect taxes and an increase in lump-sum taxes will increase welfare (see Atkinson and Stern, 1974). As we show, however, in the appendix to this chapter, in the one-consumer economy a shift away from indirect taxes toward lump-sum taxes, from an initial position, increases welfare, provided that the indirect taxes have been set optimally (given the level of lump-sum finance currently permitted). Even without optimal taxes some indirect tax exists that could beneficially be reduced in a switch to lump-sum taxes. In the many-consumer economy, the welfare effects of a shift between lump-sum taxes or transfers and indirect taxes will depend on the starting point and may be analyzed using the theory of reform described in the next section (this subject is considered in the appendix to this chapter and in chapter 4, by Deaton).

The notion underlying the first principle has been appreciated for a considerable time (by Henry George, Wicksell, Hotelling, and so on), but its application to indirect taxes and income taxes requires care. We saw in our discussion of the income tax, for example, that 100 percent marginal taxation would be indicated where the compensated elasticity of substitution between consumption and leisure is zero. This is *not* the same as a vertical supply curve for labor, which involves simply the balancing of income and substitution effects.

Again, the basic theorem of welfare economics illustrates the second principle in that distribution would be dealt with entirely through lump-sum taxes. It is illustrated, furthermore, by the theorems we have examined, where the optimal income tax was the only policy tool required when differences arose solely in earning capacity. We saw, however, that this result required other very strong assumptions concerning the structure of preferences. Thus although the target-instrument approach can point us to certain taxes, it should never delude us in a second-best world into thinking that we can forget about a target, such as distribution, after mentally allocating some tax to it. Where our instruments are imperfect, we shall have to consider all our objectives in the study of any one instrument.

The third principle is closely linked to the second. We saw that the desirability and structure of a differential system of commodity taxes depended crucially on our assumptions concerning the existence of the income tax and indeed on the type of income tax available. Taxation of necessities may be attractive where the revenue is used to provide a lump-sum grant but unattractive where no such lump-sum grant is possible. A narrow view of targets and instruments or of the policy tools available invites considerable error.

We turn now to the question of whether or not uniform taxation is desirable. We should distinguish sharply at the outset between consumer and producer taxation. The former has been the main subject of this chapter (see chapter 3 for producer taxation). We begin with a discussion of whether the taxes on final goods should be uniform. In general, the results from the many-person Ramsey analysis indicate that there is no bias in favor of uniform taxes. We saw that the rule balances two considerations: on the one hand, we exploit inelasticities in the sense of equal proportional reductions of quantities, but on the other we reduce demand less for those goods consumed by the poor. How these two effects combine will depend on the social values and the structure of demands for different individuals and groups.

The rule is modified in an important way if income taxes are allowed. We saw that in very special circumstances uniformity might be desirable. These circumstances involve, for the optimal linear income tax, for example, differences across individuals arising *only* from the wage rate, or lump-sum income, and not from preferences, a special structure of preferences (based essentially on the linear expenditure system), "minimum quantities," and marginal propensities to spend on each good that are identical across individuals regardless of income. The special nature of the conditions imply, in my judgment, that uniformity is a poor guide for developing countries. Individuals differ in many ways in their preferences, on which religion, caste, and education, for example, may have an important bearing. Furthermore, the underlying expenditure system may be an implausible representation of demands. The income tax is often an instrument of marginal importance and only partial coverage, and so we cannot assume that the poll tax or transfers have been set optimally. On the other hand, subsidized rations, public goods, and some forms of infrastructure allow transfers to certain groups. It is important to link these factors to the analysis of indirect taxation. Thus we may ask whether it is worth raising indirect taxes to finance this form of expenditure.

It is much more difficult to prescribe, however, than to identify the inadequacies of various prescriptions. The derivation of the appropriate set of commodity taxes requires information concerning patterns of complements and substitutes that is very difficult to extract from the data. Our attempts to extract it will require specifications of functional forms, which, as we saw, may have a profound effect on the recommendations. As Deaton (1981) observes: "In consequence, it is likely that empirically calculated tax rates, based on econometric estimates of parameters, will be determined in structure, not by the measurements actually made, but by arbitrary, untested (and even unconscious) hypotheses chosen by the econometrician for practical convenience" (p. 1245). Some of these difficulties are discussed further by Deaton in chapter 4, where he also shows how they relate to the theory of reform—the topic of the next section. Notwithstanding the influence of assumptions concerning functional form, we can still empirically consider the question of whether a move toward uniform taxes will improve welfare.

As we shall see in chapter 3, there is absolutely no reason to assume that uniformity of taxes on intermediate goods is desirable. In general, taxes on intermediate goods lead to inefficiencies, and there is no reason to suppose that uniform taxes lead to any less inefficiency than some arbitrary set. Taxation of intermediate goods should be avoided unless taxing a particular final good is difficult (and *its* inputs might then be taxed) or to improve the distribution of profits where this is not possible by other means.

We saw above how the marginal rate of linear income tax increases with the revenue requirement and the aversion to inequality and decreases with the elasticity of substitution between consumption and leisure. The discussion of the nonlinear tax showed us how the intuition on optimizing *functions* needs to be tutored carefully. There is no presumption, for example, of an increasing marginal rate. Indeed, the optimal schedule will, in general, show first an increasing marginal rate and then a decreasing one. In Mirrlees's calculations (1971), furthermore, the peak of the marginal rate was fairly close to the median. We should be very careful, however, to note that this behavior of the *marginal* rate should not be regarded as conflicting with any notion of the desirability of progression. Such notions should be related to the *average* rate, and it is quite possible for the marginal rate to take the required shape but for the average rate to be increasing much of the way. For this eventuality to occur, it is necessary only for the marginal to exceed the average. Where there is a uniform lump-sum grant, it is quite likely to do so over a big range. We should note, however, that, in any case, a statement concerning the relative desirability of an increasing average rate should itself be derived from a model concerning incentives and distribution and should not immediately be assumed to be obvious.

A marginal rate that at first increases and then decreases contrasts strikingly with the apparent state of affairs in the United Kingdom, where means-tested benefits give high marginal rates at the bottom and the income tax schedule shows increasing marginal rates. The practical issues are no doubt more complicated than our models indicate, but central features in public discussion are the attempt to target transfers to the poor (implying high marginal rates at the bottom) together with incentives (the so-called poverty trap). Thus the considerations of distribution and incentives lie at the heart of the debate.

Tax Reform

By "tax reform" we mean a movement away from some given status quo. We shall concentrate on marginal movements. The method was introduced formally above and will be developed and applied at a number of points in this book. An early empirical application was by Ahmad and Stern (1984)—see also chapter 11 below. We concentrate here on its relationship with calcula-

tions of optimality. Let us suppose that we have some vector of tax tools $\mathbf{t}$ in operation, the resulting level of social welfare is $V(\mathbf{t})$, and government revenue is $R(\mathbf{t})$. We can regard $V(\mathbf{t})$ as being defined by a Bergson-Samuelson social welfare function as before. We consider an increase in the ith tax t_i sufficient to raise one dollar of extra revenue. The rate of change with respect to the tax is $\partial R/\partial t_i$, hence to raise one extra dollar, we must increase the tax by $(\partial R/\partial t_i)^{-1}$. The rate of change of welfare with respect to the tax is $\partial V/\partial t_i$. We define the *fall* in welfare, λ_i, as the reduction in V consequent upon raising one more dollar by increasing the tax on the ith good.

$$\lambda_i = -\frac{\partial V}{\partial t_i} \bigg/ \frac{\partial R}{\partial t_i}. \tag{2-26}$$

We may think of λ_i as the *marginal cost in terms of social welfare of raising one more dollar from the* i*th tax*. If the marginal cost for tax i exceeds that for tax j, then it would be a beneficial reform to switch taxation on the margin from i to j. Thus, if $\lambda_i > \lambda_j$, we have a gain in welfare of $\lambda_i - \lambda_j$ from raising one more dollar via tax j and one less dollar via tax i. More generally, of any reform $\Delta\mathbf{t}$, we ask about its consequences for welfare ΔV and for revenue ΔR, and it is beneficial if $\Delta V > 0$ and $\Delta R \geqq 0$. The statistics λ_i guide us in the selection.

There is, in general, a whole collection of beneficial reforms, and we should not expect uniqueness. We will usually choose among beneficial reforms on the basis of criteria that cannot be put directly into the model. Second-best analysis in this case provides a range of desirable options and thus is far from being nihilistic or pessimistic, as it is sometimes portrayed.

The optimum is the state of affairs where no beneficial reform is possible; thus, the theories of optimality and of reform are very close. Here, optimality requires that all the λ_i are equal. Call the common value λ, then

$$\frac{\partial V}{\partial t_i} + \frac{\lambda \partial R}{\partial t_i} = 0. \tag{2-27}$$

This is precisely the first-order condition for optimality that emerges from the problem

$$\begin{array}{ll} \underset{\mathbf{t}}{\text{Maximize}} & V(\mathbf{t}) \\ \text{subject to} & R(\mathbf{t}) \geqq \bar{R}. \end{array} \tag{2-28}$$

Notice that away from the optimum there are as many marginal costs of public funds as there are tax tools, and it is misleading to speak of a unique marginal cost of public funds. The Ramsey problem, the many-person Ramsey problem, and the linear income tax are all examples of optimizing models that take the form of expression 2-28.

In work described in Ahmad and Stern (1984, and chapter 11 below), we have applied this framework to the question of tax reform in India. Thus we have approached the question of resource mobilization by asking about the marginal cost in social welfare terms of raising revenue by different means. This approach has included the comparison of taxation of different goods, of state and central taxes, and of indirect taxes and the income tax. For indirect taxes, for example, we have, at fixed producer prices

$$\frac{\partial V}{\partial t_i} = -\sum_h \beta^h x_i^h \tag{2-29}$$

and

$$\frac{\partial R}{\partial t_i} = \frac{\partial}{\partial t_i}(\mathbf{t} \cdot \mathbf{X}) = X_i + \sum_j t_j \frac{\partial X_j}{\partial t_i} \tag{2-30}$$

where β^h is the social marginal utility of income for household h and the other notation is as described above (see discussion of equation 2-8). Equation 2-29 may be derived intuitively by noting that an increase in the price of good i hits household h in money terms by the amount of x_i^h that it consumes. The number β^h (a value judgment) converts the money measure into social welfare. They are to be selected by the decisionmaker. Equations 2-29 and 2-30 give us λ_i, as in equation 2-26.

The data requirements then are a consumer expenditure survey for the x_i^h (and thus X_i), knowledge of the tax rates $\mathbf{t}$, and aggregate demand responses $\partial X_j / \partial t_i$. For many countries, some information on all these things is likely to be available. Only *aggregate* demand elasticities are necessary, and they may be estimated from time-series data. Given that tax design, not short-term demand management, is at issue here, we require, in principle, medium- or long-run elasticities. We should subject the value judgments β^h to sensitivity analysis to show how results vary in response to different specifications.

A major effort was required when we applied the theory to India to calculate the tax rates. Notice that the t_j in equation 2-30 are taxes actually levied on final goods. Thus, we need to work with actual tax collections and to calculate the effects that taxing of intermediate goods has on taxes effectively levied on final goods. We call these "effective taxes." Measuring them involves a specification of the input-output process (see chapters 3 and 11). Results are presented in chapter 11. The approach using marginal costs λ is illustrated in the appendix to this chapter in a discussion of a switch between indirect and lump-sum taxes.

Conclusions

The purpose of the chapter has been to develop and explain the main results of the modern theory of optimal taxation, to show how they might be applied to guide our judgment of tax policy, and thus to provide an introduction to some of the arguments and applications set forth in later chapters.

As we saw, the theory implies that a number of simple statements, such as "efficiency requires uniform commodity taxes" or "egalitarianism implies increasing marginal income tax rates," must be approached with great circumspection. On the other hand, we argued that the theory did yield a number of general principles that are useful in guiding the practical decisionmaker. Furthermore, the theory can be applied to the detailed calculation of possible tax reforms, as will be shown in subsequent chapters. We have concentrated on consumers and on government revenue, pushing production into the background. Production will play a more prominent role in the next chapter.

Appendix: Lump-Sum Grants and Indirect Taxes

This appendix addresses some formal aspects of the relation between lump-sum grants and indirect taxes. We know that in the one-consumer economy the best way of raising revenue is a poll tax with zero commodity taxation. In the case of many consumers, we would want differentiated lump-sum taxes or transfers and no commodity taxation. If only a poll tax is allowed, however, we will usually want to combine it with nonzero commodity taxation, and we will often find that the optimal level of a poll tax is negative (that is, it is a transfer). Such would certainly be the case if some consumers could not exist without a transfer. In this appendix we examine two issues. The first is the relationship between the structure of preferences and the indirect tax structure in the presence of a poll tax. Specifically, if we have an optimal poll subsidy, and individuals differ in the wage rate but have identical preferences, as given by the linear expenditure system (equation 2-21), then optimal indirect taxes are uniform. Second, we examine the welfare aspects of a shift toward lump-sum taxation from indirect taxation in an economy with one consumer and toward a poll tax with many consumers.

The Linear Expenditure System and Uniform Taxation

We prove the result just described, which is due to Atkinson (1977). We start from equations 2-19 and 2-20, which give the first-order conditions for optimal taxes on the assumption that the poll tax or transfer is set optimally. The condition for the optimality of the poll tax is, where $b^h = \beta^h/\lambda + \mathbf{t} \cdot (\partial \mathbf{x}^h/\partial M^h)$, and M^h is the lump-sum income of household h,

$$\bar{b} \equiv \frac{1}{H} \sum_h b^h = 1. \tag{2-A1}$$

From equations 2-19 and 2-20 we have

$$\sum_h \sum_k t_k \frac{s_{ik}^h}{H} = \bar{x}_i - x_i^* \tag{2-A2}$$

where $\bar{x}_i$ is X_i/H, the average consumption of the ith good, and x_i^* is the weighted average using weights $b^h/H\bar{b}$, which sum to one.

The LES demands for the hth consumer, with wage w^h, are (see equation 2-21)

$$x_i^h = x_i^0 + \frac{B_i}{q_i}\frac{1}{\alpha^h} \tag{2-A3}$$

where α^h, the private marginal utility of income, is given by

$$\frac{1}{\alpha^h} = M^h + w^h l_0 - \mathbf{q} \cdot \mathbf{x}^0 \tag{2-A4}$$

and M^h is the lump-sum income of household h, which includes the optimal transfer or tax. The Slutsky terms are given by

$$s_{ik}^h = \frac{B_i B_k}{\alpha^h q_i q_k}, \qquad \text{for } i \neq k$$

$$s_{ii}^h = \frac{B_i^2}{\alpha^h q_i^2} - \frac{B_i}{\alpha^h q_i^2}. \tag{2-A5}$$

Substituting into the left-hand side of equation 2-A2, we obtain

$$\frac{1}{H}\frac{B_i}{q_i}\sum_h \frac{1}{\alpha^h}\left(B^* - \frac{t_i}{q_i}\right) \tag{2-A6}$$

where

$$B^* = \sum_{k=1}^{n} \frac{B_k t_k}{q_k}.$$

The right-hand side of equation 2-A2 is

$$\frac{B_i}{q_i}(\bar{a} - a^*) \tag{2-A7}$$

where the macron and asterisk carry the same meaning as for equation 2-A2 and a_i^h is $1/\alpha_i^h$. From equations 2-A6 and 2-A7, we see immediately that t_i/q_i is independent of i, and we have uniform commodity taxation at the rate $B_0^{-1}(a^*/\bar{a} - 1)$. This argument is examined for more general preference structures in Deaton (1979) and in chapter 4 of this book.

Marginal Shifts toward Lump-Sum Taxation

We consider first the one-consumer Ramsey problem such that now the government is allowed to levy a lump-sum tax T at a fixed level that is below the revenue requirement. Thus, analogously to equation 2-4 for given T, the problem is

$$\text{Maximize by choice of } \mathbf{q}, \quad V(\mathbf{q}, w, -T)$$
$$\text{subject to} \quad R(\mathbf{t}, T) = \sum_k t_k X_k + T \geqq \bar{R} \tag{2-A8}$$

where in the indirect utility function V we have lump-sum income $-T$. We continue to assume that wage income is not taxed. Note that if all goods and services could be taxed, we could raise $\bar{R}$ in a lump-sum manner by saying that $t_i = \tau q_i$ for all goods including labor. Indirect tax revenue is then $-\tau T$ (from the budget constraint for the consumer, $q \cdot X - wl = -T$) so that lump-sum plus indirect tax revenue is $(1 - \tau)T$; we can put $\tau = \bar{R}/T - 1$ (< 0) to raise required total revenue. If we assume that one good or service (for example, labor) cannot be taxed, however, then raising $\bar{R}$ will require a combination of indirect taxes and the lump-sum tax.

Taking a Lagrange multiplier λ for the revenue constraint and forming the Lagrangean

$$\mathcal{L}(\mathbf{t}, T, \lambda) = V(\mathbf{t}, -T) + \lambda[R(\mathbf{t}, T) - \bar{R}], \tag{2-A9}$$

the first-order conditions (remember T is fixed)

$$\frac{\partial \mathcal{L}}{\partial t_i} = 0 \tag{2-A10}$$

yield equations 2-6 and 2-7 as before, with

$$\theta = -\frac{\alpha}{\lambda} + \left(1 - \sum_k t_k \cdot \frac{\partial X_k}{\partial M}\right) \tag{2-A11}$$

where M is lump-sum income (here equal to $-T$). From equation 2-7 we have

$$\sum_k \sum_i t_k s_{ik} t_i = -\theta \sum_i t_i X_i. \tag{2-A12}$$

Hence θ is a positive number, because $\Sigma_i t_i X_i > 0$ (we have assumed $T < \bar{R}$).

Consider now a shift in the optimization problem brought about by a change in T.

$$d\mathcal{L} = \sum_i \frac{\partial \mathcal{L}}{\partial t_i} dt_i + \frac{\partial \mathcal{L}}{\partial T} dT + \frac{\partial \mathcal{L}}{\partial \lambda} d\lambda. \tag{2-A13}$$

But $\partial \mathcal{L}/\partial t_i = 0$ from the first-order conditions 2-A10, and $\partial \mathcal{L}/\partial \lambda$ is zero, because we assume that the budget constraint holds with equality. Hence

$$d\mathcal{L} = \frac{\partial \mathcal{L}}{\partial T} dT. \tag{2-A14}$$

Because the budget constraint holds with equality before and after the shift in T, however, we have $d\mathcal{L} = dV$. Thus the gain in welfare from a shift toward lump-sum taxation ($dT > 0$) from a position with optimal indirect taxes is given by $\partial \mathcal{L}/\partial T$, and from equations 2-A9 and 2-A11 we have

$$\frac{\partial \mathcal{L}}{\partial T} = \theta\lambda. \tag{2-A15}$$

Hence θ is a measure, in terms of revenue, of the benefit from a switch to lump-sum taxes, and we have seen that θ is positive. This result was derived in the appendix to Atkinson and Stern (1974).

It is interesting to see how the argument may be expressed in terms of the marginal costs of funds defined earlier in this chapter (see equation 2-6). The result is that the marginal cost of funds raised by lump-sum taxation λ_T is less than that raised by commodity taxation λ_i; the benefit of the switch of a unit of revenue from commodity to lump-sum taxes $\lambda_i - \lambda_T$ is positive. Now

$$\text{(2-A16)} \qquad \lambda_T = -\frac{\partial V}{\partial T} \Big/ \frac{\partial R}{\partial T} = \frac{\alpha}{1 - \mathbf{t} \cdot (\partial \mathbf{X}/\partial M)}$$

and

$$\text{(2-A17)} \qquad \lambda_i = -\frac{\partial V}{\partial t_i} \Big/ \frac{\partial R}{\partial t_i} = \frac{\alpha X_i}{X_i + \sum_j t_j (\partial X_j / \partial t_i)}.$$

Then from the optimality of the t_i we have $\lambda_i = \lambda$ for all i, and we may write equation 2-A11, using equations 2-A15 and 2-A16, as

$$\text{(2-A18)} \qquad \lambda - \lambda_T = \frac{\theta\lambda}{1 - \mathbf{t} \cdot (\partial \mathbf{X}/\partial M)} = \frac{1}{1 - \mathbf{t} \cdot (\partial \mathbf{X}/\partial M)} \frac{\partial \mathcal{L}}{\partial T}.$$

Hence the benefit as measured by $\lambda - \lambda_T$ is the same as that measured using $\partial \mathcal{L}/\partial T$ where the extra term $1/(1 - \mathbf{t} \cdot \partial \mathbf{X}/\partial M)$ arises, because we have to raise the lump-sum tax by this amount to obtain a unit of revenue (net).

We can now examine generalizations of this analysis (1) to the case where the taxes t_i are not set optimally for given T and (2) to the case where there are many consumers. Where taxes are not set optimally, then we ask whether the marginal cost of funds from i could indeed be lower than that of lump-sum taxes for every i. Thus we ask whether it is possible that

$$\text{(2-A19)} \qquad \lambda_T > \lambda_i \qquad \text{for every } i.$$

From equations 2-A16 and 2-A17, and the Slutsky decomposition, expression 2-A19 implies

$$\text{(2-A20)} \qquad 0 < \frac{1}{X_i} \sum_j t_j s_{ji} \qquad \text{for all } i.$$

If we assume that $t_i X_i > 0$ for all i (so that both goods and factors are taxed; a negative t_k for a factor is like a tax because it signifies that the consumer price is below the producer price), then we have, on multiplying by $t_i X_i$ and summing:

$$\text{(2-A21)} \qquad 0 < \sum_i \sum_j t_j s_{ij} t_i$$

which contradicts the negative definiteness of the Slutsky matrix. Hence equation 2-A19 is not true, and there is a j such that $\lambda_j \geqq \lambda_T$. Thus even if commodity taxes are *not* optimal, there will exist some commodity tax that could be lowered beneficially with a switch to lump-sum taxation.

When there are many consumers, then we restrict attention to a poll tax, so that $R(\mathbf{t}, T) = \mathbf{t} \cdot \mathbf{X} + HT$. We may consider $\lambda_i - \lambda_T$ as before. The expressions for λ_T and λ_i become

$$\text{(2-A22)} \qquad \lambda_T = \frac{\sum_h \beta^h}{H - \mathbf{t} \cdot \sum_h (\partial \mathbf{x}^h / M^h)}$$

$$\text{(2-A23)} \qquad \lambda_i = \frac{\sum_h \beta^h x_i^h}{X_i + \sum_j t_j (\partial X_j / \partial t_i)}$$

and

$$\text{(2-A24)} \qquad \frac{\partial \mathcal{L}}{\partial T} = -\sum_h \beta^h + \lambda \left(H - \mathbf{t} \cdot \sum_h \frac{\partial \mathbf{x}^h}{\partial M^h} \right).$$

When the commodity taxes are optimal, we have $\lambda_i = \lambda$ for all i. Consider as before a marginal increase in T with adjustments in t_i to hold revenue constant from a point where $\partial \mathcal{L}/\partial t_i = 0$, that is, where $\lambda_i = \lambda$ for all i. Then, as before,

$$\text{(2-A25)} \qquad dV = d\mathcal{L} = \frac{\partial \mathcal{L}}{\partial T} dT$$

and

$$\text{(2-A26)} \qquad \frac{1}{1 - \mathbf{t} \cdot \sum_h (\partial \mathbf{x}^h / \partial M^h)} \frac{\partial \mathcal{L}}{\partial T} = \lambda - \lambda_T.$$

Notice, however, that we can no longer show that $\partial \mathcal{L}/\partial T = \theta\lambda$ as in the single-consumer case using equation 2-A12, with θ positive, because the proportional reduction in compensated demand (that is, θ for the single consumer) now depends on the good. Hence it is quite possible that $\partial \mathcal{L}/\partial T$ will be negative, so that a switch to lump-sum taxation will reduce welfare. Indeed, from equation 2-A24 we see that if β^h is very high for some h (for example, the very poor) then $\partial \mathcal{L}/\partial T$ will indeed be negative, and starting from, say, zero T, we would want to reduce it—that is, we would want to introduce a lump-sum subsidy. On the other hand, if households are so similar that the one-consumer case is a reasonable approximation, then starting from zero we would wish to increase the lump-sum tax (assuming throughout that $\bar{R} > 0$, so that we are raising rather than distributing revenue).

Notes

1. The presentation of some of the standard theory and of figure 2-1 is taken from Stern (1983*a*).
2. That is, the covariance of x_i^h and b^h divided by the product of the means of x_i^h and b^h.
3. For references to the early debate, see Atkinson and Stiglitz (1980), lecture 13, and Musgrave (1959), chap. 5.
4. Deaton emphasizes linearity, but the proof also relies on the assumption that they are identical.