DM/84/9

INTERNATIONAL MONETARY FUND

Fiscal Affairs Department

Optimum Taxation and Tax Policy

Prepared by Nicholas Stern 1/

Approved by Vito Tanzi

February 13, 1984

1. Introduction

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 are questions which 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 last fifteen years, however, have seen a tremendous growth in the formal analysis of these problems and the main purpose of this paper is to give an introduction to this newer literature. Much of it has been technical but we shall try here to give a broad understanding of the methods of approach, the type of arguments used, and the main conclusions reached. Given the precise nature of the problems and the sensitivity of many of the results to particular assumptions, we cannot, however, avoid formal details altogether. We shall try, as far as possible, to identify some fairly general and robust lessons.

The three questions posed at the beginning are at the heart of that part of the modern theory of public economics known as optimum taxation. It is important to appreciate that optimum taxation forms just one part of the modern theory of public economics, and in the next section we shall attempt to convey an impression of the main concerns of that more general theory in order to set optimum taxation in a broader context. In so doing we shall indicate some historical antecedents of optimum taxation and emphasize its foundation in the relaxation of the assumptions of classical welfare economics.

1/ Professor of Economics, University of Warwick. This paper was written while the author was a Visiting Scholar in the Fiscal Affairs Department during the Summer of 1983. He is very grateful for the helpful comments of A.B. Atkinson, A.S. Deaton, D.M.G. Newbery, and V. Tanzi, and participants at the seminar of the Fiscal Affairs Department, August 4, 1983.

In Section 3 we shall set out some of the main results of the theory of optimum taxation. We begin in Section 3(a) with commodity taxation and the well-known Ramsey rule for the one-consumer economy and then examine its extension to an economy with many consumers. Optimum income taxation, following the approach of Mirrlees, will be presented in Section 3(b), and in Section 3(c) we discuss the appropriate combination of income and commodity taxation--often discussed under the heading of the balance between direct and indirect taxation. In Section 4 we discuss the consequences of the theory for production efficiency.

Some applications of the theory to discussions of tax policy are presented in Section 5. We show that the simple principles it embodies can be used to discriminate among arguments in discussions of public policy. This is, of course, one of the main purposes of theoretical enquiry in economics--to establish which of the many possible intuitive and informal arguments are well founded. We describe briefly in Section 6 a different sort of application: this is an attempt (by Ahmad and myself) to use some of the theory in the analysis of Indian tax reform, as the welfare analysis of a small movement from a given initial position, and show its close relation to the theory of optimality. Concluding remarks are presented in Section 7.

2. The scope of modern theories and some historical antecedents 1/

a. The approach and its relations with public economics

Much of the discussion in the 19th 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 (he who benefits should pay) and those who claimed that taxation should be based on ability to pay. This last concept was itself discussed extensively in terms of which 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).

This attempt to analyze the questions of public finance in terms of a collection of principles continues and characterizes much of the literature up to the present (see, for example, Musgrave and Musgrave, 1980). The modern theories of public economics diverge strongly from much of the traditional in this respect in that they work with the criteria of classical welfare economics--usually a Bergson-Samuelson social welfare

1/ Some of the presentation of standard theory in this section is taken from my chapter entitled "Taxation for Efficiency" in Shepherd, Turk, and Silberston (1983).

function or Pareto optimality. 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 which makes one individual better off without making someone else worse off.

The unifying features of this approach provide substantial clarity and analytic power. There are, however, many interesting ethical economic issues which it leaves out. For example, the approach is essentially "consequentialist" in that policies are evaluated in terms of their consequences. One can argue that in taxation, as with other things, there are certain principles which should be observed irrespective of their consequences. An example might concern the kind of information the state should be allowed to use. Secondly, the consequences are evaluated solely in terms of changes in utility of members of the society. Again there might be aspects of the consequences of a particular policy, concerning, for example, the rights which it grants individuals, which would be excluded in this approach. For further discussion of some of these issues, see Sen and Williams (1982). Many of the questions which 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 be of considerable relevance for some aspects of social policy, for example, the question of which instruments of policy are admissible.

The use of a Bergson-Samuelson social welfare function provides clarity and a unifying theme in its portrayal of the value judgments required for normative analysis. For much of our analysis we need not be specific as to how the function is chosen and can show the consequences of using different social welfare functions embodying different ethical positions concerning, for example, income distribution. The selection of a particular policy, however, would usually involve the selection of a particular welfare function and that would be something for the policymaker. The economist may, however, be able to assist in the discussion of the choice of the social welfare function since he can show the policymakers in simple contexts the consequences of different specifications of that function. This may help in the selection of value judgments for more complicated problems.

The analysis of taxation in the modern theory proceeds therefore 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 then secondly a normative side where value judgments are introduced. Our concentration in this paper will be 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 the main textbook on the subject (Atkinson and Stiglitz, 1980) is devoted to the analysis of the consequences of taxes, and page 331 is reached before the analysis of normative issues begins. Examples of the positive issues are (i) the analysis of the consequences of income or wealth taxation for risk taking, (ii) how different forms of company taxation will affect investment and the distribution of profits, (iii) the effects of national debt and taxation on saving and growth, and so on. The application of careful and formal microeconomic theory to such questions has been a major feature of modern public economics.

It is clear that if the calculation of the consequences of policies is itself difficult, then choosing the optimum among all policies runs the risk of being intractable. For one is then searching over a set of options each of which presents analytical difficulties. Thus, the normative part of public economics has, in the main, been concerned with rather simpler models than those used for the analysis of positive questions only. For further discussion of the positive models, see Atkinson and Stiglitz (1980).

A second important area of recent research which we shall not discuss in detail concerns models of the way taxes might be determined in a nonoptimizing framework (voting, bureaucracies, political power, interest groups, and so on). Note that these nonoptimizing models also require an analysis of the consequences of taxes for the interests of different groups. The work of the "public choice" theorists (see, e.g., Buchanan and Tollison, 1972) contains some valuable insights (see also Atkinson and Stiglitz, 1980, Lecture 10). The optimizing and deterministic approaches to the analysis of taxes should be seen as complements rather than alternatives. Thus, for example, it would be interesting to compare the outcome of a deterministic or closed model with the solutions which might emerge from optimization models under different social welfare functions.

We should draw attention to a third main area of recent research which is not presented in detail here. This concerns the econometric estimation of the positive models used in public economics. It involves the empirical analysis of how people will react to different tax, pricing, or rationing schemes. And this has led to a closer integration between the theory and estimation of consumer choice and the behavior of firms, on the one hand, and the theory of public economics, on the other. After estimation one can try to use the estimated demand and utility functions to analyze the welfare effects of possible changes in policy. There have been a number of recent examples of this vertical integration of the analysis of data, economic theory, econometric skills, and policy discussion, which provide good examples of what economics can do. I will not go into detail here but refer to, for example, the Journal of Public Economics (which began in 1972), where much of the research I am discussing has been published.

A fourth area is that of computable general equilibrium models. Fortunately, it is unnecessary to go into detail here since my predecessor as a Visiting Scholar has provided a splendid survey (Shoven, 1982).

b. <u>The point of departure</u>

The modern theory of public economics takes as its point of departure the two basic theorems of welfare economics. The first of these is that a competitive equilibrium is Pareto efficient, and the second is that one can achieve a prescribed Pareto-efficient allocation as a competitive equilibrium if prices are set appropriately and lump-sum incomes are allocated to each individual to allow him to buy the consumption bundle given in the allocation at the prices which have been specified. 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 referred to as "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 <u>any</u> 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 paper we attempt to sketch the main results of that part of the investigation which concentrates on the inability to achieve a desirable set of lump-sum taxes.

The appropriate tax policies to deal with externalities have been extensively discussed in the literature (for a classic statement, see Pigou, 1947). 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, e.g., Boiteux, 1956) and, thus, our discussion of commodity taxation essentially includes also the important topic of public sector pricing. In this context, there has been valuable and interesting work in public economics on the problem of measuring marginal cost (see, e.g., Dreze, 1964).

Recall that a lump-sum tax on an individual is a payment that he cannot alter by any of his actions. 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. It is clear 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 such as earning power or wealth will be such as to prevent them from being lump sum. The individual will discover what is being measured and can usually, if he wishes, adjust that dimension. Note that lump-sum taxes are not, in general, impossible. We could have differential taxation by sex or height (assuming that there would be neither direct action to change these nor emigration). It is the achievement of desirable lump-sum taxes which causes the difficulties.

This conclusion leads in two directions. The first involves a fairly robust general notion: there is an argument in favor of taxing things which are not easily varied by individuals or firms in response to taxation. An important example would be pure rent or monopoly profits, where these can be identified. We shall shortly see an example in commodity taxation where this is embodied more formally. The second is that we need a theory which addresses the problem of taxation in a world where lump-sum taxes are not possible. This leads us directly to what is known as the theory of optimum taxation.

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

3. Optimum taxation

a. Commodity taxation

The examination of optimum taxation where lump-sum taxes are impossible has been concentrated on commodity taxation and income taxation. Analysis of the former problem goes back to Ramsey (1927), and 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 optimum income taxation was created by Mirrlees (1971, see next subsection).

The Ramsey problem is to raise a given revenue from a consumer through the taxation of the commodities he consumes 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), so we have a simple efficiency problem in that distributional considerations are ignored (a point to which we shall return).

It will be useful for the interpretation of the results from the Ramsey problem, and for further reference below, to have in front of us a brief description of the partial equilibrium approach to the question. These two pieces of analysis will be used to demonstrate the methods and develop some intuition which we shall use 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 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 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 taken as zero (producer prices are fixed so competition would drive profits to zero) and, therefore, we 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. Thus, the net loss, or deadweight loss, is the triangle ABC.

One then examines the minimization of the sum across goods of triangle ABC (i.e., total deadweight loss), subject to the constraint that the sum across goods of the rectangle ABGH (i.e., total tax revenue) is not less than a given figure. It is straightforward to show that this leads to the result that 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 price elasticity of demand for the i^{th} good.

There have been a number of calculations of such triangles in the empirical literature following the work of Harberger (1954) who applied this approach to deadweight losses from monopoly (the distance of price above marginal cost playing an analogous role to the tax). The more modern approach is to use explicit utility functions and "equivalent

1/ The presentation and discussion of Figure 1 is taken from my chapter "Taxation for Efficiency" in Shephard, Turk, and Silberston (1983).

variation," thus avoiding the unattractive assumption that the demand for a good does not depend on the prices of other goods (see, e.g., Rosen, 1978).

We now give a brief mathematical formulation of the central result in optimum commodity taxation, the so-called Ramsey rule. This dispenses with the partial equilibrium assumption concerning demands and works directly with 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 would be treated as negative purchases. It is convenient to treat the sale of labor differently from other goods and, thus, identify it separately (as, say, ℓ) in the utility function and the budget constraint. Where w is the wage faced by the consumer, where there is lump-sum income M, X is the vector of quantities transacted, and $q.X$ denotes $\Sigma_i q_i X_i$, then the individual problem may be written;

$$\underset{X,\ell}{\text{Maximize}}\ u(X,\ell) \tag{1}$$

subject to $q.X - w\ell = M$

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 $w = (1 + \tau)w_g$, where w_g 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 (1). The revenue is $\tau M/(1 + \tau)$. In the one-consumer economy with lump-sum incomes, this form of taxation would be optimum provided the revenue requirement R does not exceed M. In this case, the optimum uniform tax rate τ is given by

$$\tau/(1 + \tau) = R/M \tag{2}$$

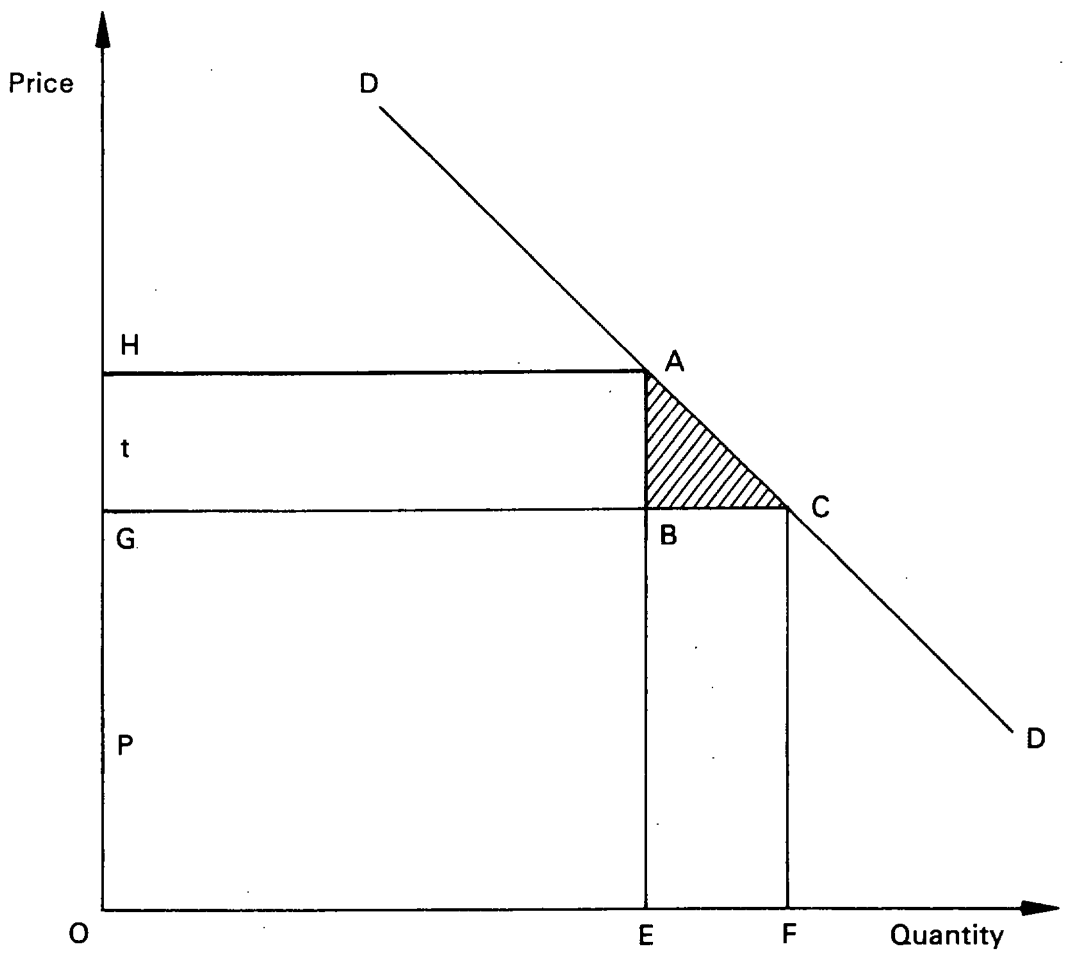
Where there are no lump-sum incomes, proportional taxes (including the wage subsidy) raise no revenue. If the revenue requirement does exceed M, then distortionary taxes (i.e., those which are not equivalent to lump-sum taxes) will be necessary.

If there are no lump-sum incomes ($M = 0$), then we may choose one good to be untaxed without loss of generality and it is convenient to choose that good to be labor. When $M = 0$ the budget constraint is simply

$$qX = w\ell \tag{3}$$

Then, for the consumer, a tax at rate τ on wage income is equivalent to raising prices to $q/(1 - \tau)$. We shall assume in what follows in this subsection that there are no lump-sum incomes and that labor is untaxed.

FIGURE 1



We consider then just one consumer whose individual demands $X(q, w)$ are a function of consumer prices only. The maximum utility an individual can achieve when facing prices q is written $V(q, w)$: this is the indirect utility function. The problem then becomes to choose t, or, equivalently, q, to maximize $V(q, w)$ (and, thus, minimize utility loss) subject to the constraint that the tax revenue $\sum_k t_k X_k$ meets the requirement $\bar{R}$. $\bar{R}$ is the value at 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 since the government can transform its revenue at prices p into whichever goods it desires. The suffix on a vector denotes the particular component: thus, t_k is the tax on the k^{th} good.

Formally, then, we have the problem

Maximize by choice of q $\quad V(q, w)$

subject to $R(t) = \Sigma t_k X_k \geqq \bar{R}$ (4)

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

$$\frac{\partial V}{\partial t_i} + \lambda \frac{\partial R}{\partial t_i} = 0 \tag{5}$$

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

$$\frac{\partial V}{\partial q_i} + \lambda (X_i + \sum_k t_k \frac{\partial X_k}{\partial q_i}) = 0 \tag{6}$$

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

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

where s_{ik} is the utility-compensated change in demand for the i^{th} good when the k^{th} price changes, and where Θ is a positive number independent of i.

An intuitive interpretation of equation (7) is as follows. We can think of $\sum_k t_k s_{ik}$ as the (compensated) change in demand for the i^{th} good as the result of the imposition of the vector of small taxes t_k. The typical term in the sum is $t_k \left.\frac{\partial X_i}{\partial t_k}\right|_{\substack{\text{constant}\\ \text{utility}}}$ 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, of course, the size of the taxes t_k is determined within the problem and we are not really justified in assuming that t_k are small. With this qualification, however, the Ramsey rule is that the proportional reduction in compensated demand as a result of the imposition of the set of taxes should be the same for all goods.

This result is an important one and provides the main insight into tax rules which arise from the theory of optimum commodity taxation. It should be emphasized that it is proportional quantity changes that are equal in this rule. Thus, crudely speaking, those quantities which are relatively insensitive to price will be taxed relatively more. It will be important in our argument which follows that this is, in general, very different from the proposition that taxation should be uniform, i.e., 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 and partial equilibrium treatment which we have just seen. The Ramsey rule provides an example of the general principle that efficient taxation is directed toward those goods which cannot be varied by consumers. Note, however, that one needs considerable care with substitutes and complements, a question which is suppressed by the partial equilibrium approach.

Given that labor was assumed untaxed and there is an endowment of time, one can interpret the Ramsey rule in terms of complementarity and substitutability of the taxed consumer goods with leisure--a notable early example was the work of Corlett and Hague (1953). Goods which are relatively complementary with leisure should bear the higher tax rate. Thus, one can show (Deaton, 1981) that if leisure is quasiseparable from all goods, then the Ramsey rule gives uniform taxation of goods.

Intuitively, quasiseparability means that all goods are equally complementary with leisure. Formally, goods i and j are quasiseparable 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 is via a proportional change in the vector (X, ℓ)). Note that the issues of complementarity, substitutability, and separability with leisure arise because there is an untaxed endowment of time and the conclusions would be expressed in terms of another good if there were a corresponding endowment.

We should note here that there is a sense on which 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 where there are fixed lump-sum incomes, this may be achieved equivalently through proportional taxation of <u>all</u> 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 <u>by itself</u> the best way to raise revenue, and indirect taxation will be required. Indeed, in the many-consumer case the optimum poll tax will often be negative, i.e., a poll subsidy. Our discussion of the Ramsey rule should therefore be seen as a development of the intuition for application in the more general case; see the discussion of the many-consumer case which follows immediately and Section 3(c) on the optimum combination of income and commodity taxes.

The Ramsey result would seem to be rather inegalitarian in that it appears to direct commodity taxation toward "necessities" which we usually think of as being fairly insensitive to price. But the formulation in terms of one consumer has explicitly ignored distributional questions. The result can, however, be generalized to many consumers in a fairly straightforward way. We simply replace $V(q, w)$ in problem (4) by the social welfare function $W(u^1, u^2, \ldots, u^H)$, where u^h is the utility function of the h^{th} individual, which we consider again as a function of consumer prices q and the wage w^h. The function $X(q, w)$ becomes $\sum_h x^h(q, w^h)$ where $x^h(q, w^h)$ is the demand function for individual h. The rule then is no longer that the proportional reduction in compensated demand should be the <u>same</u> for all goods or commodities, but the modified rule 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, I am using "the rich" here for those whose social marginal valuation of income is low. Following an argument similar to that used in the derivation of the Ramsey rule (7), one can show

$$\frac{\sum_h \sum_k t_k s^h_{ik}}{X_i} = -(1 - \bar{b}r_i) \tag{8}$$

where s^h_{ik} 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 i^{th} commodity and the net social marginal valuation of income, plus one. 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 $b^h = \beta^h/\lambda + t. \frac{\partial x^h}{\partial m^h}$ where β^h is the social marginal utility of income and λ 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 those with a high net social marginal valuation of income--see equation (19) below.

Thus, the proportional reduction of compensated demand denoted by the left-hand side of (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 points toward luxuries. The implications of (8) for tax rates will depend on the way in which these two effects combine together. Much will depend on the structure of preferences and the type of income tax tools available--as we shall see in Section 3(c).

b. Income taxation

Both Adam Smith and John Stuart Mill 1/ 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." The form "the pressure" should take was discussed extensively and often based on 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 all or that an equal proportion of utility should be sacrificed. Given a utility function (assumed the same for everyone) and one of these principles, say, equal absolute sacrifice, one can calculate a corresponding tax function. If income is Y and the tax payable is T(Y), then, given some total revenue requirement, one can

1/ For references to the early debate, see Atkinson and Stiglitz (1980), Lecture 13, and Musgrave (1959), Chapter 5.

calculate T, assuming Y is independent of the tax schedule, for each level of Y from

$$U(Y) - U(Y - T) = \text{constant} \tag{9}$$

the condition for equal absolute sacrifice. For calculations in this

framework, see Stern (1977)--one can show, for example, that if $U'(Y) = Y^{-\eta}$ then taxation is progressive (in that the marginal exceeds the average rate) for $\eta > 1$. The logarithmic or Bernoulli form corresponding to $\eta = 1$, 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" put forward by Edgeworth, has greater clarity in that it is derived 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 post-tax incomes. Thus, the marginal tax rate is 100 percent which raises the incentive question in a very stark manner. This incentive problem had been recognized very early in the discussion: for example, McCulloch (1845), Part I, Chapter IV, "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."

Given that the incentive and distribution aspects of the income tax have long been recognized, it is perhaps surprising that a model which simultaneously examined the distribution and size of the cake was not forthcoming until the article by Mirrlees (1971). This paper essentially created the subject of optimum income taxation. As we shall see, Mirrlees kept his model as simple as possible, given the issue at hand, but it is nonetheless not an easy problem and the analysis poses considerable technical difficulties. The reason is that the policy tool is the whole income tax <u>function</u>. Thus, for each income we have to specify the tax payment, and the optimization is in a space of all admissible functions. This should be contrasted with the problems usually examined in standard microtheory (for example, consumer or producer behavior) where only a finite number of variables, e.g., consumption of each type of good, is considered.

The income tax problem is considerably simplified if one confines attention to a linear tax schedule where there is a lump-sum benefit or tax combined with a constant marginal rate, and, following the Mirrlees article, there were a number of papers in which the simpler problem was examined (see Atkinson and Stiglitz, 1980, Lecture 13, for references). In our discussion of numerical results, we shall concentrate on the linear case but we shall begin by setting out the Mirrlees nonlinear problem and explaining why it takes the form it does. We shall then summarize some of the main results for the nonlinear problem. Finally, we present numerical calculations for the linear case to bring out the sensitivity to the important parameters and to compare the computed tax rates with levels we see in practice. There has been some recent work on an intermediate case with a finite number of individuals--one might interpret them as representative of certain groups--where the optimum tax schedule can be taken as piece-wise linear (see Guesnerie and Seade, 1982, and Stern, 1982).

Given that the nonlinear problem will pose difficulties, it is sensible to begin by keeping the structure as simple as is consistent with retaining the question. From this point of view the model which is 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) and, if there were no incentive problem, we have seen, in our discussion of Edgeworth above, that the marginal rate would be 100 percent. The Mirrlees model has individuals differing in only one respect, in their pretax wage or productivity, and there is only one aspect of incentives, labor supply. Thus, in the model, labor is supplied by individuals, each of whom has an identical utility function, in order to maximize utility of consumption and leisure, given 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 raising some given amount of revenue.

All individuals have the same utility function $u(c, \ell)$ 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 of being of type w.

The problem is to choose a function $g(\ )$ which relates post-tax to pretax income in order to maximize

$$\int \phi(u) f(w) dw \tag{10}$$

subject to

$$\int [w\ell - g(w\ell)]\ f(w) dw = R \tag{11}$$

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

$$c = g(w\ell). \tag{12}$$

The government revenue requirement is R which is seen as fixed for the problem (10)-(12). At a later stage we shall ask how the solution varies with different values of R. The maximand (10) is a Bergson-Samuelson social welfare function of additive form--we add $\phi(u)$ across individuals. If $\phi(u)$ represents social utility for individual of type w, then the function is utilitarian. Equation (11) represents the revenue requirement--$w\ell$ is pretax income so that $w\ell - g(w\ell)$ is the tax payment by individual type w and this is integrated or added across individuals. The constraint (12) represents the second-best nature of the problem in that it says that individuals make their own choice subject to the budget constraint, set by their wage and the government tax function. One can express it by saying that no individual would prefer the income of some other individual, taking account of the work he would have to do to earn it.

Before proceeding to results, we note some particular features of the model. First, as specified, the model is static and there is no saving. This is to keep the structure as simple as possible. From a broader perspective, we might think of ℓ as representing lifetime labor supply and c as lifetime consumption, but the treatement of c, ℓ as vectors would take us too far afield at this point. In the next subsection we shall consider a vector of different consumption goods. Secondly, equation (11) may be replaced in a general equilibrium framework by a production constraint which says that total production, a function of total effective labor $\int(\ell w)f(w)dw$, must equal total consumption $\int cf(w)dw$ plus R. Here we assume that w measures productivity so that $w\ell$ 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 (10)-(12). 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.

Thirdly, note that the constraint (12) gives the model its special structure in that it embodies the incentive constraint. Without it 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 <u>decreasing</u> in ability w if consumption and leisure are normal goods (see Mirrlees, 1979). Intuitively, high lump-sum taxes on the able lead to work being concentrated on the most productive (note that there is no difference between individuals on the consumption side). In the income tax model we assume explicitly that the government <u>cannot</u> identify one type of individual from another and measures only an individual's income (not his hours of work or wage). Thus, with this constraint, embodied in (12), we must have utility being nondecreasing in w--an individual of higher w always has the option of consuming the same as an individual of low w but doing less work.

Fourthly, one cannot guarantee that at the optimum $\ell(w) > 0$ for all individuals. Thus, it may be optimum 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) which are available are rather few. And these results themselves may not hold if we modify the model, for example, to include complementarities between different types of labor (see, e.g., Stern, 1982). The important ones in the Mirrlees model are:

(a) the marginal tax rate should be between zero and one;

(b) the marginal tax rate for the person with the highest income should be zero; and

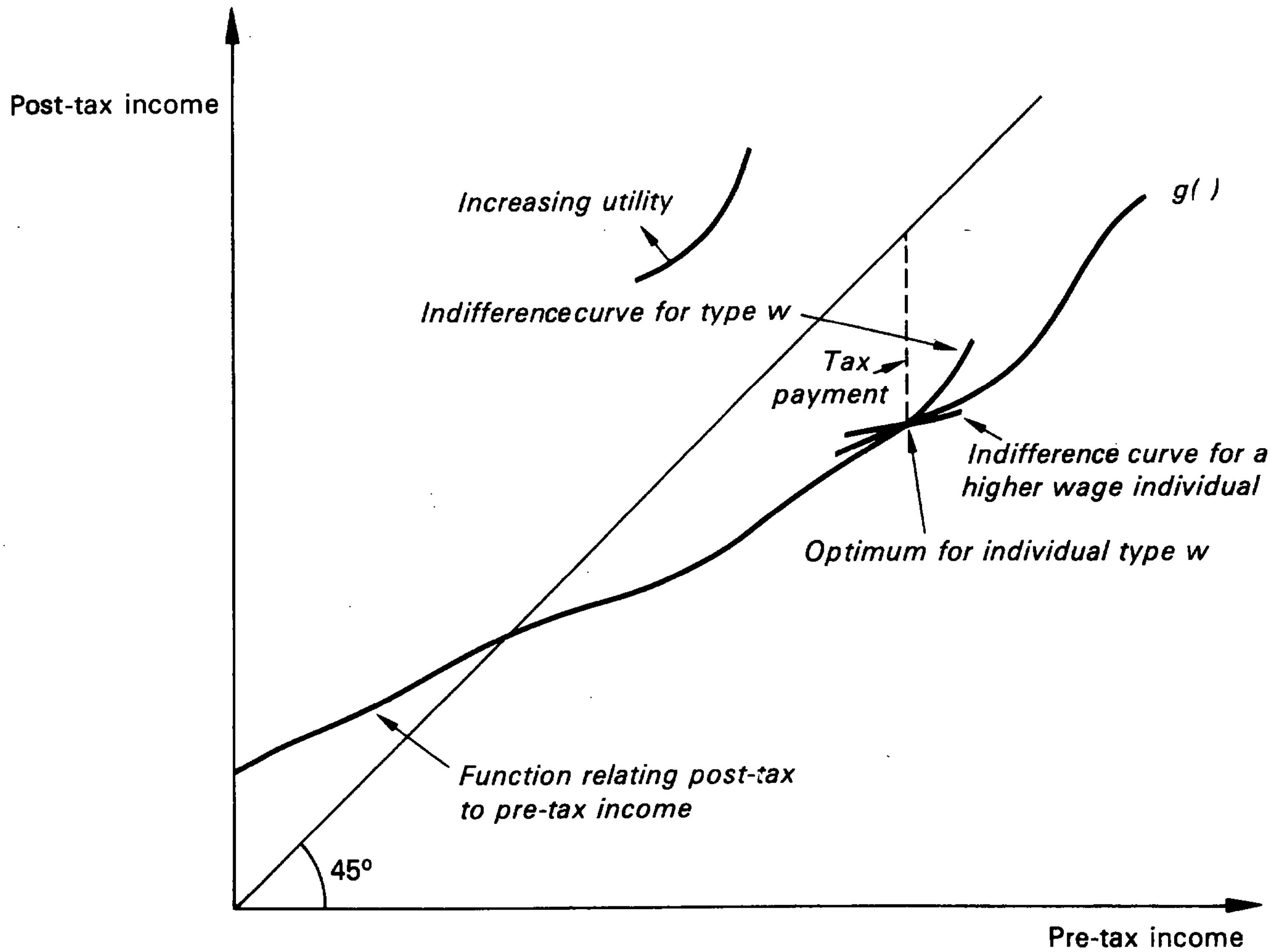
(c) if the person with the lowest w is working at the optimum, then the marginal tax rate he faces should be zero.

Formal proofs of these propositions will not be offered here but we shall 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. This would imply that the reward for the marginal hour was negative. Hence, in the model, no one would choose to work where the marginal tax rate exceeds one. Thus, we could replace any portion of the g( ) function which is downward sloping by a horizontal section without changing behavior (see Figure 2), and we can confine attention to tax schedules with marginal rates which do not exceed one.

We have illustrated the tax function and consumer choice in Figure 2. For an individual with fixed w, we can draw indifference curves in the pretax, post-tax income space since 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, since, at the given consumption level, the higher w person is doing less work and, thus, will need less extra consumption to compensate him for doing the (lower) amount of extra work required for the extra dollar. This implies, in general, that a person with higher w will locate to the right of (earn more money than) the person with lower w since, at the optimum for the lower w-person (tangency with g( )), the indifference for the higher w-person will intersect g from above (coming from the left).

The tax payment is given by the vertical difference from g( ) to the 45^o line. Note that a movement of an individual parallel to the 45^o 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^o, and, therefore, so would the indifference curve of any individual choosing that income. In this case, g( ) would be steeper than 45^o and, intuitively, an equal revenue shift of w-person in the SW direction would take him to a higher indifference curve.

FIGURE 2


Post-tax income
Increasing utility
g( )
Indifferencecurve for type w
Tax payment
Indifference curve for a higher wage individual
Optimum for individual type w
Function relating post-tax to pre-tax income
45°
Pre-tax income

We can give an intuitive argument for the second result as follows. Suppose, with some given income tax schedule, 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, he is better off. The government has lost no revenue since 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 which meets the constraints. Accordingly, the given income tax schedule could not have been optimum and the schedule that is optimum must have the property that the marginal tax rate at the top is zero. If those 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 should note that one cannot deduce that, where there is no highest income and the distribution of skills includes individuals at or above any positive skill levels, the optimum tax rate tends to zero. There are examples (see Mirrlees, 1971, and Atkinson and Stiglitz, 1980, Lecture 13) where it does not (involving the Pareto distribution). One should remember, too, that the argument assumes that there are no externalities so that making the top individual better off upsets no one. Further, the "top" may be at very high levels of income. Zero may be a poor approximation even within most of the top percentile. Nevertheless, the result is rather striking.

We 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 which has the sole effect of inducing the bottom person to do a little more work and, thus, moving a small amount along the schedule. To the first order in utility, that person is no worse off since his indifference curve was tangential to the schedule. But there is a firstorder increase in tax revenue since the marginal rate is positive. Hence, the given schedule is not optimum (for formal discussion of this and the previous result, 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 contrasts strongly with many tax-cum-social-security-systems, and we shall return briefly to this issue in Section 5.

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:

(a) the optimum tax structure is approximately linear, i.e., a constant marginal tax rate, with an exemption level below which negative tax supplements are payable;

(b) 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)); and

(c) "the income tax is a much less effective tool for reducing inequalities than has often been thought" (Mirrlees, 1971, p. 208).

Stern (1976) investigated a wider class of utility functions and, in addition, looked at sensitivity with respect to the social welfare function and the level of government revenue, but confined attention to linear taxation. He used the constant elasticity of substitution (CES) utility function

$$u(c, \ell) = [\alpha\,(1 - \ell)^{-\mu} + (1 - \alpha)c^{-\mu}]^{-1/\mu} \tag{13}$$

with welfare criterion

$$\frac{1}{(1 - \nu)} \int_0^\infty u^{(1-\nu)}(c, \ell)f(w)dw \tag{14}$$

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

$$c = (1 - t)w\ell + G \tag{15}$$

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

$$t\int w\ell f(w)dw = G + R \tag{16}$$

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

The CES in (13) has an elasticity of substitution between consumption and leisure of

$$\varepsilon = \frac{1}{1 + \mu} \tag{17}$$

One may use empirical estimates of labor supply functions to estimate ε and Stern (1976, p. 136) suggests a number around 0.4 based on estimates for married males in the United States. Where the elasticity is less than one, the labor supply function (for positive G) is forward sloping for low wages and backward 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 ε tends to 1 (μ tends to zero) and $\varepsilon = 0$ gives right angle indifference curves (zero substitution effect). One can show generally that with $\varepsilon = 0$ the optimum marginal rate is 100 percent. Note that this is zero <u>compensated</u> elasticity of labor supply and <u>not</u> inelastic labor supply.

A selection of the results is given in Table 1.

We may think of ν as analogous to the elasticity of the social marginal utility of income which is often used in analyses of measures of inequality using the Atkinson index (see Atkinson, 1970) since the utility function is homogenous degree 1 in consumption and leisure (doubling each would double utility) and is, thus, itself analogous to income. The specification of ν then completes the statement of distributional value judgments. Values of ν between 1 and 2 are quite commonly used. Dalton (1967, originally published in 1922, pp. 68-69) argued that Bernoulli's law (or utility logarithmic in income and marginal utility decreasing as the inverse of income), $\nu = 1$, "gives a rather slow rate of diminution of marginal utility" and $\nu = 2$ "as best combining simplicity and plausibility" (although he was working in the context of equal absolute sacrifice). Whether these views of ν 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 around 0.25. Hence, a revenue requirement of R of 0.05 corresponds to around 20 percent of GNP. The case $\nu = 2$, $R = 0.05$, $\varepsilon = 0.4$ gives a marginal tax rate of 54 percent. The expenditure of the 54 percent of GNP goes in 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 one considers 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 ν, and with the revenue requirement R, but decrease with ε, the elasticity of substitution.

Table 1. Stern's Calculations of Optimum Linear Tax Rates

(In percent)

ε	$\nu = 0$	$\nu = 2$	$\nu = 3$	$\nu = \infty$
R = 0 (purely redistributive tax)				
0.2	36.2	62.7	67.0	92.6
0.4	22.3	47.7	52.7	83.9
0.6	17.0	38.9	43.8	75.6
0.8	14.1	33.1	37.6	68.2
1.0	12.7	29.1	33.4	62.1
R = 0.05				
0.2	40.6	68.1	72.0	93.8
0.4	25.4	54.0	58.8	86.7
0.6	18.9	45.0	50.1	79.8
0.8	19.7	38.9	43.8	73.6
1.0	20.6	34.7	39.5	68.5
R = 0.10				
0.2	45.6	73.3	76.7	95.0+
0.4	35.1	60.5	65.1	89.3
0.6	36.6	52.0	57.1	83.9
0.8	38.6	46.0	51.3	79.2
1.0	40.9	41.7	47.0	75.6

Source: Stern (1976, Table 3).

Notes:

(1) $\nu = 0$ corresponds (roughly) to an absence of aversion to inequality in incomes and $\nu = \infty$ to the Rawlsian maxi-min.

(2) A central estimate of the elasticity of substitution ε might be 0.4.

(3) Total output in these models is around 0.25 (it is endogenous) so that R = 0.05 corresponds to government spending (excluding transfer payments) of around 20 percent of gross national product.

c. The combination of income and commodity taxes

The question of the appropriate combination of income and commodity taxation provides fertile ground for confusion. Prest (1960, p. 34) refers to the "first and best-known problem in tax analysis. This is the contention that the allocative effects of indirect taxes are inferior to those of direct taxes . . ." The contention in its simple form is mistaken since there is an excess burden or deadweight loss associated with the divergence between consumer and producer prices for labor, and, thus, the income tax, just as with other goods. A second example concerns the claim which one often hears that a switch from income tax to indirect taxes such as VAT would increase work effort. At the simple level, this is clearly false since 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 more subtle, 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."

It transpires that one can show that under certain conditions, one would want to tax income rather than goods, but it should be stressed that 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 come to a judgment as to how the obvious fact of the divergence of the world from these special conditions should influence our views on the balance between direct and indirect taxation. Thus, the subject involves difficulty in analysis and difficulty in interpretation, and one must beware of simple arguments or contentions such as the ones described.

We shall not present the details of the theorems on the optimum combination of income and commodity taxes but shall try to describe them and highlight the importance of the assumptions. There are essentially two theorems: the first deals with the case where there is a linear income tax, and the second, where 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 optimum in second-best problems since a marginal imposition of taxes from a point of view 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) since we assume that there are no sources of income other than the lump-sum grant/tax and wages.

The first-order conditions for the optimum indirect taxes are given as before by equation (8). The condition for the optimality of the lump-sum grant is that $\overline{b} = 1$, i.e., 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 (8) gives us

$$\sum_h \frac{\sum_k t_k s^h_{ik}}{X_i} = -(1 - r_i) \qquad (18)$$

where

$$r_i = \sum_h \frac{x^h_i}{X_i} \cdot \frac{b^h}{\overline{b}} \qquad (19)$$

Recall that r_i is 1 plus the normalized covariance between consumption of the i^{th} commodity by the h^{th} household and the net social marginal utility of income b^h--we thought of this as the distributional characteristic of good i. 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 (18) is zero. Indirect taxes are zero and all revenue is raised through the lump-sum grant as in the case of identical individuals. Thus, in this sense, indirect taxes are desirable <u>because</u> distributional considerations arise.

We have seen that indirect taxes appear because we are interested in distribution but this 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 play the role of raising revenue to increase the progressive lump-sum grant (or reduce the regressive tax) and the taxation of necessities may be an efficient way to do this (as in the Ramsey case). The way in which these two considerations balance depends quite critically on the form of the differences among the population and on the structure of demand functions. This is illustrated by the first of the theorems which is as follows.

If we have an optimum linear income tax, individuals differ only in the wage rate, and the direct utility function has the Stone-Geary form

$$u(x, \ell) = \sum_{i=1}^{n} B_i \log (x_i - x_i^0) + B_0 \log (\ell_0 - \ell) \tag{20}$$

then the optimum indirect taxes are uniform, i.e., the proportion of tax in consumer price (t_i/q_i) is the same for all goods. The result follows from (18) and (19) using $\bar{b} = 1$, and substituting for the specific form of the Slutsky terms derived from (20). The result was established by Atkinson (1977).

Deaton (1979) and (1981) shows that it applies in a class of cases slightly wider than the linear expenditure system--the important conditions are (i) that the Engel curves are linear and identical, 1/ i.e., for each good everyone has the same constant marginal propensity to consume and the same minimum "requirement" x^0; and (ii) weak separability (see (21)) between leisure and goods. He also shows (1979) that if a subgroup of goods satisfies these two conditions, then taxes should be uniform for the subgroup.

The second theorem states that if we have an optimum 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

$$u(x, \ell) = u[\eta(x), \ell] \tag{21}$$

where η is a scalar function, then optimum indirect taxes are uniform. Weak separability involves the marginal rate of substitution between goods being independent of labor/leisure. The proof of the theorem will not be attempted here since it involves use of the calculus of variations (where one assumes a continuous distribution of wages and integrates across wages). Intuitively, differences arise only in labor which itself separates out from the utility function. Then a flexible tax instrument which concentrates on labor income cannot be improved by indirect taxation. Note that the more sophisticated income tax in the second theorem allows a less strong assumption on preferences.

1/ Deaton emphasizes linearity but the proof also uses the assumption that they are identical.

We shall discuss the importance and interpretation of these two theorems in Section 5 but we close this section by emphasizing an important point. The taxes which emerge from optimum tax models depend critically on the combination of three sets of assumptions: (a) the form of differences between households; (b) the range of tax tools assumed to be available; and (c) the structure of preferences. These are assumptions which are made before specific parameter values, social welfare judgments, and revenue requirements are entered into the model and the results will also be sensitive to these subsequent selections.

4. Production

Up to this point we have assumed that producer prices are fixed and have put production to one side. In a competitive model producer prices will generally be independent of demand only where the nonsubstitution theorem applies (constant returns to scale, no joint production, and a single nonproduced input). The assumption of fixed producer prices allows us to concentrate on consumer welfare and government revenue but one also wants to know a little more about taxes and production outside the framework of the nonsubstitution theorem. The original Diamond-Mirrlees articles of 1971 were entitled "Optimal Taxation and Public Production" and they devoted considerable attention to the production side, although this has received less emphasis in the subsequent literature.

The title of the Diamond-Mirrlees papers immediately emphasizes one obvious but important point: taxation decisions and production are closely linked through the equilibrium of the economy and fiscal choices and physical planning should be seen as part of the same overall policy framework. For example, if production of some publicly produced and nontradable good is to be restricted, then its price (or, equivalently, the tax) should be high. Too often the fiscal and quantity sides are separated in the policy process.

A central set of results in the optimum taxation literature concerns circumstances under which production efficiency is desirable, i.e., it is a feature of the optimum. A production plan is defined to be efficient if it is impossible to have more of one good without having less of another. (If we adopt the convention that an input is a negative output then the definition covers factor inputs as well.) We shall take the definition to be synonymous with being on the frontier of the production possibility set. A necessary condition for efficiency is that the marginal rate of transformation between two goods should be the same for all enterprises where the two goods are transformed on the margin one into the other. If the enterprises are profit maximizers at fixed prices (or are required to maximize shadow profits at shadow prices), then equality of the marginal rates of transformation across enterprises requires equality of relative prices (or relative shadow prices) across enterprises.

It is important here to distinguish between efficiency of the whole productive sector of the economy and efficiency of the public sector taken by itself. We term the former aggregate productive efficiency and the latter public sector efficiency. The desirability of public sector efficiency is a very general and robust result. It says simply that if it is possible for the public sector to produce more at no extra cost in resources then it should do so. One requires only that the public sector should have a beneficial way of disposing of extra output.

Aggregate productive efficiency at the optimum can, in general, be established only where all final goods may be taxed and where the private sector is competitive and pure profits are zero. It is therefore a rather narrow result. The reason that we require zero pure profits is that without this assumption we have to consider the consequences of a reform for the incomes of profit earners. Where these profits exist, an attempt to produce extra output may increase some profits and lower others with, possibly, adverse effects on income distribution. We require all final goods to be taxed because in the absence of this assumption one may want to tax an input as means of taxing the output (which we may want to tax for reasons embodied in the models of Section 3). Broadly speaking, there are three ways of assuming away private profits for the purpose of the theorem on aggregate productive efficiency: one can assume all production is in the public sector, that all pure profits are taxed, or that there is perfect competition with constant returns to scale.

The consequences of the results concerning efficiency are important. With public sector efficiency we require that all public sector firms should face the same relative shadow prices, or market prices if financial targets are fixed in these terms. This requires more coordination of public sector production than is perhaps present in many countries. Secondly, if foreign trade at fixed prices for certain goods is part of possible public sector activity, then relative world prices for these goods give us shadow prices--foreign trade is simply one way of transforming one good into another and marginal rates of transformation in this activity should be equal to those elsewhere.

In the restrictive class of cases where aggregate efficiency is desirable then public sector shadow prices should be equal to market prices faced by private producers. Taxation of goods should fall on final goods only, for, if intermediate goods are taxed, different producers will be facing different relative prices. It is important to appreciate the circumstances under which departures from aggregate efficiency are desirable. We should tax inputs only where the taxation of final outputs is not possible or where it is necessary to improve the distribution of profit income. Thus, it is natural to question closely whether the taxation directly of the final good or the profits is possible before resorting to the substitute measure.

It should be emphasized that the discussion of efficiency in relation to taxes here includes tariffs. Where efficiency is desirable we should not have tariffs on intermediate goods. A VAT system which rebates taxes on imports by producers or a purchase tax on final goods would have this property. Notice that there is no presumption that uniform tariffs help with efficiency in this respect. The relevant question is rebating of taxes on inputs and uniformity is irrelevant to this. In general, uniformity of tariffs on intermediate goods implies inefficiency. Further, it would not lead to uniformity of taxes on final goods if these were desirable (see Sections 3 and 5) since uniform taxation of inputs becomes nonuniform on outputs through the production process, which involves different factor intensities and intensities in the use of imported goods.

5. Applications to general arguments

One of the main purposes of economic theory is to sort out correct from incorrect arguments and to help establish reliable rather than unreliable intuition. In this section we shall try to draw together from the theory some lessons of this type. We shall begin by setting out three general principles which emerge from our analysis, and then look at the question of uniformity of indirect taxation and comment briefly on the income tax.

We shall state the principles in summary form before discussing their foundation and interpretation.

a. Tax revenue is raised most efficiently by taxing goods or factors with inelastic demand or supply. Note that this abstracts from distributional questions, that inelasticity refers to compensated demands and supplies, and care should be taken with the pattern of complements and substitutes.

b. Taxation concerned with distribution and with externalities or market failures should go as far as possible to the root of the problem. Thus, for distribution one should look for the sources of inequality (e.g., land endowments or earned incomes) and concentrate taxation there. And for externalities, for example, one should attempt to tax or subsidize directly the good or activity producing the externality. Note, however, that it will often be impossible to deal completely with an issue directly and this will have very important consequences for other policies.

c. It must be recognized that it will be impossible to deal perfectly with questions of distribution and market failure directly. The former, for example, requires 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 for any particular policy we shall have to ask how it affects all our objectives--including distribution. The optimum policy for any one tax will often be very sensitive to assumptions concerning the existence and levels of other taxes.

We shall discuss the principles in turn and relate them to the preceding analysis. Our investigation started with the basic theorems of welfare economics and these establish clearly 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 points us to the first principle. The discussion of the Ramsey problem concerning indirect taxes led in the same direction but cautioned us that it was the compensated demands that were relevant. Any system of taxation will have an income effect and we distinguish among them by the "excess burden" which refers to distortions in compensated demands.

The pattern of substitutes and complements will in general be of considerable importance. For example, away from the optimum 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 which are heavily taxed. We must beware of notions of increasing marginal distortion. In a second-best world we cannot, for example, assume that a reduction in indirect taxes and an increase in lump-sum taxes will increase welfare (see Atkinson and Stern, 1974).

The notion underlying the first principle has been appreciated for a considerable time (Henry George, Wicksell, Hotelling, and so on) but its application to indirect taxes and income taxes requires care. For example, we saw in our discussion of the income tax 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. And it is illustrated by the theorems of Section 3(c) where the optimum income tax was the only policy tool required when differences arose solely in earning capacity. But we saw that this result required other very strong assumptions concerning the structure of preferences. Thus, while 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 runs the risk of considerable error.

We turn now to the question of the desirability or otherwise of uniform taxation. We should distinguish sharply at the outset between consumer and producer taxation. In Sections 3 and 4 we saw that indirect taxes should, where possible, be concentrated on final goods only. Thus, apart from special or particular arguments, we are thinking of taxes that fall on goods equally regardless of origin. This means that tariffs should be rebated on intermediate goods where it is possible to tax the final goods directly. We begin then with a discussion of whether the taxes on final goods should be uniform. In general, the results from the many-person Ramsey analysis in Section 3(a) indicate that there is no presumption 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 worse off. 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 and we saw that in very special circumstances uniformity might be desirable. These circumstances involve, for the linear income tax for example, differences in income across individuals arising only from the wage rate, a special structure of preferences (essentially the linear expenditure system), and marginal propensities to spend on each good being identical across individuals. The special nature of the conditions imply, in my judgment, that uniformity is a poor guide for developing countries. Individuals differ in many ways particularly in endowments (of, for example, land) but also in preferences where religion, caste, and education, for example, may have an important bearing. And the income tax is often an instrument of marginal importance. Further, the linear expenditure system (as Deaton, 1974, has argued persuasively) is an implausible representation of demands.

It is much more difficult to prescribe, however, than to point to the inadequacies of other prescriptions. The derivation of the appropriate set of commodity taxes requires information concerning patterns of complements and substitutes which is very difficult to extract from the data. And our attempts to do this will require specifications of functional forms which we have learned from Section 3(c) may have a profound effect on the recommendations. As Deaton (1981, p. 1245) points out: "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." One way forward is, in my view, the analysis of reform, small movements from a given starting position, which will be described briefly in the next section.

We have seen from our discussion in Section 4 that there is absolutely no presumption 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 <u>its</u> inputs might then be taxed) or to improve the distribution of profits where this is not possible by other means.

We saw in Section 3(b) 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 <u>functions</u> needs to be tutored carefully. There is no presumption, for example, of an increasing marginal rate. Indeed, the optimum schedule will, in general, show first an increasing marginal rate and then a decreasing one. And in Mirrlees' calculations (1971), the peak of the marginal rate was fairly centrally placed. We should be very careful, however, to note that this behavior of the <u>marginal</u> rate should not be seen to offend against any notion of the desirability of progression. Such notions should be related to the <u>average</u> rate and it is quite possible for the marginal rate to have the shape required but for the average rate to be increasing much of the way. For this we need only that the marginal should exceed the average and, where there is a uniform lump-sum grant, this is quite likely to be the case over a big range. But note that, in any case, a statement concerning the desirability or otherwise 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 is in striking contrast to 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. One suspects that the policy has not been designed in any systematic way.

6. <u>Tax reform</u>

By tax reform we mean a movement away from some given status quo. We shall concentrate on marginal movements. The methods will be sketched only briefly here in order to illustrate an application and possible additional applications--further discussion may be found in Ahmad and Stern (1983 a and b). Let us suppose that we have some vector of tax tools t in operation, the resulting level of social welfare is V(t), and government revenue R(t). We can think of V(t) as being defined by a Bergson-Samuelson social welfare function as before. We consider an increase in the i^{th} tax t_i sufficient to raise one dollar of extra revenue. The rate of change of revenue with respect to the tax is $\frac{\partial R}{\partial t_i}$, hence, to raise one extra dollar, we must increase the tax by $(\frac{\partial R}{\partial t_i})^{-1}$.

The rate of change of welfare with respect to the tax is $\frac{\partial V}{\partial t_i}$. 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 i^{th} good.

$$\lambda_i = -\frac{\partial V}{\partial t_i} \Big/ \frac{\partial R}{\partial t_i} \tag{22}$$

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 a beneficial reform is 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 Δt, 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 of beneficial reforms.

There is, in general, a whole collection of beneficial reforms and we should not expect uniqueness. Choice among beneficial reforms will usually be on criteria that it is not possible to 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 from which 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, thus

$$\frac{\partial V}{\partial t_i} + \frac{\lambda \partial R}{\partial t_i} = 0 \tag{23}$$

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

$$\underset{t}{\text{Maximize}}\ V(t) \qquad (24)$$

$$R(t) \geqq \overline{R}$$

The Ramsey problem, the many-person Ramsey problem, and the linear income tax are all examples of optimizing models which take the form (24).

In work described in Ahmad and Stern (1983 a and b), 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 terms of social welfare of raising revenue by different means. This has included the comparison of taxation of different goods, between state and central taxes, and between indirect taxes and the income tax. For indirect taxes, for example, we have, at fixed producer prices

$$\frac{\partial V}{\partial t_i} = -\sum_h \beta^h x_i^h \qquad (25)$$

and

$$\frac{\partial R}{\partial t_i} = \frac{\partial}{\partial t_i}(t.X) = X_i + \sum_j t_j \frac{\partial X_j}{\partial t_i} \qquad (26)$$

where β^h is the social marginal utility of income for household h and the other notation is as in Section 3. Equation (25) may be derived intuitively by noting that an increase in the price of good i hits household h in money terms by the amount x_i^h it consumes. The number β^h (a value judgment) converts the money measure into social welfare. They are to be selected by the decision maker. Equations (25) and (26) give us λ_i as in equation (22).

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

A major effort was required in our application to India to calculate the tax rates. Notice that the t_j in equation (26) are taxes actually levied on final goods. Thus, we need to work with actual tax collections and to calculate the effects of taxing intermediate goods on taxes effectively levied on final goods. We call these "effective taxes." This involves a specification of the input-output process.

We found that the method could be applied in practice and we came to a number of conclusions. First, the calculation of effective taxes provided useful information in itself. Often governments do not know the effects of their taxes on intermediate goods. In India the central excise tax is concentrated on the production of a number of basic goods. These are involved in many production processes so that the excise taxes spread out through the input-output system. This leads to some goods which are notionally subsidized being effectively taxed. Further, the resulting system is much less progressive than might be assumed from first sight. In many cases the effective taxes on domestic production are higher than on imports and higher than export rebates.

The marginal social cost of taxing different goods is quite sensitive to distributional value judgments. Cereal subsidies, for example, would be unattractive if one has little concern for inequality but more attractive otherwise. 1/ Increases in income tax have lower social cost than indirect taxes. The state sales tax seems a more attractive source of extra revenue (lower marginal social cost) than the central excise. The results were not very sensitive to changes in estimates of the aggregate demand elasticities: changes in the term $\sum_j t_j \frac{\partial X_j}{\partial t_i}$ (see (26)) seem less important than those in β^h (see (25)) over plausible ranges. Note that the consumer expenditures x_i^h are not varied when we change aggregate elasticities.

Thus, the approach can be applied and in a way that produces, in our judgment, useful results for policy.

1/ Subsidies become attractive for value judgments which involve the social marginal utility of income falling more quickly than the inverse of income--see discussion following Table 1.

Applications of the methods of modern public economics to a broad range of problems will be included in a forthcoming book to be edited by Newbery and Stern which will be published by the World Bank. 1/ In addition to developing the basic theory, applications will be included concerning the Indian tax system, agricultural pricing in Korea, energy pricing in Thailand, education in Kenya, and so on. Thus, the methods are being applied to many different problems of taxation and pricing in various countries and are likely to be applied extensively in the future.

7. Concluding remarks

A number of our general conclusions were presented in Section 5 and will not be repeated here. The purpose of the paper has been to develop and explain the main results of the modern theory of optimum taxation and to show how they might be applied to guide our judgment of tax policy.

We showed that 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 treated with great circumspection. On the other hand, we argued that the theory did yield a number of general principles (see Section 5) which are useful in guiding the practical decision maker. Further, applications to detailed calculations of possible tax reforms are possible and have been carried out in a number of contexts.

We should conclude, however, by indicating important aspects which the theories, at least up to now, have left out. First, the theories are medium term. They do not refer to short-run stabilization policy and, as yet, have not been directed toward considerations of growth. Secondly, administrative costs have been ignored. But those interested in short-run management should surely be informed by a view of where it is desirable to go in the medium term. And our discussion of administration should be influenced by our judgment of which taxes are attractive in the light of our ethical values and our application of economics.

1/ Provisional title "Modern Tax Theory for Developing Countries."

References

Ahmad, E. and N.H. Stern, "Effective Taxes and Tax Reform in India," Discussion Paper No. 25, Development Economics Research Centre, University of Warwick (1983 a).

________, "Tax Reform, Pareto Improvements and the Inverse Optimum," Discussion Paper No. 30, Development Economics Research Centre, University of Warwick (1983 b).

Atkinson, A.B., "On the Measurement of Inequality," Journal of Economic Theory, Vol. 2 (1970), pp. 244-63.

________, "Optimal Taxation and the Direct versus Indirect Tax Controversy," Canadian Journal of Economics, Vol. 10 (1977), pp. 590-606.

________ and N.H. Stern, "Pigou, Taxation and Public Goods," Review of Economic Studies, Vol. 41 (1974), pp. 119-28.

________ and J.E. Stiglitz, Lectures on Public Economics (McGraw-Hill, 1980).

Boiteux, M., "Sur la Gestion des Monopoles Publics Astreints a l'Equilibrie Budgétaire," Econometrica, Vol. 24 (1956), pp. 22-40. Translation in Journal of Economic Theory, Vol. 3 (1971), pp. 219-40.

Buchanan, J.M. and R.D. Tollison, eds., Theory of Public Choice, (University of Michigan Press, 1972).

Corlett, W.J. and D.C. Hague, "Complementarity and the Excess Burden of Taxation," Review of Economic Studies, Vol. 21 (1971), pp. 21-30.

Dalton, H., Principles of Public Finance, reprints of Economic Classics, August M. Kelley (1967).

Deaton, A.S., "A Reconsideration of the Empirical Implications of Additive Preferences," Economic Journal, Vol. 84 (1974), pp. 338-48.

________, "Optimally Uniform Commodity Taxes," Economics Letters, Vol. 2, No. 4 (1979), pp. 357-61.

________, "Optimal Taxes and the Structure of Preferences," Econometrica, Vol. 49, No. 5 (1981), pp. 1245-60.

Diamond, P.A. and J.A. Mirrlees, "Optimal Taxation and Public Production, Part I: Production Efficiency," and "Part II: Tax Rules," American Economic Review, Vol. 61 (1971), pp. 8-27 and 261-68.

Dreze, J., "Some Post-War Contributions of French Economists to Theory and Public Policy," American Economic Review, Vol. 54 (Supplement) (1964).

Feldstein, M.S., "Distributional Equity and the Optimal Structure of Public Prices," American Economic Review, Vol. 62 (1972), pp. 32-36.

Guesnerie, R. and J. Seade, "Non-Linear Pricing in a Finite Economy," Journal of Public Economics, Vol. 17 (1982), pp. 157-79.

Harberger, A.C., "Monopoly and Resource Allocation," American Economic Review, Vol. 42 (1954), pp. 77-87.

McCulloch, J.R., "A Treatise on the Principles and Practical Influence of Taxation and the Funding System" (1845), Scottish Economic Society Edition, D.P. O'Brien, ed. (Scottish Academic Press, Edinburgh, 1975).

Mirrlees, J.A., "An Exploration in the Theory of Optimum Income Taxation," Review of Economic Studies, Vol. 38 (1971), pp. 175-208.

________, "The Theory of Optimal Taxation," in Handbook of Mathematical Economics, K.J. Arrow and M.D. Intriligator, eds. (North Holland, 1979)

Musgrave, R.A., The Theory of Public Finance (McGraw-Hill, 1959).

________ and P.B. Musgrave, Public Finance in Theory and Practice, 3rd ed. (McGraw-Hill, 1980).

________ and A.T. Peacock, Classics in the Theory of Public Finance (MacMillan, 1967).

Pigou, A.C., The Economics of Welfare (MacMillan, 1947).

Prest, A.R., Public Finance, 4th ed. (Weidenfeld and Nicolson, 1960).

Ramsey, F.P., "A Contribution to the Theory of Taxation," Economic Journal, Vol. 37 (1927), pp. 47-61.

Rosen, H.S., "The Measurement of Excess Burden with Explicit Utility Functions," Journal of Political Economics, Vol. 86 (1978), S121-36.

Samuelson, P.A., unpublished memorandum for the U.S. Treasury (1951).

Seade, J., "On the Shape of Optimal Tax Schedules," Journal of Public Economics, Vol. 7 (1977), pp. 203-35.

Sen, A.K. and B. Williams, eds., Utilitarianism and Beyond (Cambridge University Press, 1982).

Shepherd, D., J. Turk, and Z.A. Silberston, eds., Macroeconomic Efficiency and Macroeconomic Performance, Philip Allan (1983).

Shoven, J.B., "Applied General Equilibrium Modelling," Departmental Memorandum, Fiscal Affairs Department, International Monetary Fund, DM/82/83 (1982).

Stern, N.H., "Welfare Weights and the Elasticity of the Marginal Valuation of Income," in Modern Economic Analysis, M. Artis and R. Nobay, eds., Proceedings of the 1976 AUTE meetings in Edinburgh, Blackwell (1977).

_______, "Optimum Taxation with Errors in Administration," Journal of Public Economics, Vol. 17 (1982), pp. 181-212.

_______, "On the Specification of Models of Optimum Income Taxation," Journal of Public Economics, Vol. 6 (1976), pp. 123-62.

Wicksell, K., "A New Principle of Just Taxation," originally published in German in 1896 and reprinted as pp. 72-118 of Musgrave and Peacock (1967).