

# ***Optimal Taxation as a Guide to Tax Policy: A Survey***

CHRISTOPHER HEADY\*

## **I. INTRODUCTION**

The purpose of this paper is to survey the field of optimal taxation. In order to provide a focus, and bearing in mind the interests of many of the readers of *Fiscal Studies*, the paper will concentrate on attempting to answer the question of whether, and to what extent, the literature on optimal taxation can provide guidance in the practical determination of tax policy. This focus means that the survey will be selective, and will neglect many intellectually interesting results that have not yet been developed to provide clear policy conclusions.

It is an interest in tax policy that has led a number of economists to undertake research on optimal taxation, and so one might expect the link between the research and the policy-making to be clear. However, the level of abstraction of much of the research, together with the extensive use of (sometimes difficult) mathematics, has caused many policy-orientated people to discount its practical value. It is dismissed as being ‘academic’, with little or no practical value. For some, the mere title of the field is off-putting: how can any aspect of public policy, with all its political constraints and administrative problems, ever hope to be optimal? Particularly, how can anything as unpopular as taxation be described in such terms?

As somebody who has contributed to the literature on optimal taxation, I would like to think that it can be helpful to practical policy-making. It has certainly been used by some economists (including myself<sup>1</sup>) who have been asked to advise governments on tax policy. However, like any theory, it has

---

\* School of Social Sciences, University of Bath.

The author would like to thank Michael Devereux and Michael Keen for their helpful comments on an earlier draft of this paper. Any remaining errors are the author's responsibility.

<sup>1</sup> For example, Heady and Mitra (1992) and Heady et al. (1992).

practical limitations and it must therefore be applied with great care. This survey is intended to indicate both the strengths and weaknesses of optimal taxation as a guide to policy-making.

The survey starts in Section II by looking at the basic ideas of optimal taxation, answering such questions as what is the point of optimisation, what is it that is being optimised and what constraints are considered? The paper then proceeds to look at the optimisation of particular aspects of the tax system: Section III deals with income taxation and Section IV with commodity taxes. Finally, Section V concludes by summarising the main lessons of optimal taxation and commenting on their practical applicability.

## **II. THE BASIC FRAMEWORK**

This section presents the basic ideas that lie behind all analysis of optimal taxation. It deals with (1) the criteria for optimality, (2) the specification of social welfare, (3) the modelling of disincentives and (4) problems of application.

### *1. The Criteria for Optimality*

Since the time of Adam Smith, and even earlier, economists have thought and written about the effects of taxation. In doing so, they have frequently tried to describe what they regarded as desirable characteristics of tax systems. Smith (1776, Book 5, Ch. 2) listed “four maxims with regard to taxes in general”:

- (i) equality: that people’s tax payments should be in proportion to their income;
- (ii) certainty: that tax liabilities should be clear and certain, rather than arbitrary;
- (iii) convenience of payment: that taxes should be collected at a time and in a manner that is convenient for the taxpayer; and
- (iv) economy in collection: that taxes should not be expensive to collect, and should not discourage business.

The second and third maxims have not been widely discussed in the economics literature, perhaps because they are self-evidently desirable. However, the ideas contained in them are frequently incorporated in statements of taxpayers’ rights.

It is the first and the last maxims that have absorbed the main interest of economists. The idea of equality has been widely discussed, and is still a major part of the evaluation of any tax policy proposal. However, Smith’s idea of equality (tax payments in proportion to income) is not the only one that has received support. Musgrave (1959, Chs 4 and 5) provides a history of differing views on what constitutes a fair distribution of the tax burden. The

administrative costs and the effects on incentives (the discouragement of business) have also been widely discussed. Taxation proposals have therefore frequently been analysed in terms of three criteria:

- (1) the need for taxes to be fair (although fairness means different things to different people);
- (2) the need to minimise administrative costs; and
- (3) the need to minimise disincentive effects.

The difficulty with having three separate criteria is that a particular policy proposal will typically satisfy one criterion but not another. For example, in choosing between the poll tax (Community Charge) and a local income tax to finance local government expenditure, most people would regard the local income tax as fairer but it would have a greater disincentive effect on labour supply than the poll tax. In order to come to a decision, it is necessary to weigh the fairness advantage of the local income tax against its disadvantage of discouraging work (not to mention taking account of the different administrative costs).

The approach of the optimal taxation literature is to use economic analysis to combine these criteria into one, implicitly deriving the relative weights that should be applied to each criterion. This is done by using the concepts of individual (or household) utility and social welfare.

Social welfare is seen as an indicator of the well-being of society and is taken to depend on the utilities of individuals. However, social welfare is not necessarily seen as simply the sum of individual utilities; it can also depend on how equally these utilities are distributed. It is typically assumed that social welfare decreases as inequality of utility increases. In this way, the concept of social welfare reflects one idea of fairness in the tax system: that taxes are fair if they reduce the degree of inequality. Thus an attempt to maximise social welfare will involve an attempt to achieve one interpretation of criterion (1).

Criterion (2) will be reflected in social welfare because higher administrative costs will require a greater amount of gross tax revenue to be collected to finance government services, thus reducing individual utilities. Criterion (3) is incorporated because the discouragement of work will distort the economy and lower people's utility and hence social welfare. In this way, all three criteria are converted to aspects of social welfare and become commensurable, and so the policy that should be chosen is the one that gives the highest level of social welfare.

This is one of the main ideas behind optimal taxation, but it is typically not carried through completely. Economists have found it very difficult to model the relationship between tax rates and administrative costs. They have therefore usually ignored administrative costs in their analysis and have concentrated on criteria (1) and (3). Effectively, they have been trying to determine the tax

systems that will provide the best compromise between equality (or fairness) and efficiency (or incentives). This neglect of administrative costs is a major shortcoming of much of the literature on optimal taxation, and is a topic to which we shall return later in this paper.

These same basic ideas have also been applied to the study of tax reform, where the aim is to identify whether specific (and typically small) tax changes will raise social welfare. There is clearly a close connection between the analysis of optimal taxation and tax reform: an optimal tax system is one in which there are no possible reforms that will increase welfare. Indeed most of the models used in tax reform analysis reflect the approach described in this paper. The only difference is that the aim is not to find the best tax system, but to find a better one. A useful exposition of tax reform analysis is given in Ahmad and Stern (1991).

The more modest aims of tax reform analysis lead to smaller informational requirements: it is only necessary to know how economic agents will respond to fairly small changes in taxes, rather than the large changes that might be involved in a move to the optimal tax structure. This is clearly an advantage, and it is worth noting that a country that follows a sequence of tax reforms that improve social welfare will eventually approach optimality. However, it is often not a good idea for a country to repeatedly change taxes and, as will be shown in this paper, there are a number of optimal tax results that are not very sensitive to the precise specification of individual economic behaviour. These results give a clear idea of some major characteristics of the desired final tax structure, whether it is approached in a series of small steps or in one large change.

None the less, it would be foolish to suppose that any country could redesign a tax system and get everything right first time. A practical approach to tax policy should combine the insights of both optimal tax theory and tax reform analysis.

## *2. The Specification of Social Welfare*

In many areas of economics, it is common to measure the well-being of people by their real after-tax incomes. However, as mentioned above, the literature on optimal taxation is based on the concept of utility. It is therefore worth considering why income is not suitable for this analysis. There are three reasons.

First, people often respond to increased taxation by working harder, to reduce the fall in their after-tax incomes. Their increased work effort clearly represents a cost to these people,<sup>2</sup> and this cost should be added to the observed reduction in real after-tax income to arrive at the total cost of the tax increase. This total

---

<sup>2</sup> If it were not a cost, they would have chosen the higher level of work effort before the tax increase.

cost would be reflected in a utility measure that took account of the disutility of work.

Second, in considering the extent of inequality, real after-tax income can be misleading. If everybody received the same hourly wage, but some worked more hours than others, there would be inequality in terms of income but the equity argument for taxing high earners more than lower earners would be weak. After all, anybody could choose to work the longer hours.<sup>3</sup> However, if the same degree of earning inequality was due to differences in wage rates, the equity argument for redistributive taxation would be considerably stronger. If utility was used instead of income as a basis for inequality measurement, the differences in work effort would be taken into account: the case of equal wage rates would show a smaller degree of utility inequality than the case with differing wage rates.

Third, when taxes are applied to consumption goods, relative prices will change and consumers will respond by changing their consumption patterns. This should result in a change in the weights used in the price index that converts nominal to real income. It is not possible to ensure that the weights change properly without knowledge of consumer preferences as represented by a utility function. It is then more convenient to use the utility function directly.

These three arguments justify the use of looking at utility functions of the form:

$$\text{Utility} = u(x_1, \dots, x_n, L) \quad (1)$$

where  $x_i$  is the consumption of good  $i$  and  $L$  is the quantity of labour supplied.<sup>4</sup>

When interest is concentrated on income taxation alone, the consumption levels of individual goods are not significant. It is only the total command over consumption goods, represented by income, that is important. The utility function can then be written as:

$$\text{Utility} = u(Y, L) \quad (2)$$

where  $Y$  is real after-tax income.

As we shall see in Sections III and IV, there are general theoretical results that are independent of the form of these utility functions. However, for quantitative applications, estimates of these functions are essential.<sup>5</sup> In the case of the simplified form (2), estimates can be obtained from studies of labour

---

<sup>3</sup> We are assuming here that there are no differences in family circumstances that affect the hours people can work.

<sup>4</sup> Although savings are not explicitly included in this utility function, they can be represented as consumption of future goods. The implications of this are discussed in Section IV(3).

<sup>5</sup> In practice, the use of statistical methods to identify the correct functional form can be very difficult. Many studies concentrate on estimating the parameters corresponding to a pre-specified functional form.

supply such as those reviewed by Blundell (1992). For the more general form (1), a complete demand system must be estimated.<sup>6</sup>

Although statistical methods can supply the form of the utility function, they cannot determine its scale: any increasing transformation of an estimated utility function would be consistent with the same observations. The scale must be fixed by the user in a way that yields the most insight into the problem being studied. In some cases, the utility function can be scaled so that utility is proportional to income or, at least, that changes in utility are proportionate to changes in income.<sup>7</sup> Then the scale can be chosen so that a one-unit increase in income (at the current prices) produces a one-unit increase in utility. This makes interpretation of results straightforward, as utility is clearly seen as real income with an adjustment for labour supplied. However, this is not always possible, with the consequence that results are more difficult to interpret.

When the individual utility functions have been determined, they must be aggregated to form social welfare. Once again, there are general theoretical results that do not depend on the method of aggregation, but quantitative results require a specific method. The simplest method, often referred to as ‘utilitarianism’, is to add the utilities:

$$\text{Social welfare} = \sum_h u^h \quad (3)$$

where  $u^h$  is the utility of individual (or household)  $h$ .

If the utility functions have been scaled to represent income (adjusted for labour supply), the measure of social welfare is simply a labour supply adjusted measure of national income.

The problem with the social welfare function (3) is that, just like national income, it takes no account of income distribution. If our idea of a fair tax system is one that reduces inequality of utility, our social welfare function must place more weight on utility gains of poor people than those of rich people. This is achieved in most studies by using the following formulation, which transforms utility:

$$\text{Social welfare} = \frac{1}{1-\varepsilon} \sum_h (u^h)^{1-\varepsilon} \quad \text{for } \varepsilon \neq 1 \quad (4)$$

$$\text{Social welfare} = \sum_h \log(u^h) \quad \text{for } \varepsilon = 1 \quad (5)$$

Concentrating first on expression (4), note that for  $\varepsilon=0$  it is the same as expression (3). So when  $\varepsilon=0$ , there is no concern for inequality. However, when

---

<sup>6</sup> For example, Ebrahimi and Heady (1988) use estimates from Blundell and Walker (1983).

<sup>7</sup> The two cases are those of homothetic and quasi-homothetic preferences, respectively.

$\varepsilon$  is positive, increases in  $u^h$  are transformed into less than proportional increases of

$$\frac{1}{1-\varepsilon}(u^h)^{1-\varepsilon}$$

This implies that less weight is attached to a given absolute increase of utility for somebody with high utility than for somebody with lower utility. The social welfare function therefore embodies a preference for equalising utility. The strength of this preference increases with the value chosen for  $\varepsilon$ . As  $\varepsilon$  approaches infinity, the extent of the preference for equality becomes so strong that only the utility level of the worst-off person has any weight in the social welfare function, representing the view proposed by Rawls (1971). Expression (4) is indeterminate for  $\varepsilon=1$  and must be replaced by (5) in that case.

The value chosen for  $\varepsilon$  has a particularly straightforward interpretation if the utility function is proportional to adjusted income. Consider two people, one with twice the adjusted income of the other. The weight placed on additional income going to the poorer person relative to that for the richer person is given by  $2\varepsilon$ . Thus, if  $\varepsilon=1$  the relative weight is 2, and if  $\varepsilon=2$  the relative weight is 4. Thought experiments of this sort allow one to think what a reasonable value of  $\varepsilon$  might be.

Before moving on to discuss disincentives, the consequences of this social welfare formulation can be illustrated by considering a case in which there are no disincentive effects: labour supply is fixed. In this case, it is straightforward to show that the optimal tax policy under social welfare function (3) would involve levying taxes so that the marginal utilities of income were equalised. In the case of social welfare function (4) it would be the marginal transformed utilities that would be equalised. If, in addition, we assumed that all utility functions were identical and everybody worked the same number of hours, both (3) and (4) imply that optimal taxation would produce a perfectly equal distribution of after-tax income.

This result demonstrates that we have departed from Adam Smith's idea of fair taxation: that taxes should be proportional to income. Instead, the objective is to minimise inequality, and in the absence of disincentive effects the inequality can be reduced to zero without any efficiency loss.

There are at least two good reasons for abandoning Adam Smith's principle of fairness: (1) keeping taxes proportional to income at all levels will cause great hardship for people at or below the poverty line and (2) as argued above, it should be utility rather than income that represents ability to pay. None the less, the idea that the aim of tax policy is to eliminate all inequality has been used in attempts to discredit optimal tax theory and requires some comments.

First, the equality result obtained above depends crucially on the assumptions of identical utility functions and equal labour supply. If people have different

needs (perhaps because of illness or numbers of children) or differ in their labour supply, equality of after-tax income is no longer desirable. Thus hard work would still be rewarded.

Second, in practice, disincentive effects prevent the equality result from being implemented. The goal of equality must be qualified by the need for incentives: this is the trade-off between efficiency and equity. The question of what would be desirable without disincentive effects therefore becomes rather irrelevant. All that is important is the wish to reduce inequality below its current level, and even that wish will be offset to some extent by a wish to maintain the efficiency that is also required to maximise social welfare.

### *3. The Modelling of Disincentives*

All taxes affect behaviour to some extent: it is simply impossible for an individual to pay a higher tax bill without reducing consumption, increasing income, reducing savings or increasing borrowing. The approach of the optimal tax literature is to model this response to taxation in a manner that is consistent with the specification of utilities discussed in the previous section and to trace through the consequences of such behaviour.

This approach can be illustrated with the example of optimal income taxation in a model where labour supply response is the only disincentive problem. In this case, the utility function (2) for each individual is used both to predict how that person will alter their labour supply when taxes are changed and to evaluate the resulting level of individual utility. The changes in labour supply will then be used to calculate the change in tax revenue, while the changes in utilities will be used to calculate the change in social welfare. The optimal tax system will be the one where it is impossible to increase social welfare without reducing overall tax revenue.

The requirement to raise a specific amount of tax revenue<sup>8</sup> is obviously fundamental, for otherwise taxes could just be reduced to zero. It has two important implications. First, it means that the solution to the optimal tax problem will depend on the size of the revenue requirement. Second, it means that the tax changes that are considered should be revenue-neutral. For example, an increase in the standard rate of income tax would allow an increase in the personal allowance so that the 'average taxpayer' will continue to pay the same amount of tax.

The importance of this last point becomes clear if we divide the effect of a tax change into an 'income effect' and a 'substitution effect', just as in the standard economic analysis of price changes. The income effect of a tax increase

---

<sup>8</sup> Most of the optimal tax literature takes the revenue requirement as given by a pre-specified level of government spending. The reasons for this level of government spending, and its possible optimisation, are not considered.

is that it reduces after-tax income and so increases the individual's labour supply, in an effort to ameliorate the reduction in consumption. The substitution effect is that the marginal return to work is reduced, thus leading to a reduction in labour supply. The fact that these two effects go in opposite directions means that the effect of an income tax increase on labour supply could be in either direction, depending on which effect is stronger. However, in revenue-neutral tax changes the average taxpayer does not have an income effect, so only the substitution effect operates for that person. Other taxpayers will experience increases or reductions in taxes, but the resulting income effects will probably approximately balance out. Thus, overall the substitution effect will prevail: an increase in the standard rate of income tax that is used to finance an increase in the personal allowance will generally reduce total labour supply.

This dominance of the substitution effect that results from revenue neutrality applies to all optimal tax problems, and leads to an emphasis on the compensated elasticities of supply and demand in the evaluation of the distortionary effects of taxation.<sup>9</sup>

Returning to the example of income taxation, why does it matter that a higher tax rate with higher personal allowances will reduce labour supply? After all, the objective is to maximise social welfare, not the size of the national income. The answer is that by choosing to work less on average, workers will have lower incomes and thus will pay less taxes. Thus a change that would have been revenue-neutral for a fixed level of labour supply will, as a result of the reduction in work, produce a revenue loss. It is this revenue loss that represents the "excess burden" of taxation. It requires an increase in tax rates to offset it; an increase that will reduce social welfare and counteract, at least in part, the gain in social welfare from the reduction in inequality that is produced by the increase in tax progressivity. The factors that determine whether the overall effect on social welfare is positive or negative will be discussed in Section III.

The approach in other areas of optimal taxation is basically similar to that described above for income taxation and labour supply. There is therefore no need to discuss them separately. However, it is worth pointing out that any modelling of disincentive effects is likely to be selective. For example, income taxation can affect other decisions apart from labour supply, such as educational choice, savings decisions and even the decision to evade taxes. Models have been constructed to look at these different disincentive effects separately or in conjunction with labour supply, but there is no model that combines them all. This is probably because the complexity of such an omnibus model would be too great to yield any useful insights.

---

<sup>9</sup> A compensated elasticity measures the strength of the substitution effect alone, indicating the response to a price change under the assumption that the individual is compensated (positively or negatively) to maintain their real income. This is appropriate because of the revenue neutrality of the tax changes being considered.

#### *4. Problems of Application*

Before proceeding, it is important to note the aim of optimal tax analysis. It is to describe the taxes that governments should set, not to explain the taxes that governments do set. One might like to think that governments do what they should, but there are a number of reasons for believing that they do not. A close relationship between the prescriptions of optimal tax analysis and the tax systems that are actually implemented should not necessarily be expected.

The idea of what should be done must, of course, be based on ethical views. In the case of taxation, the role of ethics is confined to the view that is taken about the importance of reducing inequality. In the theory, this is represented by the parameter  $\varepsilon$  in the social welfare function (4). Clearly, people can differ in their views about a reasonable value for  $\varepsilon$ , and so it is standard practice to calculate optimal taxes for several different values.

It might be thought that appropriate choice of  $\varepsilon$  could rationalise almost any tax system, but that is not correct. The studies discussed below show that variations of  $\varepsilon$  often have fairly modest effects on optimal tax rates, and typically have a very small effect on the relative size of different taxes. The range of optimal taxes that are calculated for different values of  $\varepsilon$  is both inevitable and desirable: one cannot believe that people with different views on inequality would choose the same rate of income tax.

The usefulness of the optimal tax results will depend in part on the realism of the economic models that are used in their derivation. This is not to say that the presence of any unrealistic assumption invalidates the results. Rather, any practical application of theoretical analysis requires an evaluation of whether any violation of the assumptions can be expected to alter the results significantly.

One way in which many models are unrealistic has already been mentioned: their neglect of administrative costs of tax collection. To this can be added neglect of the costs to taxpayers of compliance. These costs have usually been omitted because they do not vary continuously with the tax rates. Instead, they tend to vary with such things as the number of different rates of tax or the number of tax allowances. This makes them difficult to include in the mathematical analysis.

This does not mean that administrative and compliance costs have been completely neglected. A number of studies have used administrative costs as a reason for restricting the number of tax rates or for ruling out some taxes altogether. If numerical calculations are used, it is possible to compare optimal policies that correspond to different restrictions on the number of tax rates. The difference in the optimal values of social welfare can be converted into money terms by calculating the change in total income that would produce a similar welfare change. This sum of money then represents the gain of allowing a more complex tax system, and this could be compared with the likely additional administrative and compliance costs.

Another doubtful assumption in some, but not all, models is that the economy is otherwise undistorted. Perfect competition is usually assumed, as is the absence of environmental effects and other externalities. The dropping of these assumptions would alter the optimal pattern of taxes<sup>10</sup> but would greatly increase both the complexity of these models and their data requirements. The application of optimal tax results in situations where imperfect competition or externality problems are significant therefore requires considerable care.

Models also differ in the extent to which they recognise the diversity of households in terms of their composition and preferences. Until now, this paper has focused on individuals, but their position within households can affect their labour supply, their demand patterns and their level of utility. This fact has been recognised in some of the literature, but it has been ignored in many papers in order to clarify the main point of the discussion. Obviously, practical policy must at least take account of the different needs of different demographic groups.

Finally, it should be noted that optimal tax analysis has concentrated on personal income taxes and commodity taxes, including taxes on international trade. It has not dealt with company taxation, capital gains tax or inheritance taxes.<sup>11</sup> The reason for this is probably that the effects of these taxes on behaviour and utility are less well understood than the effects of personal income tax and commodity taxes. Also, issues of administrative costs and enforcement are considered as more important for these taxes. This presents an extra difficulty in devising a suitable mathematical formulation of the tax design problem.

### **III. OPTIMAL INCOME TAXATION**

The first analysis of income taxation using the methodology described in Section II was by Mirrlees (1971). This paper stimulated a number of further studies, each analysing the same basic model of the economy, often referred to as the 'Mirrlees model'.

The two fundamental assumptions of the Mirrlees model are: (1) the only disincentive effect of taxation is on the number of hours supplied by each worker and (2) differences between the wages of different workers are produced by differences in their fixed productivities. The fixed productivities of assumption (2) imply that relative pre-tax wages are fixed.

The Mirrlees paper considered the problem of designing an optimal non-linear income tax, in which the marginal tax can vary as income rises. However, it is also interesting to consider the design of simpler tax systems with a constant

---

<sup>10</sup> Some of the effects of introducing imperfect competition are discussed in Myles (1989a).

<sup>11</sup> To the extent that these taxes are simply seen as taxes on savings, they are covered by the analysis discussed in Section IV(3). However, these taxes also have an effect on the choice between different financial assets, and the full consequences of that have not been modelled.

marginal rate: an optimal linear income tax. It is also important to consider the implications of changing the two key assumptions. These three topics are now discussed in turn.

### *1. Non-Linear Income Taxation*

Much of the discussion in Mirrlees (1971) is technically difficult but the main ideas can be understood without going into the technicalities.

The fundamental policy issue is whether it would be a good idea to increase the rate of income tax and use the proceeds to fund an increase in tax allowances, thus reducing after-tax income inequality.

Because we are dealing with non-linear taxation, it is possible to consider the effect of changing the marginal tax rate over a short range without changing the marginal tax rate at any other incomes. The people with incomes below the range of tax increase will be unaffected by this change. People with income within the range will experience an income effect from the higher taxes which would tend to increase their labour supply. However, they will also experience a substitution effect tending to reduce their labour supply because of the reduced reward for additional work. Because the increased tax rate only applies to a small proportion of their income, the size of the reduction in after-tax income will be small and so the income effect will generally be less than the substitution effect. Therefore, the labour supply of this group will be reduced and they will therefore probably pay less tax.

People with incomes above the range will not experience a substitution effect because their marginal tax rate has not changed. However, they will pay more tax because of the higher rate applied to some of their income. This will have an income effect, causing them to work more and so pay yet more tax.

Overall, there are three effects of the tax increase on tax revenue and welfare: (i) the tax payments of people with the increased marginal rate will probably fall, (ii) the tax payments of people with income above the range of increase will rise and (iii) the utility levels of both groups affected by the tax increase will fall.

If the net effect of (i) and (ii) is negative, there is no extra revenue available to fund an increase in tax allowances and so the increase in the marginal tax rate is clearly not desirable. This is most likely to occur if either effect (i) is large, because of a high compensated elasticity of labour supply, or effect (ii) is small, because the number of people above the range of increase is small.

If the net effect of (i) and (ii) is positive, the revenue gain from the tax increase must be weighed against the utility loss of effect (iii). This can be done by calculating the welfare gain that would be produced by using the additional revenue to increase tax allowances. It is then a matter of weighing the utility loss to higher-income people against the utility gain to lower-income people.

Overall, therefore, the net effect on social welfare will depend on four factors:

- (1) the compensated elasticity of labour supply: a high elasticity will mean that the net revenue gain is either small or negative, so the tax increase is less likely to increase social welfare;
- (2) the degree of concern for inequality, as represented by  $\varepsilon$  in equation (4): the higher is  $\varepsilon$ , the smaller is the relative weight placed on the utility losses of the losers from the tax increase, and so the tax increase is more likely to increase social welfare;
- (3) the degree of income inequality: a high level of inequality implies a greater income difference between the (relatively poor) gainers and the (relatively rich) losers from the tax change, implying that a greater relative weight should be attached to the gains, and that the tax increase is more likely to increase social welfare;
- (4) the proportion of the population above the range of the tax increase: the higher is this proportion, the greater is the amount of gain to the poorest, and so the tax increase is more likely to increase social welfare.

One implication of factor (4) that has attracted considerable attention in the theoretical literature is that the marginal income tax rate for the person with the highest income should be zero. This is because there is no extra revenue to be obtained by raising it above zero, and so no reason to distort that person's labour supply decision.

The argument really needs to be more complicated because the government cannot know in advance the precise level of the highest income. A full analysis involves consideration of the probability distribution of incomes, and the result becomes less sharp: the marginal tax rate should approach zero as incomes become very high, provided that the probability density declines at a sufficiently fast rate.

However, from a practical policy point of view, the importance of this result does not lie in its precise form: tailoring the very top of the income tax schedule to minimise the disincentive effects for a very small number of people can hardly be seen as a major policy issue.<sup>12</sup> Rather, its importance lies in it being a counter-example to the widespread belief that a redistributive government must use an income tax schedule with increasing marginal rates. Here is a case where, however strong is the wish to increase the utility of the poor, the marginal tax rate declines at the top of the income distribution.

This theoretical result can only be derived for the top of the income distribution. However, Mirrlees (1971) also calculated complete optimal income tax schedules for some specific numerical examples. These show a very gentle decline in the marginal tax rate over most of the income distribution. The decline

---

<sup>12</sup> 14 This point is reinforced by calculations of optimal income tax schedules reported in Tuomala (1990), which show substantial positive marginal rates even for people in the top 1 per cent of the income distribution.

is so slight that the optimal income tax schedule can be approximated fairly well by a tax system with a constant marginal rate: a linear income tax.<sup>13</sup>

This finding has considerable practical significance because it suggests that there is no conflict between theoretical optimality and administrative convenience. All the advantages of a single marginal rate, most notably the ability to accurately withhold taxes from multiple sources, can be achieved with no worsening of the inescapable trade-off between equity and efficiency.

It is important to realise that a linear income tax can achieve significant redistribution. Personal tax allowances that are a substantial fraction of average income can produce considerable progressivity in the average tax rate over most of the income distribution. The constant marginal rate implies that the degree of progressivity declines (but is still positive) at high income levels. But there are so few people at those incomes that additional progressivity will yield little or no extra revenue.

## *2. Linear Income Taxation*

The conclusion that optimal income tax schedules are approximately linear has allowed investigators to concentrate on analysing optimal linear taxation. This has the important practical advantage of greatly simplifying the calculations needed for numerical examples.

The most important study is that of Stern (1976), who took considerable care in selecting realistic functional forms and parameter values. The study uses the same basic structure as the Mirrlees model and reports the optimal tax rates that correspond to a range of values for the compensated elasticity of labour supply, the degree of concern for inequality and the size of the government's revenue requirement.

Table 1 presents a small part of Stern's results. It shows the optimal tax rates corresponding to three different values of  $\epsilon$ , for each set of assumptions about the compensated elasticity of labour supply and the size of the government's revenue requirement. Case 1 corresponds to Stern's best estimate of labour supply elasticity and a reasonable value for the government's revenue requirement. The remaining cases alter one parameter at a time, keeping the other constant. Thus Case 2 reduces the value of the labour supply elasticity and Case 3 raises it, while the revenue requirement is kept constant at the value for Case 1. Similarly, Case 4 reduces the revenue requirement and Case 5 raises it, keeping the labour supply elasticity at the value for Case 1.

---

<sup>13</sup> Tuomala (1990) shows that more rapidly declining marginal rates can be desirable when different parameter values are assumed. However, none of the numerical examples supports the widespread use of increasing marginal tax rates.

TABLE 1  
Optimal Tax rates for Varying Parameters

	$\varepsilon = 2$	$\varepsilon = 3$	$\varepsilon = \infty$
Case 1	54%	59%	87%
Case 2	68%	72%	94%
Case 3	45%	50%	80%
Case 4	48%	53%	84%
Case 5	61%	65%	90%

Note: See text for description of cases.

The results confirm the effect of the first two factors considered in the discussion of non-linear income taxation and Stern states that unreported calculations also confirm the third effect. The original fourth effect no longer applies as the marginal tax rate is constant, but can be replaced by the effect of the government's revenue requirement. Thus the optimal marginal rate of income tax is higher for:

- lower values of the compensated elasticity of labour supply;
- higher values of  $\varepsilon$ , the degree of concern for inequality;
- greater inequality in pre-tax wages;
- higher government revenue requirement.

The results also confirm another important aspect of the analysis of optimal non-linear income taxation: for some parameter values, the revenue effect of raising the tax rate is negative. This implies that there is a limit to how high the tax rate should be raised, however strong is the concern for inequality. This limit depends crucially on the value of the compensated elasticity of labour supply. This shows that observable market behaviour can limit the role of the ethical views in determining tax policy: we can all agree that some tax rates would be too high.

None the less, differences in concern for inequality can generate a substantial range of optimal marginal tax rates, especially when one notes that smaller values of  $\varepsilon$  would produce lower optimal tax rates. For reasonable parameter values, this range includes the tax rates currently used in most OECD countries.<sup>14</sup> One cannot say that one of these rates is better than another without taking a firm view on how much concern should be shown for inequality.

<sup>14</sup> As the Mirrlees model does not explicitly include any other taxes, any comparison between model results and actual rates should take account of all taxes that vary with income, not just the formal "income tax". In the UK, this would include VAT and other sales taxes, so that the model-equivalent marginal income tax for the typical taxpayer would include both the standard rate of income tax and the percentage of additional income that would go in sales taxes. The question of whether National Insurance contributions should be included turns on

One important feature of linear income taxation is that people with income below the tax exemption level will receive a transfer payment from the government.<sup>15</sup> In other words, the linear income tax incorporates a negative income tax system of social security. This is a feature that differs quite markedly from the practice of most OECD countries, in which social security benefit entitlement is much more complex and the implicit marginal tax rate on benefit recipients is frequently very much higher than the rate applied to non-recipients.

### *3. Changing the Assumptions*

Although the Mirrlees model has held centre stage in the study of optimal income taxation, there have been attempts to analyse the consequences of altering the basic model. Tuomala (1990) provides a thorough discussion of many of these alterations. Two interesting examples are the work of Allen (1982) and Atkinson (1973).

Allen (1982) considers the effects of dropping the assumption of fixed relative wages, but otherwise maintains the Mirrlees assumptions. Two types of worker are considered and their relative wages are assumed to depend on the relative supply of the two types of labour, the wage of one type declining as its supply increases. If, in the absence of taxation, one type of worker is paid less than the other, the optimal non-linear income tax will be quite different from the Mirrlees case. The poorer group will face a positive marginal tax rate, while the richer group will face a negative marginal tax rate. The reason for this is that it will reduce the supply of low-paid workers and increase the supply of high-paid workers, thus reducing the extent of pre-tax inequality. The size of this effect will clearly depend on how difficult it is to substitute one type of worker for another.<sup>16</sup>

This result is certainly provocative, although some of its features depend on there being just two distinct groups with an income gap between them.<sup>17</sup> None the less, Carruth, Heady and Ulph (1983) show that the result can be extended to a model with a large number of different types of worker. In that case the marginal tax rate at the top of the income distribution is negative, rather than the zero of the Mirrlees model. However, it is difficult to establish the form of the complete tax schedule in a model as complicated as this. Moreover, there is little evidence on the crucial question of how difficult it is to substitute between different types of worker. This means that further work is needed before an

---

whether they are viewed as a tax or as savings. In the UK, the connection between marginal contributions and level of benefit is so weak that they should probably be regarded as a tax.

<sup>15</sup> This characteristic of optimal income tax schemes is not confined to the linear case: the non-linear calculations in Mirrlees (1971) also included transfer payments to low-income individuals.

<sup>16</sup> In more technical terms, it will depend on the inverse of the elasticity of substitution between different types of labour.

<sup>17</sup> The income gap allows the tax schedule to be manipulated to make the higher-paid group pay more tax despite their negative marginal rate.

assessment can be made as to whether relative wage effects significantly affect the practical policy implications of the Mirrlees model.

Atkinson (1973) looks at a different type of disincentive. In his model, hours of work are fixed but the efficiency of labour depends on the educational level of the worker. Receiving education is costly and so an income tax can have a disincentive effect on educational choice. Atkinson restricts his attention to the optimal linear income tax. Although the model is concerned with a different dimension of labour supply choice — quality rather than quantity — the mathematical structure is similar to that of Stern's analysis of linear income taxation in the Mirrlees model. It is therefore not surprising that the range of optimal tax rates obtained for reasonable parameter values is similar to that obtained by Stern.

Neither of these two variations on the basic optimal income tax model provides us with much practical guidance. The most important practical result from the Mirrlees model is that the optimal income tax schedule is approximately linear. Atkinson's model does not address this issue, as it imposes linearity from the start. Allen's model casts some doubt on the desirability of linear income taxation, but there is insufficient evidence of the degree of tax non-linearity that it would imply in practice. Clearly there is room for more work here.

Finally, it is worth noting that most of the optimal income tax literature does not deal with one important practical policy issue: the extent to which demographic characteristics of households should be reflected in the tax system. For example, how should the presence of children affect a household's tax liability? This sort of question has, however, been addressed in the literature on optimal commodity taxation, and it is to this topic that we now turn.

#### **IV. OPTIMAL COMMODITY TAXATION**

The literature on optimal commodity taxation is mainly concerned with the design of final sales taxes, such as value added tax and the excise duties on alcohol, tobacco and petrol. However, it has also dealt with the taxation of intermediate goods and international trade, and can be used to analyse the taxation of savings.

##### *1. Final Sales Taxes*

The first analysis of optimal sales taxes was undertaken by Ramsey (1927) and considerably predates the literature on optimal income taxation. It focused on a rather different question. Instead of looking at the trade-off between equity and efficiency, it analysed the problem of designing sales taxes to raise a given amount of revenue at the least possible distortionary cost in a single-person economy (or, equivalently, an economy with many identical people).

To a certain extent, this is not a serious problem. If there is no inequality, there is no reason to avoid the use of a poll tax, which would have no distortionary cost. It is only a concern for the regressive impact of a poll tax that leads to the desirability of using distortionary taxation. None the less, the results of Ramsey and of Corlett and Hague (1953) in the single-person context provided useful insights for the subsequent analysis of optimal sales taxes in an economy with inequality.

Ramsey showed that, when only a very small amount of revenue had to be raised, the taxes should produce equal proportional reductions in the consumption of each good. He then showed that this result continued to hold, even for substantial revenue requirements, if there were no income effects and if the demand curves for the goods were linear. These conditions are most unlikely to hold in practice but, as is shown in Atkinson and Stiglitz (1980, p. 372), this 'equal proportional reductions' rule can be expressed as a generally applicable mathematical condition that optimal taxes should satisfy.

Unfortunately, this condition does not provide a direct indication of which goods should be most heavily taxed. Greater intuition can be obtained by making an additional assumption: that the demand for each good is independent of the prices of other goods. Using this assumption, Ramsey derived the 'inverse elasticity rule', that goods with more price-inelastic demands should be taxed more heavily. This rule only strictly applies under the assumption of independent demands and needs considerable revision when income inequality is taken into account. However, the rule has wide influence and its basic rationale — that the taxation of inelastic goods yields more revenue because demand only falls a little — is probably partly responsible for the high taxation of alcohol, tobacco and petrol all over the world.

Corlett and Hague (1953) approached the issue of tax design from a different perspective. Instead of asking which pattern of taxes would be optimal, they looked at a situation where there are two consumption goods taxed at the same rate and asked whether efficiency could be improved by introducing some non-uniformity (raising the tax on one good and lowering the tax on the other). They showed that, if the goods differed in their degree of complementarity or substitutability with leisure, efficiency could be improved by increasing the tax rate on the good that was most complementary (or least substitutable) with leisure and reducing the tax rate on the other good.

The intuition behind this result is as follows. A uniform tax on the two consumption goods is effectively the same as an income tax (ignoring savings). The distortionary effect is therefore one of discouraging labour supply or encouraging leisure. An increase in tax on a good that is complementary with leisure will discourage the consumption of leisure, increase labour supply, and so partially offset the original distortion.

Clearly, if uniform taxation were optimal, the introduction of non-uniformity would not improve efficiency. The Corlett and Hague result therefore tells us

that uniform taxes are optimal if all goods have the same degree of complementarity or substitutability with leisure. It also suggests a result which was later demonstrated by Diamond and Mirrlees (1971): that, in this two-good case, optimal taxation involves placing a heavier tax on the good that is most complementary to leisure.

The Corlett and Hague result was obtained in a model that is basically the same as that used by Ramsey and so one should expect the two results to be consistent, as indeed they are. The relationship between the two results is explored in Heady (1987), using a diagrammatic analysis. All that need be noted here is that in the 'inverse elasticity' case (with independent demands), it can be shown that the good which is most complementary to leisure will also be the good with the most inelastic demand curve. Thus the two results pick the same good to be most heavily taxed.

The next major step in the development of the theory of optimal commodity taxation came with the analysis of an economy with inequality by Diamond and Mirrlees (1971). They showed that the introduction of distributional considerations alters the equal proportional reductions rule substantially. The most significant alteration was that goods which are consumed particularly heavily by the poor should experience a lower-than-average proportional reduction. The extent of the differentiation in proportional reductions would depend on the degree of concern for the poor, as represented by our parameter  $\epsilon$ , and the extent of differences in consumption patterns between the rich and the poor.

In the case of independent demands, the Diamond and Mirrlees result shows that the optimal tax rate on a good should depend not only on the inverse of its price elasticity of demand but also on its income elasticity, which indicates how the budget share of a good changes as income rises. The significance of this modification can be appreciated when one notes that many goods with low price elasticities also have low income elasticities: the demand for goods that are regarded as necessities will not be very responsive to changes in either price or income. For these goods, the efficiency argument for high taxation must be balanced against the distributional argument for low taxation. The question arises of whether differential taxation really is a good idea, and this has been the focus of much recent research on optimal sales taxes.

The analysis of whether differential sales taxation is desirable was encouraged by two further considerations. First, the costs of administration and compliance are much lower if sales taxes are uniform. Second, if we ignore the issue of savings (which will be taken up later), a uniform sales tax accompanied by a uniform payment to all households is equivalent to a linear income tax with a suitably chosen exemption level and marginal tax rate.<sup>18</sup> Therefore, if uniform

---

<sup>18</sup> The exemption level would have to be chosen so that people with no income would receive the same real payment from the government under the two schemes. The equivalent marginal income tax rate would be lower

sales taxes are optimal, the choice of the relative proportions of revenue to be raised by income tax and sales tax need only take account of the effect on savings and of administrative considerations, such as collection costs and the need to minimise tax evasion.

The major results on whether differential sales taxes are desirable in an economy where households differ only in their incomes and not in their underlying preferences are shown in Atkinson and Stiglitz (1980, Ch. 14). An important aspect of their analysis is the role of the uniform payment to all households (or the income tax exemption level). If all goods are normal, in the sense of being consumed in larger quantities by people with higher incomes, the poor will always benefit more by an increase in the uniform payment than by the same amount of money being used to reduce the sales tax on a particular good: the reduction in sales taxes will benefit the rich more because they buy more of the good.

As empirical studies in the UK have failed to find any categories of goods that are not normal, this argument implies that the government's redistributive goals are best achieved by an appropriate choice of the uniform payment. Therefore, the issue of whether to have differential sales taxes is really one of efficiency, provided that the uniform payment is set optimally, and we are back with the Corlett and Hague question of whether differential sales taxes will reduce the disincentive effect on labour supply of an income tax. As explained above, the answer depends on differences in the degree of complementarity between individual goods and leisure, and Atkinson and Stiglitz show that the condition for uniform taxation to be optimal is that the uniform payment is set optimally and that there is weak separability between goods and leisure.<sup>19</sup> This condition means that households with different hours of work but the same income (because of different wage rates) will choose the same quantities of consumer goods.

The intuition here is essentially the same as for the Corlett and Hague result. Consider a good that is more heavily consumed by households that take more leisure, golf-clubs for example. Additional taxation of such a good is essentially a tax on leisure and will ameliorate the effect of the tax on work. On the other hand, if one could identify goods consumed by those who work more (convenience foods, perhaps), a lower rate of tax on them would encourage work effort.

An obvious difficulty that arises in attempting to apply the Atkinson and Stiglitz result to a country like the UK is that their model ignores differences in

---

than the uniform sales tax rate. For example, a 100 per cent sales tax halves a person's real disposable income and so is equivalent to a 50 per cent income tax rate.

<sup>19</sup> This condition is slightly different from the Corlett and Hague result, because of the presence of the uniform payment. It should also be noted that the Atkinson and Stiglitz result requires that the Engel curves (the relationship between a household's demand for a good and its income) be linear, but this is not required if there is an optimal non-linear income tax.

preferences between households that might arise from differences in demographic characteristics. This is particularly significant because the arguments in favour of VAT zero-rating for some goods is that they form a large part of the budget of particular demographic groups. Thus the zero-rating of food and children's clothing is justified by the observation that families with large numbers of children are particularly prone to poverty and spend a high proportion of their budgets on these items.

Deaton and Stern (1986) extend the Atkinson and Stiglitz result to an economy with different demographic groups and show that uniform taxation is still desirable if preferences are weakly separable, provided that households in each demographic group receive an optimally chosen payment which is uniform within each group but differs between groups. The idea here is that the redistribution between groups is accomplished most efficiently by the use of direct payments to households, leaving the sales tax rates to deal with problems of efficiency.

Ebrahimi and Heady (1988) develop the Deaton and Stern analysis and apply numerical analysis to look at the question of whether it would be better to abolish the zero-rating of food and use the additional funds to finance an increase in child benefit. The numerical results confirm the Deaton and Stern theoretical result under their assumptions, but the effect of relaxing these assumptions is also investigated. Some of the results are shown in Table 2.

TABLE 2  
**Optimal Sales Taxes with Child Benefit ( $\epsilon = 1$ )**

	(1)	(2)	(3)	(4)
Tax on energy	30%	41%	28%	35%
Tax on food	34%	40%	32%	37%
Tax on clothing	52%	41%	55%	42%
Tax on other goods	44%	40%	49%	46%
Lump sum (per week)	£28	£28	£57	£56
Child benefit (per week)	£22	£22	0	0

Notes: Column (1) is based on the estimates from Blundell and Walker (1983).  
 Column (2) is the same as column (1) but separability had been imposed.  
 Column (3) is the same as column (1) but there is no child benefit.  
 Column (4) is the same as column (2) but there is no child benefit.

All of the results in Table 2 correspond to a zero government revenue requirement (above that needed to pay for the child benefit and personal tax allowances). The personal tax allowances have been converted into equivalent lump-sum payments, and so we have the equivalent of a negative income tax. However, there is no explicit income tax in the model: all the revenue is raised by the sales taxes, which accounts for their high rates, but could equivalently be

raised by an income tax and a lower level of sales taxes. That would have no effect on the degree of non-uniformity of the sales taxes, which is what concerns us here.

Column (1) represents the optimal sales tax pattern combined with optimally set lump-sum payments and child benefit. It is based on demand system estimates that allow for non-separability and so the existence of some non-uniformity is not surprising. It shows that available empirical estimates of the degree of non-separability do justify a small amount of differential taxation, even if child benefit is set optimally. However, the welfare loss of imposing uniformity is equivalent to only 0.04 per cent of GNP and could well be outweighed by the administrative problems of non-uniformity.

Column (2) shows the effects of altering the demand system estimates to impose separability. The resulting optimal taxes are almost uniform, but not quite because there are two of Deaton and Stern's requirements that have not been met. First, there is a condition that the Engel curves should have the same slope for all households. Second, we have imposed a uniform child benefit for each child, while Deaton and Stern allow child benefits to be set separately for families with different numbers of children. However, the non-uniformity is so small as to suggest that these two theoretical conditions are of little practical importance.

Columns (3) and (4) show the optimal taxes when there is no child benefit. These results show that the desirability of uniform taxation depends crucially on the optimal setting of child benefit. If it is not set optimally, substantial non-uniformity can be justified even if the weak separability condition is satisfied.

Table 2 also provides an illustration of how optimal tax theory can be used to calculate the special tax treatment given to different demographic groups, a point that was found to be missing in the review of optimal income tax in Section III. Child benefit is obviously not a tax provision, but its effects are identical to a child tax allowance under a linear negative income tax system.<sup>20</sup>

The optimal child benefit reported in Table 2 is substantially higher than the current value in the UK. However, this result should be treated with some care. As explained in Ebrahimi and Heady, this result is based on the assumption that while children are costly to raise, they provide no benefits to the household. This is obviously untrue, for otherwise no household would choose to have children. The proper setting of child benefit must involve more complex issues than the simple economic cost of child-raising. These would include issues of parental responsibility, and the extent to which children should be allowed to suffer from the fertility decisions of their parents.

---

<sup>20</sup> The equivalent child tax allowance is one which has the same budgetary cost as the child benefit. The linearity of the tax system and its extension to households below the tax threshold (the "negative income tax") is necessary to ensure that, like the child benefit, the net benefit of the tax allowance is the same for all households.

This suggests that optimal tax analysis cannot settle all issues related to the treatment of different demographic groups. However, it does not compromise the importance of the tax uniformity result: direct payments are more effective than non-uniform sales taxes in achieving distributional goals. It is only when direct payments are impracticable that non-uniform taxes should be employed.

Although the optimal tax literature has addressed the issue of differential sales taxes that might be sought on distributional grounds, it has not really confronted the issue of whether the high rates of taxation on alcohol, tobacco and petrol are optimal. In terms of the analysis in this section, these high rates of tax can only be justified if it could be shown that these goods are very strongly complementary to leisure. In fact, the estimation of consumer demand functions for these goods involves particular statistical problems and they are often excluded from the complete demand system estimates that are used to inform calculations of optimal tax rates. There is therefore no definitive answer as to whether the high rates of tax can be justified in terms of complementarity to leisure.

In fact, my guess is that such high rates of tax could not be justified in these terms alone. The calculated optimal taxes on goods that have been estimated to be particularly complementary to leisure have not been anything like as high. The justification, if there is one, for these high rates of tax must be found elsewhere: either in terms of the externalities that the consumption of these goods impose on other people (including possible costs to the National Health Service) or on the basis of a paternalistic concern for the consumer's health. As observed in Section II(4), these concerns have not yet been integrated into the literature on optimal taxation.

## *2. The Taxation of Intermediate Goods and International Trade*

In addition to introducing distributional considerations into the theory of optimal commodity taxation, Diamond and Mirrlees (1971) demonstrated an important result about the desirability of production efficiency. They showed that even though considerations of income distribution might justify the use of distortionary taxes on the supply of factors of production or the consumption of final goods, they did not justify any distortion to the way in which production is organised.

The basic idea behind this result is that each household's level of welfare depends on the prices it receives for the labour and other factors of production that it sells and on the prices it pays for the goods it consumes. It is these prices, therefore, that determine the distribution of utility. If the government is setting all taxes optimally, it is able to control all of these prices independently of the prices that firms face in trades between themselves. For example, if the producer price of cars was to rise, the government could prevent the consumer price from rising by reducing the sales tax on cars. Thus there is no improvement in social

welfare that would result from a manipulation of producer prices; all that is required of the production sector of the economy is that it should be as efficient as possible, and it is a standard result in economics that this efficiency is maximised if there is no taxation on trade between firms.

This production efficiency result has wide implications. It is straightforward to see that it implies that turnover taxes are inefficient and should be replaced by VAT or final sales taxes. It also implies that public sector enterprises should attempt to maximise profits at market prices, in just the same way as private companies. Finally, and most controversially, it implies that a country that has no monopoly or monopsony power in world trade (a “small country”) should not tax either imports or exports.<sup>21</sup> The idea here is that international trade can be seen as just another production activity, converting exports into imports, and so should not be distorted. If substantial revenues are raised from import duties, they should be replaced by domestic sales taxes on the same goods at the same rates. These will yield at least as much revenue and will not distort production. If the resulting imports result in the reduction in the wage of people with particular skills, then it is more efficient to make transfer payments to those workers than to distort production.

A result as powerful as this must clearly rely on quite far-reaching assumptions. It certainly assumes that all markets are functioning efficiently and that unemployment would not result from the removal of trade barriers. Obviously, in practice, the removal of trade barriers or the exposure of nationalised industries to full market pressures might well cause problems of adjustment. The result is telling us that, once these short-term problems are out of the way, the final level of social welfare will be higher. However, it cannot tell us whether the short-term loss is outweighed by the long-term gain. This would depend on the precise nature of the adjustment process and the rate at which future benefits were discounted in comparison to current costs.

Another assumption that is required to produce this result is that private sector firms do not earn pure after-tax profits, something that would be guaranteed if there was perfect competition and constant returns to scale. That is not to say that they do not earn any accounting profits. Rather, the requirement is that they do not earn any more profit than is required to attract the equity capital in the firm. This assumption is needed because otherwise changes in producer prices will affect pure profits, and hence household utility. The assumption is almost certainly violated in the real world, but it is hard to know how important this is because of the difficulty of measuring pure profits.<sup>22</sup>

---

<sup>21</sup> This does not of course alter the “optimal tariff” argument for countries to restrict trade to exploit monopoly or monopsony power.

<sup>22</sup> Imperfect competition is one way in which profits can arise, and the implications of this for the taxation of intermediate goods are discussed in Myles (1989b).

The final assumption is much more significant: the assumption that the government is able to alter taxes on individual goods to precisely manipulate their prices. This is very difficult for the government to do, and the example above of compensating workers who lose out from trade liberalisation illustrates this. It would be very difficult to identify exactly which workers have suffered and calculate the precise amount of compensation needed.

The violation of this assumption opens the possibility of justifying import tariffs in terms of the need to protect particular vulnerable groups of workers. Heady and Mitra (1987) use numerical analysis to investigate the size of the tariffs that could be justified. The optimal tariffs they find are low (usually less than 20 per cent) but could be significant in their impact. However, these arguments could not justify many of the tariffs that are actually observed. They are frequently unjustifiably high and/or provide benefit to the rich rather than the poor.

The difficulties of setting optimal taxes, or any taxes at all, are greatest in underdeveloped countries. It is often nearly impossible to directly collect taxes from subsistence agriculture or the informal urban sector. In such cases, trade taxes may be the only way of collecting sufficient revenue and, indeed, underdeveloped countries are usually much more reliant on the revenue from trade taxes than are the OECD countries. Tax issues for underdeveloped countries are discussed in Newbery and Stern (1987).

Despite the practical objections raised here to the blanket application of the production efficiency result, its practical importance must not be overlooked. It is often easy to provide some example of an assumption that is not quite satisfied but, at least for countries like the UK, it provides a very useful bench-mark that is widely applicable. It is best to presume the desirability of production efficiency, and place the burden of proof on the people who are arguing for an exception.

### *3. The Taxation of Savings*

So far in this paper, the issue of savings has been ignored, except to say that its existence produces the only difference in effect between a linear income tax and a uniform sales tax. However, the effect of taxation on savings is a major policy issue and there have been a number of tax changes in recent years that have been designed to encourage savings.

The issue is whether people should be taxed on their full income, including income from capital, or on their expenditure on goods and services. A tax on expenditures could be implemented either as a uniform sales tax, such as VAT, or by extending the current sheltering of certain types of savings under the income tax. There have been moves of both types in the UK over the last 15 years: the progressive reduction in income tax rate, accompanied by increases in the rate of VAT, and the introduction of various tax-sheltered forms of savings.

Income taxation taxes both the income that is saved and the subsequent return on that saving, something that is often referred to as the “double taxation of savings”. Expenditure taxation either allows the interest to be tax-exempt or postpones the taxation on the amount saved until it is spent, at which point both the savings and the return are taxed. Under income taxation, the saver’s return on the savings is less than the rate paid by the borrower. Under expenditure taxation, the saver’s return is equal to the rate paid by the borrower. Thus the income tax can be seen as discouraging saving, while the expenditure tax does not discourage it.

This makes the expenditure tax seem better. However, because it exempts savings, an expenditure tax does not raise as much revenue as an income tax with the same nominal rate. To be revenue-neutral, the expenditure tax must be levied at a higher rate. The question is whether the disadvantages of the higher rate outweigh the advantages of not discouraging savings.

The analysis of this question can become extremely difficult, as it involves people making savings decisions that will have effects for a long time into the future. However, Atkinson and Stiglitz (1980, Ch. 14) show how the theory of optimal commodity taxation can shed some light on the question. They divide a person’s life into two periods: work and retirement. The person earns wage income in the first period and divides the proceeds between consumption in the two periods. The total consumption in each period can be viewed as one composite good. Viewed in this way, expenditure taxation is equivalent to an equal tax rate on the two goods, while income taxation is equivalent to a higher tax rate on retirement consumption. The analysis in Section IV(1) suggests that which is better depends on whether consumption in retirement is more complementary with leisure in the work period than is consumption in the work period. If it is, income tax is better. If the two types of consumption are equally complementary with leisure (if there is weak separability), the expenditure tax is better.

This analysis has not settled the issue, partly because it is clearly a gross simplification of reality and partly because there is no conclusive evidence about the structure of people’s intertemporal preferences. However, this analysis does show that the theory of optimal commodity taxation can provide some insight into the issues involved. Unfortunately, it has not progressed far enough to yield practical policy conclusions.

Even if we cannot determine the rate at which savings should be taxed, can we presume that all forms of savings should be taxed at the same rate? This is an important practical issue because we observe a wide range of tax treatments for different forms of savings: bank and building society deposits, equities, pension funds.

This is not an issue that the optimal tax literature has addressed directly, but it could be argued that such variations in tax treatment result in different firms facing different costs of capital. If that is the case, the differential tax treatment

violates the production efficiency result discussed above, which requires all firms to face the same prices for all inputs and outputs.

This argument has a strong appeal, and it is hard to see any justification for the wide range of tax treatments for different types of savings. However, the financial markets are so complex that it is often difficult to trace through the effects of all these different tax treatments and demonstrate the way in which resources are being misallocated. This is clearly an area that deserves attention in future research.

## V. CONCLUSIONS

The organising theme of this survey has been the assessment of the contribution that the literature on optimal taxation can make to the formulation of practical tax policy. It is therefore appropriate to conclude by summarising that contribution.

The most important, and perhaps surprising, conclusion from the literature on optimal income taxation is that the optimal schedule can often be approximated by a linear income tax. There is certainly no reason to expect the optimal marginal rate to increase at higher income levels: any wish to increase tax progression is balanced by the fact that there are few people to pay those higher taxes. This conflicts with the practice of most countries, but perhaps the conflict arises from a lack of general awareness about how strongly redistributive linear taxation can be.

The other important conclusion about income taxation is that although a fairly wide range of optimal marginal tax rates correspond to a reasonable range of key parameter values, these optimal rates are influenced just as strongly by estimates of (objective) labour supply behaviour as by the (subjective) degree of aversion to inequality.

Turning to indirect taxes, the most significant conclusion is that there is little or no reason to have differential sales tax rates on distributional grounds, and little reason to have them on efficiency grounds either. However, these results leave out the considerations required to provide a proper analysis of the excise duties on alcohol, tobacco and petrol. The setting of these rates depends on matters outside the standard optimal taxation literature, such as externalities and paternalism.

The other major result on commodity taxation is the desirability of production efficiency, implying the superiority of VAT over turnover taxes, the need to run nationalised industries on market principles and the desirability of free international trade. This result depends on some fairly strong assumptions, but forms a useful bench-mark for policy decisions.

Finally, the literature on optimal tax has also been unable to resolve the important policy issue of whether we should have an income tax or an

expenditure tax (whether savings should be tax-sheltered) although it has illuminated the issues involved.

The overall message that comes through from this is that optimal taxation can help in some, but not all, areas of tax policy. What is perhaps surprising is that, despite its general disregard for administrative and compliance costs, it has produced results that do not conflict significantly with administrative considerations. Also, although value judgements necessarily come into decisions about rates of tax, there are a number of important tax design matters (for example, linearity of the income tax or uniformity of the sales tax) that are quite independent of such judgements.

## REFERENCES

- Ahmad, E. and Stern, N. (1991), *The Theory and Practice of Tax Reform in Developing Countries*, Cambridge: Cambridge University Press.
- Allen, F. (1982), 'Optimal linear income taxation with general equilibrium effects on wages', *Journal of Public Economics*, vol. 17, pp. 135–44.
- Atkinson, A. B. (1973), 'How progressive should income tax be?', in M. Parkin and A. R. Nobay (eds), *Essays in Modern Economics*, London: Longman.
- and Stiglitz, J. E. (1980), *Lectures on Public Economics*, London: McGraw-Hill.
- Blundell, R. W. (1992), 'Labour supply and taxation: a survey', *Fiscal Studies*, vol. 13, no. 3, pp. 15–40.
- and Walker, I. (1983), 'Limited dependent variables in demand analysis: an application to modelling family labour supply and commodity demand behaviour', University of Manchester, Discussion Paper in Econometrics no. 126.
- Carruth, A., Heady, C. J. and Ulph, D. (1983), 'Optimal income taxation and comparative advantage', mimeo.
- Corlett, W. J. and Hague, D. C. (1953), 'Complementarity and the excess burden of taxation', *Review of Economic Studies*, vol. 21, pp. 21–30.
- Deaton, A. and Stern, N. (1986), 'Optimally uniform commodity taxes, taste differences and lump-sum grants', *Economics Letters*, vol. 20, pp. 263–6.
- Diamond, P. A. and Mirrlees, J. A. (1971), 'Optimal taxation and public production: I and II', *American Economic Review*, vol. 61, pp. 8–27 and 261–78.
- Ebrahimi, A. and Heady, C. J. (1988), 'Tax design and household composition', *Economic Journal*, Conference Papers, vol. 98, no. 390, pp. 83–96.
- Heady, C. J. (1987), 'A diagrammatic approach to optimal commodity taxation', *Public Finance*, vol. 42, pp. 250–63.
- and Mitra, P. K. (1987), 'Distributional and revenue raising arguments for tariffs', *Journal of Development Economics*, vol. 26, pp. 77–101.
- and — (1992), 'Taxation in decentralizing socialist economies: the case of China', World Bank, Policy Research Working Paper no. WPS820.
- , Pearson, M., Rajah, N. and Smith, S. (1992), *Report on the Czechoslovak Government's Tax Proposals*, London: Institute for Fiscal Studies.
- Mirrlees, J. A. (1971), 'An exploration in the theory of optimum income taxation', *Review of Economic Studies*, vol. 38, pp. 175–208.

- Musgrave, R. A. (1959), *The Theory of Public Finance*, New York: McGraw-Hill.
- Myles, G. D. (1989a), 'Ramsey tax rules for economies with imperfect competition', *Journal of Public Economics*, vol. 38, pp. 95–115.
- (1989b), 'Imperfect competition and the taxation of intermediate goods', *Public Finance*, vol. 44, pp. 62–74.
- Newbery, D. and Stern, N. (1987), *The Theory of Taxation for Developing Countries*, Oxford: Oxford University Press.
- Ramsey, F. P. (1927), 'A contribution to the theory of taxation', *Economic Journal*, vol. 37, pp. 47–61.
- Rawls, J. (1971), *A Theory of Justice*, Cambridge, Mass.: Harvard University Press.
- Smith, A. (1776), *An Enquiry into the Nature and Causes of the Wealth of Nations*, Cannan edition, London: Methuen, 1904.
- Stern, N. H. (1976), "On the specification of models of optimum income taxation", *Journal of Public Economics*, vol. 6, pp. 123–62.
- Tuomala, M. (1990), *Optimal Income Tax and Redistribution*, Oxford: Clarendon Press.