

PARETO EFFICIENT AND OPTIMAL TAXATION AND THE NEW NEW WELFARE ECONOMICS

JOSEPH E. STIGLITZ*

Princeton University

1. Introduction¹

For more than a hundred years, economists have attempted to show that progressive taxation can be justified on more fundamental principles. Among the earliest of such attempts was that of Edgeworth (1868, 1897), who tried to show that utilitarianism (combined with two other assumptions) implied progressivity. His argument was simple: he postulated that all individuals had the same utility of income function, and that they exhibited diminishing marginal utility. Since social welfare was simply the sum of the utility functions of all individuals, it immediately followed that the decrease in social welfare from taking a dollar (or a pound) away from a rich person was less than the decrease in social welfare from taking a dollar away from a poor person.

With the advent of the New Welfare Economics in the 1930s, interest in optimal taxation waned: interpersonal utility comparisons of the kind employed in utilitarianism were viewed to be inadmissible. Whether taxes should be progressive might be an appropriate question for a moral philosopher, but not for

*All workers in this area owe an intellectual debt to James Mirrlees. I should like to acknowledge this debt as well as my indebtedness to my several collaborators who have greatly influenced the formulation of my ideas: Partha Dasgupta, Richard Arnott, Raaj Sah, Bob Britto, Steve Slutsky, John Hamilton, and especially A.B. Atkinson. I am also indebted to Alan Auerbach, Yungoll Yun and B. Salanie who provided comments on an earlier draft.

Financial support from the National Science Foundation is gratefully acknowledged.

¹Although this chapter is primarily a survey of existing literature, several sections are based on until now unpublished work, and several sections contain new proofs of previously published results. New results are reported in Section 8, on optimal taxation where those who are more productive in "outside" employment are also more productive in activities at home; Section 13.1, on the taxation of capital income in imperfect capital markets; Section 14, on altruism and inheritance taxation; and Section 15, on commitment. New proofs are contained in Section 5, on random taxation; in Section 3, on partial pooling; in Section 9, on Pareto efficient taxation with general equilibrium effects; and in Section 2, on characterizing the Pareto efficient tax structure.

an economist. Economists, in their role as economists, should limit themselves to identifying Pareto efficient allocations, and to finding Pareto inefficiencies and showing how these could be eliminated.

Many policy economists found this a too circumscribed view of their role. In almost all of the choices which they faced, some individuals were better off, others worse off. If they could not make some interpersonal comparisons, they would not be in a position to contribute much to policy debates.

This provided the setting for the re-examination of the design of tax structures in the 1970s, at first making use of utilitarian social welfare functions, but then broadening the analysis to investigate the consequences of a wider class of social welfare functions. The major advance over Edgeworth's earlier work [due to Mirrlees (1971)] was the recognition of the incentive effects associated with taxation;² there was a trade-off between equity and efficiency considerations, a trade-off which Edgeworth had ignored. Mirrlees' calculations (though only an example) provided less support, however, for the advocates of progressivity than they had hoped. With what seemed reasonable hypotheses concerning individual utility functions, the optimal tax schedule looked close to linear.

Those brought up within the doctrines of the New Welfare Economics felt distinctly uneasy about these developments; the analysis seemed so dependent on the implicit or explicit interpersonal utility comparisons.

This chapter is concerned with what the *New New Welfare Economics* has to say about the design of tax structures. The New New Welfare Economics is distinguished by two features: First, unlike much of the "old" New Welfare Economics and unlike the earlier Utilitarian analysis, it does not assume that the government has at its disposal the information required to make lump-sum redistributions; it identifies who is able to pay higher taxes at least partly by using endogenous variables, like income or consumption.³ It is this limitation on the information of the government which results in taxation being distortionary, and which gives rise to the trade-off between equity and efficiency.

Granted that the government has imperfect information, on the basis of which it can redistribute income, what is the best that can be done? The New New Welfare economics attempts to answer this question and to describe the *Pareto efficient tax structures*, i.e. the tax structures which get the economy to the utilities possibilities schedule, given the limitations on the government's information and other limitations on the government's ability to impose taxes.

Just as the earlier New Welfare Economics argued that it is essential first to identify the set of Pareto efficient allocations, so that one could separate out efficiency considerations from the value judgements associated with choices

² The presence of trade-offs between equity and efficiency considerations had, of course, long been recognized, and there had even been some formal modelling [see Fair (1971)].

³ It may, of course, also use exogenous variables, like age.

among Pareto efficient points, so too does the New New Welfare Economics. There are some properties that all Pareto efficient tax structures have, whereas other properties may be specific to particular Pareto efficient tax structures, e.g. those which would be chosen with a utilitarian social welfare function. Economists may make an important contribution by identifying Pareto efficient tax structures, but it still remains the case that many of the critical choices necessitate interpersonal trade-offs, choices among alternative Pareto efficient allocations. The use of Social Welfare Functions can provide a useful way of thinking systematically about these trade-offs; one of the objectives of this chapter is to show how this has been done, and some of the dangers and pitfalls that arise in doing so.

The topics which are the subject of this chapter include some of the areas in which there has been the most active research during the past fifteen years. We cannot and do not provide a complete survey; what we attempt to do is to put this literature in perspective, to provide some of the key parts of the analysis, and to refer the reader to more detailed surveys or articles dealing with particular topics.

The chapter is divided into three parts. Part I discusses Pareto efficient taxation in a one-commodity world. The central question is: How progressive should the tax structure be? Part II is concerned with the circumstances under which it is desirable to impose commodity taxation at different rates on different commodities and to tax the return to capital. We conclude in Part III with some general observations concerning extensions and interpretations of the general theory.

Part I: One-commodity neoclassical models

2. The basic model

The basic issues can be illustrated by a simple model of an economy in which there are two types of individuals and a single produced commodity; we shall assume that type 2 individuals are more productive than type 1 individuals, i.e. their output per unit of time spent working is larger. Assume that all individuals have the same utility function,

$$U(C_i, L_i),$$

where C_i is the consumption of an individual of type i (this may be a whole vector of consumption goods, possibly dated), and L_i is his labor supply.⁴ An

⁴Throughout, we assume that utility functions are "well behaved".

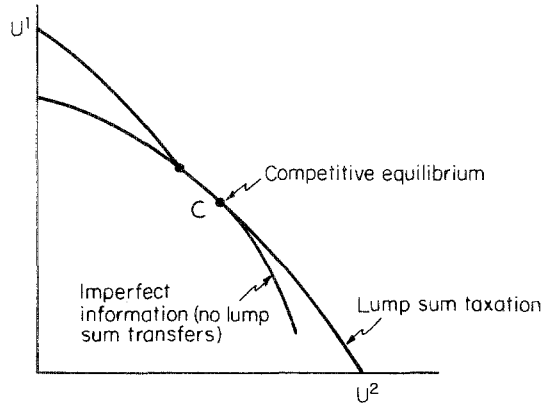


Figure 2.1. Utility possibilities schedule with and without lump-sum transfers.

individual of type i has a productivity (output per hour) of w_i . Thus, in the absence of taxation, the individual's budget constraint would simply be

$$C_i = Y_i = w_i L_i, \quad (1)$$

where Y_i is the individual's income. By imposing lump-sum taxes (or subsidies), M_i , on individuals of type i (so the individual's budget constraint is now

$$C_i = Y_i - M_i = w_i L_i - M_i), \quad (1')$$

so that the revenues raised by taxes on one group are redistributed as lump-sum payments to the other group:

$$M_1 N_1 + M_2 N_2 = 0, \quad (2)$$

where N_i is the number of individuals of type i , we can trace out the whole utilities possibilities schedule (see Figure 2.1).

2.1. Optimal lump-sum taxes

The *optimal* redistributive tax depends on the social welfare function. Two social welfare functions which have been extensively employed are the utilitarian,

$$W = \sum U^i, \quad (3a)$$

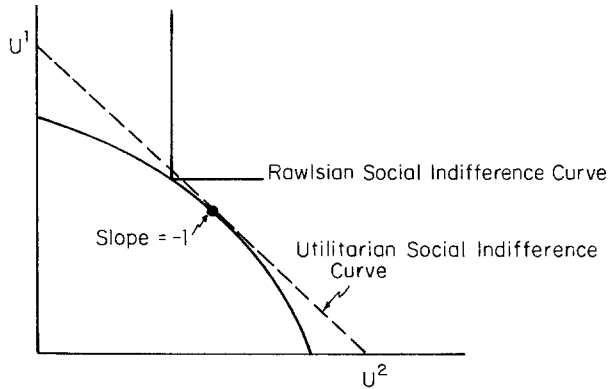


Figure 2.2. With different social welfare functions, different points on the utilities possibilities schedule will be chosen.

and the Rawlsian,

$$W = \min\{U^i\}, \tag{3b}$$

where U^i is the utility of the i th type of individual. The former is maximized at the point on the utilities possibilities schedule where the slope is -1 ; the latter by the intersection of the utilities possibilities schedule with a 45° line (Figure 2.2).

Analytically, the utilitarian solution with lump-sum taxation is characterized by the equality of the marginal utility of consumption:⁵

$$\partial U^1 / \partial C_1 = \partial U^2 / \partial C_2; \tag{4}$$

while the marginal rate of substitution between goods and leisure is equal to the individual's marginal product (wage):

$$w_i \partial U^i / \partial C_i + \partial U^i / \partial L_i = 0. \tag{5}$$

This, in turn, means that *so long as leisure is a normal good, the more able individual is actually worse off*. (This assumes, of course, that we can make interpersonal comparisons. The utilitarian solution has the property that the high

⁵Assume output $Q = F(N_1 L_1, N_2 L_2)$, a function of the inputs of the two types of labor. We form the Lagrangian:

$$\mathcal{L} = N_1 U^1 + N_2 U^2 + \mu [F(N_1 L_1, N_2 L_2) - (C_1 N_1 + C_2 N_2)].$$

Equations (4) and (5) follow from the first-order conditions.

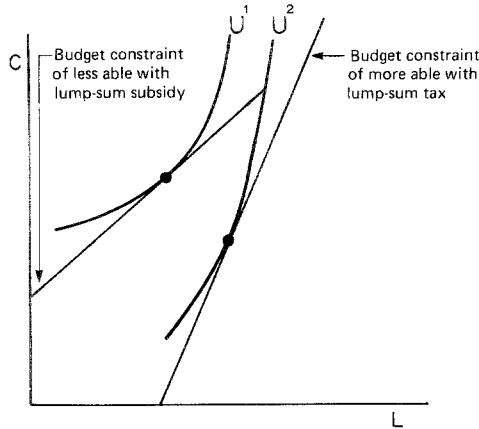


Figure 2.3. Utilitarian optimum with lump-sum taxes: the more able work more and are worse off than the less able.

ability individuals would prefer the $\{C, L\}$ allocation of the low ability individual.) This can be seen most easily in the case of a separable utility function

$$U_i = u(C_i) - v(L_i). \quad (6)$$

For then, (4) ensures that all individuals, regardless of their ability, receive the same level of consumption; but then the efficiency condition (5) implies that those with a higher productivity have a higher marginal disutility of work, i.e. that they provide a greater labor supply (see Figure 2.3). (This seems in accord with the Marxian dictum, “from each according to his abilities, to each according to his needs”). Notice that the more able individuals are actually better off in the Rawlsian solution than they are under utilitarianism.

2.2. Imperfect information

The previous analysis assumed that the government had perfect information about who was of type 2, and who was of type 1. Clearly, if the government could not tell, the more able individuals would (under the utilitarian solution) have no incentive to come forward to claim their greater ability; they would know that, if they were to do so, they would actually be worse off. The government, in its choice of tax structures, must recognize these limitations on its information. In our example with only two groups, the government has a choice of two kinds of tax structures: (a) those in which the government is unable to distinguish (ex post

or *ex ante*) among individuals; this is called a *pooling equilibrium*;⁶ and (b) those in which the more able and the less able can (*ex post*) be identified as a result of the actions undertaken by the different groups; this is called a *self-selection equilibrium*.⁷ To attain a self-selection equilibrium, the two groups have to be given choices, such that the first group prefers one choice, the second group the other choice. (These are called the self-selection constraints.)⁸

We assume, however, that the individual's labor input is not observable, just as his productivity is not observable; his before-tax income,

$$Y_i = w_i L_i,$$

is, however, observable. (Obviously, if both before-tax income and labor input were observable, then one could immediately infer the individual's productivity, w_i ; more generally, even if hours on the job is observable, effort is not.)⁹ Thus, it may be possible to differentiate among individuals on the basis of their choices of the observable variables $\{C, Y\}$. We seek to characterize the Pareto efficient pairs of $\{C_i, Y_i\}$. To do this, we rewrite the i th individual's utility function in terms of the observable variables $\{C_i, Y_i\}$:

$$V^i(C_i, Y_i) \equiv U(C_i, L_i w_i / w_i) = U(C_i, Y_i / w_i). \quad (7)$$

It is useful to note three properties of this derived utility function. First, increases in Y (income) decrease utility (keeping C , consumption, fixed), because to attain the increases in income, more labor must be expended. Second, the more productive individual has a flatter indifference curve. This follows from observing that the marginal rate of substitution,

$$(\partial C^i / \partial Y^i)_{\bar{V}} = -[\partial V^i / \partial Y^i] / [\partial V^i / \partial C^i] = -[(\partial U^i / \partial L^i) / w^i] / [\partial U^i / \partial C^i].$$

⁶This vocabulary was first introduced, in the context of the analysis of how competitive markets distinguish among individuals with different characteristics, by Rothschild and Stiglitz (1976).

⁷It is also referred to as a *separating equilibrium*, because the observed actions of the individuals separates them according to types; or as a *revealing equilibrium*, because their actions reveal who they are.

⁸Since, in self-selection equilibria, individuals' types can be identified, one could, alternatively, have asked the individual what type he was, with the understanding that the consequences for the individual are precisely those associated with the self-selection equilibrium. Similarly, in the pooling equilibrium, if the individual were told that nothing would depend on what statement he made concerning his abilities, he would have no incentive to lie. Thus, the set of possible equilibria correspond to those which could be attained if individuals are simply asked what their ability is, and are provided with incentives not to lie. This (loosely speaking) is referred to as the *revelation principle*. See R. Myerson (1983).

The self-selection constraints are sometimes referred to as the *incentive compatibility constraints*.

⁹In jobs where individuals punch time clocks, presumably hours worked is observable. But note that were the government to base its tax on the individual's wage rate, firms would have an incentive to collude with the workers in "cheating" on the official hours worked. Such practices are already common in circumstances where bureaucratic regulations impose a maximum on compensation per hour.

If leisure is a normal good, then at any given set of values of $\{C, Y\}$, labor will be lower for the more productive individual, and hence the marginal rate of substitution between consumption and labor will be lower, i.e.

$$(\partial C^i / \partial L^i)_{\bar{v}} = - [\partial U^i / \partial L^i] / [\partial U^i / \partial C^i]$$

is smaller. Higher ability individuals have flatter indifference curves (in $\{C, Y\}$). (See Figure 2.4.)

Finally, we note that when individuals maximize their utility in the absence of taxation, they set their marginal rate of substitution between consumption and labor equal to the wage, i.e.

$$(\partial C^i / \partial L^i)_{\bar{v}} = - [\partial U^i / \partial L^i] / [\partial U^i / \partial C^i] = w^i,$$

from which it immediately follows that in the absence of taxation

$$\begin{aligned} (\partial C^i / \partial Y^i)_{\bar{v}} &= - [\partial V^i / \partial Y^i] / [\partial V^i / \partial C^i] \\ &= - [(\partial U^i / \partial L^i) / w^i] / [\partial U^i / \partial C^i] = 1; \end{aligned}$$

the slope of the indifference curve in $\{C, Y\}$ space is unity.

The first basic result on *Pareto efficient tax structures* is this: with two groups, Pareto efficiency requires a self-selection equilibrium; the government will, in fact, wish to differentiate among individuals on the basis of their observed incomes [Stiglitz (1982b)].¹⁰ This result is not general: with many groups, though a complete pooling equilibrium is not consistent with Pareto efficiency, a partial pooling equilibrium (one in which several different types take the same action) may be.¹¹

The set of Pareto efficient self-selection equilibria is easy to characterize: though the government in general sets a whole tax function (which in turn determines the relationship between after-tax and before-tax income), when there are only two groups, only two points on this schedule will be chosen. We simply

¹⁰ This result is easy to see. Assume there were a pooling equilibrium, E , as depicted in Fig. 2.4. Any point lying between the two indifference curves will, together with E , separate. For instance, if the government were to offer H and E , the more able would choose H and the less able E . Assume that the slope of the more able indifference curve through E is flatter than 1. Then there exists points (such as H) slightly above the more able individual's indifference curve, which will be chosen only by the more able, and which increase government revenue (since consumption increases less than income.) The pair $\{H, E\}$ Pareto dominates E . (If the slope of the more able indifference curve through E is steeper than 1, then there exists a point below E , in the shaded area, which will separate and which will generate more government revenue.)

¹¹ The essential assumption for this result is that the indifference curve between consumption and pre-tax income is unambiguously steeper for the less able.

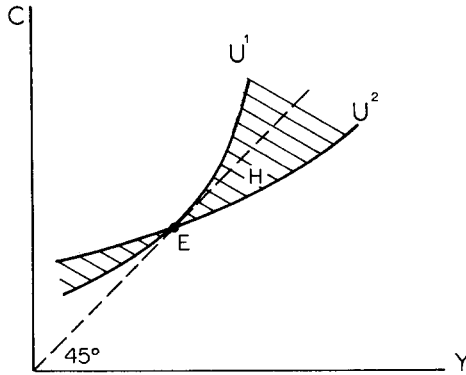


Figure 2.4. More able individuals have flatter indifference curves.

need to characterize those two points, or equivalently, to characterize the {consumption, income} decisions of the two groups (Figure 2.5).

Formally, then, the set of Pareto efficient allocations is described by the solution to

$$\max_{\{C_1, C_2, Y_1, Y_2\}} V^2(C_2, Y_2) \tag{8}$$

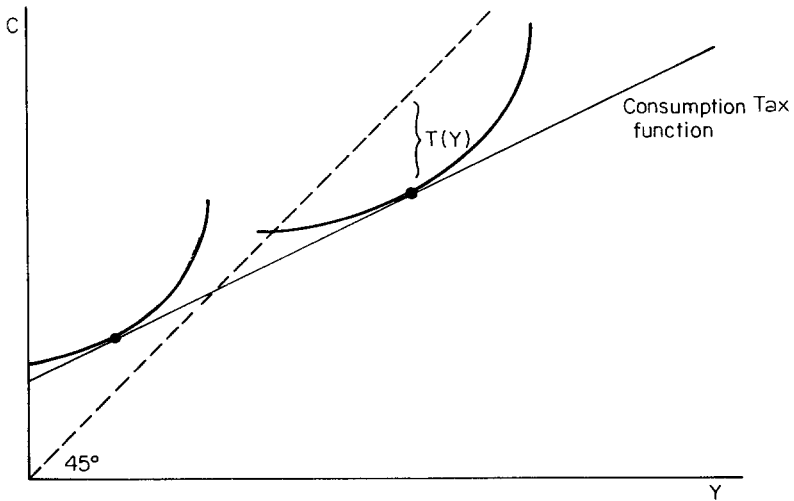


Figure 2.5. Although the government may set a whole tax function, which in turn determines the relationship between after tax and before tax income, when there are only two groups, only (at most) two points will be chosen.

s.t.

$$V^1(C_1, Y_1) \geq \bar{U}^1, \quad (9)$$

$$\left. \begin{aligned} V^2(C_2, Y_2) &\geq V^2(C_1, Y_1) \\ V^1(C_1, Y_1) &\geq V^1(C_2, Y_2) \end{aligned} \right\}, \text{ the self-selection constraints,} \quad (10)$$

$$(11)$$

$$R = (Y_1 - C_1)N_1 + (Y_2 - C_2)N_2 \geq \bar{R}, \text{ the revenue constraint} \quad (12)$$

(where R is government revenue, \bar{R} is the revenue requirement).¹²

The Lagrangian for this maximization problem may be written as

$$\begin{aligned} \mathcal{L} = & V^2(C_2, Y_2) + \mu V^1(C_1, Y_1) + \lambda_2 (V^2(C_2, Y_2) - V^2(C_1, Y_1)) \\ & + \lambda_1 (V^1(C_1, Y_1) - V^1(C_2, Y_2)) + \gamma [(Y_1 - C_1)N_1 + (Y_2 - C_2)N_2 - \bar{R}]. \end{aligned} \quad (13)$$

The first-order conditions for this problem are straightforward:

$$\frac{\partial \mathcal{L}}{\partial C_1} = \mu \frac{\partial V^1}{\partial C_1} - \lambda_2 \frac{\partial V^2}{\partial C_1} + \lambda_1 \frac{\partial V^1}{\partial C_1} - \gamma N_1 = 0, \quad (14a)$$

$$\frac{\partial \mathcal{L}}{\partial Y_1} = \mu \frac{\partial V^1}{\partial Y_1} - \lambda_2 \frac{\partial V^2}{\partial Y_1} + \lambda_1 \frac{\partial V^1}{\partial Y_1} + \gamma N_1 = 0, \quad (14b)$$

$$\frac{\partial \mathcal{L}}{\partial C_2} = \frac{\partial V^2}{\partial C_2} + \lambda_2 \frac{\partial V^2}{\partial C_2} - \lambda_1 \frac{\partial V^1}{\partial C_2} - \gamma N_2 = 0, \quad (14c)$$

$$\frac{\partial \mathcal{L}}{\partial Y_2} = \frac{\partial V^2}{\partial Y_2} + \lambda_2 \frac{\partial V^2}{\partial Y_2} - \lambda_1 \frac{\partial V^1}{\partial Y_2} + \gamma N_2 = 0. \quad (14d)$$

¹²Notice that this problem is just the dual to the standard problem of a monopolist attempting to differentiate among his customers [Stiglitz (1977) and (forthcoming)]. There, the problem was to maximize profits (corresponding to R here), subject to utility constraints on each of the two types of individuals (that he be willing to purchase the good, referred to as the reservation utility or individual rationality constraint) and subject to the self-selection constraints. The Lagrangian which we form to analyze the two problems is identical.

Here the constraint on the individual's utility arises because we wish to ensure him a certain level of utility; in the standard monopoly problem, the constraint arises because the transaction is voluntary, and if the monopolist does not offer him the level of utility that he could have obtained without trading with the monopolist, he will choose not to deal with him.

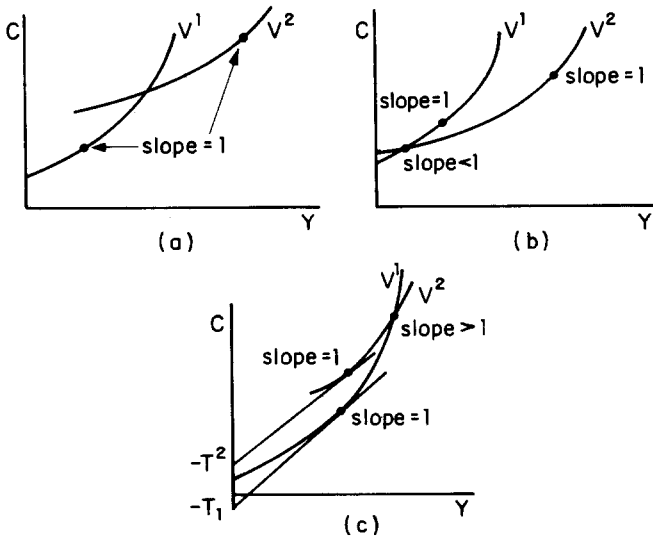


Figure 2.6. (a) First-best taxation fully revealing. (b) First-best taxation not fully revealing: Pareto optimal taxation entails a positive marginal tax rate on low ability and a zero marginal tax rate on high ability. (c) First-best taxation not fully revealing: Pareto optimal taxation involves a zero marginal tax rate on low ability and a negative marginal tax rate on high ability.

It is easy to see that, under our assumptions concerning the relative slopes of the indifference curves, there are three possible regimes:

- (i) $\lambda_1 = 0, \lambda_2 = 0$ [Figure 2.6(a)];
- (ii) $\lambda_1 = 0, \lambda_2 > 0$ [Figure 2.6(b)]; and
- (iii) $\lambda_2 = 0, \lambda_1 > 0$ [Figure 2.6(c)].

That is, at most one of the two self-selection constraints is binding. Moreover, it is also easy to show that $\mu > 0$, i.e. the constraint on the utility level of the low ability individuals is binding.

The case where $\lambda_1 = \lambda_2 = 0$ is illustrated in Figure 2.6(a). With first-best taxation, the equilibrium is fully revealing. This is always the case at the competitive equilibrium. Hence, if the social welfare function chooses the competitive allocation (point C in Figure 2.1) or a point near it, neither self-selection constraint will be binding.¹³

¹³The same result obviously holds if a uniform lump-sum tax is imposed. (See Figure 2.7.)

The “normal” case, on which most of the literature has focused, is that where $\lambda_1 = 0$ and $\lambda_2 > 0$. This arises whenever the government wishes to improve the welfare of the poor significantly relative to what they would have obtained at the competitive allocation (with uniform lump-sum tax to raise the requisite revenue). With a utilitarian objective function ($\mu = 1$) (or indeed any concave social welfare function) and separable utility functions it can be shown that this is the only possibility.¹⁴ But, more generally, the possibility that $\lambda_1 > 0$ and $\lambda_2 = 0$ cannot be ruled out. The case with $\{\lambda_1 > 0 \text{ and } \lambda_2 = 0\}$ has the property that if lump-sum taxation were feasible, the lump-sum tax imposed on the low ability individual would exceed that on the high ability [Figure 2.6(c)].

We can again characterize the utilities possibilities schedule, this time under the constraint that lump-sum taxes are not feasible. The new utilities possibilities schedule coincides with the old over a range, but elsewhere is interior. The range over which the two coincide corresponds to the situations where $\lambda_1 = \lambda_2 = 0$. (See Figure 2.1.)

2.3. Implementing the optimal allocations by means of a tax structure

In the preceding subsection we described the Pareto efficient resource allocation. We posed the problem as if the government confronted the individual with a choice of two “bundles”. These allocations can easily be implemented by means of a tax schedule, as illustrated in Figure 2.8. The tax schedule must pass through the points $\{C_i, Y_i\}$, $i = 1, 2$; and elsewhere must lie below the indifference curves through C_i, Y_i . Given such a tax schedule, individual i will clearly choose the point $\{C_i, Y_i\}$.

If the tax schedule were differentiable,

$$T = T(Y),$$

(so that $C = Y - T(Y)$) the individual’s first-order condition would be

$$-\frac{\partial U^i / \partial L_1}{\partial U^i / \partial C_i} = (1 - T') w_i, \quad (15)$$

or

$$-\frac{\partial V^i / \partial Y_i}{\partial V^i / \partial C_i} = (1 - T').$$

¹⁴The issue of which of the self-selection constraints is binding has played a major role in the debate over the relevance of the theory of implicit contracts with asymmetric information; with separable utility functions and risk-neutral firms, one obtains over-employment (all firms wish to pretend to be in the good state). (See Azariadis and Stiglitz (1983); or Stiglitz (1986b). The analysis is exactly parallel to that presented here.)

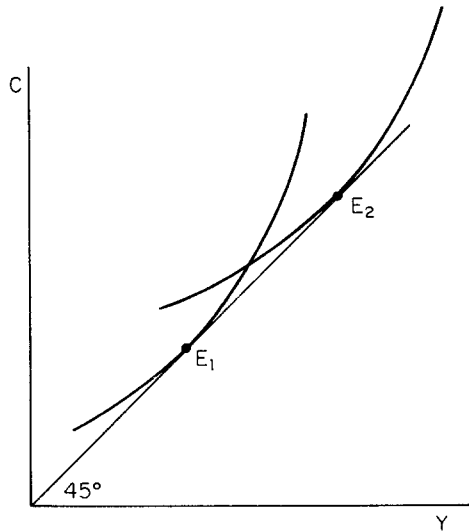


Figure 2.7. At the competitive allocation, the self-selection constraints are never binding (with no taxation, $C = Y$).

The left-hand side is the individual's marginal rate of substitution; the right-hand side is the after-tax marginal return to working an extra hour.

It is clear from Figure 2.8 that the implementing tax schedules will not, in general, be differentiable. Still we shall refer to

$$1 + \frac{\partial V^i / \partial Y_i}{\partial V^i / \partial C_i} = 1 + \frac{1}{w_i} \frac{\partial U^i / \partial L_i}{\partial U^i / \partial C_i} \tag{16}$$

as the marginal tax rate.¹⁵

2.4. The optimal tax structure with $\lambda_2 > 0, \lambda_1 = 0$

Dividing (14d) by (14c) we immediately see that

$$\frac{-\partial V^2 / \partial Y_2}{\partial V^2 / \partial C_2} = -\frac{\partial U^2 / \partial L_2}{\partial U^2 / \partial C_2} \frac{1}{w_2} = 1 \tag{17a}$$

¹⁵There exist optimal tax structures for which

$$\frac{1}{w_i} \frac{\partial U^i / \partial L_i}{\partial U^i / \partial C_i} + 1$$

is the left-handed derivative of the tax function at $Y = w_i L_i$.

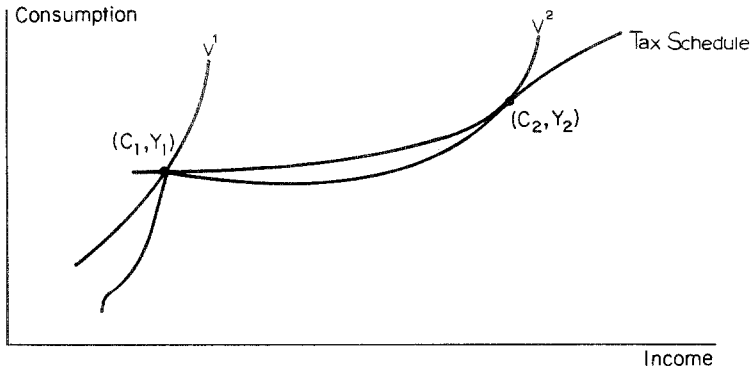


Figure 2.8. Implementing the allocation $\{(C_1, Y_1), (C_2, Y_2)\}$ by means of a tax system.

i.e. *the marginal tax rate faced by the more able individual is zero* (although the average tax rate is positive). [This corresponds to the result noted earlier by Sadka (1976) and Phelps (1973).] The argument for this can be easily seen diagrammatically. Assume that the government offered a pair of separating contracts, $\{E_1, E_2\}$, as in Figure 2.9, but in which the slope of the more able's indifference curve (in terms of $\{C, Y\}$) at $\{E_2\}$ differed from unity. Then, keeping $\{E_1\}$ fixed, and moving along the more able's indifference curve, one obtains alternative pairs of separating contracts; if the slope at $\{E_2\}$ was less than one, by increasing Y_2 , one increases 2's consumption by less than income, and hence government revenue is increased; the original allocation could not have been Pareto efficient; similarly if the slope at $\{E_2\}$ is greater than unity, one can increase government revenue by reducing Y_2 .

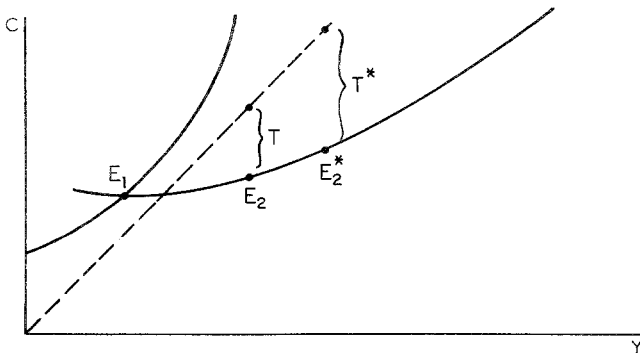


Figure 2.9. Pareto efficiency requires a zero marginal tax rate on the more able: moving from $\{E_2\}$ to $\{E_2^*\}$ increases revenue, but leaves the utility of all individuals unchanged.

Dividing (14b) by (14a):

$$-\frac{\partial V^1/\partial Y_1}{\partial V^1/\partial C_1} = \frac{-1}{w_1} \frac{\partial U^1/\partial L_1}{\partial U^1/\partial C^1} = \frac{-\lambda_2 \partial V^2/\partial Y_1 + \gamma N_1}{\lambda_2 \partial V^2/\partial C_1 + \gamma N_1} < 1. \quad (17b)$$

To see this, define

$$\alpha^i = -\frac{\partial V^i/\partial Y_1}{\partial V^i/\partial C_1}$$

and

$$v = \frac{\lambda_2 \partial V^2/\partial C_1}{N_1 \gamma}$$

Then (17b) can be rewritten as

$$\alpha^1 = \frac{1 + v\alpha^2}{1 + v} = \alpha^2 + \frac{1 - \alpha^2}{1 + v}.$$

Since, by assumption, $\alpha^1 > \alpha^2$, it therefore follows that

$$\alpha^2 < \alpha^1 < 1.$$

We immediately see that *the marginal tax rate faced by the less able individual will be positive*.¹⁶

2.5. The optimal tax structure with $\lambda_1 > 0$, $\lambda_2 = 0$

Exactly the same kinds of arguments as used in the previous section can be employed to establish that if $\lambda_1 > 0$ and $\lambda_2 = 0$, the *marginal* tax rate faced by the more able individual is negative; self-selection requires that they work more than they would in a non-distortionary situation. For the rest of this chapter, we focus our attention on the “normal” case with $\lambda_1 = 0$ and $\lambda_2 > 0$.

Our analysis has thus revealed a second property of Pareto efficient tax structures: *at least one of the groups must face no distortionary taxation*. The result holds regardless of the number of groups in the population.

¹⁶This corresponds to the result noted earlier by Mirrlees for the case of a continuum of types.

2.6. Equity–efficiency trade-offs

This analysis makes clear the nature of the equity–efficiency trade-offs that are inherent in redistributive tax policies. There is a “distortion” associated with redistributing any significant amount of resources from the more able to the less able. The distortion arises from the self-selection constraint. The magnitude of the distortion depends in part on the amount of revenue that the government is attempting to raise for public goods. Note that one feasible policy, a uniform lump-sum tax on all individuals, is among the set of Pareto efficient tax structures. It is, however, one which does not seem acceptable to most modern governments. [It is sometimes assumed that lump-sum taxes are not feasible. Uniform lump-sum taxes are clearly feasible (provided they are not too large). What is *not* feasible are lump-sum taxes at different rates for individuals of different abilities; this is not feasible because the government does not have the requisite information to implement such a tax.]

3. Continuum of individuals

Though most of the central insights in the theory of Pareto efficient (and optimal) taxation can be gleaned from the simple model presented above, the earliest analyses of optimal income tax structures employed a model in which there was a continuum of individuals [Mirrlees (1971)]. A rigorous analysis of that problem entails a large number of mathematical niceties, which are discussed by Mirrlees (1976, 1985). Here, we ignore these and follow the approach of Atkinson and Stiglitz (1976) using Pontryagin’s technique.

It is assumed that the distribution of individuals by ability is given by $F(w)$. The problem of the government is to maximize

$$\int U(C(w), Y(w)/w) dF(w) \quad (18)$$

subject to the national income constraint,

$$\int [C(w) - Y(w)] dF(w) = 0, \quad (19)$$

and subject to the self-selection constraints, which here take on the simple form that utility must increase with w , which we write (making use of the first-order condition for utility maximization)

$$dU/dw = -U_L Y/w^2. \quad (20)$$

Here, we follow Atkinson and Stiglitz (1980) in analyzing a slightly more general form of the problem, employing a social welfare function

$$\int G(U) dF(w). \quad (21)$$

Utilitarianism involves the special case of $G \equiv 1$. The Rawlsian social welfare function is $G'(U) = 0$, for $U > U_{\min}$; $G(U) = U$, for $U = U_{\min}$. Egalitarian social welfare functions entail $G'' \leq 0$, the social marginal utility of income is non-increasing in utility. Form the Hamiltonian,

$$H = [G(U) - \gamma(C - Y)](dF/dw) - \lambda dU/dw, \quad (22)$$

and let U be our “state variable” (like capital in an intertemporal problem), and $L(w)$ be the control variable (like investment in the typical optimal accumulation problem) – knowing L and U determines $Y(w) = wL(w)$ and $C(w)$ [from the equation $U(w) = U(C(w), L(w))$]. Now take the derivative of H with respect to L :

$$\partial H / \partial L = \gamma \{ w - (dC/dL)_{|U} \} f + \lambda \varepsilon^* U_L / w = 0, \quad (23)$$

where

$$\varepsilon^* = \partial \ln(dU/dw) / \partial \ln L = \partial \ln - (LU_L/w) / \partial \ln L, \quad (24)$$

where $f = dF/dw$, the density function, and where $(dC/dL)_{\bar{U}} = -U_L/U_C$. In the case of a separable utility function, we thus obtain:

$$\varepsilon^* = 1 + LU_{LL}/U_L. \quad (24a)$$

Again making use of the individual’s first-order condition, we can rewrite (23) as

$$T' / (1 - T') = U_C \lambda(w) \varepsilon^* / \gamma f w, \quad (25)$$

where T' is the marginal tax rate.

Thus (25) implies that the marginal tax rate depends on four factors:

(1) The marginal tax rate is lower, the larger the fraction of the population f at the particular income level that pays that marginal tax rate; one does not want to impose large distortions where there are many people.

(2) The marginal tax rate is lower, the smaller is $\lambda(w)$, the shadow price on the self-selection constraint. Analyzing $\lambda(w)$ is a rather complicated matter. For the

Rawlsian social welfare function, it is monotonically decreasing. For the utilitarian social welfare function, it increases and then decreases. The one general result that holds (for ability distributions with a finite range) is¹⁷ that at the extreme, the marginal tax rate on the highest income (ability) individual should be zero. This result corresponds to the result noted earlier in the two-group model. It is often cited as one of the few general, qualitative propositions to arise from the theory of optimal taxation, but in fact, as we shall see, it is not a very general result. The intuition behind the result is simple: one of the advantages of increasing the marginal tax rate at a particular income level is that the *average* tax rate of those at higher income levels is increased (keeping their marginal tax rates unchanged). The distortion (dead weight loss) arising from any tax is associated with the marginal tax rate; thus increasing the average tax rate on an individual while keeping the marginal tax rate the same increases revenues without increasing the deadweight loss.

(3) The marginal tax rate is lower (other things being equal) the higher the productivity of the group; that is, the loss in output from distortionary taxation is more important from groups whose output per unit of labor is higher.

(4) The marginal tax rate is lower the higher is the labor supply response. To see this, we consider the special case of a separable utility function (so $U_{CL} = 0$). In that case, it is easy to show that the compensated elasticity of labor (in the absence of taxation) is just ϵ^* .

4. Some mathematical problems

The maximization problems with which we have been concerned here are not well behaved. Even though the individual utility functions are, the self-selection constraints are not (in general.) This has several consequences. First, of course, there may be allocations which satisfy the first-order conditions but are not the global optimum (and indeed may be a local minimum).

Secondly, if the government can implement random tax policies, it may wish to do so (see below). But even if the government cannot, or does not, wish to

¹⁷Note that the derivative of the Hamiltonian with respect to the state value is equal to minus the derivative of the associated shadow price with respect to the running variable:

$$-(\partial\lambda/\partial w) = \partial H/\partial U = (G' - \gamma(\partial C/\partial U)_L)f + \lambda(L/w)U_{LC}(\partial C/\partial U)_L.$$

For simplicity, we assume a separable utility function. This enables us to integrate to obtain

$$-\lambda(w)/\gamma = \int_{w^*}^{w^{\max}} \{1/U_C - (G'/\gamma)\} dF,$$

using the fact that at $w = w_{\max}$, $\lambda(w) = 0$.

For a general discussion of the conditions under which this limiting result obtains, see Sadka (1976) and Seade (1977).

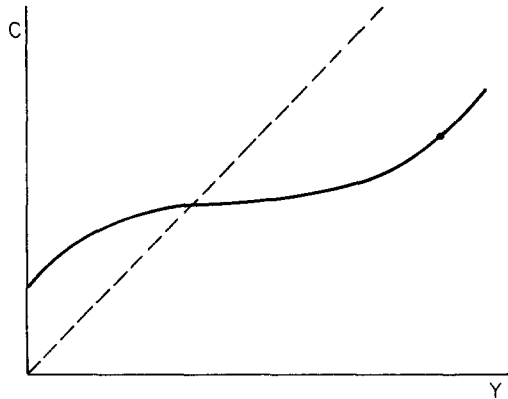


Figure 4.1. There must be a convex region to the function relating consumption to income. This induces at least a small randomization.

implement random tax policies, it must be concerned lest individuals randomize, unless, of course, the government can prohibit such gambling (which seems unlikely, since that would require that it differentiate between “naturally” occurring risks and artificially created risks).¹⁸ Recall our earlier result that the marginal tax rate on the highest income individual is zero. This, combined with the fact that his average tax rate is positive, and, if the tax system is at all redistributive, the fact that the average tax rate on poor individuals must be negative, implies that there must be a region in which the function, giving after-tax income as a function of before-tax income, must be convex (see Figure 4.1). If individuals are not too risk averse, they will, accordingly, attempt to randomize; indeed, an employer who offered any individual whose before-tax income was in the convex region a random pay could increase his profits. Unfortunately, there does not exist a correct analysis of Pareto efficient (or optimal) tax structure taking into account this class of responses to the tax structure. Note that this problem does not arise in the analysis of linear tax structures (Section 6 below).

4.1. Non-differentiability

In the analysis of a continuum of types we have implicitly assumed that the optimal tax function will be differentiable. It may well not be. There is an

¹⁸Even if it could prohibit artificially created gambles, it would simply encourage individuals to undertake naturally occurring gambles, even if they were not quite actuarially fair.

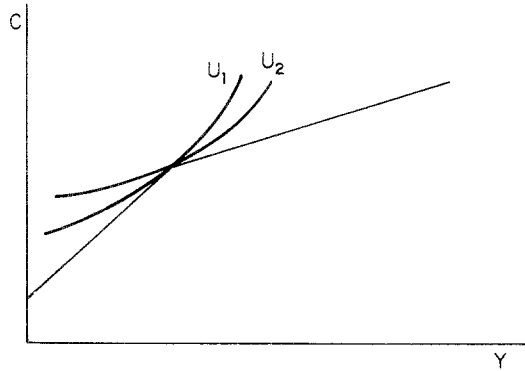


Figure 4.2. Points of non-differentiability in the tax schedule induce bunching (partial pooling).

interesting interpretation to these points of non-differentiability. Note in Figure 4.2 that if the consumption function is not differentiable, there will be a range of types of individuals who will not be distinguishable, i.e. who will choose precisely the same point (have the same income). Thus, it may be optimal to have a partially pooling equilibrium, and this corresponds precisely to the circumstances under which the tax function is not differentiable.

To see why with three or more groups partial pooling may be desirable, turn to Figure 4.3, where there are just three groups. Assume we fix the utility of the high ability (type 2) and the low ability (type 0). As we raise U^1 , the revenue we obtain from that group decreases (since we are moving along U^2 's indifference curve, and below E_2 , its slope is less than unity). But the revenue we raise from group 0

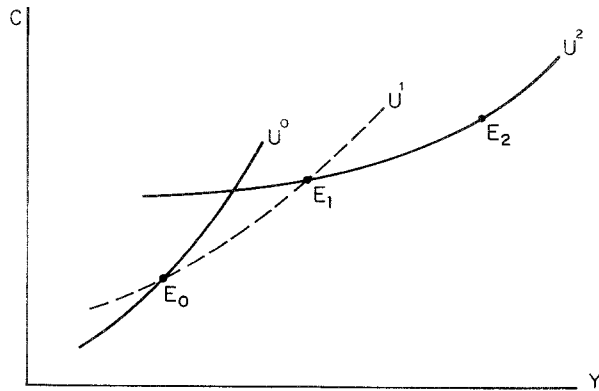


Figure 4.3. Desirability of partial pooling.

may be increased, if the slope of U^0 is less than unity at E_0 . Clearly, if there are relatively few type 1's, Pareto efficiency requires a pooling equilibrium in which types 0 and 1 cannot be differentiated.

We can use Pontryagin's technique to analyze Pareto efficient tax structures with non-differentiabilities, and derive conditions under which partial pooling is desirable.¹⁹

4.2. Discontinuities

Even if we preclude randomization (gambling), since the tax function is at least partially convex, and the indifference curves are convex, it is possible that there be multiple tangencies between the indifference curve and the function relating after-tax to before-tax income. Under these circumstances, small changes in the tax (consumption) function could cause seemingly large changes in revenues. However, if we assume a continuum of individuals, and if we give the government the right to assign a given proportion of any group which is indifferent between two allocations to each of the allocations, then it is clear that this presents no serious problem to the analysis.²⁰

5. Generalizations: Random taxation²¹

This chapter is concerned with the analysis of Pareto efficient tax structures. Whether a given tax structure can or cannot be improved upon depends crucially on the set of admissible tax structures. Indeed, this is perhaps one of the central lessons to emerge from the literature during the past two decades. [See also Dasgupta and Stiglitz (1971).]

Unfortunately, there is no agreement about how to define the set of admissible tax structures. Presumably, this should be related to the technology of tax collection; to the costs to the government of monitoring, or to the private sector of compliance (or of non-compliance); and to the accuracy with which the government can monitor the relevant variables. And these costs are likely to differ across countries, and to change over time. Most of the analyses, however, simply begin by postulating the admissible tax structure.

There are those who contend that the tax structures analyzed so far are overly restrictive, and those who contend that they are not restrictive enough. We take up each of these allegations in the next two sections.

¹⁹See Stiglitz (1977) for an analysis in the context of the monopoly insurance problem.

²⁰See Mirrlees (1971, 1976, 1985) for a detailed discussion of the mathematical problems posed by the optimal tax problem.

²¹This section draws heavily upon Arnott and Stiglitz (1986) and Stiglitz (1982b).

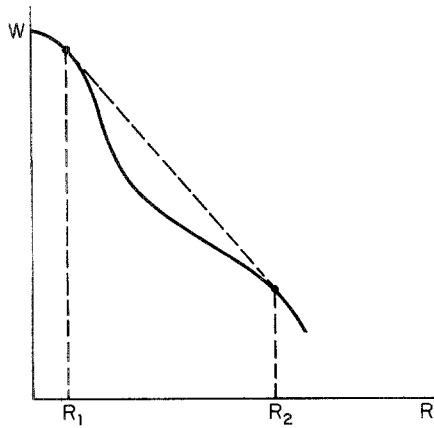


Figure 5.1. Welfare non-concave function of R . For revenues between R_1 and R_2 ex ante randomization is desirable.

Those who contend that we have been overly restrictive in our analysis point out that we should at least consider the possibility of random taxes.²²

It turns out, in fact, that under a variety of situations, random taxation is desirable. This can take on two forms. In *ex ante randomization* the government randomly assigns individuals to one of two tax schedules; if the tax is purely redistributive, the net revenue raised by one, S , is equal to the net loss of the other. More generally, the desirability of randomization can be seen most simply in the case where the government maximizes a utilitarian social welfare function. We can then plot average social welfare as a function of the revenue raised; if there is a convex region, then for some values of R (in the diagram, between R_1 and R_2) ex ante randomization is desirable. Arnott and Stiglitz derive conditions under which this can occur (see Figure 5.1).

In *ex post randomization*, the individual is not told what his tax schedule is until after he has announced his ability type. This kind of randomization may be a way of loosening the self-selection constraints. If, for instance, the more able individuals were much more risk averse than the less able, by randomizing the taxes paid by the less able the government makes it less attractive for the more able to attempt to pretend to be less able. This in turn implies that the distortionary tax imposed on the poor may be reduced. This can be seen most simply by considering Figure 5.2, where we choose two points on the poor individual's indifference curve yielding the same average government revenue as the original point E_1 . This randomization has no effect on the poor or on government revenue. But as risk aversion increases, the utility of the rich

²²Other more general tax structures include those which make taxes a function of wages. If wages are a function of effort one could not obtain a first-best outcome, even were wages observable. See also Part II below.

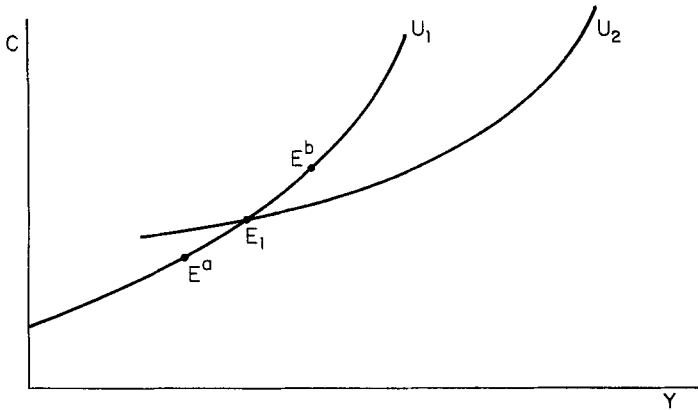


Figure 5.2. With sufficiently risk-averse individuals, ex post randomization weakens the self-selection constraint.

approaches the utility associated with the worst of the outcomes (from his perspective); hence we can effect a Pareto improvement.²³

What are we to make of these results? Though they might indeed provide a rationale for the seeming capriciousness with which taxes are sometimes imposed, we doubt whether such a policy would ever be openly argued – and voted for – in a democratic society. Our results show that the principle of horizontal equity – that otherwise identical individuals should be treated the same – may not only not be derivable from a utilitarian social welfare function, but that that principle may in fact be inconsistent with the principle of Pareto efficiency (at least on an ex ante expected utility basis).

I suspect that it is because of the possibilities of abuse – the belief that a random tax would not in fact be random, that the die would be loaded in favor of those in political power – that a proposal for tax randomization would meet such opposition.²⁴

Still, the analysis of random taxation is of interest for several reasons. First, it serves to remind us of the difficulties of defining the set of admissible taxes, and that our models may be – indeed probably are – leaving out some important considerations in determining tax structures.

Secondly, some of the properties of optimal tax structures may reflect an attempt to introduce randomization surreptitiously. Assume, for instance, that there were two types of labor, denoted M and W , which in all production and consumption characteristics were identical. In those circumstances in which

²³I am indebted to Steve Slutsky for this argument.

²⁴One important exception to the general presumption against randomization is in law enforcement. It is often argued that efficient law enforcement entails large fines imposed rarely. Another exception is provided by the draft lottery.

ex ante randomization is desirable, the optimal tax will “tell” us to impose different tax structures on the two groups. Most of us would say that that should not be allowed. But what if the two groups are slightly different? How different should they be to make differential taxation admissible? The theory gives us no guidance on this issue.

This issue arises even more forcefully in Part II, where we consider commodity taxation. Consider two groups of individuals who are identical except that one prefers chocolate ice cream, the other vanilla ice cream. In those cases where tax randomization is desirable, we will be told to tax vanilla and chocolate at different rates. But is this any more acceptable than randomizing taxes?

6. Limitations: Restricted taxation

So far we have assumed that there are no restrictions on the type of consumption (wage) tax which the government can impose. The tax schedules may be highly non-linear, with marginal tax rates increasing over some intervals and decreasing over others. The tax schedules which typically emerge from these analyses are far more complex than those currently employed in the United States. Curiously enough, a major focus of tax reform discussions in the United States has been simplifying the tax structure, by having a single marginal rate (the flat tax, or linear income tax). Non-linear tax structures present problems with income averaging. The unit of taxation becomes important – under the current U.S. tax system there are strong incentives for income shifting. Moreover, decreasing marginal rates²⁵ provides incentives for gambling (for firms to pay random wages).²⁶ Non-linear tax schedules open up large opportunities for tax arbitrage, with associated inequities and inefficiencies. One of the major advantages of a flat rate tax is that it enables income to be taxed at source, thus reducing considerably the administrative and compliance costs of the income tax.²⁷

We noted earlier that the underlying problem in the analysis of income taxation is an information problem: if the government could perfectly tell who is of what ability, it could impose lump-sum redistributive taxes. But the government can only make inferences about who is of what ability by the income they receive. The problems we have just been discussing can also be viewed as information problems. The government cannot perfectly and costlessly observe the income or consumption of each individual; if it could, there would be no compliance or administrative costs associated with running a non-linear tax system.

²⁵Recall the standard result that the marginal rate for the highest income individual should be zero; hence, the optimal tax schedule (as usually defined) entails decreasing marginal rates over at least some interval.

²⁶See the discussion above in Section 4.

²⁷For an excellent discussion of the advantages of a flat tax, see Hall and Rabushka (1983).

Opposing these objections to a highly non-linear tax are the possible increases in social welfare from the greater redistribution that can be achieved through a non-linear tax. If there were no administrative costs, etc. it is easy to show that a non-linear tax Pareto dominates a linear tax. But there is a growing consensus in the United States that the problems associated with very non-linear tax schedules outweigh these gains in redistribution.²⁸ Whether there would be significant gains in going from a flat rate tax to a two- or three-tier schedule remains a question of on-going research.

A number of economists have analyzed the optimal linear tax.²⁹ This problem is a fairly simple one: a tax schedule is described by an equation of the form

$$C = I + (1 - t)Y, \quad (26)$$

where I is the lump-sum payment to each individual, and t is the marginal tax rate. The government looks for that feasible tax structure which maximizes social welfare, that is, it solves

$$\max \int G(U(C(w), L(w))) dF(w) = \int G(v(w(1 - t), I)) dF(w) \quad (27)$$

subject to

$$R + \int \{C(w) - L(w)w\} dF(w) = R + I - \int tL(w)w dF(w) \leq 0, \quad (28)$$

where R is the revenue which the government is required to raise for financing public goods (may be zero) and where we have expressed individual utilities in terms of the indirect utility function, giving utility as a function of the after-tax wage and lump-sum transfer. The condition simply says that the aggregate value of expenditures on public goods plus private consumption equals the aggregate value of output. Note that the objective function is the same as that employed in our earlier analysis; the only difference is that whereas before, any tax function was admissible, now we are restricted to linear tax functions.

Three general results can be obtained:

(a) If $R \leq 0$, then the optimal linear tax entails $I > 0$. The marginal deadweight loss from a very small tax is negligible, but the gain in redistribution is positive.

(b) If R is large enough, then $I < 0$; indeed, since in general an increase in a proportional tax beyond a certain point will actually decrease revenues, if R exceeds the maximal revenue that can be raised through a proportional tax, the government will have to supplement it with a uniform lump-sum levy. More generally, the greater is R , the smaller is I .

²⁸These views are reinforced by Mirrlees' (1971) calculations suggesting that the optimal tax structure is close to linear.

²⁹See, in particular, Dixit and Sandmo (1977), Sheskinski (1972), and Stiglitz (1976b). The discussion here follows closely along the lines of the last of these papers.

(c) The optimal tax formula can be written in a remarkably simple form:

$$t/(1-t) = -\text{cov}[b, Y]/\int Y\epsilon_{LL} dF, \quad (29)$$

where

$b = G'\alpha/\gamma + tw(\partial L/\partial I)$, the net social marginal valuation of income,³⁰

$\alpha = \partial U/\partial C$, the marginal utility of income,

γ = the marginal value of a dollar to the government (the Lagrange multiplier associated with the revenue constraint),

$\text{cov}(b, Y) = E(b - b)(Y - Y)$, the covariance between b (defined above) and Y ,
and

ϵ_{LL} = compensated elasticity of labor supply.³¹

³⁰The first term is the direct effect of giving an extra dollar to an individual (normalized by the value of a dollar to the government), while the second term is the change in revenue to the government resulting from the fact that the individual will alter his labor supply as a result.

³¹To obtain this result, we form the Lagrangian:

$$\mathcal{L} = \int [G + \gamma(twL - I - R)] dF, \quad (30)$$

and differentiate with respect to I and t to obtain:

$$\int [G'\alpha + \gamma(tw(\partial L/\partial I) - 1)] dF = 0 \quad (31)$$

and

$$\int [G'(\partial U/\partial t) + \gamma(wL + tw(\partial L/\partial t))] dF = 0. \quad (32)$$

The first condition can be written simply as

$$Eb = 1.$$

Equation (29) follows from (32), making use of the Slutsky relation,

$$\partial L/\partial t = -w(\partial L/\partial w)_U - wL(\partial L/\partial I),$$

the fact that

$$\partial U/\partial t = -\alpha wL,$$

and the definition of $\epsilon_{LL} = [w(1-t)/L](\partial L/\partial w)_U$, to obtain:

$$\int wL(G'\alpha - \gamma\{1 - tw(\partial L/\partial w)_U/L - tw(\partial L/\partial I)\}) dF = 0,$$

or, dividing through by γ and using the fact that $Eb = 1$,

$$\int Y(b - Eb + t\epsilon_{LL}/(1-t)) dF = 0.$$

The covariance may be seen as a *marginal* measure of inequality. Consider, in particular, the case of a small tax, so $b \approx G'\alpha/\gamma$. If we gave a lump-sum payment to everyone, financed by a tax at rate t , $I = t\bar{Y}$, where \bar{Y} is mean income. The increment in social welfare is thus $\int tG'\alpha(\bar{Y} - Y) dF = -t\gamma \text{cov}(b, Y)$.

Thus, our basic formula says that the tax rate should be greater, the greater this measure of marginal inequality, and the smaller the (weighted average) compensated elasticity of labor supply. The marginal measure of inequality will be greater, the more egalitarian the social welfare function, i.e. the larger is $-G''U/G'$.

7. Limitations: The push-pin versus poetry controversy

While the New Welfare Economics emphasized the impossibility of interpersonal utility comparisons, much of the recent work in optimal tax theory has simply ignored these concerns. By positing that all individuals have the same indifference curves between leisure and consumption, they could simply assert that all had the same utility function; there seemed a “natural” way of making interpersonal utility comparisons. Three problem were thus ignored.

First, there are many alternative ways of representing the same family of indifference maps. Each representation would yield a different optimal income tax. The literature has developed no persuasive way for choosing among these alternative representations.

Secondly, when indifference curves cross, as in Figure 7.1, then one must decide which of individual 2's indifference curves corresponds to $U^1 = \alpha$.

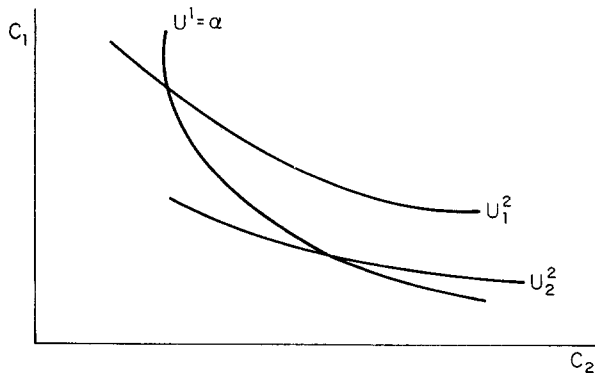


Figure 7.1. When indifference curves cross, there is no obvious way of ascertaining which of individual 2's indifference curves correspond to $U^1 = \alpha$.

curve of individual 1 that we have labeled α . This choice too will have a critical effect on the desired income tax rate.

Thirdly, conventional models have assumed that individuals who are more productive in the workplace have the same capacity for enjoyment of goods. The relationship between these two capacities troubled the early utilitarians. The fact that there was no resolution of the ensuing controversy does not mean that the issues should be ignored. In particular, the consequences can be seen most forcefully within the strand of modern economic theory which has emphasized the role of home production. Assume that individuals do not consume the goods they produce directly, but rather transform purchased goods into true consumption goods. Thus, for simplicity, assume that there is a single ultimate good which the consumer is consuming (e.g. book-reading), with $U = U(Q)$. The output of this good, Q , is a function of the input of time (leisure) and of goods (C): the functional relationship will differ from individual to individual, depending on the abilities of individuals in household production.

Such a model has important – and probably not completely acceptable – implications within a utilitarian framework. Assume that Q is Cobb–Douglas and normalise the available time for market plus home production at unity:

$$Q = C^a(1 - L)^{1-a}h_i = w_i^a h_i L^a(1 - L)^{1-a}.$$

Then, in a free market economy, all individuals will supply labor equal to

$$L = a.$$

The marginal utility of income of an individual of market productivity w_i and home productivity h_i is equal to

$$MU_i = aU'(Q)C_i^{a-1}(1 - L_i)^{1-a}h_i = a^a(1 - a)^{1-a}U'(Q)h_i w_i^{a-1}.$$

Letting home productivity depend on market productivity,

$$h_i = h(w_i), \quad d \ln h / d \ln w = x,$$

and differentiating, we obtain:

$$\frac{d \ln MU_i}{d \ln w_i} = (a - 1) + \frac{h'}{h} w_i + \frac{U''}{U'} Q \left[a + \frac{h' w_i}{h} \right] \leq 0,$$

as

$$\rho \geq \frac{a + x - 1}{a + x},$$

where ρ is the elasticity of marginal utility. Assume, for instance, that h is proportional to w , so $x = 1$. Then, so long as $\rho < a/(a + 1)$, there should be redistribution towards individuals with higher wages. Although h_i is in principle observable (by detailed observation of the individual within the home), there is, in this example, no way we can infer from observing individual's behavior in the market place (i.e. his labor supply) his home productivity. It would appear, then, that one cannot derive, on the basis of the utilitarian ethic, a general case in favor of redistribution from those who have high wages in the market to those with low wages. Such an argument requires specific assumptions concerning the rate at which marginal utility diminishes, the relationship between home and market productivity, and characteristics of the home production process. Yet, there is more widespread agreement that there ought to be redistribution from the more able to the less able than this utilitarian analysis would suggest. These considerations – as well as the practical considerations entailed in deciding upon the appropriate cardinalization within the utilitarian framework – make the utilitarian approach questionable as a guide to policy.

8. Limitations: General equilibrium effects of taxation

Traditional tax theory has been much concerned with the incidence of taxation, with the extent to which taxes are shifted. There has been concern, for instance, that taxes on doctors and lawyers simply result in higher fees. This central issue of the *incidence* of taxation was completely ignored in the early analyses of optimal income taxation; it was simply assumed that the before-tax wage of an individual would be unaffected by the imposition of a tax. Several recent papers have called attention to the potentially important consequences of these general equilibrium effects.³² Consider, for instance, a modified version of the two-group model presented in the first section. There, we assumed that the productivity (output per man hour) of each group was fixed. Now, assume that the two groups provide different kinds of labor services (skilled versus unskilled labor). The relative wage will then depend on the relative supplies of the two types of labor. If the labor supply of the skilled is relatively elastic, and of the unskilled relatively inelastic, then a uniform marginal tax rate on both might increase the before-tax relative wage of the skilled; to offset this, one might wish to have a lower marginal tax rate on the skilled.

Indeed, Stiglitz (1982b) has shown that (in the standard case where it is the high ability self-selection constraint which is binding) *Pareto efficient taxation requires that the marginal tax rate on the most able individual should be negative,*

³²See Allen (1982), Stern (1982), Feldstein (1973), and Stiglitz (1982b, 1985b).

except in the limiting case where the two types of labor are perfect substitutes, in which case it is zero.

The intuition behind this result is simple. Since there is a deadweight loss from redistributing income, it is always desirable to incur some small deadweight loss to change the before-tax distribution of income in a desirable manner.

Stiglitz also shows that while the marginal tax rate on the less able is always positive, its magnitude depends on the elasticity of substitution; the smaller the elasticity of substitution, the larger the marginal tax rate. The government increasingly relies on the general equilibrium incidence of the tax, the change in the before-tax relative wages, to redistribute income.

Formal proof.

These results can be seen more formally as follows. Assume that output is a function of the supply of hours by each of the two types:

$$Q = F(N_1L_1, N_2L_2) = L_1N_1f\left(\frac{N_2L_2}{N_1L_1}\right), \quad (33)$$

where F exhibits constant returns to scale. If each factor receives its marginal product,

$$\begin{aligned} w_2 &= \frac{\partial F}{\partial (N_2L_2)} = f'(n); \\ w_1 &= \frac{\partial F}{\partial (N_1L_1)} = f(n) - nf'(n) = g(n), \end{aligned} \quad (34)$$

where $n = N_2L_2/N_1L_1 = N_2Y_2w_1/N_1Y_1w_2$.

We can now write the Lagrangian:³³

$$\begin{aligned} \mathcal{L} &= V^2 + \mu V^1 \\ &\quad + \gamma (F(N_1Y_1/w_1, N_2Y_2/w_2) - C_1N_1 - C_2N_2 - \bar{R}) \\ &\quad \quad \quad \text{(resource constraint)} \\ &\quad + \lambda_2 (V^2(C_2, Y_2; w_2) - V^2(C_1, Y_1; w_2)) \\ &\quad \quad \quad \text{(self-selection constraint)} \\ &\quad + \phi (w_1 - g(n)), \end{aligned} \quad (35)$$

³³The proof presented here is rather different from that presented in Stiglitz (1982b), which re-expressed the problem in terms of the control variables L_i and C_i . The proof here uses the fact that $\phi > 0$, i.e. welfare would be raised if unskilled workers were paid a wage in excess of their marginal product.

where we have expressed the resource constraint making use of the aggregate production function; we have appended the constraint on the value of w_1 ; and we focus on the case where the upper self-selection constraint is binding. Then straightforward differentiation yields:

$$\frac{\partial \mathcal{L}}{\partial C_2} = \frac{\partial V^2}{\partial C_2} + \lambda^2 \frac{\partial V^2}{\partial C_2} - \gamma N_2 = 0, \quad (36a)$$

$$\frac{\partial \mathcal{L}}{\partial Y_2} = \frac{\partial V^2}{\partial Y_2} + \lambda^2 \frac{\partial V^2}{\partial Y_2} + \gamma N_2 - \phi g'n/Y_2 = 0. \quad (36b)$$

Dividing

$$-\frac{\partial V^2/\partial Y_2}{\partial V^2/\partial C_2} = 1 - \frac{\phi g'n}{\gamma N_2 Y_2} \quad (37)$$

that is, unless $f'' = 0$, (unless the elasticity of substitution is infinity) the more productive individual should work beyond the point where the slope of his indifference curve (in C, Y space) is unity, that is, he should face a *negative* marginal tax rate.

9. Other limitations on the general model

Two other limitations on the general model should be noted. First, we have assumed that income is a non-stochastic function of effort. Views about the desirability of progressive taxation are often related to views concerning the extent to which differences in income are due to differences in effort, to differences in ability, or to differences in luck.³⁴

Secondly, we have assumed that income is monitored perfectly. In fact, measured income can be viewed as a noisy signal of true income. Using the techniques developed by Radner and Stiglitz,³⁵ it should be possible to show that if it is a noisy enough signal, then it will not be desirable to differentiate among individuals on that basis (using, say, a utilitarian social welfare function). The gains in redistribution will be more than offset by the costs.³⁶ [See also Stern (1982).]

³⁴For a brief discussion on optimal taxation with stochastic income, see Stiglitz (1982b), Varian (1980), Eaton and Rosen (1980), and Diamond, Helms and Mirrlees (1980).

³⁵See Radner and Stiglitz (1984).

³⁶Another line of research is concerned with ascertaining conditions for the optimal allocation of resources to monitoring.

Table 10.1
Utilitarian and maxi-min solutions of optimal income tax.^a

Level of w	Utilitarian		Maxi-min	
	Tax rates		Tax rates	
	Average	Marginal	Average	Marginal
Case I ^b				
w_0	—	—	—	—
Median	6	21	10	52
Top decile	14	20	28	34
Top percentile	16	17	28	26
Case II ^b				
w_0	—	—	—	—
Median	-13	19	-6	50
Top decile	4	19	22	33
Top percentile	7	17	25	25

^aSource: Atkinson (1972, Table 3).

^bIn Case I the revenue requirement is positive, in Case II it is negative (the government disposing of the profits of public sector production).

10. Numerical calculations

We have seen that only limited qualitative results concerning the nature of the optimal tax schedule have been derived. Several economists have attempted to calculate the optimal tax schedule, making particular assumptions concerning the distribution of abilities, the social welfare function, and individuals' utility functions. Not surprisingly, the results seem to vary greatly with the particular assumptions employed. The elasticity of labor supply is obviously critical in determining the extent of our concern with incentive effects, and unfortunately, there is not agreement among economists about the magnitude of this. The more egalitarian the social welfare function, the larger the role to be played by redistributive taxation. As an example, Mirrlees calculated the optimal income tax for a utilitarian social welfare function, assuming the special utility function

$$U = \frac{1}{2} \log C + \frac{1}{2} \log(1 - L),$$

and a lognormal distribution for w . The schedule he obtained was remarkably close to linear. By contrast, Atkinson analyzed the optimal tax schedule for the same problem, assuming a Rawlsian social welfare function. He obtained increasing average but decreasing marginal tax rates.³⁷

³⁷See Atkinson (1972). For other numerical calculations, see Stern (1976).

Welfare gains

One of the major thrusts of the recent research in optimal taxation has been its concern with the incentive effects of redistributive taxation. Given an egalitarian social welfare function, there are gains from a more equal distribution of income, but there are losses from the distortions associated with the income tax. The question is: How significant are the gains relative to the costs? The answer clearly depends on at least three critical assumptions: how elastic the (compensated) labor supply is; how averse to inequality society is; and how large a share in national income government expenditure is. For a utilitarian social welfare function, with “reasonable” estimates of the compensated labor supply function, Stiglitz (1976b) suggested that the welfare gains to be achieved by the optimal linear income tax were minimal, without counting the administrative costs associated with the tax.

Part II: Pareto efficient taxation with many commodities and many periods**11. Alternative tax bases**

So far we have focused our attention on the analysis of an optimal wage income tax. Other tax and expenditure policies have often been employed as redistributive devices. Most governments impose taxes on such luxuries as perfumes, and many governments provide subsidies for such necessities as food.

There was a widespread feeling that the government should concentrate its redistributive policies within the income tax. Using excise taxes to redistribute simply added additional distortions. Clearly, if labor were inelastically supplied, then an optimally designed wage tax would be all that is required. But labor is not inelastically supplied, and, as Ramsey (1927) forcefully established more than fifty years ago, counting the number of distortions is no way to do welfare analysis. If we have an optimally designed income tax, should we supplement it with excise taxes? The heuristic argument, that by doing so we impose less of a burden on the income tax, seems persuasive, but unfortunately is incorrect.

The answer turns out to depend simply on whether the utility function is separable between the vector of consumption goods and work, i.e. on how changes in leisure affect individual marginal rates of substitution between different consumption goods. *If the utility function is separable between consumption and leisure ($U = u(C) - v(L)$, where C is a vector), then all Pareto efficient tax structures entail no taxation of commodities.*

11.1. Formal analysis

The result can be seen most simply in the context of our two-group model. If we now interpret C as a vector, the Lagrangian for this problem is identical to that formulated earlier (Section 2.2), except the government budget constraint is now written as

$$R \leq N_1 Y_1 + N_2 Y_2 - \sum_j (C_{1j} N_1 + C_{2j} N_2). \quad (38)$$

If we now differentiate the Lagrangian with respect to C_{ij} , we obtain [see the analogous equations (12) and (14)]:

$$\frac{\partial \mathcal{L}}{\partial C_{1j}} = \mu \frac{\partial V^1}{\partial C_{1j}} - \lambda_2 \frac{\partial V^2}{\partial C_{1j}} + \lambda_1 \frac{\partial V^1}{\partial C_{1j}} - \gamma N_1 = 0, \quad (39a)$$

$$\frac{\partial \mathcal{L}}{\partial C_{2j}} = \frac{\partial V^2}{\partial C_{2j}} + \lambda_2 \frac{\partial V^2}{\partial C_{2j}} - \lambda_1 \frac{\partial V^1}{\partial C_{2j}} - \gamma N_2 = 0, \quad (39b)$$

$$\frac{\partial \mathcal{L}}{\partial Y_1} = \mu \frac{\partial V^1}{\partial Y_1} - \lambda_2 \frac{\partial V^2}{\partial Y_1} + \lambda_1 \frac{\partial V^1}{\partial Y_1} + \gamma N_1 = 0, \quad (39c)$$

$$\frac{\partial \mathcal{L}}{\partial Y_2} = \frac{\partial V^2}{\partial Y_2} + \lambda_2 \frac{\partial V^2}{\partial Y_2} - \lambda_1 \frac{\partial V^1}{\partial Y_2} + \gamma N_2 = 0. \quad (39d)$$

We again focus on the case where $\lambda_1 = 0$ and $\lambda_2 > 0$: only the second self-selection constraint is binding. From (39a)–(39d) we obtain:

$$\frac{\partial V^2 / \partial C_{2j}}{\partial V^2 / \partial C_{2k}} = 1, \quad - \frac{\partial V^2 / \partial C_{2j}}{\partial V^2 / \partial Y_2} = 1, \quad (40a)$$

$$\frac{\partial V^1 / \partial C_{1j}}{\partial V^1 / \partial C_{1k}} = \frac{N_1 \gamma + \lambda_2 \partial V^2 / \partial C_{1j}}{N_1 \gamma + \lambda_2 \partial V^2 / \partial C_{1k}}. \quad (40b)$$

Equation (40a) yields the familiar result that there should be no distortionary taxation on the individual with the highest ability. The interpretation of (40b) is, however, somewhat more subtle. Consider first the case where individuals have separable utility functions between leisure and goods, i.e.

$$\frac{\partial^2 U^i}{\partial C_{ij} \partial L_i} = 0, \quad \text{all } i, j. \quad (41)$$

Since we assume that individuals have the same indifference curves (in $\{C, L\}$ space), at the same consumption vector

$$\partial V^2/\partial C_{1j} = \partial V^1/\partial C_{1j}, \quad (42a)$$

$$\partial V^2/\partial C_{1k} = \partial V^1/\partial C_{1k}, \quad (42b)$$

and equation (40b) becomes:

$$\frac{\partial V^1/\partial C_{1j}}{\partial V^1/\partial C_{1k}} = 1. \quad (43)$$

We have thus established our basic result: If leisure and goods are separable, there should be no commodity taxation. It should be noted that in this analysis we allow tax functions which are not only non-linear functions of consumption, but are also not separable, i.e. the marginal rate imposed on the consumption of commodity j may depend not only on the consumption of commodity j but on other commodities as well.

If the utility function is not separable, we obtain:

$$\mu \left(\frac{\partial V^1}{\partial C_{1j}} - \frac{\partial V^1}{\partial C_{1k}} \right) = \lambda_2 \left(\frac{\partial V^2}{\partial C_{1j}} - \frac{\partial V^2}{\partial C_{1k}} \right) \quad (44)$$

or

$$\begin{aligned} \frac{\partial V^1/\partial C_{1j}}{\partial V^1/\partial C_{1k}} - 1 &= \frac{\lambda_2}{\mu} \frac{\partial V^2/\partial C_{1k}}{\partial V^1/\partial C_{1k}} \left(\frac{\partial V^2/\partial C_{1j}}{\partial V^2/\partial C_{1k}} - 1 \right) \\ &= \frac{\lambda_2}{\mu} \frac{\partial V^2/\partial C_{1k}}{\partial V^1/\partial C_{1k}} \left(\frac{\partial V^2/\partial C_{1j}}{\partial V^2/\partial C_{1k}} - \frac{\partial V^1/\partial C_{1j}}{\partial V^1/\partial C_{1k}} \right) \\ &\quad + \frac{\lambda_2}{\mu} \left(\frac{\partial V^1/\partial C_{1j}}{\partial V^1/\partial C_{1k}} - 1 \right) \frac{\partial V^2/\partial C_{1k}}{\partial V^1/\partial C_{1k}} \\ &= \frac{\frac{\lambda_2}{\mu} \frac{\partial V^2/\partial C_{1k}}{\partial V^1/\partial C_{1k}} \left(\frac{\partial V^2/\partial C_{1j}}{\partial V^2/\partial C_{1k}} - \frac{\partial V^1/\partial C_{1j}}{\partial V^1/\partial C_{1k}} \right)}{1 - \frac{\lambda_2}{\mu} \frac{\partial V^2/\partial C_{1k}}{\partial V^1/\partial C_{1k}}}. \end{aligned}$$

Thus, whether commodity j should be taxed or subsidized relative to k depends on whether the more able individual's marginal rate of substitution of j for k exceeds that of the low ability person, or conversely.

In the case we have focused on, where the low ability and high ability individuals have the same utility function, the relative values of the marginal rate of substitution depend simply on how the amount of leisure affects the marginal rate of substitution. The critical parameter that has to be estimated, then, for determining the structure of commodity taxes is

$$\frac{d \ln U_k / U_j}{d \ln L}.$$

The result that, with separability, only an income tax is needed, which seemed so surprising at first, becomes entirely understandable within this framework; if the two groups of individuals have the same indifference curves (locally) between two commodities we cannot use the differential taxation as a basis of separation; if they differ, we can. By taxing the commodity which the more able individual values more highly in the lower ability individual's package, we make the lower ability individual's "package" less attractive to him. (Since in this model both groups have identical utility functions, the only difference in the evaluation of a given consumption bundle arises from the differences in leisure which they enjoy at any given level of income.) We thus can tax the higher ability individual more heavily without having him trying to "disguise" himself as a low ability person.

11.2. Ramsey versus Atkinson–Stiglitz Pareto efficient taxation

The results we have just obtained stand in marked contrast to standard results on the design of commodity tax structures.³⁸ In the simple form (where the compensated demand for each commodity depends only on its own price, and where producer prices are fixed) optimal taxation entails tax rates which are inversely proportional to the elasticity of demand; in the more general case, optimal taxation calls for an equal percentage reduction of consumption (again along the compensated demand curves) for all commodities.³⁹ The factors which are relevant for optimal Atkinson–Stiglitz pricing are completely different; there, all

³⁸ The list of contributors to this literature is enormous. For a comprehensive survey, see Chapter 2 by Auerbach in this Handbook. Besides the classic papers by Ramsey (1927) and Boiteux (1956), important contributions by Diamond and Mirrlees (1971), Baumol and Bradford (1970), Dasgupta and Stiglitz (1971), Atkinson and Stiglitz (1972, 1976), Diamond (1975), Meade (1955), Corlett and Hague (1953) may be noted.

³⁹ Both results assume all individuals are identical, so that distributional considerations can be ignored. See below.

that is relevant is how the marginal rate of substitution between two commodities changes as the amount of leisure changes.

This poses a problem: Which theory provides the appropriate bases for the design of tax policy? How can these alternative approaches be reconciled?

Ramsey's analysis (and much of the subsequent work which developed from it) was based on an artificial problem: he assumed that all individuals were identical, but that the government could not impose a lump-sum tax. As we noted earlier, there is really no reason the government cannot impose a *uniform* lump-sum tax; what the government cannot do is levy lump-sum taxes which vary according to the individual's ability to pay. But if all individuals are identical, then a uniform lump-sum tax is the optimal tax. Our analysis has assumed not only that the government can impose a uniform lump-sum tax, but that it can impose a non-linear consumption (wage) tax. Giving the government this additional instrument is what changes the nature of the optimal commodity tax structure.

Virtually all developed countries impose some income tax, which makes Ramsey's analysis inapplicable to those countries. On the other hand, for a variety of reasons, many less developed countries do not impose a very effective income tax. For these, Ramsey's analysis might seem appropriate; but Ramsey (and the subsequent work cited in footnote 38) assumed a neoclassical model of the economy, a model which is particularly ill-suited for less developed countries. In a later section, we describe briefly recent work attempting to extend this kind of analysis to less developed countries.

11.3. Redistributive commodity taxes with optimal linear income taxes

In the previous section we noted the limited role to be played by redistributive commodity taxation when there is an optimal consumption tax. Is that still true if there is an optimal *linear* consumption (or income) tax?

The answer is provided by the generalized Ramsey Rule [Atkinson and Stiglitz (1976)]:

$$\sum \sum t_i S_k^h / HC_k = \theta_k, \quad (45)$$

where S_k^h is the h th household's compensated price derivative of the i th commodity with respect to the k th price; and

$$\theta_k = \sum \text{cov}(C_k^h, b^h) / HC_k \bar{b}, \quad (46)$$

the normalized covariance between C_k^h and b^h . Generally, goods which are consumed by the rich will have a high (absolute) value of θ_k . It is a measure of the extent to which the good is consumed by those with low values of social

marginal utility of income. H is the total number of households; C_k is the average consumption of commodity k and HC_k is the aggregate consumption of commodity k . b^h is the net marginal social utility of income, as defined earlier.

Equation (45) says that the percentage reduction in consumption of each commodity (along the compensated demand curve) should be proportional to its distributional characteristic θ_k . This stands in contrast to the standard Ramsey result that the percentage reduction should be the same for all commodities.⁴⁰

Equation (45) has some simple and remarkable implications. With linear demand curves derived from a utility function of the form

$$2d_i(Q_i - \bar{Q}_i)^2 - aL,$$

it is easy to show that (45) implies that there should be no commodity taxation (Atkinson and Stiglitz (1976, 1980)). Deaton (1979) has shown that this holds for more general classes of utility functions, including the frequently estimated linear expenditure system. Another special case is that originally investigated by Ramsey, where there is constant marginal utility of income and separable demand functions. When wage differences are the only sources of inequality, this implies that we can write the demand curves as

$$C^h = C_i(q_i/w^h),$$

where w^h is the wage of the h th individual. Moreover,

$$b^h = 1/\gamma w^h,$$

where, it will be recalled, γ is the marginal utility of income to the government. Now, the optimality condition (45) becomes:

$$t_k \sum (C'_k/w^h) = \sum_h (b^h - \bar{b})(C_k^h - C_k). \quad (48)$$

Let $v^h = 1/w^h$. Taking a Taylor series approximation to the right-hand side, we obtain (setting, without loss of generality, $\bar{v} = 1$, and expressing all aggregate

⁴⁰In fact, both results are special cases of the more general result of Atkinson and Stiglitz (1976):

$$\sum \sum t_i S_k^h / HC_k = \bar{b} \theta_k - (1 - \bar{b}). \quad (47)$$

With an optimal linear income tax, $\bar{b} = 1$. If all individuals are identical (as in Ramsey's original analysis), $\theta_k = 0$.

results in per capita terms):⁴¹

$$C'_k \sigma_v^2 q_k / \gamma$$

while taking a Taylor series approximation to the left-hand side yields:

$$t_k \sum (C'_k / w^h) \sim t_k C'_k \{1 + [-\varepsilon_k - m_k \varepsilon_k / 2] \sigma_v^2\},$$

where $\varepsilon_k = -C''_k q / C'_k$, the elasticity of demand, and $m_k = C'''_k q / C''_k$. Thus,

$$t_k / q_k \sim \sigma_v^2 / \{1 + [-\varepsilon_k - m_k \varepsilon_k / 2] \sigma_v^2\} \gamma.$$

Notice that for the quadratic utility function, $\varepsilon_k = m_k = 0$, so uniform taxation is again optimal. More generally, even to a first-order approximation, the relative tax rates will depend not only on the elasticity of demand, but also on m_k , a parameter which is hard to estimate. Assuming that the demand curves exhibit constant elasticity has, in particular, strong implications for the structure of optimal commodity taxes. Finally, note that t_k does not vary inversely with the elasticity of demand, even for constant elasticity demand functions.

In short, if there is an income tax, the role for indirect taxation is greatly changed, and if commodity taxes are imposed, the optimal structure of these taxes may depend on parameters other than those that it depended on in the absence of an income tax, parameters for which it seems intrinsically hard to obtain reliable estimates; and even in those special cases where the only demand parameter on which the structure of taxes depends is the elasticity of demand, it does not vary simply inversely with the demand elasticity.

11.4. Other forms of restricted taxation

In the previous subsection we noted the important consequences of restrictions on taxation: the optimal commodity tax structure depended critically on whether there was or was not an income tax, and whether, if there was an income tax, it was restricted to being linear.

Another important example where restrictions on taxation have an important implication for tax structure arises when the government cannot impose 100 percent profits taxes. No government does this, perhaps because of the difficulty of differentiating between true profits (in the sense that economists use that term) and return to capital or to entrepreneurship. With 100 percent profits taxes, the

⁴¹ $b^h = v^h / \gamma$, $(b^h - \bar{b}) = (v^h - 1) / \gamma$; $C_k^h \approx C_k(q_k) + C'_k(q_k)(v^h - 1)q_k$.
 $\sigma_v^2 \equiv E(v - \bar{v})^2$.

commodity tax structure depends simply on properties of the demand curves, and there should be no taxation of producers (which would interfere with the production efficiency of the economy).⁴² The former result conflicts with Ramsey's original prescription, where optimal tax rates depended on both the demand and supply elasticities. The reconciliation was provided by Stiglitz and Dasgupta (1971), who showed that if there were not 100 percent profits taxes, producer taxation may be desirable, and optimal commodity taxes would depend on supply elasticities. Ramsey had implicitly assumed no profits taxes.

12. Implications for capital taxation

Consumption at different dates is, of course, like consumption of different commodities. Thus, the Atkinson–Stiglitz results provide the suggestion of an answer to the long-standing controversy on whether interest income should be taxed. Early advocates of a consumption tax (taxing consumption only is equivalent to exempting interest) emphasized that it was both equitable (individuals should be taxed on what they take out of society – their consumption – and not what they contribute – their income⁴³) and more efficient (since it did not introduce the additional intertemporal distortion.)⁴⁴ Ramsey's analysis of optimal taxation should have made the latter argument suspect even before Kaldor put it forward: fewer distortions are not necessarily better than more. Indeed, Ramsey's analysis suggested that whether consumption at later dates should be taxed at a higher or lower rate should depend on the elasticity of demand for consumption at later dates relative to the elasticity at earlier dates.

The Atkinson–Stiglitz analysis provided a framework within which both the equity and efficiency issues could be addressed. They showed that *if there was separability between consumption (at all dates) and work, then a consumption tax was Pareto efficient.*

Still, that did not close the matter, for theirs was a partial equilibrium model. We noted earlier (Section 8) that general equilibrium effects of tax policy must be taken into account when tax policy can affect factor prices. Partial equilibrium models cannot address the effects of taxation on savings and investment, and hence on the intertemporal distribution of income. These questions have been addressed by two sets of models.

⁴²This result was first established by Diamond and Mirrlees (1971). For a simple interpretation of that result see Stiglitz and Dasgupta (1971).

⁴³This is the position most associated with Fisher (1937).

⁴⁴See, for instance, Kaldor (1955).

12.1. Overlapping generations models

Stiglitz (1985b) has extended the Atkinson–Stiglitz analysis using overlapping generations. He asked: What can we say in that context about the set of Pareto efficient tax structures? Three results stand out:

(1) If relative wages between the skilled and unskilled workers is fixed (independent of the capital stock), then the partial equilibrium results remain valid: provided there is separability in the utility functions, Pareto efficiency requires that there be a consumption tax.⁴⁵

(2) If relative wages between the skilled and unskilled workers changes with the level of capital accumulation, the government will, in general, wish either to have an interest income tax or subsidy. The intuitive reason is analogous to that presented earlier concerning why, when there are these general equilibrium effects, it would be desirable for the government to impose a negative marginal tax rate on the more productive individuals. Here, if increasing the level of capital increases the income of the more able relative to the less able (for

⁴⁵The Lagrangian for this problem can be written

$$\mathcal{L} = \sum_t \left\{ \sum_i \mu_{it} U_{it} + \lambda_{2,t} (U^2(C_{2t}, L_{2t}) - U^2(C_{1t}, \hat{L}_{1t})) + \gamma_t \left[K_{t+1} - K_t + \sum_i C_{it,t} + \sum_i C_{it-1,t} - F_t(K_t, E_t) \right] \right\},$$

where $Q = F(K, E)$ is the aggregate production function, E the number of efficiency units of labor ($E = \nu N_1 L_1 + N_2 L_2$, where ν is the ratio of the productivity of the two types of laborers), where γ_t is the Lagrange multiplier for the resource constraint at time t ; where $C_{it,t}$ is the consumption of the i th type in the t th generation the first year of its life; $C_{it,t+1}$ is its consumption in the second year of its life; and, as in the standard overlapping generations model, we assume individuals live for only two periods, working only in the first. $C_{it} \equiv \{C_{it,t}, C_{it,t+1}\}$.

Note that we have written the self-selection constraint in a slightly different form than in our earlier analysis. We define \hat{L}_{1t} as the labor that the more able would have to supply, in order to imitate the less able (recall that we are focusing on the case when it is the more able's self-selection constraint that is binding). Clearly, $\hat{L}_{1t} = L_{2t}\nu$.

We obtain from the first-order conditions for C_{it} and L_{it} that

$$\frac{\partial V_t^i / \partial C_{it,t}}{\partial V_t^i / \partial C_{it,t+1}} = \frac{\gamma_t}{\gamma_{t+1}},$$

i.e. the marginal rate of substitution must equal the ratio of the shadow prices of resources; but differentiating the Lagrangian with respect to K , we obtain:

$$\gamma_{t+1}(1 + F_K(K_{t+1}, E_{t+1})) = \gamma_t,$$

from which the indicated result immediately follows.

instance, if unskilled labor and capital are substitutes, while skilled labor and capital are complements) then the government may wish to discourage capital accumulation, and it will use tax policy to do this.⁴⁶

(3) If the government cannot completely control the level of capital stock, through debt or social security policy,⁴⁷ as seems to be the case, then the government will, in general, wish to use tax policy, e.g. interest income subsidies, if it wishes to encourage savings, taxes if it wishes to discourage accumulation.

12.2. Infinitely lived individuals

The other model which has been intensively investigated is that of infinitely lived individuals [Judd (1985), Chamley (1980), Stiglitz (1985b)]. The first two of these studies are extensions, in this context, of the standard Ramsey analysis, and the same criticisms can be raised: if individuals are identical, there is no reason not simply to impose a uniform lump sum tax. Still, their result that *asymptotically, there should be no interest income tax*, is of interest.⁴⁸

⁴⁶The analysis follows by writing the constant returns to scale production function $Q = F(K, E_1, E_2)$ where $E_i = L_i N_i$, with the two types of laborers being imperfect substitutes. With this modification, the Lagrangian remains the same as that presented in footnote 45 except that now, we note that, ν , relative wages, are a function of K . This means that the first-order condition for K now becomes:

$$\gamma_{t+1}(1 + F_K) = \gamma_t + \lambda_{2,t+1} \frac{\partial U^2}{\partial L_{it}} (C_{1t+1,t+1}, C_{1t+1,t+2}, L_{it+1}) \frac{\partial \nu}{\partial K}.$$

The other first-order conditions remain unchanged. It immediately follows that the marginal rate of substitution is now not equal to the marginal rate of transformation.

⁴⁷See Atkinson and Stiglitz (1980) for a discussion of how this could be done.

⁴⁸The result is in one sense stronger than the earlier results for the overlapping generation model, and in one sense weaker. It is stronger, in that it does not require separability; it is weaker in that it is only valid asymptotically. (The earlier results characterized all Pareto efficient tax structures, and held at each moment of time.)

The simplest way to see these results is to write the Lagrangian

$$\begin{aligned} \mathcal{L} = & \sum U_t \delta^t + \beta_t [U_{C_t} - \delta(1 + r_t)U_{C_{t+1}}] \text{ (f.o.c.)} \\ & + \gamma_t [F(K_t, L_t) - (C_t + K_{t+1} - K_t) - R_t] \text{ (resource constraint)} \\ & + \phi [w_0 L_0 - C_0 + \sum (w_t L_t - \sum C_{it}) / \pi(1 + r_t)] \text{ (budget constraint)} \end{aligned}$$

where we have embedded the individual's first-order condition for allocating income among different periods; δ is the pure rate of time preference and where ϕ is the Lagrange multiplier associated with

But, as in the Atkinson–Stiglitz analysis, as soon as there are general equilibrium effects of taxation, this result no longer holds; that is to say, if there are two types of individuals, and the government cannot offset the changes in relative wages induced by tax policy, then it will wish to take these into consideration; even asymptotically, an interest income tax or subsidy will, in general, be desirable.⁴⁹

the individual's budget constraint. The first-order conditions are thus

$$\frac{\partial \mathcal{L}}{\partial C_t} = \delta' U_{C_t} - \gamma_t + \beta_t U_{C C_t} - \beta_{t+1} \delta(1+r_t) U_{C C_{t+1}} - \frac{\phi}{\pi(1+r_t)} = 0,$$

$$\frac{\partial \mathcal{L}}{\partial K_t} = \gamma_t(1+F_K) - \gamma_{t-1} = 0.$$

In the steady state,

$$\delta(1+r_t) = 1, \quad U_{C_t} = U_{C_{t+1}}, \quad \beta_t = \beta_{t+1}$$

so

$$\frac{\gamma_{t+1}}{\gamma_t} = \frac{1}{1+r_t} = \frac{1}{1+F_K}.$$

The consumer rate of interest converges to the producer rate of interest (the marginal rate of transformation); there is no interest income taxation.

⁴⁹This result is, in some sense, more surprising than the earlier result, because it implies that, viewed as of time 0, the ratio of consumer and producer prices approaches (asymptotically) either zero or infinity, although each price itself is also approaching zero.

To derive this result, we assume the government must impose uniform taxation on wages and interest. This imposes two additional sets of constraints:

$$1+r_{1t} = 1+r_{2t}$$

and

$$\frac{w_2}{w_1} = \frac{F_{L2}}{F_{L1}} = \nu(K, E_1, E_2).$$

We reformulate the Lagrangian, appending these two additional constraints. Letting m_t and ψ_t be the Lagrange multipliers associated with the two constraints, the derivative of the Lagrangian with respect to K is now

$$\gamma_{t+1}(1+F_K) = \gamma_t + \psi_t \nu_K,$$

and it immediately follows that the government will wish to take into account the effect of a change in capital accumulation on the distribution of before-tax income. If capital and unskilled labor are substitutes, then the government will wish to discourage the accumulation of capital, and will thus impose an interest income tax.

13. Imperfect capital markets

Throughout this discussion (as in most of the rest of the literature) we have assumed a perfect capital market, i.e. that all investors obtain the same return. There is considerable evidence that this is not true. The reasons for this are not hard to find: there is imperfect information concerning who are good investors.

In an imperfect capital market, then, individuals will be characterized not only by their labor productivity, but also by their capital productivity.⁵⁰ We then will wish to use self-selection mechanisms not only to determine who among the population are more productive laborers but also who are more productive investors. Assume, for simplicity, that individuals differed only in their investment productivity. The results [Stiglitz (1985b)] are parallel to those obtained earlier: the government will wish to impose an interest income tax on the less able investors, but no interest income tax on the more able.

14. Altruism and inheritance taxation

In the life-cycle model, it is assumed that individuals do not care at all for their children. One interpretation of the model with infinitely lived individuals is that there is a dynastic utility function: the parent cares about his child, the child about his child, and so on.

In the simplest form, we can assume that the utility of the t th generation is just

$$U_t = U(C_t, L_t) + \alpha U_{t+1},$$

where α is the weight assigned to the utility of the children. Then, successive substitution yields the utility function facing the current generation:

$$U_t = \sum_T \alpha^T U(C_{t+T}, L_{t+T}).$$

In this interpretation the only savings are bequests, and hence a tax on savings is equivalent to a tax on bequests.

The analysis of Pareto efficient taxation is little affected by the introduction of these interdependencies in utilities: we are assumed to keep the utility levels of all but one type of individual fixed. Thus, if we fix the utility of all future

⁵⁰In a perfect capital market, the most productive investors will invest all of society's resources, and thus the return to capital will be the same for all.

generations, then maximizing the utility of the current generation is equivalent to maximizing the utility he derives from his direct consumption, which is precisely the problem we have previously solved.

To answer, however, the question about whether we wish to impose an inheritance tax, we must first answer some difficult questions concerning the formulation of the social welfare function. If each generation's utility – including their valuation of future generation's utility – enters the social welfare function, we obtain:

$$W = \sum_t U_t \delta^t = \sum_t \frac{\alpha^t ((\delta/\alpha)^{t+1} - 1)}{\delta/\alpha - 1} U(C_t, L_t),$$

where we have weighted future generation's total utility by a factor δ . Notice now that there is a discrepancy between the way the government values future utility and how individuals do. *Since giving increases the utility both of the giver and the receiver, it is doubly blessed in our social welfare function, and the government will seek to encourage it, through a bequest subsidy.*

15. Commitment

In the intertemporal problems we have analyzed so far, we have assumed that the government can commit itself to a tax structure. The government will, however, in general be able to identify who the more able are from actions that they take early in life. Thus, if it were not committed to a given tax structure, the government would use this information to impose a lump-sum tax on such individuals later on in their life. Recently, Britto, Hamilton, Slutsky and Stiglitz (1986) have investigated the consequences of assuming that the government cannot commit itself. Then individuals know that if they reveal who they are early in their life, they will be confronted with a lump-sum tax. The government must take this into account in designing its tax structure. Britto et al. show that in general, the tax structure is characterized by three phases: in the first, there is a pooling equilibrium, i.e. all individuals are treated identically; then there is a period in which separation occurs, i.e. there is distortionary taxation; after that, the government imposes a lump-sum tax. Because individuals know that there will be a lump-sum tax later, it is harder to induce the more able to reveal (through their actions) who they are; the effective distortion in the one period in which the identification occurs is greater. Indeed, in some cases, where discount rates are low enough, with infinitely lived individuals, the only feasible equilibria

may entail pooling. (When there are many groups, there will be partially separating/partially pooling equilibria.)

Part III: Concluding comments

16. Optimal taxation in non-neoclassical economies

The analysis of optimal tax structures prior to 1980 focused on models of competitive (neoclassical) economies. Pigou (1947) had earlier emphasized the use of taxation to correct market distortions (such as externalities). The optimal tax structure thus might be markedly different in an economy with monopolies, externalities, and other imperfections. One of the important lessons to emerge from the literature on optimal taxation was the important consequences of interdependencies in demand; corrective taxation needs to take these interdependencies into account as well. Arnott and Stiglitz (1986) provide an example of how this may be done; economies in which there are moral hazard problems, i.e. in which the provision of insurance affects the probability of the occurrence of the insured-against event, are almost always constrained Pareto inefficient;⁵¹ there always exist a set of taxes and subsidies which can make everyone better off. They analyze the structure of the optimal set of corrective taxes.

Similarly, when there are monopolies, expressions for the optimal tax can be derived which can be thought of as consisting of two parts; in addition to the standard terms, there are those correcting the monopoly distortions.

Less developed economies

We have also emphasized throughout our analysis the importance of identifying the set of admissible tax instruments. This is likely to depend on the structure of the economy; it is much easier to monitor many transactions in developed economies than in less developed countries. Thus, few LDCs have an effective income tax, and the taxation of wage income in the rural sector appears to be a virtual impossibility. Accordingly, the analysis of the design of tax structures in LDCs differs in two important respects from the analysis we have presented earlier: (a) the set of tax instruments is different; and (b) there may be important distortions in the economy, e.g. wages may be set at above market-clearing levels

⁵¹The result on the constrained Pareto inefficiency of the economy with imperfect information and incomplete risk markets is in fact more general. See Greenwald and Stiglitz (1986).

in the urban sector.⁵² In a series of papers, Braverman, Sah, and Stiglitz⁵³ have investigated the structure of optimal taxes and prices under a variety of assumptions concerning the structure of the LDC economy. In their analyses they have attempted to characterize Pareto improvements, policies which would make everyone better off, as well as policies which would increase social welfare, under weak assumptions concerning the structure of the social welfare function.

17. Concluding remarks

The literature which we have surveyed in this chapter has, I believe, cast problems of the design of tax structure into a new perspective. It has emphasized the informational limitations on the government. And it has attempted to ascertain what economic theory can say about the design of tax structures without imposing welfare judgments, i.e. to identify *Pareto efficient* tax structures. As in so many other areas of economic analysis, there are perhaps more negative results than positive ones. Which taxes are not admissible has been shown to be a crucial determinant of the rates at which the remaining, feasible taxes should be set. Thus, a critical difference between Ramsey's earlier analysis and more recent analyses, such as that of Diamond and Mirrlees, is that Ramsey assumed that there were no profits taxes, while Diamond and Mirrlees assumed that there were no profits (or, equivalently, profits were taxed at 100 percent). Neither is a good assumption for most countries. And both analyses assumed the absence of an income tax.

Though traditional arguments against commodity and interest income taxation, based on counting the number of distortions, are not persuasive, the theory of self-selection shows that the nature of commodity taxation should depend simply on how the marginal rate of substitution between two commodities is affected by the amount of leisure; in the central case of separability, no commodity taxes should be imposed.

Although this analysis has shown that the widely held view that in the second-best world nothing can be said is incorrect – we have obtained some quite general qualitative propositions – we have also shown that those qualitative results are, in turn, quite dependent on the precise assumptions concerning the set of admissible taxes. The theory itself has, however, had little to say about the determinants of the set of admissible taxes.

⁵²This may not, however, be simply because of unions or other institutional considerations. It may be because the productivity of workers depends on the wage they receive, and thus firms find it profit maximizing to set wages at above market clearing levels. [See Stiglitz (1974, 1982a).]

⁵³See in particular, Braverman, Sah and Stiglitz (1982), Sah and Stiglitz (1984, 1986, 1987), and Braverman, Ahn and Hammer (1983). The latter study entails an empirical investigation of the consequences of different tax structures for the Korean economy. Other studies investigating the consequences of different tax structures for other countries, using the same basic framework, are presently underway at the World Bank.

Though optimal consumption tax structures are, in general, non-linear, the presence of non-linearities increases the administrative problems associated with the tax system. Whether the gains are worth the cost remains unproven. The suggestion contained in some of the numerical calculations, namely that the overall welfare gains from optimal taxation (using a utilitarian welfare criterion) are small and that the optimal tax structure may be close to linear, indicates that it may not be unreasonable to focus attention on linear tax structures. Moreover, Pareto efficient non-linear structures have some properties that may make them politically unacceptable: the marginal tax rate on the highest income individuals is negative, except in the limiting case where the elasticity of substitution between laborers of different skills is infinite, in which case the marginal tax rate on the highest income individual should be zero.

The focus on the incentive effects of taxation has, however, provided key insights: it has enabled us to identify the parameters which determine the magnitude of the deadweight losses associated with any tax system. The trade-offs on which this literature has focused in the design of the tax structure are at the core of many of the recent policy debates. Negative income tax proposals typically increase marginal tax rates for some individuals while reducing it for others, and an assessment of these proposals requires a balancing of the efficiency and equity effects on these different groups.

On the other hand, many of the central policy issues transcend these models, and the disparity between these may provide suggestions for future directions for research. We have already referred to one set of such issues: those related to administrative problems. The information problems associated with monitoring and assessing income to capital may be no less important than those with which we have been concerned here. The view that by encouraging savings and investment growth will be enhanced and thus all individuals, including the poor, will be better off, is partly reflected in our analysis of the general equilibrium effects of taxation, but only partly so. In particular, we have paid little attention to *R&D*, to entrepreneurship and risk-taking, or to the consequences of imperfect capital markets.

Though the work discussed here suggests that there are distinct advantages to imposing differential tax rates on different groups, whose labor supply characteristics differ (and that ex ante expected welfare may even be increased by imposing differential tax rates on groups whose characteristics do not differ), popular views of fairness seem inconsistent with these results produced by the utilitarian calculus.⁵⁴

Though utilitarianism has often been thought to provide a basis for choosing among alternative Pareto efficient tax structures, to do so requires choosing a

⁵⁴We have, moreover, consistently ignored any political economy considerations. These might suggest that constitutional restrictions on the set of admissible taxes may be welfare enhancing, for instance when the rates and terms at which those taxes are imposed are determined by majority voting.

particular cardinalization, and no persuasive basis for choosing among alternative cardinalizations has been provided.⁵⁵ Moreover, most analyses, while assuming that individuals differ in their productivity in the market place, ignore differences in home productivity. When these are taken into account, the case for redistributive taxation becomes even weaker.

Thus, while the original goal of this line of research, which was to provide a “scientific” basis for arguing for a progressive tax structure and for analyzing the design of tax systems, has not been achieved – and does not seem achievable – important insights have been gleaned, which should enable governments to make better choices of tax policies in the future.

References

- Allen, F., 1982, Optimal linear income taxation with general equilibrium effects on wages, *Journal of Public Economics* 17, 135–143.
- Arnott, R. and J.E. Stiglitz, 1985, Randomization with asymmetric information: A simplified exposition, Institute for Economics Research Discussion Paper No. 594 (Queen's University).
- Arnott, R. and J.E. Stiglitz, 1986, Moral hazard and optimal commodity taxation, *Journal of Public Economics*.
- Atkinson, A., 1972, Maxi min and optimal income taxation, paper presented at the Budapest Meeting of the Econometric Society.
- Atkinson, A., 1973, How progressive should income tax be?, in: M. Parkin and A.R. Nobay, eds., *Essays in modern economics* (Longman, London).
- Atkinson, A., 1977, Optimal taxation and the direct versus indirect tax controversy, *Canadian Journal of Economics* 10, 590–606.
- Atkinson, A. and A. Sandmo, 1977, The welfare implications of personal income and consumption taxes, *The Economic Journal*.
- Atkinson, A. and J.E. Stiglitz, 1972, The structure of indirect taxation and economic efficiency, *Journal of Public Economics* 1, 97–119.
- Atkinson, A. and J.E. Stiglitz, 1976, The design of tax structure: Direct versus indirect taxation, *Journal of Public Economics* 6, 55–75.
- Atkinson, A. and J.E. Stiglitz, 1980, *Lectures in public economics* (McGraw-Hill, New York).
- Azariadis, C. and J.E. Stiglitz, 1983, Implicit contracts and fixed price equilibria, *Quarterly Journal of Economics* (Supplement) 1–22.
- Baumol, W.J. and D.F. Bradford, 1970, Optimal departures from marginal cost pricing, *American Economic Review* 60, 265–283.
- Berry, B., 1973, *The liberal theory of justice* (Oxford).
- Blum, W. and H. Kalven, 1963, *The uneasy case for progressive taxation* (University of Chicago Press, Chicago).
- Boiteux, M., 1956, On the management of public monopolies subject to budgetary constraints, *Journal of Economic Theory* 3, 219–240.

⁵⁵We have also not entered into the debate concerning the appropriateness of the utilitarian approach on more fundamental philosophical grounds. We have noted Rawls' views suggesting that utilitarianism is insufficiently equalitarian. Brian Berry, while disagreeing with much of Rawls' work, remarks, "... Rawls, in taking utilitarianism as the main rival to his own account, is flogging an almost dead stalking-horse."

One can perhaps best view the utilitarian calculus as simply providing a convenient way of summarizing the effects of a policy change [see Stiglitz (1986a)].

Notice that our results on *Pareto efficient tax structures* are not liable to most of the criticisms leveled against the utilitarian approach.

- Boskin, M. and E. Sheshinski, 1978, Optimal income redistribution when individual welfare depends on relative income, *Quarterly Journal of Economics*, 92, 589–602.
- Braverman, A., C.Y. Ahn and J.S. Hammer, 1983, Government deficit reduction and alternative agricultural pricing policies in Korea, unprocessed mimeo (The World Bank, Washington, D.C.).
- Braverman, A., R. Sah and J.E. Stiglitz, 1982, The town-versus-country problem: Optimal pricing in an agrarian economy, presented to a conference at the World Bank.
- Brito, D. and W.H. Oakland, 1977, Some properties of the optimal linear tax, *International Economic Review*, 407–424.
- Britto, D., J. Hamilton, S. Slutsky and J.E. Stiglitz, 1986, Information and multi-period optimal taxation with self-selection, *Mimeo* (Princeton University).
- Carruth, A.A., 1982, On the role of the production and consumption assumptions for optimum taxation, *Journal of Public Economics*, 17, 145–156.
- Chamley, Christophe, 1980, Optimal intertemporal taxation and the public debt, *Cowles Foundation Discussion Paper* 554.
- Cooter, R. and E. Helpman, 1974, Optimal income taxation for transfer payments under different social welfare criteria, *Quarterly Journal of Economics*, 88, 656–670.
- Corlett, W.J. and D.C. Hague, 1953, Complementarity and the excess burden of taxation, *Review of Economics Studies* 10, 295–336.
- Dasgupta, P. and J.E. Stiglitz, 1974, Benefit-cost analysis and trade policies, *Journal of Political Economy* 82, 1–33.
- Deaton, A.S., 1979, Optimal taxes and the structure of preferences, *Economic Letters*.
- Deaton, A. and N. Stern, 1986, Optimally uniform commodity taxes, taste differences, and lump-sum grants, *Economics Letters*, 20, 263–266.
- Diamond, P.A., 1975, A many-person Ransley Tax Rule, *Journal of Public Economics* 4, 283–299.
- Diamond, P., L. Helms and J. Mirrlees, 1980, Optimal taxation in a stochastic economy: A Cobb–Douglas example, *Journal of Public Economics*, 14, 1–29.
- Diamond, P.A. and J.A. Mirrlees, 1971, Optimal taxation and public production I–II, *American Economic Review* 61, 8–27, 261–278.
- Dixit, A.K. and A. Sandmo, 1977, Some simplified formulae for optimal income taxation, *Scandinavian Journal of Economics* 79, 417–423.
- Eaton, J. and H. Rosen, 1980, Labor supply, uncertainty, and efficient taxation, *Journal of Public Economics* 14, 365–374.
- Edgeworth, F.Y., 1868, *Mathematical physics*.
- Edgeworth, F.Y., 1897, The pure theory of taxation, *Economic Journal* 7, 46–70, 226–238 and 550–571 (reprinted in Edgeworth, 1925).
- Fair, R., 1971, The optimal distribution of income, *Quarterly Journal of Economics*, 85, 551–579.
- Feldstein, M.S., 1973, On the optimal progressivity of the income tax, *Journal of Public Economics* 2, 357–376.
- Fisher, I., 1937, Income in theory and income taxation in practice, *Econometrica* 5, 1–55.
- Greenwald, B. and J.E. Stiglitz, 1986, Externalities in economies with imperfect information and incomplete markets, *Quarterly Journal of Economics*, May, 229–264.
- Guesnerie, R., 1975, Production of the public sector and taxation in a simple second best model, *Journal of Economic Theory* 10, 127–156.
- Guesnerie, R. and J. Seade, 1982, Nonlinear pricing in a finite economy, *Journal of Public Economics*, 17, 157–180.
- Hall, R.E. and A. Rabushka, 1983, *Low tax, simple tax, flat tax* (McGraw-Hill, New York).
- Helpman, E. and E. Sadka, 1978, The optimal income tax: Some comparative statics results, *Journal of Public Economics*, 9, 383–394.
- Judd, K.L., 1985, Redistributive taxation in a simple perfect foresight model, *Journal of Public Economics* 28, 59–83.
- Kaldor, N., 1955, *An expenditure tax* (Allen and Unwin, London).
- Lerner, A.P., 1944, *The economics of control* (Macmillan, New York).
- Meade, J.E., 1955, *Trade and Welfare: Mathematical Supplement* (Oxford University Press, Oxford).
- Mirrlees, J., 1971, An exploration in the theory of optimum income taxation, *Review of Economics Studies* 38, 175–208.
- Mirrlees, J., 1976, Optimal tax theory: A synthesis, *Journal of Public Economics* 6, 327–358.

- Mirrlees, J. 1985, The theory of optimal taxation, in: K.J. Arrow and M.D. Intriligator, eds., *Handbook of Mathematical Economics* (Elsevier Science Publishers B.V.) vol. 3.
- Myerson, R.B., 1983, Incentive compatibility and the bargaining problem, *Econometrica*, 47, 175–208.
- Ordovery, J. and E. Phelps, 1979, The concept of optimal taxation in an overlapping generations model of capital and wealth, *Journal of Public Economics*, 12, 1–26.
- Phelps, E.S., 1973, The taxation of wage income for economic justice, *Quarterly Journal of Economics* 87, 331–354.
- Pigou, A.C., 1947, *A study in public finance*, 3rd ed. (Macmillan, London).
- Radner, R. and J.E. Stiglitz, 1984, A nonconcavity in the value of innovation, in: M. Boyer and R. Khilstrom, eds., *Bayesian models in economic theory* (Elsevier Science Publications).
- Ramsey, F.P., 1927, A contribution to the theory of taxation, *Economic Journal* 37, 47–61.
- Rawls, J., 1971, *A theory of justice* (Harvard University Press, Cambridge, Mass.).
- Rawls, J. 1974, Concepts of distributional equity: Some reasons for the maximum criterion, *American Economic Review, Papers and Proceedings*, 64, 141–146.
- Rothschild, M. and J.E. Stiglitz, 1976, Equilibrium in competitive insurance markets, *Quarterly Journal of Economics* 90, 629–650.
- Sadka, E., 1976, On income distribution, incentive effects and optimal income taxation, *Review of Economic Studies* 43, 261–268.
- Sah, R. and J.E. Stiglitz, 1984, The economics of price scissors, *American Economic Review* 74, 125–138.
- Sah, R. and J.E. Stiglitz, 1986, Taxation and pricing of agricultural and industrial goods in developing economies, forthcoming in: D. Newbery and N. Stern, eds., *Modern tax theory for developing countries*.
- Sah, R. and J.E. Stiglitz, 1987, Price scissors and the structure of the economy, *Quarterly Journal of Economics*.
- Seade, J.K., 1977, On the shape of optimal tax schedules, *Journal of Public Economics* 7, 203–236.
- Sheshinski, E., 1972, The optimal linear income tax, *Review of Economic Studies*, 39 297–302.
- Slemrod, J. and S. Yitzhaki, 1983, On choosing a flat-rate income tax system, *National Tax Journal*, 36, 31–44.
- Slemrod, J., Do we know how progressive the income tax should be?, *National Tax Journal*, 36, 361–369.
- Stern, N., 1976, On the specification of models of optimum income taxation, *Journal of Public Economics* 6, 123–162.
- Stern, N., 1982, Optimum taxation with errors in administration, *Journal of Public Economics*, 1982, 181–211.
- Stiglitz, J.E., 1974, Alternative theories of wage determination and unemployment in L.D.C.'s: The labor turnover model, *Quarterly Journal of Economics*, LXXXVII, 194–227.
- Stiglitz, J.E., 1975, Information and economic analysis, in: Parkin and Nobay, eds., *Current economic problems* (Cambridge University Press, Cambridge) 27–52.
- Stiglitz, J.E., 1976a, Prices and queues as screening devices in competitive markets, IMSSS Technical Report No. 212 (Stanford University).
- Stiglitz, J.E., 1976b, Simple formulae for the measurement of inequality and the optimal linear income tax, IMSSS Technical Report No. 215 (Stanford University).
- Stiglitz, J.E., 1977, Monopoly, non-linear pricing and imperfect information: The insurance market, *Review of Economic Studies*, 44, 407–430.
- Stiglitz, J.E., 1982a, Alternative theories of wage determination and unemployment: The efficiency wage model, in: M. Gersovitz, C.F. Diaz-Alejandro, G. Ranis and M. Rosenzweig, eds., *The theory and experience of economic development: Essays in honor of Sir Arthur W. Lewis* (George Allen & Unwin, London) 78–106.
- Stiglitz, J.E., 1982b, Self-selection and Pareto efficient taxation, *Journal of Public Economics* 17, 213–240.
- Stiglitz, J.E., 1982c, Utilitarianism and horizontal equity: The case for random taxation, *Journal of Public Economics* 18, 1–33.
- Stiglitz, J.E., 1983, Some aspects of the taxation of capital gains, *Journal of Public Economics* 21, 257–294.
- Stiglitz, J.E., 1985a, The general theory of tax avoidance, *National Tax Journal* 38, 325–338.

- Stiglitz, J.E., 1985b, Inequality and capital taxation, IMSSS Technical Report No. 457 (Stanford University).
- Stiglitz, J.E., 1986a, Economics of the public sector (W.W. Norton, New York).
- Stiglitz, J.E., 1986b, Theories of wage rigidities, in: J.L. Butlerewicz, K.J. Koford and J.B. Hiller, eds., Keynes' economic legacy (Preager Publishers, New York).
- Stiglitz, J.E. [forthcoming], The wage-productivity hypothesis: Its economic consequences and policy implications, paper presented to the American Economic Association, December 1982 and forthcoming in: M. Boskin, ed., Essays in honor of A. Harberger.
- Stiglitz, J.E. and P. Dasgupta, 1971, Differential taxation, public goods, and economics efficiency, *Review of Economic Studies* 38, 151-197.
- Varian, H., 1980, Redistributive taxation as social insurance, *Journal of Public Economics* 14, 49.