

SELF-SELECTION AND PARETO EFFICIENT TAXATION

Joseph E. Stiglitz

Econometric Research Program
Research Memorandum No. 290

September 1981

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, New Jersey

ABSTRACT

Self-Selection and Pareto Efficient Taxation

by

Joseph E. Stiglitz

This paper analyzes the set of Pareto efficient tax structures. The formulation of the problem as one of self-selection not only shows more clearly the similarity between this problem and a number of other problems (such as the optimal pricing of a monopolist) which have recently been the subject of extensive research, but also allows the derivation of a number of new results. We establish (i) under fairly weak conditions, randomization of tax structures is desirable; (ii) if different individuals are not perfect substitutes for one another in production, then the general equilibrium effects -- until now largely ignored in the literature -- of changes in the tax structure may be dominant in determining the optimal tax structure; in particular if the relative wage of high ability and low ability individuals depends on the relative supplies of labor, the optimal tax structure entails a negative marginal tax rate on the high ability individuals, and a positive marginal tax rate on the low ability individuals (the magnitude of which depends on the elasticity of substitution); (iii) if individuals differ in their preferences, Pareto efficient taxation may entail negative marginal tax rates for high incomes; while (iv) if wage income is stochastic, the marginal tax rate at the upper end may be 100%.

Our analysis thus makes clear that the main qualitative properties of the optimal tax structure to which earlier studies called attention are not robust to these attempts to make the theory more realistic.

Department of Economics
Dickinson Hall
Princeton University

Self-Selection and Pareto Efficient Taxation

Joseph E. Stiglitz

1. Introduction

It is now widely recognized that the optimal income tax problem is one of a number of closely related problems, in which one agent (a government, a monopolist, a firm) attempts to differentiate among ("screen") a set of other agents. It does this by means of a self-selection mechanism; it confronts individuals with a set of choices, and individuals with different characteristics (preferences) make different selections from the set. Their choices thus reveal information about their characteristics. Although the discrimination may be perfect, it will not in general be costless; to induce self selection requires structuring the choice set in such a way that the conventional efficiency conditions (e.g. equating marginal rates of substitution) will not be satisfied. The problem of the government (the monopolist, the employer, etc.) is to design "efficient" self-selection mechanisms; to put it somewhat loosely, they seek to structure the choice sets to reveal the desired information at the minimum cost.

In this paper, we explicitly formulate the optimal tax problem as one of self-selection. The government would like to differentiate between low ability and high ability individuals. If it could identify them costlessly, it would impose differential lump sum taxes. It

* Financial support from the National Science Foundation is gratefully acknowledged. I am indebted to F. Allen, D. Pierce, R. Lindsey, and R. Arnott for helpful comments. An earlier version of this paper was presented at the NBER-CEME Conference on Information and Game Theory, Northwestern University, October, 1980.

can, however, only observe differences in earned income. It seeks a tax structure which leads the more able to reveal that they are more able by earning a higher income (rather than pretending to be less able and enjoying more leisure). And it seeks to do this in the most efficient manner. Our formulation not only allows us so see more clearly the similarity between this problem and a number of other problems which have recently been the subject of extensive research, but it also allows us to generalize the conventional results, enabling us to show clearly that most of the qualitative properties that have been derived are properties not only of utilitarian tax structures (of the kind studied, e.g. by Mirrlees (1971) and Atkinson-Stiglitz (1980)), but of any pareto optimal tax structure.

Moreover, we are able to provide a new, and we think clearer, interpretation of the result (Atkinson-Stiglitz (1976)) that, with an optimal income tax, if the utility function is separable between leisure and consumption commodities, then there should be no commodity taxes. For self-selection mechanisms to work, the individuals must have different indifference curves. We show that the condition of separability is equivalent to the condition that the indifference curves (between say commodity 1 and commodity 2) are identical.

Finally, and perhaps most important, we are able to derive several new results.

First, in the literature on self-selection, it has been shown that randomization may serve as an effective screening device (Stiglitz (forthcoming)). High ability individuals always have the alternative of working less and enjoying a lower level of consumption. The tax structure must be designed in such a way that the high ability individuals are willing to "disclose" their ability by earning higher incomes. If high ability individuals are more risk averse than low ability individuals (in a sense to be defined precisely in the paper), by randomizing the taxes imposed on low ability individuals, the high-leisure, low consumption alternative of pretending to be a low ability individual becomes less attractive. The low ability individuals, if they are risk averse, obviously are worse off as a result of the randomization; but the ability to differentiate between high and low ability more easily may allow us to lower the average tax rate imposed on the low ability individuals; and under certain circumstances, we can lower it enough that they are no worse off. Perhaps more striking, we can show that we can do this at the same time as raising total revenue. Thus, this analysis extends the earlier results of Atkinson and Stiglitz (1976) and Stiglitz (1976) on random taxation to show that randomization may be desirable for a much less restricted set of tax structures than was considered in those papers (earlier analyses were essentially confined to linear tax structures).

The second major set of new results relate to extending optimal income taxation to a simple general equilibrium model.¹ Most of the

¹ After this paper was finished, my attention was called to Section 3 of N. Stern's paper, "Optimum Taxation with Errors in Administration," where some similar results are derived.

earlier literature limited itself to analyzing the optimal income tax under the assumption that individuals' relative before tax wages were exogenously determined. The individuals were perfect substitutes for one another. Recently, F. Allen (1980) has shown that such results may be very misleading. He examined optimal linear income taxes, in a two class model in which the relative marginal productivities were endogenous. He showed, in particular that the general equilibrium effects may be dominant in determining the design of the tax structure. Indeed, under not implausible conditions, it was possible for the optimal tax structure to be regressive, even for a Rawlsian social welfare objective function.

This paper extends his results by considering optimal tax structures (i.e. we do not restrict ourselves to linear tax structures) in the simplest possible general equilibrium model. We obtain two important results:

- a) The widely discussed property of the optimal tax structure, that the most able individual faces a zero marginal tax rate, is only true if all individuals are perfect substitutes; in all other cases, the highest ability individual should face a negative marginal tax rate.
- b) The tax which should be imposed on the less able individual depends on the elasticity of substitution between the two types of laborers, which determines the general equilibrium effects of taxation.

Previous analyses of optimal income tax structures have made two further restrictive assumptions (besides that all individuals are perfect substitutes in production): (a) that the preferences of all individuals are identical; and (b) that income is a deterministic function of

effort. We do not provide here a general characterization of the optimal tax structure with heterogenous individuals and stochastic income. But what we can show, using slight modifications of our basic two group model, is that either modification necessitates serious alteration in the optimal tax structure: in the former case, at the upper end the marginal rate is negative, while in the latter it is 100% (rather than zero, as in the conventional story).

2. Pareto Efficient Taxation: The Simplest Case

We begin our discussion with the simplest possible model, in which there are only two types of individuals, differing in ability but having the same utility function. (This, as we shall see, is not critical for most of the results we shall obtain.) The i th individual faces a before tax wage (output per hour) of w_i , and thus, in the absence of taxation, his budget constraint is simply

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

where

C_i = the i th individual's consumption

L_i = number of hours worked by i th individual

(L_i could equally well be interpreted as being effort.) Neither w_i nor L_i is separately observable, but¹

¹ The subsequent discussion will make it clear that our results are equally applicable to the more general specification

$$(2') \quad Y_i = \psi_i(L_i), \quad \psi_i' > 0, \quad \psi_i'' \leq 0.$$

(2') is a more appropriate specification for the interpretation of the model where "L" is investment in education.

(2) $Y_i = w_i L_i$, the i th individual's income

is observable. The i th individual receives utility from consuming goods, and disutility from work:

(3) $U^i = U^i(C_i, L_i)$

$$\frac{\partial U^i}{\partial C_i} > 0, \quad \frac{\partial U^i}{\partial L_i} < 0 \quad \text{and} \quad U \text{ is quasi-concave.}^1$$

Assume now the government imposes a tax as a function of income

(3) $T_i = T(Y_i)$.

The individual's consumption now is his income minus his tax payments

(4) $C_i = Y_i - T(Y_i)$.

The individual maximizes his utility subject to his budget constraint

(5) $\max_{\{C_i, L_i\}} U^i(C_i, L_i)$

$$\text{s.t. } C_i \leq w_i L_i - T(w_i L_i),$$

yielding the first order condition (assuming differentiability, etc.)

$$(6) \quad \frac{\partial U^i / \partial L_i}{\partial U^i / \partial C_i} = -w_i(1-T')$$

The LHS is the individual's **marginal rate** of substitution. The RHS is the after-tax marginal return to working an extra hour.

¹ In our analysis of randomized taxes, we make the stronger assumption that U is concave. Assumptions about concavity are obviously important in the analysis of optimal utilitarian tax structures (Stiglitz, 1976) but play no role in the analysis of Pareto efficient tax structures.

It will turn out in the sequel that the optimal tax structure (with a finite number of groups) is never differentiable. We shall refer to

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

as the marginal tax rate.¹

In many self-selection problems, it turns out to be useful to write the utility function in terms of the observable variables: Here we assume Y_i and T_i (and hence C_i) are the only observables. Hence, we write²

$$(7) \quad U = U^i(C_i, \frac{Y_i}{w_i}) \equiv V^i(C_i, Y_i; w_i)$$

For later reference, we note that

$$(7') \quad \frac{\partial V^i}{\partial C_i} = \frac{\partial U^i}{\partial C_i}, \quad \frac{\partial V^i}{\partial Y_i} = \frac{\partial U^i}{\partial L_i} \frac{1}{w_i}, \quad \frac{\partial V^i}{\partial w_i} = - \frac{\partial U^i}{\partial L_i} \frac{Y_i}{w_i^2} = - \frac{\partial V^i}{\partial Y_i} \frac{Y_i}{w_i}$$

¹ There 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 $Y = w_i L_i$.

² For simplicity, we shall often write $V^i(C_i, Y_i)$ rather than $V^i(C_i, Y_i; w_i)$.

$$-\frac{\partial v^i / \partial Y_i}{\partial v^i / \partial C_i} = 1 - T',$$

from which it follows, in the first best optimum, with only lump sum taxation,

$$-\frac{\partial v^i / \partial Y_i}{\partial v^i / \partial C_i} = 1.$$

Note that an increase in Y lowers utility, because to attain it, the individual must forgo more leisure; and it lowers the utility of the less able by more, since they must forgo more leisure (for a given increase in Y).

Even if all individuals have the same utility of consumption-and-leisure functions, their utility of consumption-and-before tax income functions will differ. In Figure 1, it is clear that individuals of higher ability have flatter indifference curves: the increase in consumption that is required to compensate an individual for a given increase in before tax income is smaller for the more able, since to obtain the given increase in before tax income he needs to forgo less leisure. In the subsequent discussion we shall assume that individual 2 is the more able individual.

Formulated that way, we can see that income will provide us with a basis of self-selection: individuals with different abilities will make different choices of (C, Y) pairs, since they have different indifference curves.

The problem of the government concerned with Pareto efficiency is to maximize the utility of, say, individuals of type 2, subject

to (a) individuals of type 1 having at least a given level of utility and (b) raising a given amount of revenue. It does this by offering two $\{C, Y\}$ packages, one of which will be chosen by the first group, the other of which will be chosen by the second group.¹

Formally, the government

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

$$(9) \text{ s.t. } v^1(C_1, Y_1) \geq \bar{U}^1$$

$$(10) \quad v^2(C_2, Y_2) \geq v^2(C_1, Y_1)$$

} the self-selection constraints

$$(11) \quad v^1(C_1, Y_1) \geq v^1(C_2, Y_2)$$

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

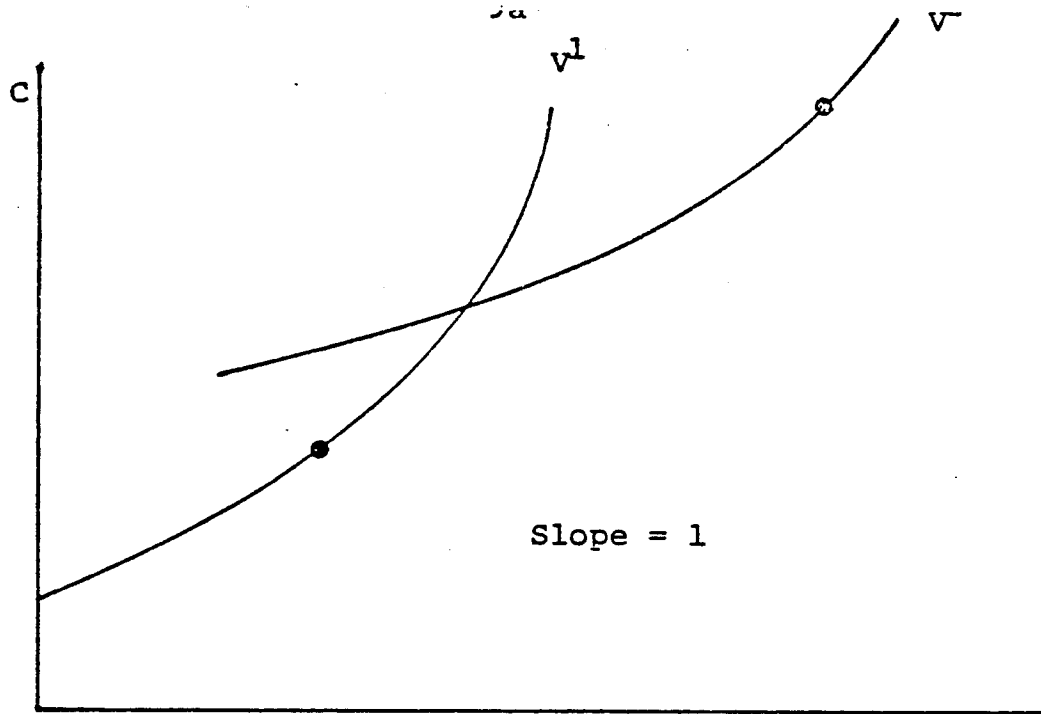
(where R is government revenue, \bar{R} is the revenue requirement, N_i the number of individuals of type i).²

The Lagrangian for this maximization problem may be written

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

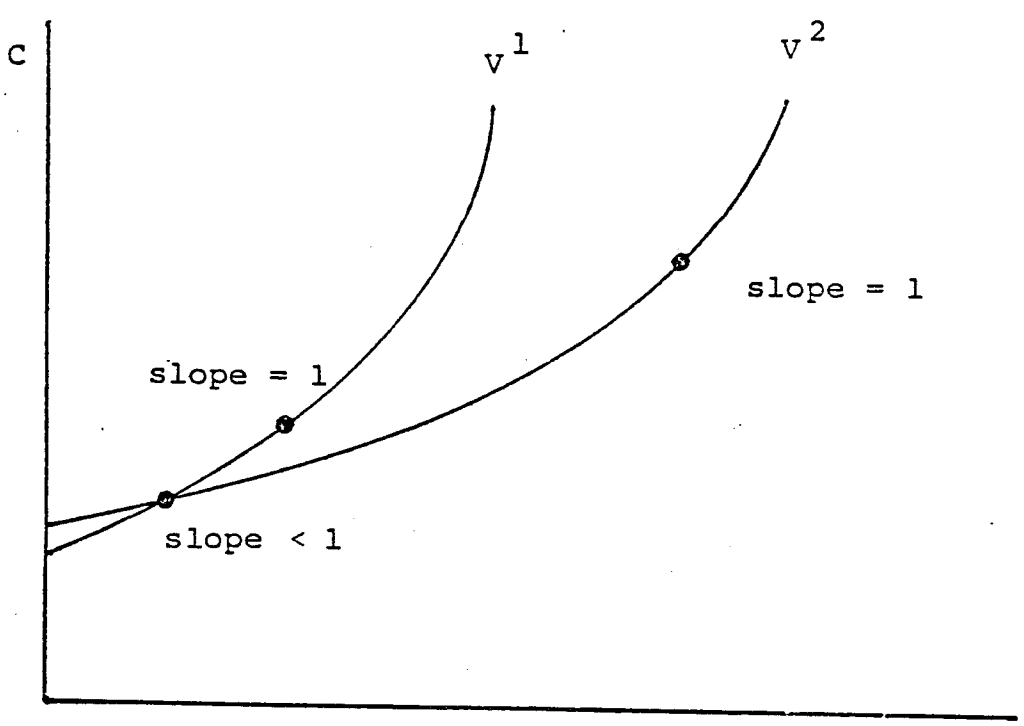
¹ Obviously, the government can offer a continuum of $\{C, Y\}$ packages (i.e. an entire tax function), but at most two will be chosen, and therefore we need be concerned with at most two.

² Notice 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 and subject to the self-selection constraints. The Lagrangian which we form to analyze the two problems is identical.



Y

Figure 1a
 First Best Taxation Fully Revealing



Y

Figure 1b
 First Best Taxation Not Fully Revealing: Pareto
 Optimal Taxation Entails Positive Marginal Tax
 Rate on Low Ability Zero Marginal Tax Rate on
 High Ability. High Ability (Type "2") Individuals
 Have Flatter Indifference Curves.

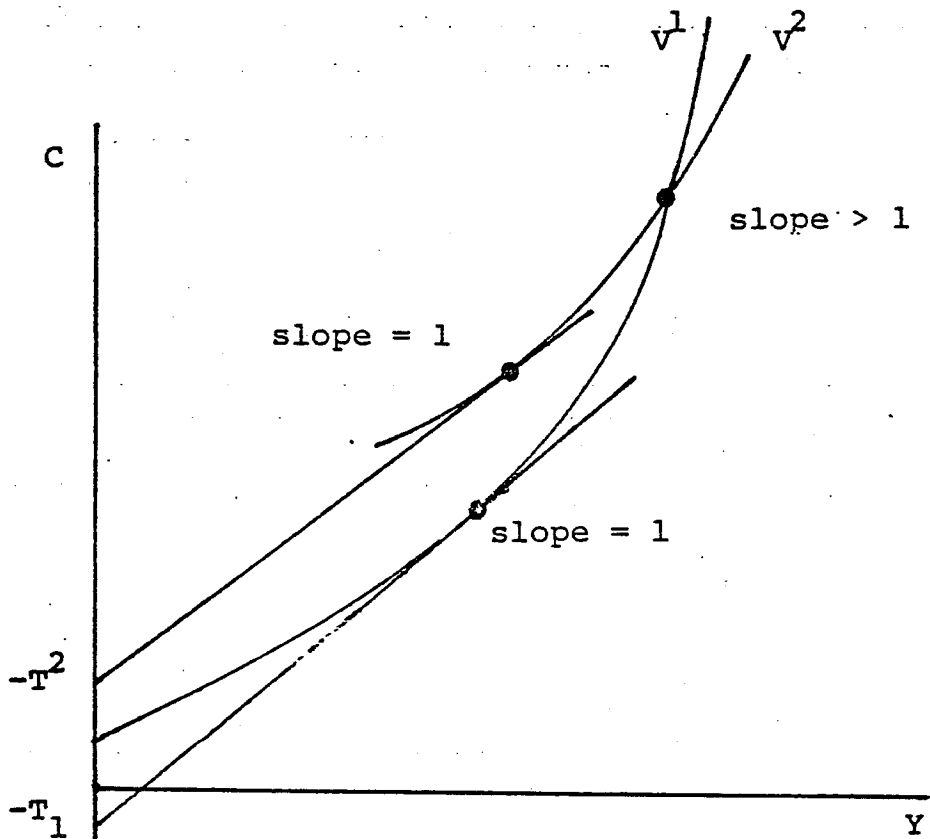


Figure 1c

First Best Taxation Not Fully Revealing
Pareto Optimal Taxation Involves Zero
Marginal Tax Rate on Low Ability, Negative
Marginal Tax Rate on High Ability.

The first order conditions for this problem are straightforward:

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

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

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

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

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 1a)
- (ii) $\lambda_1 = 0, \lambda_2 > 0$ (Figure 1b)
- (iii) $\lambda_2 = 0, \lambda_1 > 0$ (Figure 1c)

i.e. at most one of the two self-selection constraints is binding. Moreover, it is also easy to show that $\mu > 0$, 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 1a. With first best taxation, the equilibrium is fully revealing.

The "normal" case, on which most of the literature has focused, is that where $\lambda_1 = 0, \lambda_2 > 0$. 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. (See Arnott, Hosios, and Stiglitz (1980).) But more generally, the possibility that $\lambda_1 > 0, \lambda_2 = 0$ 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 1c.)

2.1 The Optimal Tax Structure with $\lambda_2 > 0, \lambda_1 = 0.$

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

$$(15a) \quad - \frac{\partial V^2 / \partial Y_2}{\partial V^2 / \partial C_2} = - \frac{\partial U^2 / \partial L_2}{\partial U / \partial C_2} \cdot \frac{1}{w_2} = 1,$$

the marginal tax rate faced by the more able individual is zero.

(This corresponds to the result noted earlier by Sadka (1976) and Phelps (1973).)

Dividing (14b) by (14a),

$$(15b) \quad - \frac{\partial V^1 / \partial Y_1}{\partial V^1 / \partial C_1} = \frac{1 - \lambda_2 (\partial V^2 / \partial Y_1) / N_1 \gamma}{1 + \lambda_2 (\partial V^2 / \partial C_1) / N_1 \gamma} < 1.$$

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 (15b) 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.¹

2.2 The Optimal Tax Structure with $\lambda_1 = 0, \lambda_2 > 0$.

Exactly the same kinds of arguments as used in Section 2.1 can be employed to establish that if $\lambda_1 = 0, \lambda_2 > 0$, the marginal tax rate faced by the less able individual is zero, while 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 paper, we focus our attention on the "normal" case with $\lambda_1 > 0, \lambda_2 = 0$.

2.3 Endogenous Wages. In the previous discussion, we assumed wages were fixed. It is easy to incorporate general equilibrium effects. Changes in $\{C_i, Y_i\}$ affect the wages. This not only has a direct effect on welfare, but also has an effect on the self-selection constraints which needs to be taken into account.

¹ This corresponds to the result noted earlier by Mirrlees for the case of a continuum of types.

Assume that output is a function of the supply of hours by each of the two types:

$$(16) \quad Q = F(N_1 L_1, N_2 L_2) = L_1 N_1 f\left(\frac{N_2 L_2}{N_1 L_1}\right)$$

where F exhibits constant returns to scale. If each factor receives its marginal product,¹

$$(16a) \quad w_2 = \frac{\partial F}{\partial (N_2 L_2)} = f'(n), \quad w_1 = \frac{\partial F}{\partial (N_1 L_1)} = f(n) - n f'(n)$$

where $n = N_2 L_2 / N_1 L_1$.

We can thus solve for n and hence wages as a function of Y_1 and Y_2 :²

$$(17) \quad \frac{Y_1}{Y_2} = \frac{w_1 L_1}{w_2 L_2} = \frac{f - n f'}{n f'} \frac{N_2}{N_1}$$

We then write $w_1(Y_1, Y_2)$, $w_2(Y_1, Y_2)$, substitute back into our Lagrangian and differentiate.

In this particular case, it turns out to be easier if we take as our control variables L_1 and L_2 . This necessitates

¹ We still assume the government cannot distinguish between the different kinds of labor (although, presumably, the firms who employ the two kinds of laborers can). This is a quite plausible assumption.

² (17) can be inverted provided the elasticity of substitution is not equal to unity.

a reformulation of our self-selection constraints. We must choose $\{L_1, C_1, L_2, C_2\}$ so that the more able do not wish to pretend to be less able. The labor input required of the more able to attain the same income (which, it should be recalled, is the only observable variable) as the less able is much smaller. As before, we let w_1 be the wage of the first group, Y_1 its income, and L_1 its labor input. Then for the second group to have the same income as the first group requires a labor input of

$$\hat{L}_2 = \frac{L_1 w_1}{w_2} = \frac{L_1 (f - n f')}{f'} \equiv L_1 \phi \left(\frac{L_2}{L_1} \right), \quad \phi' > 0.$$

As the ratio of L_2/L_1 increases, w_2/w_1 decreases, so the required labor input of individual two, L_2 , to obtain the same income one has, increases. Thus we can rewrite the self-selection constraints as

$$U^2(C_2, L_2) \geq U^2(C_1, L_1 \phi(\frac{L_2}{L_1}))$$

$$U^1(C_1, L_1) \geq U^1(C_2, L_2 / \phi(L_2/L_1)).$$

Writing the revenue constraint as

$$F(N_1 L_1, N_2 L_2) - N_1 C_1 - N_2 C_2 - \bar{R} \geq 0$$

we form the Lagrangian

$$\begin{aligned} \mathcal{L} = & U^2(C_2, L_2) + \mu U^1(C_1, L_1) + \gamma (F(N_1 L_1, N_2 L_2) - N_1 C_1 - N_2 C_2 - \bar{R}) \\ & + \lambda_2 (U^2(C_2, L_2) - U^2(C_1, L_1 \phi(\frac{L_2}{L_1}))) \\ & + \lambda_1 (U^1(C_1, L_1) - U^1(C_2, L_2 / \phi(L_2/L_1))). \end{aligned}$$

We obtain first order conditions analogous to those derived earlier (for the case $\lambda_1 = 0$, $\lambda_2 > 0$):

$$(18a) \quad \frac{\partial \mathcal{L}}{\partial C_1} = \mu \frac{\partial U^1}{\partial C_1} - \lambda_2 \frac{\partial U^2}{\partial C_1} - \gamma N_1 = 0,$$

$$(18b) \quad \frac{\partial \mathcal{L}}{\partial L_1} = \mu \frac{\partial U^1}{\partial L_1} - \lambda_2 \frac{\partial U^2 [C_1, L_1 \phi]}{\partial L_2} (\phi - \frac{L_2}{L_1} \phi') + \gamma F_1 N_1 = 0,$$

$$(18c) \quad \frac{\partial \mathcal{L}}{\partial C_2} = \frac{\partial U^2}{\partial C_2} + \lambda_2 \frac{\partial U^2 [C_2, L_2]}{\partial C_2} - \gamma N_2 = 0,$$

$$(18d) \quad \frac{\partial \mathcal{L}}{\partial L_2} = \frac{\partial U^2}{\partial L_2} + \lambda_2 (\frac{\partial U^2 [C_2, L_2]}{\partial L_2} - \frac{\partial U^2 [C_1, L_1 \phi]}{\partial L_2} \phi') + \gamma F_2 N_2 = 0.$$

Dividing (18d) by (18c), we obtain

$$-\frac{\partial U^2/\partial L_2}{\partial U^2/\partial C_2} = F_2 - \frac{\lambda_2}{\gamma N_2} \frac{\partial U^2}{\partial L} \phi' \geq F_2 \equiv w_2 \text{ as } \phi' \geq 0$$

If the two types of labor are not perfect substitutes, then the marginal tax rate on the most able individual should be negative.

Dividing (18b) by (18a) we obtain¹, denoting the elasticity of

substitution by σ

$$(19) \quad -\frac{\partial U^1/\partial L_1}{\partial U^1/\partial C_1} = \frac{F_1 + \frac{\lambda_2}{\gamma N_1} \frac{\partial U^2}{\partial C_1} \left(-\frac{\partial U^2/\partial L_1}{\partial U^2/\partial C_1} \phi\right) \left(1 - \frac{1}{\sigma}\right)}{1 + \lambda_2 \frac{\partial U^2}{\partial C_1} / \gamma N_1} <$$

$$\frac{F_1 - \lambda_2 \frac{\partial U^2}{\partial L_1} \phi / N_1 \gamma}{1 + \lambda_2 \frac{\partial U^2}{\partial C_1} / N_1 \gamma} < F_1$$

The first inequality follows from $\sigma \geq 0$, and the second follows by the same argument used to establish inequality (15b), (using (7')), (19) and (15) are in fact identical when $\sigma = \infty$).

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.

¹ $\frac{\phi' L_2/L_1}{\phi} = -\frac{d \ln w_1/w_2}{d \ln n_1/n_2} = \frac{-f''fn}{f'(f - nf')} \equiv 1/\sigma.$

2.4 Utilitarian Optimal Taxes

We have analyzed here Pareto efficient taxation.

Most of the earlier optimal tax literature assumed a much stronger objective function: the government wished to maximize a utilitarian objective function, i.e. in the present context, it

$$\max U^1 N_1 + U^2 N_2 \equiv W$$

subject to the self selection and revenue constraints. If we write down the Lagrangean expression for this problem, it is identical to (13), with one minor difference: while in (13), we specified \bar{U}^1 , and μ , the lagrange multiplier associated with the constraint, was one of the variables to be determined in the analysis; here it is as if we knew the value of the Lagrange multiplier ($\mu = N_1/N_2$); we can solve for the value of \bar{U}^1 which corresponds to this particular value of the Lagrange multiplier. With this slight modification, all of the earlier analysis becomes directly applicable to this problem.

Alternatively, suppose we represent consumers utility by a monotone (but not necessarily concave) transform of the utility function U

$$U^{*i} = \phi(U^i).$$

Then, in the first order conditions describing the optimal tax structure, wherever we previously had U_j^i , we now have $\phi' U_j^i$. Since ϕ' can take on any positive value, it is clear that the first order conditions describing Pareto efficient taxation and those that describe the utilitarian tax structure for an appropriately specified ϕ function are equivalent.

We can calculate the maximized value of social welfare associated with any value of \bar{R} (the revenue requirement). Even though U is concave, $W(\bar{R})$ may not be. (See Figure 2.)¹ In that case average social welfare may be increased by raising R_1 per capita from a fraction of the population, R_2 from the remainder, e.g. by randomizing the tax schedules imposed on the population.² In the next section, we discuss a quite different kind of randomization, where randomization is effectively serving as part of the self-selection mechanism.

¹ This possibility was originally noted by Stiglitz (1976) for the case of linear tax schedules.

² A simple example illustrating this, in the present context, is provided by the family of indifference curves of Figure 3. This has two critical properties. For each level of L , there is a saturation level of consumption $C(L)$. For $\{C, L\}$ smaller than the critical level, indifference curves are straight lines with a slope of $w_2\beta$, with $\frac{w_1}{w_2} < \beta < 1$. β is chosen to ensure that, in the optimal tax structure, type 1 individuals are idle. (This simplifies the example.)

Thus the optimal tax problem can be represented as

$$\max_{\{C_1, C_2, Y_1, Y_2\}} w = (C_2 - Y_2\beta)N_2 + C_1N_1$$

subject to

$$(20a) \quad C_2 - Y_2\beta \geq C_1$$

$$(20b) \quad (Y_2 - C_2)N_2 - C_1N_1 \geq \bar{R}$$

where we have made use of the fact that $Y_1 = 0$ and because of our assumption about satiation at $C = C(L)$, we set

$$C_2(Y_2) \equiv C(Y/w_2)$$

At the optimum both the revenue constraint and the self-selection constraint will hold with equality. Substituting (20a) into (20b), we obtain

(footnote 2 continued)

$$(21) \quad \bar{R} = (Y_2 - C_2(Y_2))N_2 - [C_2(Y_2) - Y_2^\beta]N_1$$

Differentiating (21), we obtain

$$\frac{dY_2}{d\bar{R}} = \frac{1}{(1-C_2')N_2 - (C_2' - \beta)N_1}$$

Hence

$$\frac{\partial W}{\partial \bar{R}} = N_1 \frac{dC_1}{d\bar{R}} + N_2 \frac{(C_2' - \beta)}{(1 - C_2')N_2 - (C_2' - \beta)N_1}$$

$$\begin{aligned} \frac{\partial \ln \partial W / \partial \bar{R}}{\partial \bar{R}} &= C_2'' \left[\frac{1}{C_2' - \beta} + \frac{1}{(1 - C_2')\frac{N_2}{N} - (C_2' - \beta)\frac{N_1}{N}} \right] \frac{dY_2}{d\bar{R}} \\ &= \frac{C_2''(1 - \beta)N_2/N^2}{(C_2' - \beta)[(1 - C_2')\frac{N_2}{N} - (C_2' - \beta)\frac{N_1}{N}]^2} \end{aligned}$$

which can be either positive or negative. Although in our example, we have let utility be a linear (rather than strictly concave) function of C and Y, for levels below saturation, it is clear the result would still obtain provided U is not too concave.

(end footnote)

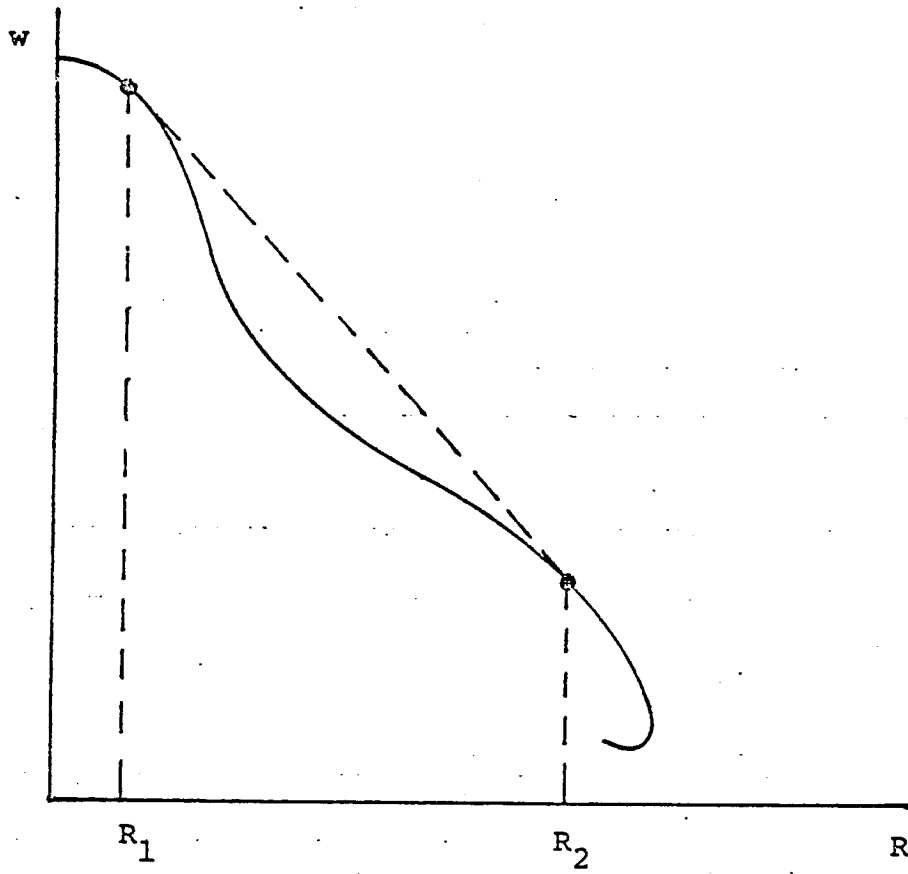


Figure 2. Welfare non-concave function of R :
for revenues between R_1 and R_2
randomization is desirable.

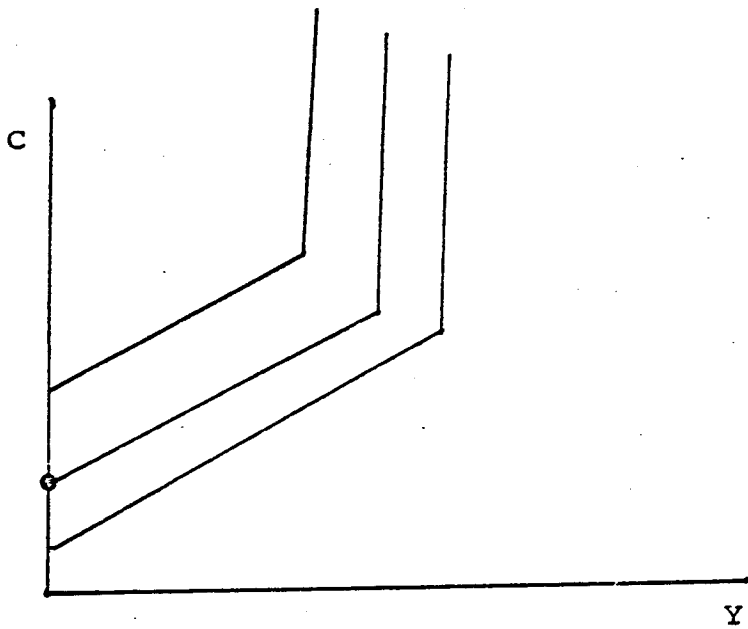


Figure 3

3. Desirability of Randomization

In this section we derive conditions under which randomization of taxes is desirable. As we noted in the preceding section, there are a number of different kinds of randomization. There, we considered the desirability of ex ante randomization -- randomizing the tax before the individual has chosen a level of effort. In my earlier paper (Stiglitz, 1976) I analyzed the desirability of ex post randomization, randomizing the tax after the individual has chosen his level of effort (although the individual chose his level of work effort knowing that the tax he would face would be random). The analysis in this section represents a generalization of these earlier results in two ways: first, while the earlier analysis was restricted to linear tax functions, here we are not so restricted; secondly, by employing non-linear tax structures, we can, in effect, obtain not only a randomization of C , for given Y , but also a randomization of $\{C, Y\}$ packages. That is, we allow the individual either to declare that he is among the more able, in which case we confront him with a tax schedule which generates $\{C_2^*, Y_2^*\}$; or to declare that he is among the less able, in which case he will be confronted with, say, one of two tax schedules, leading to $\{C_1^*, Y_1^*\}$ or $\{C_1^{**}, Y_1^{**}\}$. $\{C_1^*, Y_1^*, C_1^{**}, Y_1^{**}, C_2^*, Y_2^*\}$ must be chosen so that the more able person has a higher utility with $\{C_2^*, Y_2^*\}$ than his expected utility with the random tax scheme.

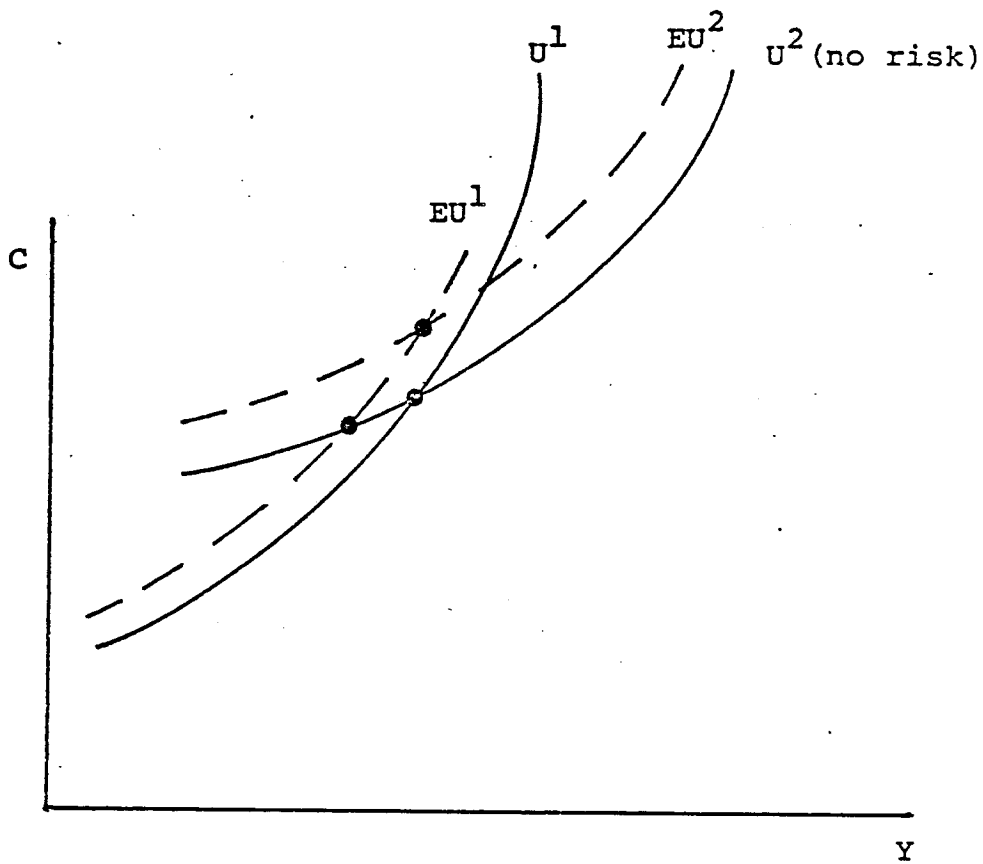


Figure 4

Randomization Increases \bar{C} , by Less Than it
Increases \bar{Y}_1 .

As in other similar screening (or principal agent) problems, the objective of randomization is to increase the effectiveness of screening (or, to put it another way, to reduce the welfare loss associated with the self-selection constraints.) Randomizing the tax imposed on the low ability group lowers its welfare, for any given average tax rate.¹ To leave them at the same level of expected utility, we must, at each Y , increase the mean consumption, as illustrated in figure 4. At the same time, the maximum mean consumption we can provide to the low ability group, for each level of Y , and still have the upper ability group choose the point $\{Y_2^*, C_2^*\}$, is raised by a sufficient amount that the "separating" points may entail a higher \bar{Y}_1 and a higher average level of consumption, \bar{C}_1 ; but it is possible that \bar{C}_1 has increased by less than \bar{Y}_1 , so that the government revenue is increased.

For randomization to be desirable, attitudes toward risk of the more able must differ from those of the less able.

¹ We can show that we do not wish to randomize the tax rates imposed on the high ability individuals. This follows from the same arguments that establish that the marginal tax rate to be imposed on these individuals should be zero. If the regime in which

$$\lambda_1 = 0, \quad \lambda_2 > 0,$$

it is the low ability individuals' tax which should not be randomized.

In this paper, we have assumed that the more able and the less able have the same utility function; but since, at $\{C_1, Y_1\}$ they enjoy different amounts of leisure, their attitudes towards risk may still differ. Indeed, it is apparent that individuals' attitudes towards variability in C may well differ from that attitude towards variability in Y ; more generally, attitudes towards risk depends on how C and Y vary together. We establish that there will frequently be some way of randomizing which improves welfare.

To see what conditions are required for randomization, let

$$(22a) \quad Y_1^* = \bar{Y}_1 + \lambda h, \quad C_1^* = \bar{C}_1 + h$$

$$(22b) \quad Y_1^{**} = \bar{Y}_1 - \lambda h, \quad C_1^{**} = \bar{C}_1 - h$$

with \bar{C}_1 and \bar{Y}_1 chosen to satisfy

$$(23a) \quad v^2(C_1^*, Y_1^*) + v^2(C_1^{**}, Y_1^{**}) \leq 2\bar{v}^2 \equiv 2v(C_2^*, Y_2^*)$$

$$(23b) \quad v^1(C_1^*, Y_1^*) + v^1(C_1^{**}, Y_1^{**}) \geq 2\bar{v}^1$$

The first constraint is the self-selection constraint. The random tax must yield individual 2 an expected utility lower than he obtains at C_2^*, Y_2^* , the second assures us that EV^1 is not lowered by randomization. (v^1 is the utility level attained by group 1 with the optimal non-random tax structure.) For each value of λ and h we can solve (23) for \bar{C} and \bar{Y} . Randomization will be desirable if there exists a value of (λ, h) such that government revenue exceeds that without randomization.

$$(24) \quad \bar{Y}_1(\lambda, h) - \bar{C}_1(\lambda, h) \geq \bar{Y}_1(0, 0) - \bar{C}_1(0, 0)$$

Rather than make the discrete comparison entailed by (24), we calculate $\frac{d\bar{Y}_1}{dh} - \frac{d\bar{C}_1}{dh}$. Differentiating (23) we obtain

$$\begin{bmatrix} sv_C^2 & sv_Y^2 \\ sv_C^1 & sv_Y^1 \end{bmatrix} \begin{bmatrix} d\bar{C} \\ d\bar{Y} \end{bmatrix} = - \begin{bmatrix} DV_C^2 + \lambda DV_Y^2 \\ DV_C^1 + \lambda DV_Y^1 \end{bmatrix} dh$$

where

$$(26a) \quad sv_j^i \equiv v_j^i(C_1^*, Y_1^*) + v_j^i(C_1^{**}, Y_1^{**})$$

$$(26b) \quad DV_j^i \equiv v_j^i(C_1^*, Y_1^*) - v_j^i(C_1^{**}, Y_1^{**})$$

Hence,¹ letting $M^i = DV_C^i + \lambda DV_Y^i$, $MRS^i = - \frac{\partial v^i / \partial Y_1}{\partial v^i / \partial C_1}$,

$$(27) \quad \frac{d(\bar{Y}_1 - \bar{C}_1)}{dh} \Big|_{\bar{h}=0} = \frac{\frac{M^2}{\partial v^2 / \partial C_1} (1 - MRS^1) - \frac{M^1}{\partial v^1 / \partial C_1} (1 - MRS^2)}{2(MRS^2 - MRS^1)}$$

¹ Using Cramer's rule,

$$\frac{d(\bar{Y} - \bar{C})}{dh} = \frac{(sv_C^1 + sv_Y^1)(DV_C^2 + \lambda DV_Y^2) - (sv_C^2 + sv_Y^2)(DV_C^1 + \lambda DV_Y^1)}{sv_C^2 sv_Y^1 - sv_C^1 sv_Y^2}$$

Dividing the numerator and demoninator by $sv_C^1 sv_C^2$ and letting $h \rightarrow 0$, we obtain (27).

At $h = 0$, $M^1 = M^2 = 0$, and hence $\frac{d(\bar{Y} - \bar{C})}{dh} = 0$.

Hence we need to calculate¹

$$(28) \quad \frac{d^2(\bar{Y} - \bar{C})}{dh^2} \Big|_{h=0} = \{ (V_{CC}^2 + 2\lambda V_{CY}^2 + \lambda^2 V_{YY}^2) \frac{(1 - MRS^1)}{\partial V^2 / \partial C_1} - (V_{CC}^1 + 2\lambda V_{CY}^1 + \lambda^2 V_{YY}^1) \frac{(1 - MRS^2)}{\partial V^1 / \partial C_1} \} / (MRS^2 - MRS^1)$$

From our earlier analysis we know that $MRS^2 < MRS^1 < 1$. Hence, randomization is desirable if, for some value of λ ,²

$$(29) \quad \frac{V_{CC}^2 + 2\lambda V_{CY}^2 + \lambda^2 V_{YY}^2}{V_C^2 + V_Y^2} < \frac{V_{CC}^1 + 2\lambda V_{CY}^1 + \lambda^2 V_{YY}^1}{V_C^1 + V_Y^1}$$

To see that (29) may easily be satisfied, assume $U = u(C) - v(L)$. Then (29) can be rewritten as

$$\frac{\lambda^2 v''/w_2^2 - u''}{u' - v'/w_2} > \frac{\lambda^2 v''/w_1^2 - u''}{u' - v'/w_1}$$

If $\lambda = 0$, this will be satisfied if

$$\frac{v'(Y_1/w_2)}{w_2} > \frac{v'(Y_1/w_1)}{w_1},$$

which is impossible. $(\frac{dv'(Y/w)/w}{dw} = -\frac{1}{2}[\frac{v''Y}{w} + \dots] < 0)$

If $\lambda = \infty$ and $\eta \equiv \frac{v''L}{v'}$ this will be satisfied if

¹Footnotes on page 22a.

In this paper, we have assumed that the more able and the less able have the same utility function; but since, at $\{C_1, Y_1\}$ they enjoy different amounts of leisure, their attitudes towards risk may still differ. Indeed, it is apparent that individuals' attitudes towards variability in C may well differ from that attitude towards variability in Y ; more generally, attitudes towards risk depends on how C and Y vary together. We establish that there will frequently be some way of randomizing which improves welfare.

To see what conditions are required for randomization, let

$$(22a) \quad Y_1^* = \bar{Y}_1 + \lambda h, \quad C_1^* = \bar{C}_1 + h$$

$$(22b) \quad Y_1^{**} = \bar{Y}_1 - \lambda h, \quad C_1^{**} = \bar{C}_1 - h$$

with \bar{C}_1 and \bar{Y}_1 chosen to satisfy

$$(23a) \quad v^2(C_1^*, Y_1^*) + v^2(C_1^{**}, Y_1^{**}) \leq 2\bar{v}^2 \equiv 2v(C_2^*, Y_2^*)$$

$$(23b) \quad v^1(C_1^*, Y_1^*) + v^1(C_1^{**}, Y_1^{**}) \geq 2\bar{v}^1$$

The first constraint is the self-selection constraint. The random tax must yield individual 2 an expected utility lower than he obtains at C_2^*, Y_2^* , the second assures us that EV^1 is not lowered by randomization. (v^1 is the utility level attained by group 1 with the optimal non-random tax structure.) For each value of λ and h we can solve (23) for \bar{C} and \bar{Y} . Randomization will be desirable if there exists a value of (λ, h) such that government revenue exceeds that without randomization.

¹ We use the fact that

$$\frac{d^2(\bar{Y}_1 - \bar{C}_1)}{dh^2} \Big|_{h=0} = \frac{M^2 \frac{d(1 - MRS^1)/\partial V^2/\partial C_1}{dh} - M_1 \frac{d(1 - MRS^2)/\partial V^1/\partial C_1}{dh}}{MRS^2 - MRS^1} + \frac{d(\bar{Y}_1 - \bar{C}_1)}{dh} \frac{d \ln(MRS^2 - MRS^1)}{dh} + \frac{(1 - MRS^1)}{V^2/C_1} \frac{dM^2}{dh} - \frac{(1 - MRS^2)}{V^1/C^1} \frac{dM^1}{dh}$$

At $h = 0$, $M^i = 0$, and hence the first two terms are zero.

We then calculate

$$\frac{dM^i}{dh} = v_{CC}^i + 2\lambda v_{CY}^i + \lambda^2 v_{YY}^i.$$

Substituting, we obtain (28).

² Consumption randomization (corresponding to, say, random enforcement of the tax laws) is desirable if (29) is satisfied when $\lambda = 0$.

$$\frac{\eta(L_2)}{\frac{1}{MRS^2} - 1} > \frac{\eta(L_1)}{\frac{1}{MRS^1} - 1}$$

a sufficient condition for which is that η' be sufficiently negative.

Since at $w_1 = w_2$, both sides of (29) are identical, randomization is desirable provided for some λ the derivative of

$$\frac{V_{CC} + 2\lambda V_{CY} + \lambda^2 V_{YY}}{V_C + V_Y}$$

with respect to w is negative. To obtain interpretable results, we express the derivatives of V in terms of the derivative of the underlying utility function (using 7').

Randomization is desirable if

$$\frac{U_{CCL} + \frac{2\lambda U_{CL}}{w} \left[\frac{U_{LLC}}{U_{LC}} + 1 \right] + \frac{\lambda^2 U_{LL}}{w^2} \left[\frac{U_{LLL}}{U_{LL}} + 2 \right]}{U_{CC} + \frac{2\lambda U_{LC}}{w} + \frac{\lambda^2 U_{LL}}{w^2}} > \frac{U_{CL} + \frac{U_L}{w} \left[\frac{U_{LL}}{L} + 1 \right]}{U_C + \frac{U_L}{w}}$$

which may easily be satisfied.

In this section, we have shown how randomization may enable a weakening of the self-selection constraints, and therefore an increase in expected utility. Finally, we note that it may be desirable to employ both kinds of randomization we have discussed. If the maximized value of expected utility, employing the optimal randomization of tax schedules, is not a concave function of the revenue raised, maximizing the sum of expected

utilities will entail randomization of the sets of tax schedules, one out of which will be imposed randomly on an individual who declares he is of low ability.¹

¹ The desirability of both kinds of randomization was originally discussed in Stiglitz (1976). He analyzed the conditions under which a random tax was imposed both before and after effort was decided upon. In the present context, the latter corresponds to a randomization of C , for a given Y . The analysis was limited to linear tax structures, while here we employ highly non-linear structures.

4. Desirability of Differentiation

We noted in our introduction that there was a cost to differentiating among different individuals. It is not obvious, in the context of say a utilitarian social welfare function, that it is always desirable to differentiate, or to differentiate completely if there are many groups. In the general screening literature, equilibria in which we cannot infer perfectly the characteristics of the individuals are referred to as pooling equilibria (Rothschild-Stiglitz, 1976), and equilibria in which we can are referred to as separating equilibria. It can be shown that pooling equilibria can arise in a variety of circumstances (Stiglitz, 1977). In the present context, if the income tax schedule results in both low ability and high ability individuals having the same income and consumption, then the equilibrium is a pooling equilibrium. If each ability group enjoys a different income, then we have a separating equilibrium. Here, we show (i) if there are two groups, and the more productive group's indifference curves have a flatter slope in $\{C, Y\}$ space then differentiation is desirable; (ii) if more productive groups have indifference curves with a slope in $\{C, Y\}$ space which at some point is the same as that of the less productive group, then a pooling equilibrium cannot be ruled out; (iii) if there are **three** or more groups, then pooling among a subset may well be desirable.

To see the first result, assume the government imposed a tax schedule so that everyone worked hard enough so that before tax income was Y^* . (Clearly the less able worked harder than the more able, but the government could not observe the level of effort or hours.) Consequently, all had the same level of consumption, C^* . The two groups are "pooled" together. See Figure 5. Any point (\hat{C}, \hat{Y}) in the shaded area generates a separating equilibrium (i.e. one group prefers $\{\hat{C}, \hat{Y}\}$ to $\{C^*, Y^*\}$, the other group prefers $\{C^*, Y^*\}$ to $\{C, Y\}$). Any point along the lower envelope of 1 and 2's indifference curves separates. Moreover, the level of welfare of each group in the separating equilibrium is the same as at the pooling point P. We need to see what happens to government revenue. If, at P,

$$(31a) \quad \left. \left(\frac{dC}{dY} \right) \right|_{\bar{U}^2} < 1,$$

by offering a point such as A, we "separate." This increases government revenue, since the required increase in 2's consumption is less than the increase in his output (before tax income). Similarly, if at P,

$$(31b) \quad \left. \left(\frac{dC}{dY} \right) \right|_{\bar{U}^1} > 1$$

a point such as B separates, and the reduction in consumption exceeds the reduction in income: government revenue thus increases. Since

$$(32) \quad \left. \left(\frac{dC}{dY} \right) \right|_{\bar{U}^1} > \left. \left(\frac{dC}{dY} \right) \right|_{\bar{U}^2}$$

if (31a) is not true, i.e.

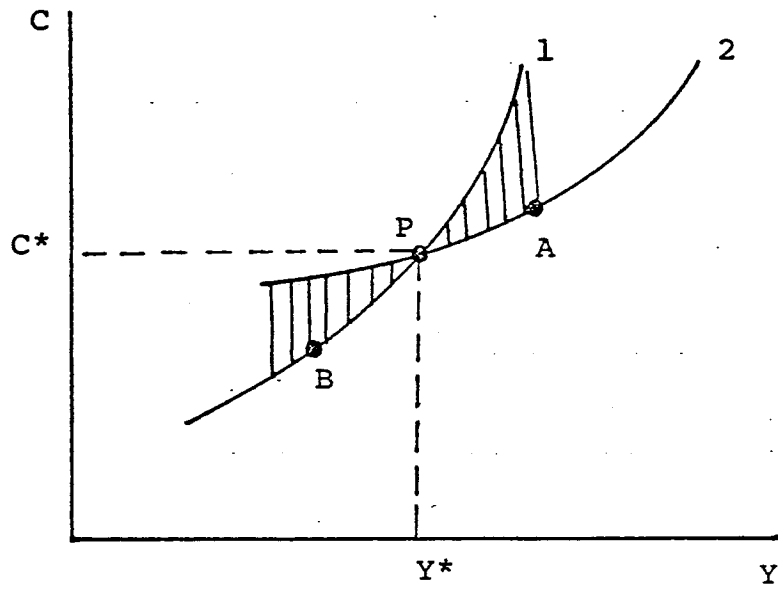


Figure 5

Pareto Efficiency Requires Differentiation

$$(33) \left(\frac{dC}{dY} \right)_{\bar{U}^2} > 1$$

then,

$$(34) \left(\frac{dC}{dY} \right)_{\bar{U}^1} > 1.$$

Thus, there always exists a separating contract which increases revenue and leaves utilities of all individuals unchanged. The only pareto efficient tax structures entail separation.

The same argument obviously holds if the less productive individuals always have flatter indifference curves, but this is not a particularly plausible assumption.

In figure 6 we illustrate what happens if the different types of individuals have different preferences, such that the more able s' indifference curve is not always flatter than the less ables! The point P is a point of tangency. The shaded area represents the set of {C, Y} points which together with P, separate the two groups. But clearly, it is possible (although presumably not likely) that

$$(35) \quad \frac{dC}{dY} = 1.$$

Figure 7 illustrates the result that with three or more groups, partial pooling may be desirable. Two points are offered, E_1 and E_2 , with E_1 chosen by the high ability group, E_2 , by the two low ability groups. The points which separate 2 and 3 are those which lie between their indifference curves; but those

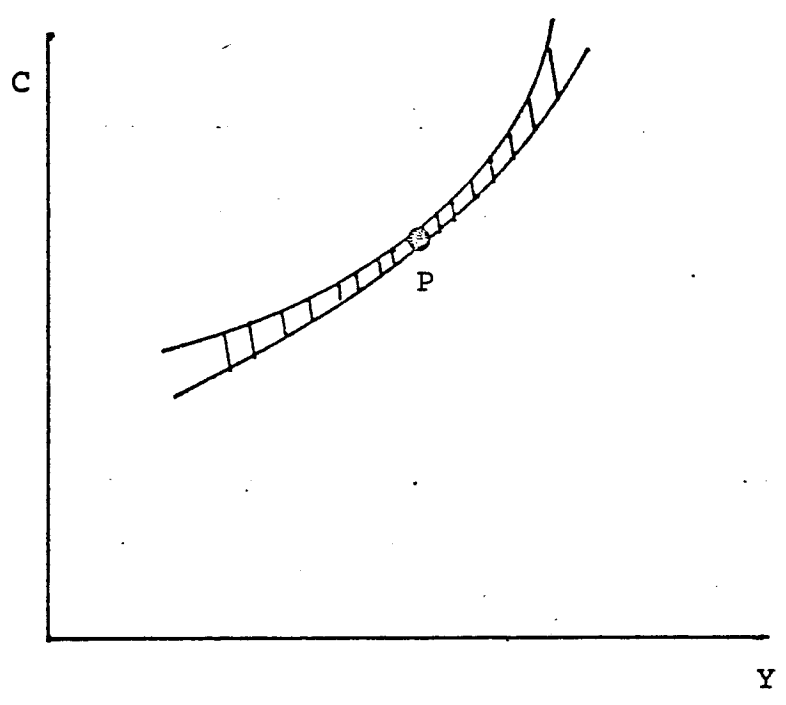


Figure 6
Pooling Equilibrium

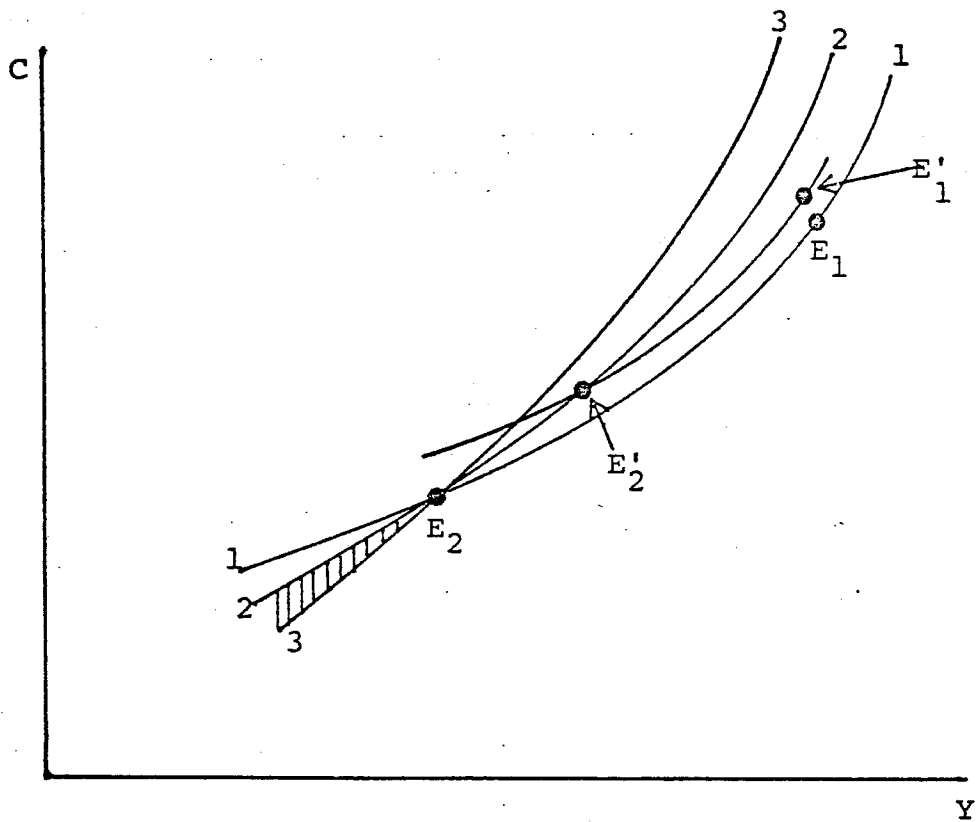


Figure 7

Partial Pooling May be Desirable

which separate 2 and 3 and also separate 1 are only those which lie between 2 and 3 below 1's indifference curve (the heavily shaded area). Thus if at E_2 ,

$$(36) \quad \left(\frac{dC}{dY}\right)_{U^3} < 1$$

clearly, we cannot keep everyone on their same indifference curves and increase government revenue.¹

The same argument obviously holds if we have a continuum of types. This analysis provides some insights into the results noted earlier (Mirrlees (1971), Stiglitz (1977)) that the optimal tax structure with a continuum of individuals will not, in general, be differentiable. There may well be "kinks" in the optimal tax structure. Individuals with different marginal rates of substitution obtain exactly the same income (Figure 8).

¹ This does not, of course, prove that the $\{E_1, E_2\}$ constitutes an efficient tax structure. It may be possible² to raise revenue and increase 1's utility level. If (32a) is true, it is clear that

$$\left(\frac{dC}{dY}\right)_{U^2} < 1.$$

Hence, by offering a new set of points $\{E'_1, E'_2\}$ as illustrated in the figure, we can separate, and increase government revenue collected from individuals of type 2. At the same time, we decrease the revenue collected from individuals of the highest ability (recall, that efficient taxation implies that there is no distortionary taxation on the highest ability individual) and hence as we increase their welfare, we decrease work and increase consumption; government revenue collected from him thus must decrease. Whether total revenue collected increases or decreases thus depends on the relative number of individuals of the two types.

Finally, figure 9b illustrates a case with 2 ability groups and 2 taste groups. In each ability group, there are some individuals who dislike working more than others; their indifference curves (in $\{C, Y\}$ or $\{C, L\}$ space) are accordingly steeper. The important characteristic is that in $\{C, Y\}$ space, the indifference curves of a high ability lazy worker and a low ability industrious worker may intersect several times, as illustrated in figure 9a. This may occur even with an additive utility function. Indeed an individual with wage w_2 and indifference curves $U^2(C, L)$ and an individual with wage w_1 and indifference curves of the form $U^1(C, L) \equiv U^2(C, Lw_1/\sqrt{w_2})$ have identical indifference tax curves in $\{C, Y\}$ space; there is thus no way of differentiating between them with an income tax.

There may be no other ways of differentiating among these individuals; for instance, these individuals do have different levels of consumption of leisure. Although we cannot observe their levels of consumption of leisure, we may be able to observe their purchases of goods which are complements of leisure, and use this as a basis of inferring their ability.

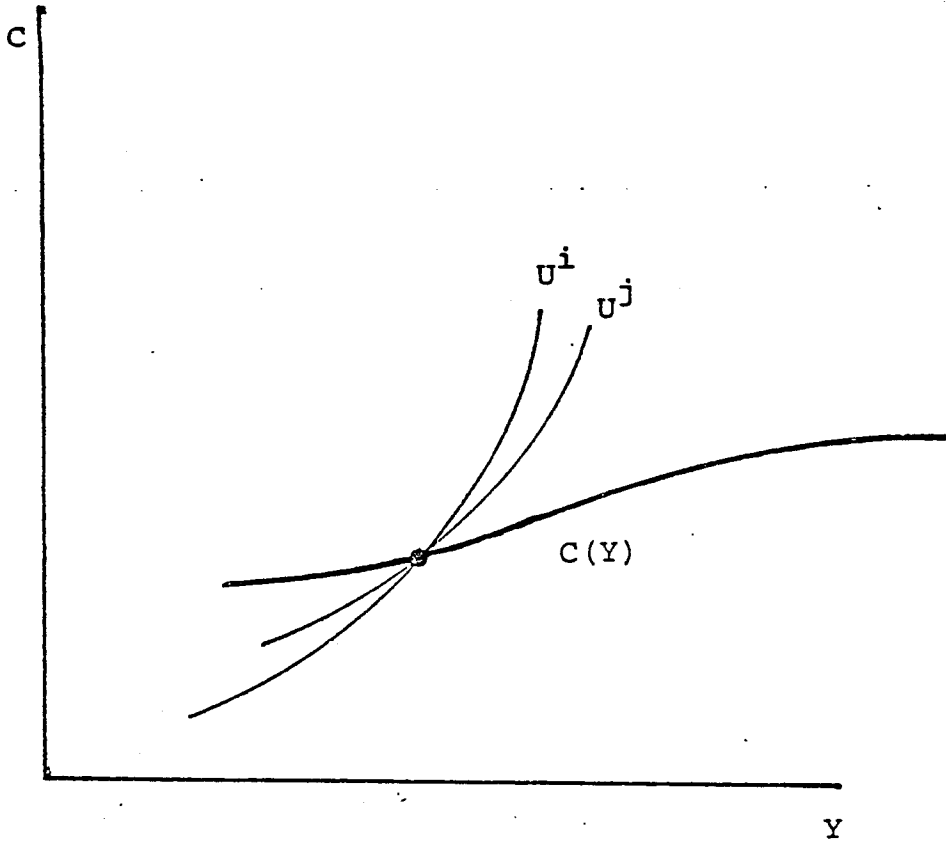


Figure 8

Kinked Optimal Tax Structure

5. Pareto Efficient Taxation with Different Tastes

The framework we have developed allows us to obtain some simple but interesting results on the structure of Pareto efficient taxation with two or more taste groups.

We assume that **some individuals are more averse to work than others**. For simplicity, we assume there are three groups, two high ability types and a single low ability type.

We wish to establish three propositions: First, it is always Pareto efficient to differentiate on the basis of tastes if one group is always more averse to work than the other (so the slope of its indifference curve is always steeper); we should never "pool" the two high ability groups together. The "ability to pay" principle of the determination of taxes is, in this sense, inconsistent with the principle of Pareto efficiency. Second, Pareto efficient taxation often will entail hyper-regressivity, i.e. marginal rates which are less than zero. Third, if individuals differ in tastes as well as abilities, then complete differentiation will not, in general, be possible.

The first proposition is equivalent to the proposition established in the preceding section that when abilities differed, Pareto efficiency is desirable. What was critical in that argument was that the indifference curves in $\{C, Y\}$ space differ. In Figure 9 we have assumed that the government offers two points, E_1 and E_2 , with both of the high ability groups (denoted by U^1 and U^2) at E_1 . By the same kind of reasoning used earlier, clearly any point between the two indifference curves separates, and either

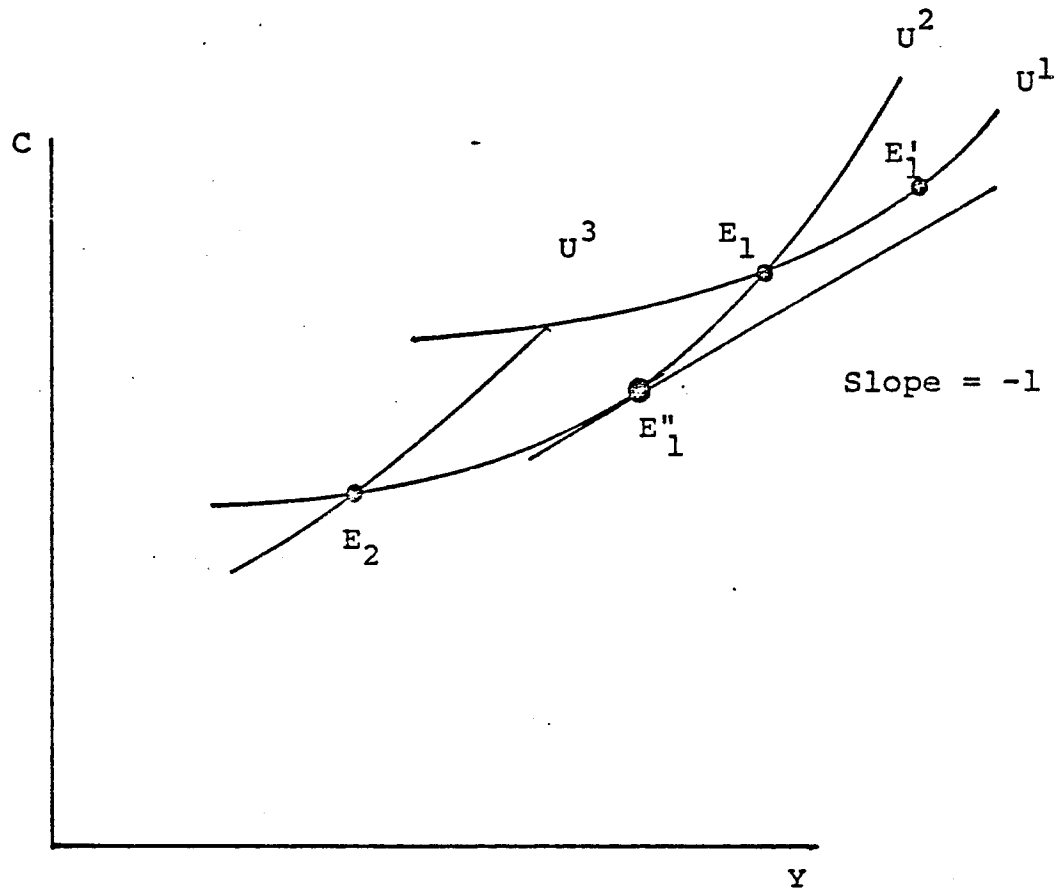


Figure 9

U^1 and U^2 Both High Ability, But Different Tastes: Differentiation is Pareto Efficient and Optimal Tax Structure and Entail Hyper-Regressivity.

$$\left(\frac{dC}{dY}\right)_{\bar{U}^1} < 1$$

or

$$\left(\frac{dC}{dY}\right)_{\bar{U}^2} > 1$$

(or both); hence there exist points which increase government revenue and leave every individual's utility unaffected. Indeed, the efficient set of points for this example denoted $\{E_1', E_1'',$ and $E_2\}$ are such that the marginal rate paid by both of the two upper ability groups is zero. We have drawn through E_1'' a line with a slope of 45° . In the figure, it passes below E_1' . This implies that the increment in consumption in moving from E_1'' to E_1' exceeds the increment in income, i.e. the mean marginal rate over that interval may be negative; on average, there may be regressive taxation at the upper end of the distribution.¹

The third proposition -- the impossibility of differentiating completely -- follows immediately from the observation that we can only differentiate on the basis of differences in indifference curves in $\{C, Y\}$ space. Individuals of high ability and high aversion to risk may thus be indistinguishable from individuals of low ability and low aversion to risk.

¹ Note that in this case the lump sum tax that would be imposed on group 1 is lower than that imposed on group 2 with first best taxation. Lump sum taxation with a utilitarian objective function would entail equating the marginal utility of consumption of the two groups. This will not, in general, imply equal lump sum taxes. The individual who is less averse to work may face a higher lump sum tax.

Indeed an individual with wage w_2 and indifference curves $U^2(C,L)$ and an individual with wage w_1 and indifference curves of the form $U^1(C,L) \equiv U^2(C, Lw_1/w_2)$ have identical indifference tax curves in $\{C,Y\}$ space; there is thus no way of differentiating between them with an income tax.

There may be no other ways of differentiating among these individuals; for instance, these individuals do have different levels of consumption of leisure. Although we cannot observe their levels of consumption of leisure, we may be able to observe their purchases of goods which are complements of leisure, and use this as a basis of inferring their ability. We examine this possibility in Section 7.

6. Stochastic Income

In sections 2.3 and 5 we noted two instances where the optimal tax structure entailed negative marginal tax rates at higher incomes.

This result should not, however, be stressed too much; a second modification allowing income to be stochastic, leads to just the opposite result: marginal rates of 100%.

Assume that an individual who works L receives an income of

Obviously, this two group model is much over simplified; just as in the conventional optimal income tax problem we could infer the individual's ability by his income, so too here; although we have introduced a stochastic element to his income, we can still infer perfectly the individual's ability from his income. More generally, however, we will not be able to distinguish perfectly a low ability lucky individual from a high ability unlucky individual. This makes the design of the optimal tax structure with stochastic income far more difficult (and more interesting) than the deterministic case upon which the analysis has thus far focused. But so long as there is a finite number of groups (or even a continuum, with a finite range) then if the probability distribution of incomes is bounded, the highest incomes observed will always be received by the highest ability individuals who are lucky. Optimal taxation entails 100% taxation at the margin.

The unreasonableness of this result arises from the assumption that individuals have no control over the stochastic elements in their income stream. A tax structure which imposed 100% taxation at the margin at the top would have peculiar (and probably undesirable) incentive effects with respect to risk taking.

Obviously, this two group model is much over simplified; just as in the conventional optimal income tax problem we could infer the individual's ability by his income, so too here; although we have introduced a stochastic element to his income, we can still infer perfectly the individual's ability from his income. More generally, however, we will not be able to distinguish perfectly a low ability lucky individual from a high ability unlucky individual. This makes the design of the optimal tax structure with stochastic income far more difficult (and more interesting) than the deterministic case upon which the analysis has thus far focused. But so long as there is a finite number of groups (or even a continuum, with a finite range) then if the probability distribution of incomes is bounded, the highest incomes observed will always be received by the highest ability individuals who are lucky. Optimal taxation entails 100% taxation at the margin.

The unreasonableness of this result arises from the assumption that individuals have no control over the stochastic elements in their income stream. A tax structure which imposed 100% taxation at the margin at the top would have peculiar (and probably undesirable) incentive effects with respect to risk taking.

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

If we now differentiate the Lagrangian with respect to C_{ij} , we obtain (see the analogous equations (13) and (14))

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

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

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

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

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

$$(40a) \quad \frac{\partial v^2 / \partial C_{2j}}{\partial v^2 / \partial C_{2k}} = 1$$

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

(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.

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

If we now differentiate the Lagrangian with respect to C_{ij} , we obtain (see the analogous equations (13) and (14))

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

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

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

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

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

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

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

(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.

$$= \frac{\frac{\lambda_2}{\mu} \frac{\partial V^2 / \partial C_{1k}}{\partial V^1 / \partial C_{1k}} \left[\frac{\partial V^2 / \partial C_{1j}}{\partial V^2 / \partial C_{1k}} - \frac{\partial V^1 / \partial C_{1j}}{\partial V^1 / \partial C_{1k}} \right]}{1 - \frac{\lambda_2}{\mu} \frac{\partial V^2 / \partial C_{1k}}{\partial V^1 / \partial C_{1k}}}$$

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

Thus 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 the leisure which they enjoy at any given level of income.) We thus can

$$= \frac{\frac{\lambda_2}{\mu} \frac{\partial v^2 / \partial c_{1k}}{\partial v^1 / \partial c_{1k}} \left[\frac{\partial v^2 / \partial c_{1j}}{\partial v^2 / \partial c_{1k}} - \frac{\partial v^1 / \partial c_{1j}}{\partial v^1 / \partial c_{1k}} \right]}{1 - \frac{\lambda_2}{\mu} \frac{\partial v^2 / \partial c_{1k}}{\partial v^1 / \partial c_{1k}}}$$

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

Thus 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 the leisure which they enjoy at any given level of income.) We thus can

if not, he should tax or subsidize one commodity relative to a second depending on whether the individuals who consume more have higher or lower marginal rate of substitution between the two commodities.

It should also be obvious that although we have limited our attention to the problem of optimal taxation, the problem of the optimal pricing of a public utility is precisely the same problem. The only distinction that arises, at least in some cases, is that the public utility is in general allowed to control only a subset of the prices. If we assume that the other prices are fixed, then we can form a Hicksian composite commodity (called "other goods"), and the determination of the total outlay (charge for the package of services supplied by the public utility) determines the amount of the "other good" available to the individual. With these modifications, (interpreting "Y" now as "other goods") the earlier analysis is directly applicable to the problem at hand.

Moreover, if relative prices of the "other goods" are not fixed, then we can modify the analysis of the multi-product case, in the same way that we earlier modified our analysis of the income tax with endogenous wages with parallel results: now, even for the most able individual, we will wish to impose distortionary taxation (charge distortionary prices).

if not, he should tax or subsidize one commodity relative to a second depending on whether the individuals who consume more have higher or lower marginal rate of substitution between the two commodities.

It should also be obvious that although we have limited our attention to the problem of optimal taxation, the problem of the optimal pricing of a public utility is precisely the same problem. The only distinction that arises, at least in some cases, is that the public utility is in general allowed to control only a subset of the prices. If we assume that the other prices are fixed, then we can form a Hicksian composite commodity (called "other goods"), and the determination of the total outlay (charge for the package of services supplied by the public utility) determines the amount of the "other good" available to the individual. With these modifications, (interpreting "Y" now as "other goods") the earlier analysis is directly applicable to the problem at hand.

Moreover, if relative prices of the "other goods" are not fixed, then we can modify the analysis of the multi-product case, in the same way that we earlier modified our analysis of the income tax with endogenous wages with parallel results: now, even for the most able individual, we will wish to impose distortionary taxation (charge distortionary prices).

- the optimal tax structure; the marginal rate on the most able individual is always negative; on the less able individuals it is positive, and its magnitude depends on the elasticity of substitution;
- (c) if different individuals have different attitudes towards leisure, the tax structure may be regressive in the upper tail;
 - (d) if income is stochastic, the limiting marginal tax rate may be 100%.

The main qualitative properties of earlier analyses of the optimal tax structure are clearly not robust to these attempts to make the theory more "realistic." On the one hand, our analysis makes it clear that ~~there~~ is much more to be done. Until a more general theory is developed, none of the qualitative results can be accepted as a basis of policy. On the other hand, the extreme sensitivity of the results to the changes in the assumptions suggests that results which are sufficiently clear and robust to form the basis of policy may well not be obtained; rather the objective of future research should perhaps be the clarification of the important dimensions of choice (risk taking, effort, etc.) affected by the income tax structure and the trade-offs which emerge.

- the optimal tax structure; the marginal rate on the most able individual is always negative; on the less able individuals it is positive, and its magnitude depends on the elasticity of substitution;
- (c) if different individuals have different attitudes towards leisure, the tax structure may be regressive in the upper tail;
 - (d) if income is stochastic, the limiting marginal tax rate may be 100%.

The main qualitative properties of earlier analyses of the optimal tax structure are clearly not robust to these attempts to make the theory more "realistic." On the one hand, our analysis makes it clear that ~~there~~ is much more to be done. Until a more general theory is developed, none of the qualitative results can be accepted as a basis of policy. On the other hand, the extreme sensitivity of the results to the changes in the assumptions suggests that results which are sufficiently clear and robust to form the basis of policy may well not be obtained; rather the objective of future research should perhaps be the clarification of the important dimensions of choice (risk taking, effort, etc.) affected by the income tax structure and the trade-offs which emerge.

REFERENCES

- Allen, F., 1980, Optimal income taxation with general equilibrium effects on wages, mimeo, Oxford.
- 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 on public economics, New York and London: McGraw-Hill.
- Arnott, R., A. Hosios, and J. E. Stiglitz, 1980, Implicit contracts, labor mobility and unemployment, Princeton University mimeo, paper presented at NSF-NBER CEME Conference on Implicit Contracts and Fixed-Price Equilibria, Princeton University, December.
- Mirrlees, J., 1971, An exploration in the theory of optimum income taxation, *Review of Economic Studies* 38, 175-208.
- Phelps, E. S., 1973, The taxation of wage income for economic justice, *Quarterly Journal of Economics* 87, 331-354.
- 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.
- Stern, N., 1979, Optimum taxation with errors in administration, Warwick University mimeo.
- Stiglitz, J. E., 1976, Utilitarianism and horizontal equity: the case for random taxation, IMSSS Technical Report No. 214, Stanford University.

Stiglitz, J. E., 1977, Monopoly, non-linear pricing and
imperfect information: the insurance market, Review of
Economic Studies, October, 407-430.

Stiglitz, J. E., (forthcoming), Information and economic
analysis, Oxford University Press.