NBER WORKING PAPER SERIES

SELF-SELECTION AND PARETO EFFICIENT TAXATION

Joseph E. Stiglitz

Working Paper No. 632

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

February 1981

Financial support from the NSF is gratefully acknowledged. I am indebted to F. Allen 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. The research reported here is part of the NBER's research program in Taxation. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #632
February 1981

Self-Selection and Pareto Efficient Taxation

ABSTRACT

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 optimal pricing of a monopolist) which have recently been the subject of extensive research, but also allow 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, then the general equilibrium effects -- until now largely ignored in the literatures -- of changes in the tax structure may be dominant in determining the optimal tax structure; in particular if relative wages 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.

Professor Joseph E. Stiglitz
Department of Economics
Princeton University
Princeton, New Jersey 08544

(609) 452-4014

Self-Selection and Pareto Efficient Taxation

Joseph E. Stiglitz

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 satisifed. The problem of the government (the monoplist, 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. This formulation not only allows us to 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

---

Financial support from the NSF is gratefully acknowledged . I am indebted to F. Allen 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.

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 four new results.

First, in the literature on self-selection, it has been shown that randomization may serve as an effective screening device (Stiglitz (1981)). 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 the desirability of random taxation to show that randomization may characterize a much less restricted set of tax structures (the 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 earlier literature limited itself to analyzing the optimal income tax under the assumption that individual's relative productivities 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 <u>linear</u> 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 even possible for the optimal tax structure to be regressive, even for a Rawlsian social welfare objective function.

---

[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.

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, 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) They have assumed that the preferences of all individuals are identical; and (b) They have assumed 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 one case, we show that at the upper end, the marginal tax rate is 100% (rather than zero, as in the conventional story) while in the other case, we show that, at the upper end, the marginal rate is negative.

1. Pareto Efficient Taxation: The Simplest Case

We begin our discussion with the simplest possible model, in which there are only two individuals, differing in their ability but having the same utility function (this, as we shall see, is not critical for most of the results we shall obtain). The ith 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) $C_i = w_i L_i$

where

C_i = the ith individual's consumption

L_i = number of hours worked by ith individual

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

(2) $Y_i = w_i L_i$, ith individual's income

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

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

$$\frac{\partial U^i}{\partial C_i} > 0 , \quad \frac{\partial U^i}{\partial L_i} < 0 .$$

His indifference curve is depicted in Figure 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) $\max U^i(C_i, L_i)$

$\quad$ s.t. $C_i \leq w_i L_i - T(w_i L_i)$

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

(6) $\dfrac{\partial U_i / \partial L_i}{\partial U_i / \partial C_i} = -w_i(1-T')$.

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

The optimal consumption-leisure of the individual before and after taxes is depicted in Figures 1a and 1b.

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 write[1]

(7) $U = U^i(C_i, \dfrac{Y_i - T_i}{w_i}) \equiv \hat{U}^i(C_i, Y_i; w_i)$.

---

[1] For simplicity, we shall often write $\hat{U}^i(C_i, Y_i)$ rather than $\hat{U}^i(C_i, Y_i; w_i)$.

Even if all individuals have the same utility of consumption-and-leisure functions, their utility of consumption-and-before tax income will differ. It is clear that individuals of higher ability have, in Figure 2, flatter indifference curves: the increase in consumption that is required for a given increase in before tax income is smaller, since to obtain the given increase in before tax income they need to forego much less leisure.

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 now easily stated. It wishes to maximize the utility of say, individual 2, subject to (a) individual 1 having at least a given level of utility and (b) the constraint that it raises 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.

Formally, the government

(8) max $\hat{U}^2(C_2, Y_2)$

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

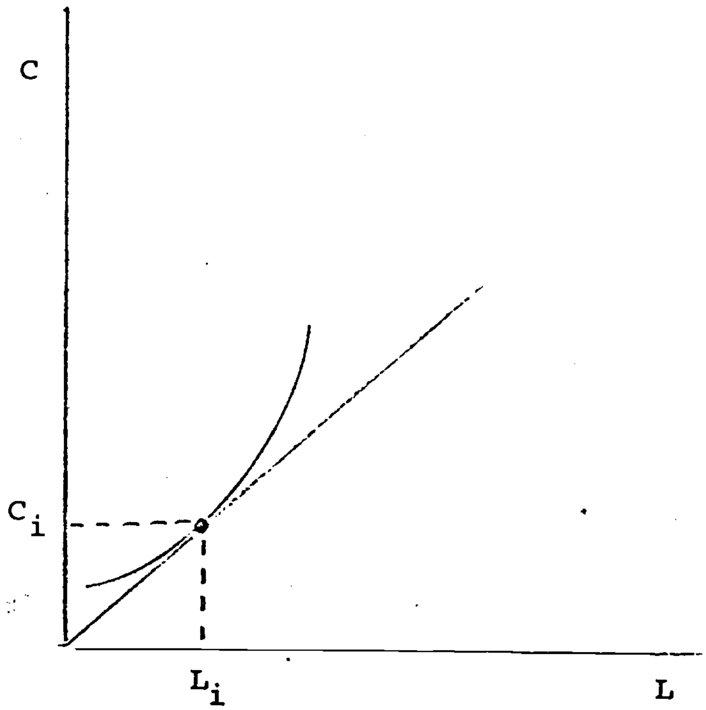
(10) $\hat{U}^2(C_2, Y_2) \geq \hat{U}^2(C_1, Y_1)$

(11) $\hat{U}^1(C_1, Y_1) \geq \hat{U}^1(C_2, Y_2)$

} the self-selection constraints.

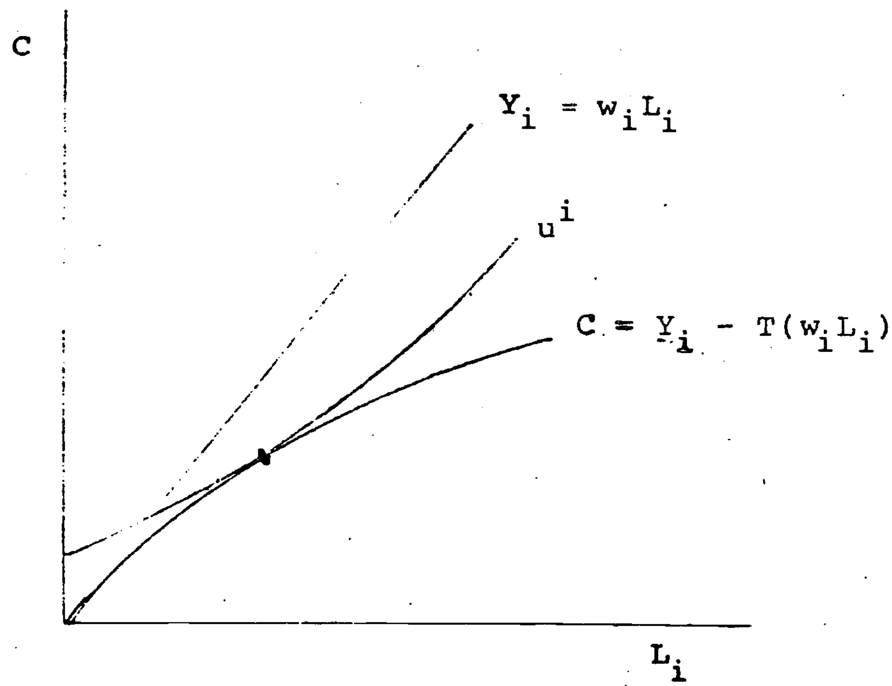
(12) $R = (Y_1 - C_1)N_1 + (Y_2 - C_2)N_2 \geq \bar{R}$, the revenue constraint

(where R is government revenue, $\bar{R}$ is the revenue requirement, N_i the number of individuals of type i). Notice that this problem



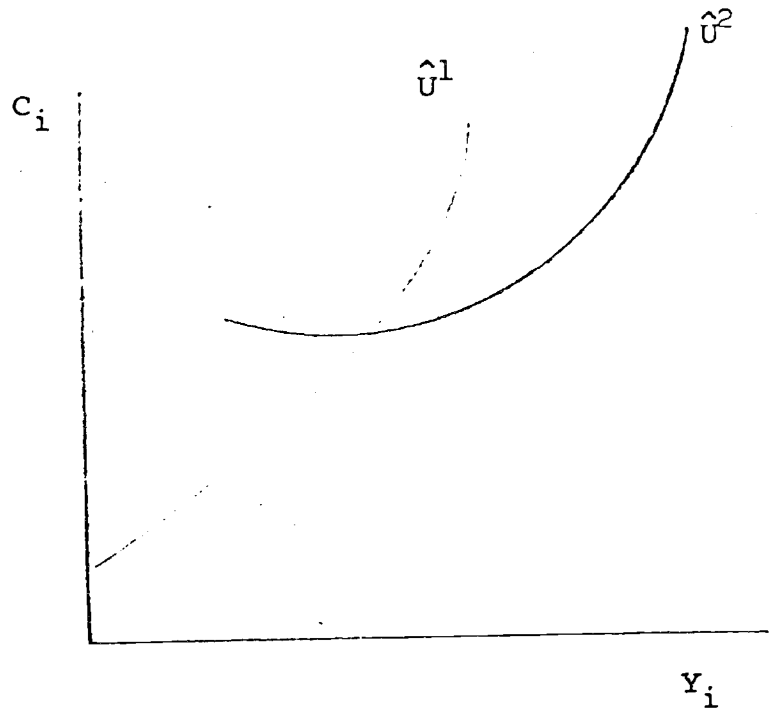
Individual's Indifference Curve between Consumption and Leisure and The Before Tax Budget Constraint.

Figure 1a



After Tax Budget Constraint and Consumption Leisure Choice

Figure 1b



Individuals of Higher Ability Have Flatter Difference Curves

Figure 2

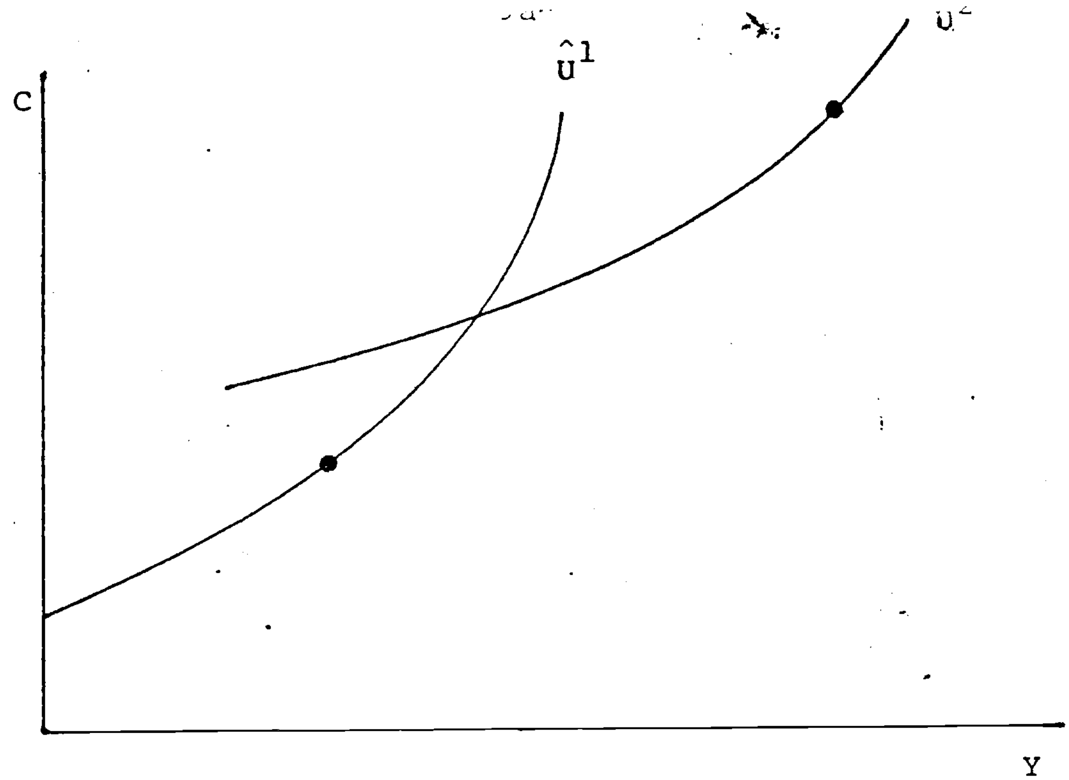


Figure 3a
First Best Taxation Fully Revealing

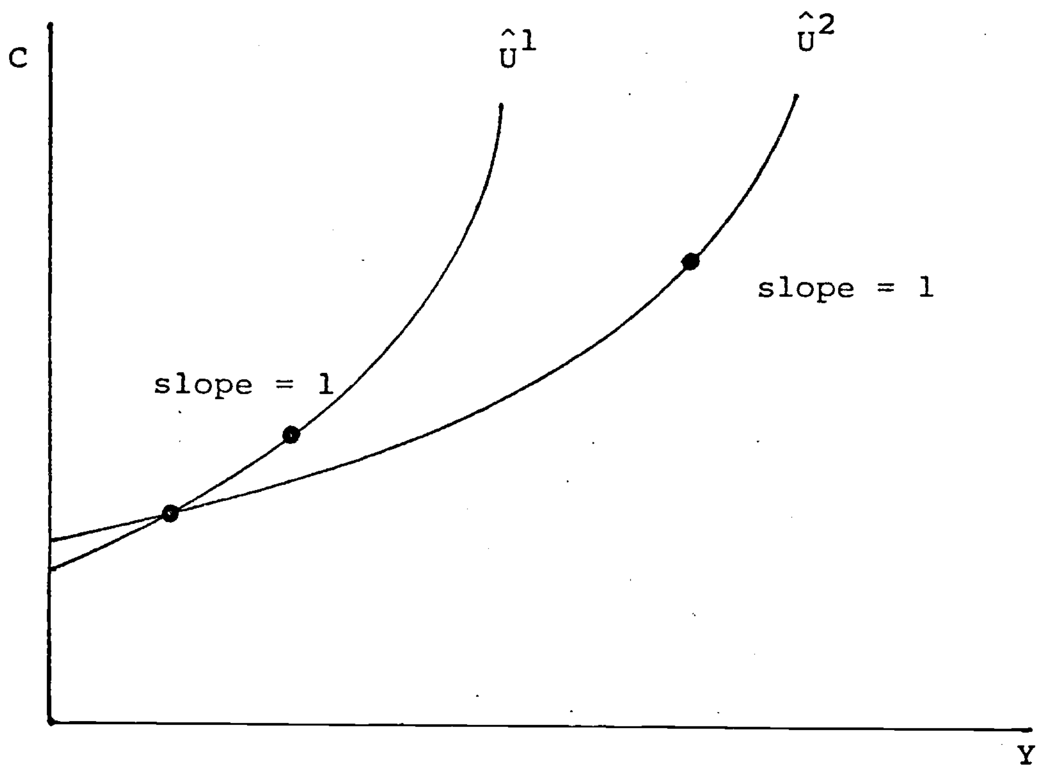


Figure 3b
First Best Taxation Not Fully Revealing: Pareto Optimal Taxation Entails Positive Tax Rate

is just the dual to the standard problem of a monopolist attempting to differentiate among his customers (Stiglitz 1977, 1981).

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 constraint. The Lagrangian which we form to analyze the two problems is identical:

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

The first order conditions for this problem are straightforward:

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

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

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

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

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

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

or

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

i.e. only one of the two self-selection constraints is binding provided that, with first best taxation, the equilibrium is not fully revealing. (In the two group case, it is possible that the first best tax structure is fully revealing, as illustrated in figure 3a.) Moreover, it is also easy to show that $\mu > 0$, the constraint on the utility level of the low ability individuals is binding.

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 ($\mu = 1$) and separable utility functions it can, for instance, 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.

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

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

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

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

Dividing (14b) by (14a) ,[1]

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

---

[1] The individual maximizes

$$\hat{U}(Y - T(Y), Y)$$

where $T(Y)$ is the tax function. Hence

$$-\frac{\partial \hat{U}}{\partial Y}\Big/\frac{\partial \hat{U}}{\partial C} = 1 - T'(Y) < 1 \quad \text{implies} \quad T' > 0.$$

Define

$$\alpha^i = -\frac{\partial\hat{U}^i/\partial Y_i}{\partial\hat{U}^i/\partial C_i}$$

and

$$\nu = \frac{\lambda_2 \partial\hat{U}^2/\partial C_2}{N_1 \gamma}$$

Then (15b) can be rewritten as

$$\alpha^1 = \frac{1 + \nu\alpha^2}{1 + \nu}$$

from which it follows that (Figure 3c) either

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

or

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

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

$$\alpha^2 < \alpha^1 < 1,$$

We immediately see that <u>the marginal tax rate faced by the less able individual will be positive; it will be greater the smaller the proportion of low ability individuals there are in the population</u>.

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. (See Figure 3d.) For the rest of this paper, we focus our attention on the "normal" case with $\lambda_1 > 0$, $\lambda_2 = 0$.

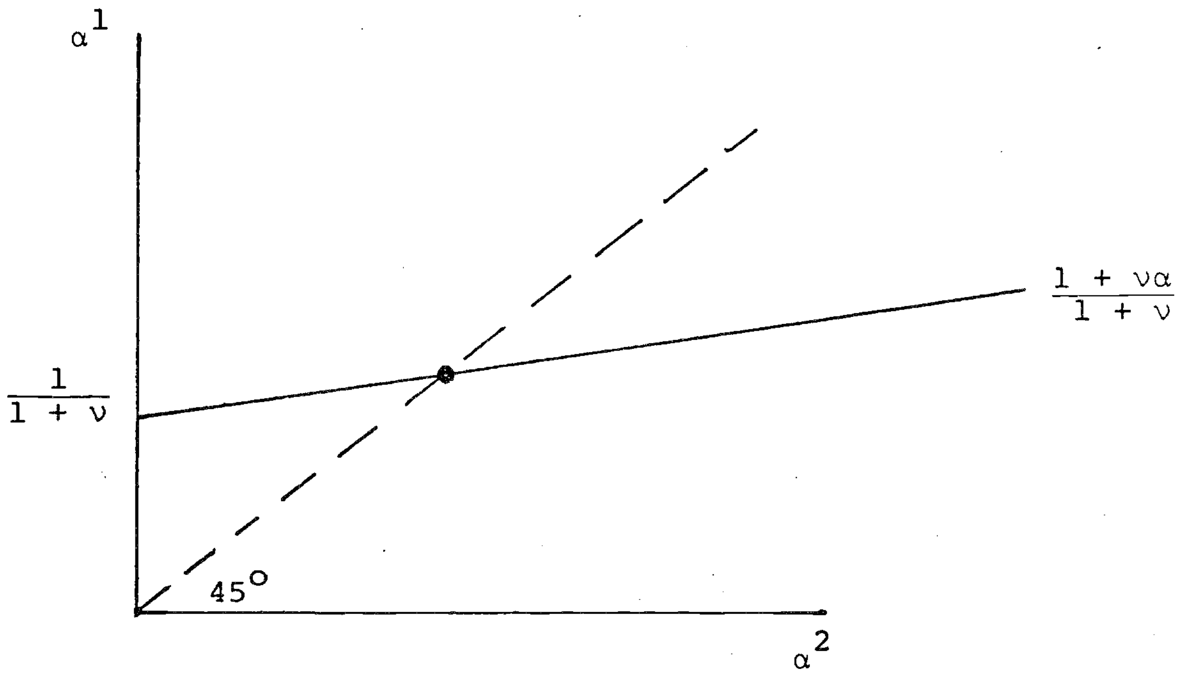


Figure 3c

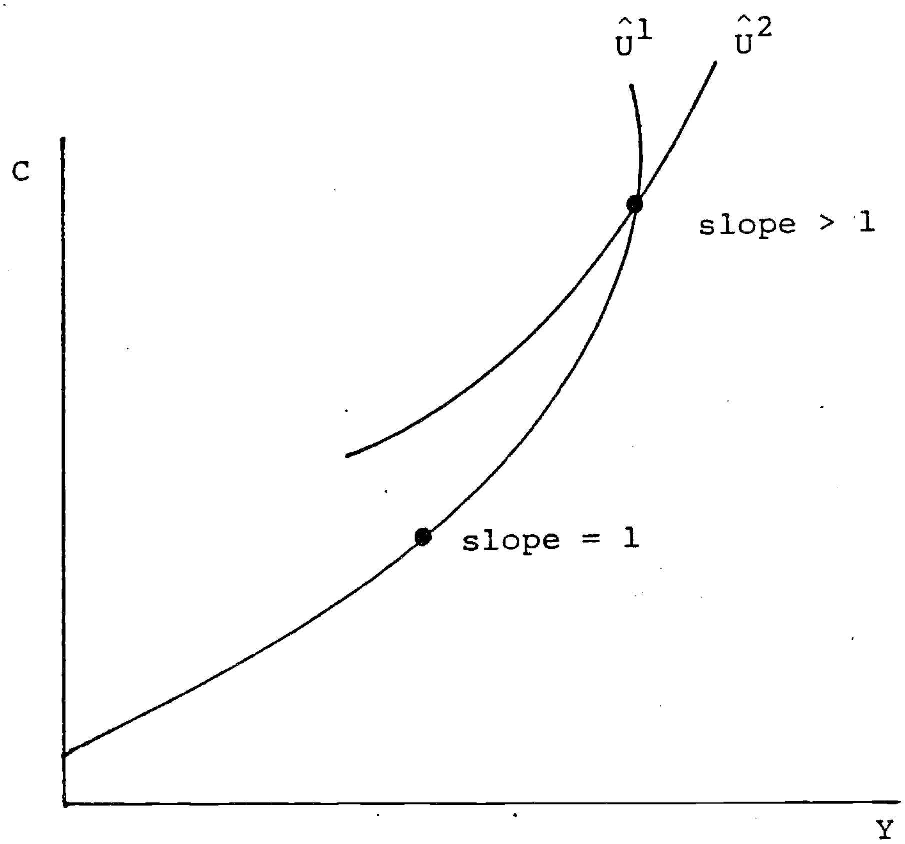


Figure 3d

3. Desirability of Randomization

In this section we derive conditions under which randomization of taxes is desirable. 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.) It is easy to establish that it is never

desirable to randomize the tax imposed on the upper income individual. Randomizing the tax (the after tax income) enjoyed by the low ability group lowers their welfare, at the same average tax rate. To leave them at the same level of expected utility, we must, at each Y, increase the mean consumption, as illustrated in figure 4a′. 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" contract entails a higher Y and a higher average level of consumption, $\overline{C}_1$; and it is possible that $\overline{C}_1$ has increased by less than Y_1, so that the government revenue is increased.

The derivation of the necessary and sufficient conditions for the desirability of a small amount of randomization are straightforward.

Let

$$C_1 = \bar{C}_1 + \Delta \qquad \text{with probability} \quad .5$$
$$C_1 = \bar{C}_1 - \Delta \qquad \text{with probability} \quad .5$$

The equilibrium separating contract generating utility levels U^1 and U^2 to the two groups is defined by the pair of equations[1]

(16a) $$\bar{U}_1 = \frac{\hat{U}^1(\bar{C}_1 + \Delta, Y_1) + \hat{U}^1(\bar{C} - \Delta, Y_1)}{2} ,$$

(16b) $$\bar{U}_2 = \frac{\hat{U}^2(\bar{C}_1 + \Delta, Y_1) + \hat{U}^2(\bar{C} - \Delta, Y_1)}{2} .$$

Government revenues are

(17) $$R = (Y_1 - \bar{C}_1)N_1 + (Y_2 - C_2)N_2 \cdot$$

We simply need to calculate, as we increase Δ , whether Y_1 increases more or less rapidly than C_1 ; in the former case randomization is desirable.

Formally we solve (16a) for

(18) $$Y_1 = \phi(\bar{C}_1, \Delta) \cdot$$

Substituting into (16b), and differentiating we obtain

(19) $$-\frac{d\bar{C}_1}{d\Delta} = \frac{\left[\frac{\partial\hat{U}^2}{\partial C_1}(\bar{C}_1 + \Delta, Y_1) - \frac{\partial\hat{U}^2(\bar{C} - \Delta, Y_1)}{\partial C_1}\right] + \frac{\partial Y_1}{\partial\Delta}\left[\frac{\partial\hat{U}^2(\bar{C}_1 + \Delta, Y_1)}{\partial Y_1} - \frac{\partial\hat{U}^2(\bar{C} - \Delta, Y_1)}{\partial Y_1}\right]}{\left[\frac{\partial\hat{U}^2(\bar{C}_1 + \Delta, Y_1)}{\partial C_1} + \frac{\partial\hat{U}^2(\bar{C} - \Delta, Y_1)}{\partial C_1}\right] + \frac{\partial Y_1}{\partial C_1}\left[\frac{\partial\hat{U}^2(\bar{C}_1 + \Delta, Y_1)}{\partial Y_1} + \frac{\partial\hat{U}^2(\bar{C} - \Delta, Y_1)}{\partial Y_1}\right]}$$

[1] In the following discussion, we drop the subscript of $\bar{C}_1$, where there is no ambiguity as a result.

where

(20a) $$-\frac{\partial Y_1}{\partial \Delta} = \left(\frac{\partial \hat{U}^1(\bar{C}_1 + \Delta, Y_1)}{\partial C_1} - \frac{\partial \hat{U}^1(\bar{C}_1 - \Delta, Y_1)}{\partial C_1}\right) \Bigg/ \left(\frac{\partial \hat{U}^1(C_1 + \Delta, Y_1)}{\partial Y_1} + \frac{\partial \hat{U}^1(C_1 - \Delta, Y_1)}{\partial Y_1}\right),$$

(20b) $$\frac{\partial Y_1}{\partial \bar{C}} = -\left(\frac{\partial \hat{U}^1(\bar{C}_1 + \Delta, Y_1)}{\partial C_1} + \frac{\partial \hat{U}^1(\bar{C}_1 - \Delta, Y_1)}{\partial C_1}\right) \Bigg/ \left(\frac{\partial \hat{U}^1(C_1 + \Delta, Y_1)}{\partial Y_1} + \frac{\partial \hat{U}^1(C_1 - \Delta, Y_1)}{\partial Y_1}\right).$$

Using (18), we calculate

(21) $$\frac{dY_1}{d\Delta} = \frac{\partial Y_1}{\partial \Delta} + \frac{\partial Y_1}{\partial \bar{C}_1} \frac{dC_1}{d\Delta}$$

so

(22) $$\frac{dY_1}{d\Delta} - \frac{d\bar{C}_1}{d\Delta} = \frac{\partial Y_1}{\partial \Delta} + \frac{d\bar{C}_1}{d\Delta}\left(\frac{\partial Y_1}{\partial \bar{C}_1} - 1\right).$$

Let

(23a) $$\bar{U}_1^i = \frac{\partial \hat{U}^i(\bar{C}_1 + \Delta, Y_1)}{\partial C_1} + \frac{\partial \hat{U}^i(\bar{C}_1 - \Delta, Y_1)}{\partial C_1}$$

(23b) $$\Delta U_1^i = \frac{\partial \hat{U}^i(\bar{C}_1 + \Delta, Y_1)}{\partial C_1} - \frac{\partial \hat{U}^i(\bar{C}_1 - \Delta, Y_1)}{\partial C_1}$$

and define $\bar{U}_2^i$ and ΔU_2^i as the corresponding derivatives with respect to Y. Then we can write

(24) $$\frac{1}{N_1}\frac{dR}{d\Delta} = \frac{\Delta U_1^2 - \frac{\Delta U_1^1}{\bar{U}_2^1}\Delta U_2^2}{\bar{U}_1^2 - \frac{\bar{U}_1^1}{\bar{U}_2^1}\bar{U}_2^2} \cdot \left(\frac{\bar{U}_1^1}{\bar{U}_2^1} + 1\right) - \frac{\Delta U_1^1}{\bar{U}_2^1}$$

It is immediate that

(25) $$\left.\frac{dR}{d\Delta}\right|_{\Delta = 0} = 0$$

and

(26) $$\left.\frac{d^2R}{d\Delta^2}\right|_{\Delta = 0} = \frac{U_{11}^2 \tau/1-\tau}{U_1^2\left(1 - \frac{MRS^1}{MRS^2}\right)} - \frac{U_{11}^1}{U_1^1} MRS^1$$

$$= \left(\rho^2 \frac{MRS^2/MRS^1}{MRS^2 - MRS^1} MRS^1 + 1 - \rho^1\right)\frac{MRS^1}{C_1}$$

where

(27a) $$MRS^i = -\frac{\partial\hat{U}^i/\partial C_1}{\partial\hat{U}^i/\partial Y_1}$$

(27b) $\frac{\tau}{1-\tau} = MRS^1 - 1$, marginal tax rate on incomes on individual 1

(27) $\rho^i = \frac{U_{11}^i C_1}{U_1^i}$, measure of relative risk aversion

Thus, the desirability of random taxation depends on three factors:

(i) the magnitude of the marginal distortion imposed by the non-random tax (the effective marginal tax rate)

(2) the differences in relative risk aversion

and

(3) the differences in the marginal rates of substitution (the slopes of the indifference curves) in $\{C, Y\}$ space).

We have described the conditions under which a particular (but natural) kind of randomization is desirable: the individual is told his tax liability only after he has filled in his tax form; he makes his work decision, of course, before he knows what his tax will be.

The government could, however, have randomized its tax schedules prior to the individual undertaking his work decision. 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.

To see what conditions are required for randomization, let

$$Y_1^* = Y_1 - p, \quad C_1^* = C_1 - n$$

$$Y_1^{**} = Y_1 + p, \quad C_1^{**} = C_1 + h$$

with

$$\hat{U}^2(C_1 - n, Y_1 - p) + \hat{U}^2(C_1 + h, Y_1 + p) \overset{(\leq)}{=} \bar{U}^2 \tag{28a}$$

$$\hat{U}^1(C_1 - n, Y_1 - p) + \hat{U}^1(C_1 + h, Y_1 + p) \overset{(\geq)}{=} \bar{U}^1. \tag{28b}$$

The first constraint is the self-selection constraint, the second assures us that $E\hat{U}^1$ is not lowered by randomization.

Randomization is desirable provided $n > h$. Differentiating (28) we obtain

(29)
$$\begin{bmatrix} -\hat{U}^2_1 & \hat{U}^2_1 \\ -\hat{U}^1_1 & \hat{U}^1_1 \end{bmatrix} \begin{bmatrix} dn \\ dh \end{bmatrix} = \begin{bmatrix} \hat{U}^2_2(C_1 - n, Y_1 - p) - \hat{U}^2_2(C_1 + h, Y_1 + p) \\ \hat{U}^1_2(C_1 - n, Y_1 - p) - \hat{U}^1_2(C_1 + h, Y_1 + p) \end{bmatrix} dp$$

Thus

(30)
$$\frac{d(n-h)}{dp} = \frac{-\Delta\hat{U}^2_2 \cdot \Delta\hat{U}^1_1 + \Delta\hat{U}^1_2 \cdot \Delta\hat{U}^2_1}{D} .$$

Differentiating the numerator twice with respect to p, we obtain at $p = 0$

$$4 \{ -\hat{U}^2_{21} \hat{U}^1_{11} + \hat{U}^1_{21} \hat{U}^2_{11} \}$$

Thus, recalling the definition of $\hat{U}^i$, provided

$$\frac{U_{122}}{U_{121}} \neq \frac{U_{12}}{U_{11}}$$

randomization is desirable.[1]

---

[1] For randomization to be desirable, all we require is that there exists some $\{p,h,n\}$ satisfying (28) such that $n > h$. n and h do not need to be positive.

We have thus shown how, under fairly weak conditions, randomization may enable a weakening of the self-selection constraints, and therefore an increase in expected utility. There is a quite different kind of randomization noted in Stiglitz (1976) where the maximized value of (expected) utility is observed to be a (locally) convex function of revenue raised when distortionary taxation is imposed. In that case, utilitarianism requires randomization (ex post horizontal inequity). A simple example illustrating this, in the present context, is provided by the family of indifference curves of Figure 4b . This has two critical properties. For each level of Y(work), there is a saturation level of consumption C(Y). For {C,Y} smaller than the critical level, indifference curves are straight lines; for convenience, we assume they have a slope of β.

Thus the optimal tax problem can be represented as

$$\max \quad U = (C_2 - Y_2\beta)N_2 + C_1N_1$$

subject to

$$C_2 \leq C_2(Y_2), \; C_2 \geq C_1 + Y_2\beta$$

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

where

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

$$C_1 = C_2(Y_2) - Y_2\beta$$

(It is easy to show, for this problem, that $Y_1 = 0$.) Since the self-selection constraint will be binding

$$U = C_1N$$

Since

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

$$\frac{\partial U}{\partial \bar{R}} = N \frac{dC_1}{dR} = N \frac{(C_2' - \beta)}{(1 - C_2')N_2 - (C_2' - \beta)N_1} .$$

Hence

$$\frac{\partial \ln \partial U/\partial \bar{R}}{\partial \bar{R}} = C_2''[\frac{1}{C_2' - \beta} + \frac{1}{(1 - C_2')\frac{N_2}{N} - (C_2' - \beta)\frac{N_1}{N})]}] \frac{dY_2}{d\bar{R}}$$

$$= \frac{C_2''(1 - \beta)N_2/N}{(C_2' - \beta)[(1 - C_2')\frac{N_2}{N} - (C_2' - \beta)\frac{N_1}{N_1})]^2}$$

which can be either positive or negative.[1] Thus, rather than raising $\bar{R}$ from the population in a "uniform" manner, it pays to divide the population arbitrarily into two groups, raising $\bar{R} - \Delta$ from one-half, $\bar{R} + \Delta$ from the other. (See Figure 4c.)

---

[1] 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.

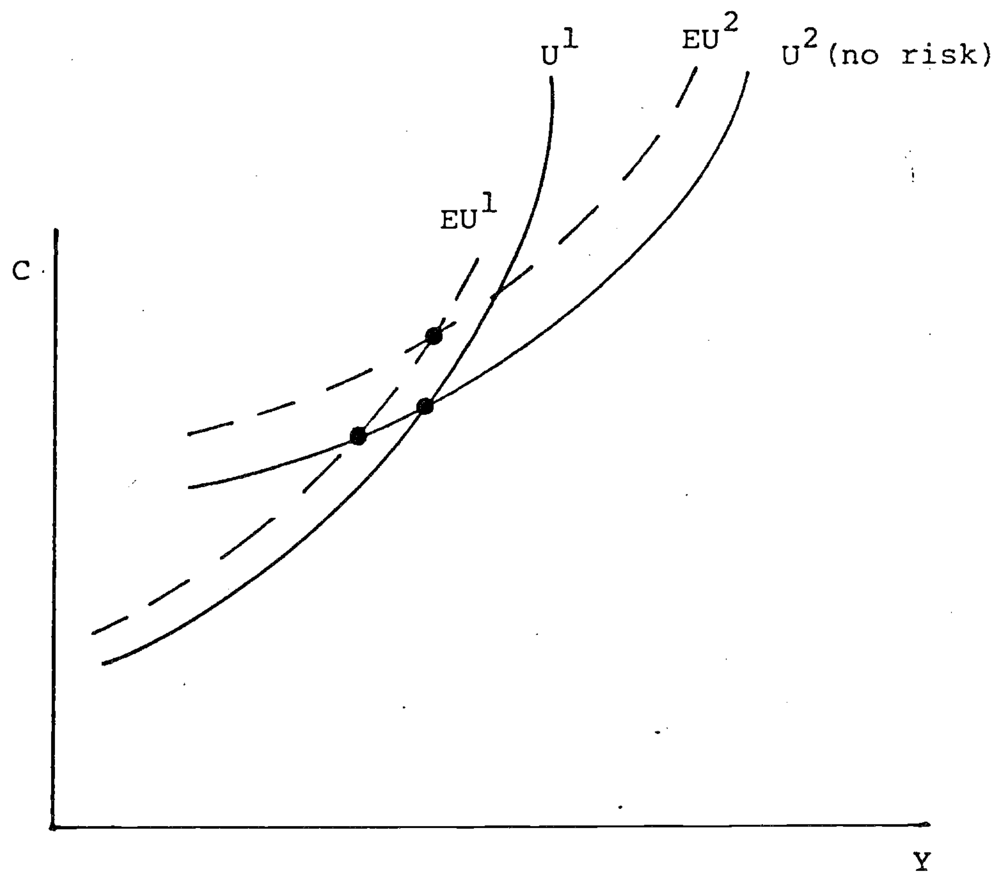


Figure 4a

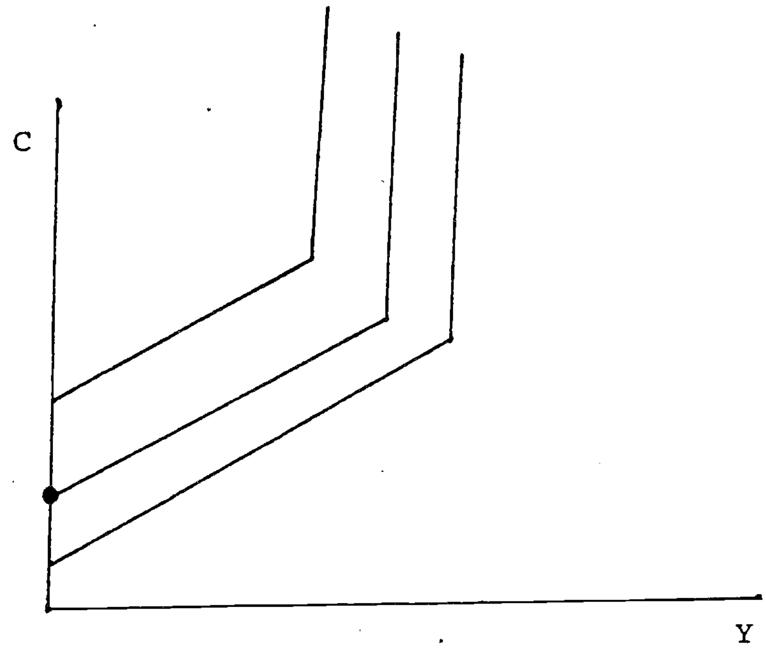


Figure 4b

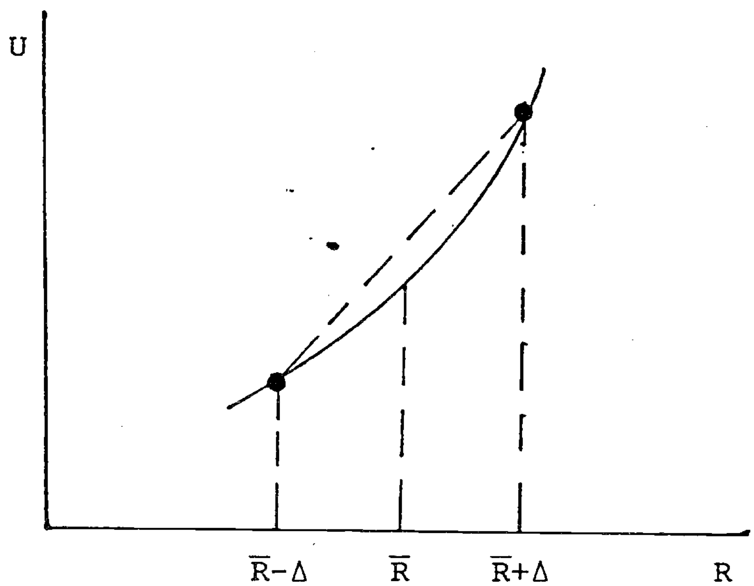


Figure 4c

3. Utilitarian Optimal Taxes

The previous sections focused on 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

$$\max \quad U^1N_1 + U^2N_2$$

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 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 had U_j^i , we now have and $\Phi' U_j^i$; similarly for the second individual. Since Φ' can take on any value, it is clear that the first order conditions describing pareto efficient taxation and those that describe utilitarian tax structure for appropriately specified Φ function are equivalent.

5. 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 individuals who are different are treated the same (and in which, as a result, we cannot infer perfectly the characteristics of the individuals) are referred to as pooling equilibria (Rothschild-Stiglitz, 1976), and it can be shown that pooling equilibria can arise in a variety of circumstances (Stiglitz, 1977). Here, we show (i) if there are two groups, and the more productive groups indifference curves have a flatter slope in {C, Y} space then differentiation is desirable;
(ii) if more productive groups have 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;

(iv) if individuals differ in tastes as well as abilities, then complete differentiation will not, in general, be possible.

To see the first result, we have depicted in figure 5 a case with two groups of individuals "pooled" together. Any point in the shaded area generates a separating equilibrium, and any point along the lower envelope of 1 and 2's indifference curves separates and leaves the welfare of each group unaffected. We need to see what happens to government revenue. If

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

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

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

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

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

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

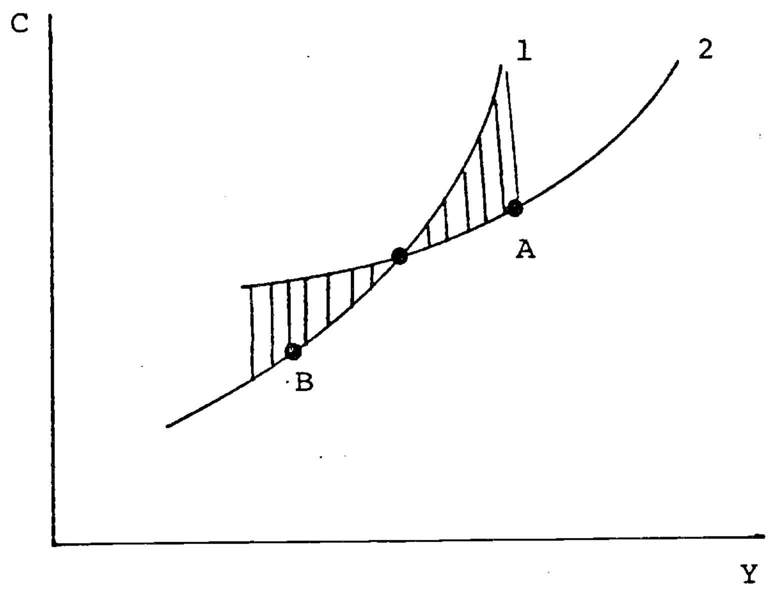


Figure 5

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

then,

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

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 that

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

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

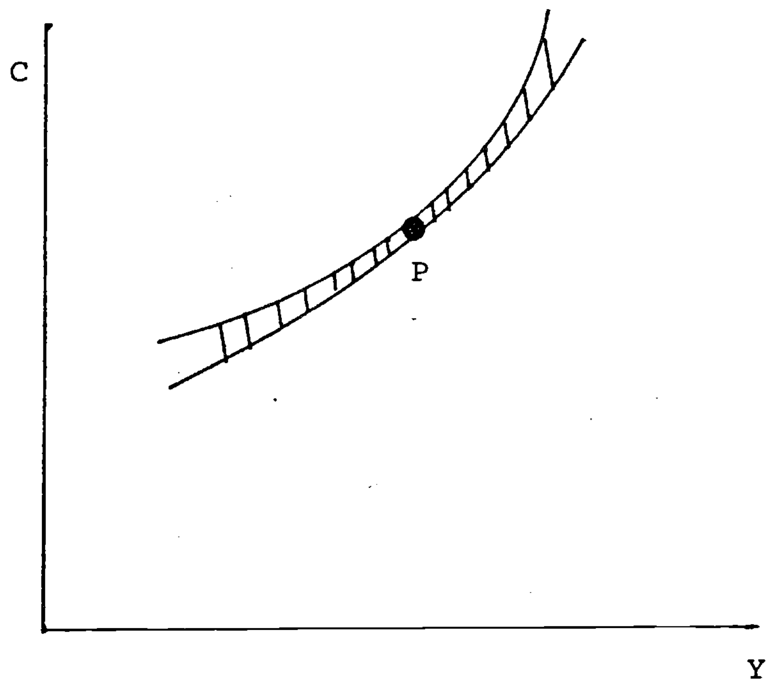


Figure 6

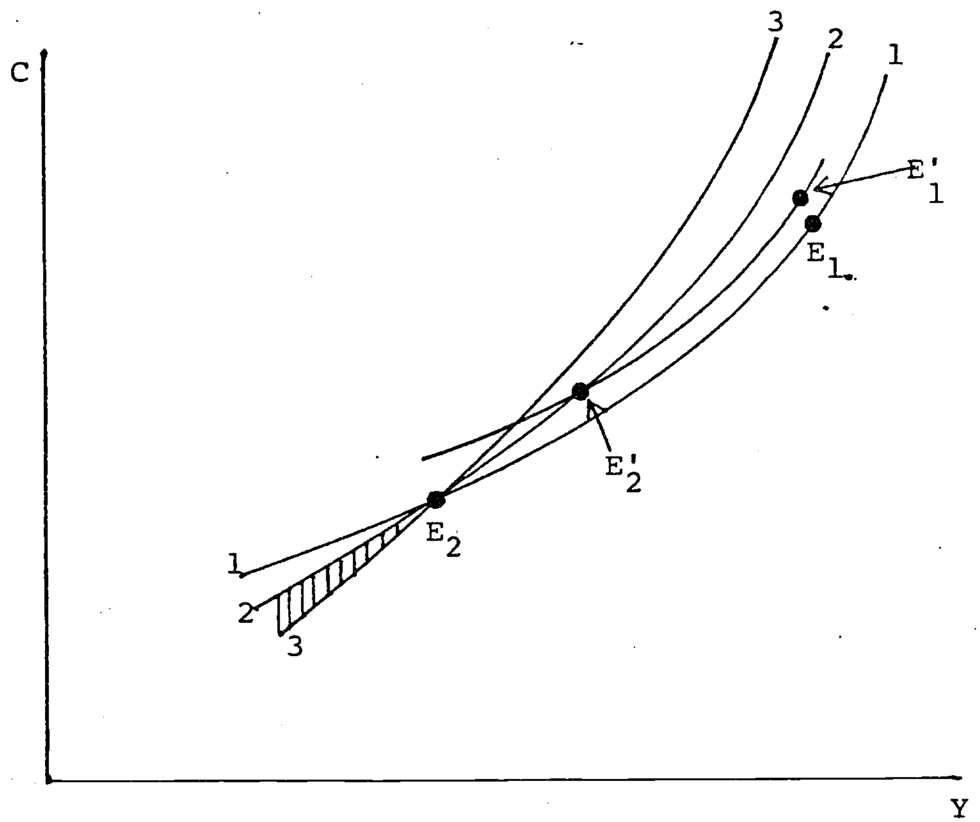


Figure 7

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,

(32a) $$\left(\frac{dC}{dY}\right)_{U^3} < 1$$

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

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

(32b) $$\left(\frac{dC}{dY}\right)_{U^2} < 1 .$$

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.

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, which have the property that individuals with different marginal rates of substitution obtain exactly the same income (Figure 8) .

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,

$$U = u(C) - v(L)$$

$$= u(C) - v(Y/w_.) .$$

The marginal rate of substitution at any point is just

$$\left(\frac{dC}{dY}\right)_{\bar{U}} = \frac{v'}{u'} \frac{1}{w} .$$

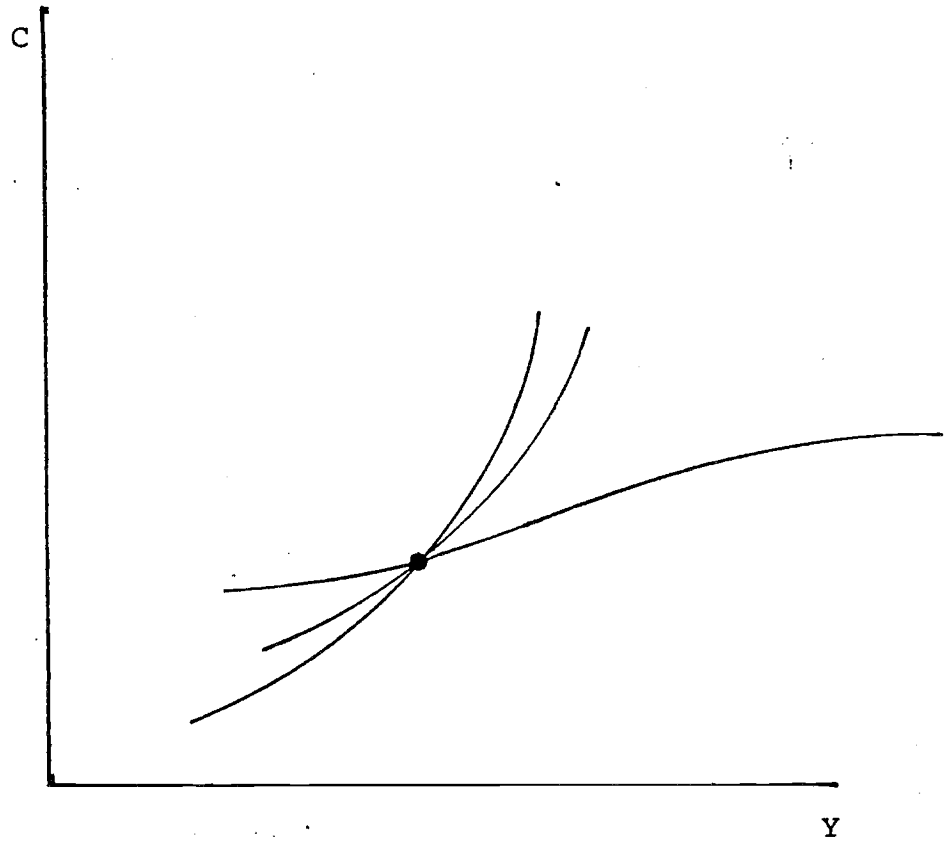


Figure 8

Kinked Optimal Tax Structure

Thus

$$\left(\frac{dC}{dY}\right)_{U^1} - \left(\frac{dC}{dY}\right)_{U^2} = \frac{1}{u'}\left[\frac{v'}{w_1} - \frac{v'}{w_2}\right]$$

But

$$\frac{w_1 Y}{v'} \left.\frac{d[v'/w_1 - v'/w_2]}{dY}\right|_{v'/w_1 = v'/w_2} = \frac{v''(L_1)L_1}{v'(L_1)} - \frac{v''(L_2)L_2}{v'(L_2)} \gtrless 0 \; .$$

Thus, the higher ability individual could have a greater aversion to work, but if the elasticity of his marginal aversion to work function is less than that of the low ability individual, for high enough income levels, his indifference curve is flatter than the low ability individual's indifference curve. At low levels of income, however, his indifference curve is steeper. This is the case illustrated in the figure 9a . The converse case is also clearly possible.

The result that their indifference curves could intersect several times (or indeed may completely coincide) should not be surprising. An individual who has a productivity of k times another, and an individual who receives "disutility" of work of 1/k times another are indistinguishable on the basis of their indifference curves in {C,Y} space; it is their indifference curves in {C,Y} space which provide the basis of the self-selection mechanism. Note that there may be 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

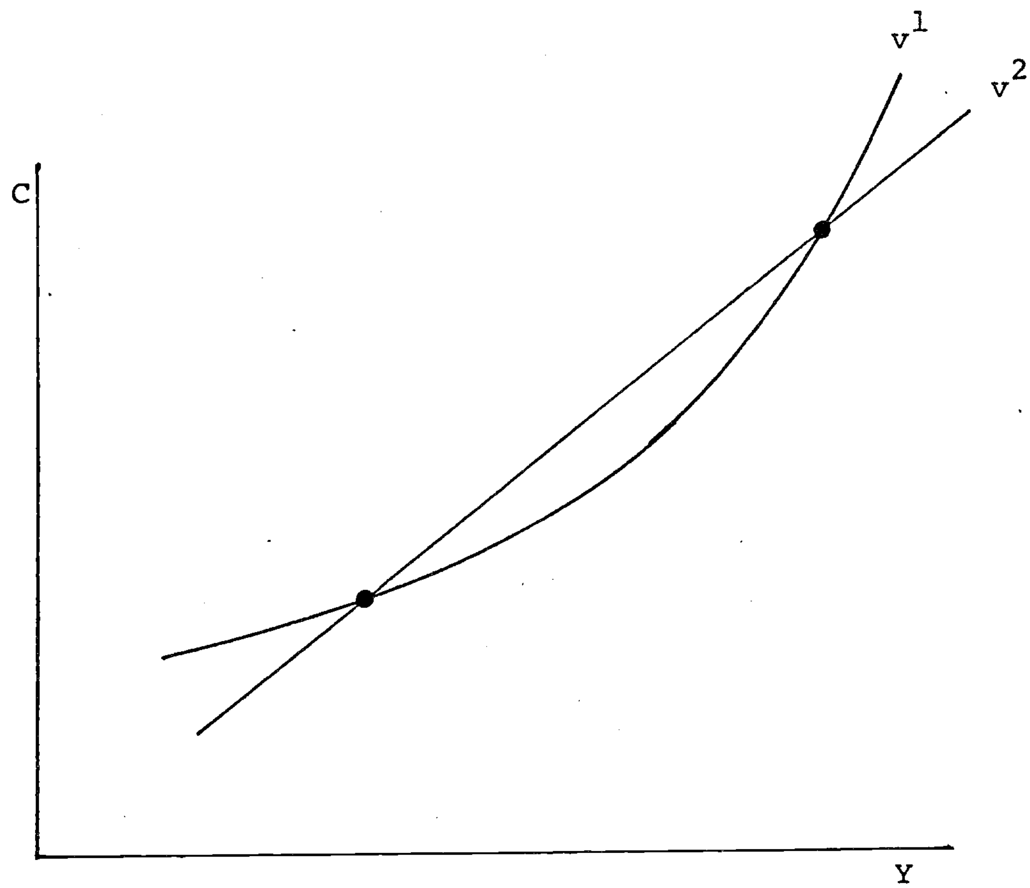


Figure 9a

which are complements of leisure, and use this as a basis of inferring their ability. We shall examine this possibility in greater detail in Section 6 .

In Figure 9b we show a tax structure with three points, E_1, E_2 and E_3; the high ability low aversion to work individuals choose E_1; the high ability, high aversion to work individuals choose E_2; the two low ability individuals choose E_3. Note that there is no point near E_3 which separates (i.e. lies between the two indifference curves) and also lies below the indifference curve of the high ability-high aversion to work individual.

This, of course, is not particularly disturbing. One might want to argue that one does not wish to differentiate between individuals on the basis of their attitudes towards work, only on their ability (but see Section 8 below). But now, let us reinterpret our result: let U^2 be the low ability low aversion to work individual, and U^3 be the high ability high aversion to work individual. (From our previous calculations we know that this is a possible configuration.) Then our analysis shows that efficient taxation may entail treating high ability lazy individuals identically to low ability hard working individuals.[1]

---

1 Although we have only established the inability to differentiate locally, it is easy to extend the arguments to show that the equilibrium may be Pareto efficient. Let $\{C^A, Y^A\}$ and $\{C^B, Y^B\}$ represent the two nearest points to E_3 which separate. Then we require

$$C^3 - C^A < Y^3 - Y^A$$
$$C^B - C^3 > Y^B - Y^3.$$

Moreover, if types 3 and 4 are small relative to types 1 and 2, any movement within the shaded area which improves U^2's welfare (and possibly, as a consequence U^1's welfare) will decrease revenue.

(34) $\max U^2(C_2, L_2)$

(34a) s.t. $U^1(C_1, L_1) \geq \bar{U}^1$

(34b) $F(N_1L_1, N_2L_2) - N_1C_1 - N_2C_2 \geq \bar{R}$

and subject to the self selection constraints. These require a little care to write down correctly. The government, it must be recalled, does not observe L_1 and L_2. It only observes Y_1 and Y_2. If the wage of the ith group is w_i, if the first group has an income of Y_1 and labor input of L_1, for the second group to have the same income as the first group requires a labor input of

$$\hat{L}_2 = \frac{L_1 w_1}{w_2} .$$

Using (33), we can rewrite this as

$$\text{(35)} \quad \hat{L}_2 = L_1 \left(\frac{f - \frac{N_2 L_2}{N_1 L_1} f'}{f'} \right) \equiv L_1 \, \varphi \left(\frac{L_2}{L_1} \right), \quad \varphi' > 0$$

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

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

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

We form the Lagrangian

$$\mathcal{L} = U^2(C_2, L_2) + \mu U^1(C_1, L_1) + \gamma(F(N_1L_1, N_2L_2) - N_1C_1 - N_2C_2 - \bar{R})$$

$$+ \lambda_2(U^2(C_2, L_2) - U^2(C_1, L_1\varphi(\frac{L_2}{L_1})))$$

$$+ \lambda_1(U^1(C_1, L_1) - U^1(C_2, L_2/\varphi(L_2/L_1))) .$$

As before, we can easily show $\lambda_1 = 0$, $\mu > 0$. We obtain first order conditions analogous to those derived earlier:

(36a) $$\frac{\partial \mathcal{L}}{\partial C_1} = \mu \frac{\partial U^1}{\partial C_1} - \lambda_2 \frac{\partial U^2}{\partial C_1} - \gamma N_1 = 0$$

(36b) $$\frac{\partial \mathcal{L}}{\partial L_1} = \mu \frac{\partial U^1}{\partial L_1} - \lambda_2 \frac{\partial U^2}{\partial L_1} (\phi - \frac{L_2}{L_1} \phi') + \gamma F_1 N_1 = 0$$

(36c) $$\frac{\partial \mathcal{L}}{\partial C_2} = \frac{\partial U^2}{\partial C_2} + \lambda_2 \frac{\partial U^2}{\partial C_2} - \gamma N_2 = 0$$

(36d) $$\frac{\partial \mathcal{L}}{\partial L_2} = \frac{\partial U^2}{\partial L_2} + \lambda_2 (\frac{\partial U^2}{\partial L_2} - \frac{\partial U^2}{\partial L_2} \phi') + \gamma F_2 N_2 = 0$$

Dividing (36d) by (36c), we obtain

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

If the two types of labor are not perfect substitutes, then the marginal tax rate on the most able individual should be negative. Dividing (36b) by (36a) we obtain

$$(37) \qquad -\frac{\partial U^1/\partial L_1}{\partial U^1/\partial C_1} = \frac{F_1 + \lambda_2 \frac{\partial U^2}{\partial C_1}\left(-\frac{\partial U^2/\partial L_1}{\partial U^2/\partial C_1}\phi\right)\left(1 - \frac{1}{\sigma}\right)}{1 + \lambda_2 \frac{\partial U^2}{\partial C_1}}$$

where we have made use of the fact that

$$\varphi - \frac{L_2}{L_1}\varphi' = \frac{f - \frac{N_2 L_2}{N_1 L_1} f'}{f'} + \left(\frac{N_2 L_2}{N_1 L_1}\right)^2 \frac{f''}{f'} + \frac{f - \frac{N_2 L_2}{N_1 L_1} f'}{[f']^2} \frac{N_2 L_2}{N_1 L_1} f''$$

$$= \frac{f - \frac{N_2 L_2}{N_1 L_1} f'}{f'} + \frac{f}{f'^2} \frac{N_2 L_2}{N_1 L_1} f''$$

$$= \frac{w_1}{w_2}\left(1 - \frac{1}{\sigma}\right).$$

where σ = the elasticity of substitution.

Since (assuming $w_1 < w_2$)

$$-\frac{1}{w_2}\frac{\partial U^2/\partial L_2}{\partial U^2/\partial C_2} < -\frac{\partial U^1/\partial L_1}{w_1 \partial U^1/\partial C_1}$$

from (37),

$$-\frac{\partial U^1/\partial L_1}{\partial U^1/\partial C_1} < \frac{F_1 + (\lambda_2 \partial U^2/\partial C_1) \qquad \phi\left(-\frac{\partial U^2/\partial L_1}{\partial U^2/\partial C_1}\right)}{1 + \lambda_2 \partial U^2/\partial C_1}$$

$$< \frac{F_1 + \lambda_2 \, \partial U^2/\partial C_1 \quad \left(-\frac{\partial U^1/\partial L_1}{\partial U^1/\partial C_1}\right)}{1 + \lambda_2 \partial U^2/\partial C_1}$$

i.e.,

$$(38) \qquad -\frac{\partial U^1/\partial L_1}{\partial U^1/\partial C_1} < F_1,$$

the low ability individual faces a positive marginal tax rate, the magnitude of which depends on the elasticity of substitution.

7. Simultaneous Taxation of Income and Commodities

Our earlier discussion suggested that if not only income, but also the levels of consumption of various commodities were observable, that the government might want to base its taxation on these variables as well. Consumption of luxuries is often thought to be a better indicator of well-being than reported income.

This problem can easily be analyzed within our framework. We now let the individual's utility be a function of a whole vector of consumption goods,

$$\underline{C}_1 = \{C_{11}, C_{12}, C_{13}, \ldots.\}$$

$$\underline{C}_2 = \{C_{21}, C_{22}, C_{23}, \ldots.\} \quad .$$

For simplicity, we assume that each of the goods costs one unit of labor to produce (this is just a choice of units). The individual is given a choice of two "packages"; now each involves a vector of consumption goods and a level of before tax income. The government must choose these packages to maximize individual 1's utility, subject to individual 2 obtaining a given level of utility, and subject to the self section and budget constraints.

If we now interpret C as a vector, the Lagrangian for this problem is identical to that formulated earlier, except for the term in the budget constraint. The government budget constraint is now written

$$R = N_1Y_1 + N_2Y_2 - (N_1\underline{C}_1 + N_2\underline{C}_2) \cdot e$$

where e is the unit vector, i.e.

$$R \geq N_1Y_1 + Y_2N_2 - \sum_j (C_{1j}N_1 + C_{2j}N_2) \quad .$$

If we now differentiate the Lagrangian with respect to C_{ij}, we obtain

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

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

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

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

It is easy to show, as before, that $\lambda_1 = 0$: only the second self-selecting constraint is binding. From (39a) and (39b), we obtain

(40a) $$\frac{\partial U^2/\partial C_{2i}}{\partial U^2/\partial C_{2k}} = 1$$

$$\text{(40b)} \quad \frac{\partial U^1/\partial c_{1j}}{\partial U^1/\partial c_{1k}} = \frac{N_1\gamma-\lambda_2\partial U^2/\partial c_{1j}}{N_1\gamma-\lambda_2\partial U^2/\partial c_{1k}}$$

(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.

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

Then

$$\text{(42a)} \quad \partial U^2/\partial c_{1j} = \partial U^1/\partial c_{1j}$$

$$\text{(42b)} \quad \partial U^2/\partial c_{1k} = \partial U^1/\partial c_{1k}$$

and (40b) becomes

$$\text{(43)} \quad \frac{\partial U^1/\partial c_{1j}}{\partial U^1/\partial c_{1k}} = 1.$$

<u>If leisure and goods are separable, there should be no commodity taxation</u>. If they are not, we obtain

$$\text{(44)} \quad \frac{\partial U^1}{\partial c_{1j}} - \frac{\partial U^1}{\partial c_{1k}} = \lambda_2 \left(\frac{\partial U^2}{\partial c_{1j}} - \frac{\partial U^2}{\partial c_{1k}}\right)$$

or

$$\text{(45)} \quad T_{jk} \equiv \frac{\partial U^1/\partial c_{1j}}{\partial U^1/\partial c_{1k}} - 1 = \lambda_2 \frac{\partial U^2/\partial c_{1j}}{\partial U^1/\partial c_{1k}} \left(1 - \frac{\partial U^2/\partial c_{1k}}{\partial U^2/\partial c_{1j}}\right)$$

$$= \frac{\lambda_2 \dfrac{\partial U^2/\partial c_{1k}}{\partial U^1/\partial c_{1j}}}{1+\lambda_2 \dfrac{\partial U^2/\partial c_{1k}}{\partial U^1/\partial c_{1j}}} \left(\frac{\partial U^1/\partial c_{1j}}{\partial U^1/\partial c_{1k}} - \frac{\partial U^2/\partial c_{1j}}{\partial U^2/\partial c_{1k}} \right)$$

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, we make the lower ability individuals "package" less attractive to him. We thus can tax the higher ability individual more heavily without having him trying to "disguise" himself as a low ability person.

Earlier, we remarked that, since the analysis of the discriminating monopolist and of pareto efficient taxation were formally identical, we could borrow results originally obtained in one area to the other. Here, we note that the result we have just obtained has immediate implications for the pricing policy of a multiproduct monopolist. If the individual's utility function is separable in "other goods" and the goods purchased from the monopolist, then the monopolist should charge relative prices of the different commodities equal to the marginal production costs;

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 analysis 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 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 single product case, with parallel results: now, even for the most able individual, we will wish to impose distortionary taxation (charge distortionary prices).

8. 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. As in Section 5 , 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 two propositions: First, it is always Pareto efficient to differentiate on the basis of tastes; we should never "pool" the two high ability groups together. Secondly, Pareto efficient taxation often will entail regressivity, i.e. marginal rates which are less than zero.

To see the first proposition, turn to Figure 10 where we have assumed that the government offers two contracts, E_1 and E_2, with both of the high ability groups at E_1. By the same kind of reasoning used earlier, clearly any point between the two indifference curves separates, and either

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

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 contracts 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 are zero. We have drawn through E_1' a line with a slope of 45^o. 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 is negative; on average, there is regressive taxation at the upper end of the distribution.

9. Stochastic Income

This result on the structure of the optimal income tax should not, however, be taken too seriously, 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

$$(w_i + \Delta)L$$

with probability .5 and

$$(w_i - \Delta)L$$

with probability .5. Assume, moreover, that he cannot insure the risk. As before, w and L are unobservable; only income is observable. The optimal tax structure now requires a specification of "two packages" as before, but the packages are more complicated. By deciding on a level of effort (L) the individual is essentially "purchasing" a lottery. The structure of the tax structure determines the pay-offs on the lottery. Thus, the government will specify four consumption-income

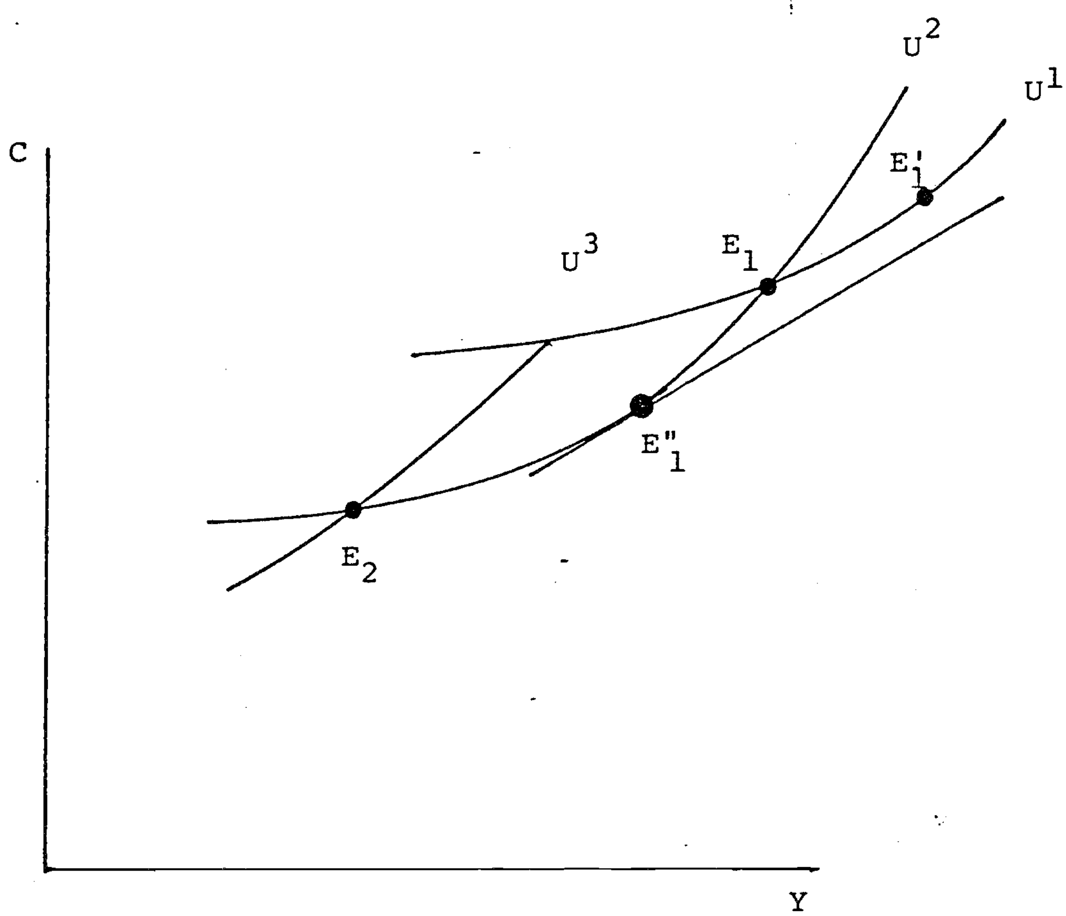


Figure 10

points, denoted $\{C_{1L},Y_{1L},C_{1H},Y_{1H},C_{2L},Y_{2L},C_{2H},Y_{2H}\}$ with the property that (expected) government income is maximized, subject to the self-selection constraints and subject to the (expected) utility constraints for each of the two types. The problem is thus formally identical to that discussed earlier.

We will, accordingly, not set up the problem, but we shall borrow one result from our earlier analysis: the "package" offered to the high ability individuals must be "non-distortionary," i.e. it maximizes the revenue obtained from him subject to the utility constraint. But if the individual is risk averse, this implies that he must receive the same consumption in the two states. But this, in turn, implies that the marginal tax rate on incomes in excess of Y_{2L} (letting individual 2 be the high ability individual) is 100%.

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 are a finite number of groups (or even a continuum, with a finite range) 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. Such a tax structure would have peculiar (and probably undesirable) incentive effects with respect to risk taking.

10. Concluding Comments

This paper has examined the structure of Pareto efficient taxation. Although we have greatly simplified the standard treatment, by focusing on the special case where there are only two groups, we have been able to obtain considerable insight into the determinants of the optimal structure of taxation. In particular, we have been able to show that assumptions that were previously taken to be merely simplifying turn out to play a central role in determining the optimal structure of taxation:

(a) if tax rates can be randomized, they should be under a variety of circumstances

(b) if different individuals are not perfect substitutes for one another, then the general equilibrium effects -- until now ignored in the literature -- of changes in the tax structure are dominant in determining the optimal tax structure; the marginal rate on the most able individual is always negative; on the less able individuals it is positive.

(c) if different individuals have different attitudes towards leisure, the tax structure, in the upper tail, may be regressive;

(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

F. Allen (1980), "Optimal Income Taxation with General Equilibrium Effects on Wages," mimeo., Oxford.

A. Atkinson and J. E. Stiglitz (1976), "The Design of Tax Structure: Direct versus Indirect Taxation," Journal of Public Economics, Vol. 6, pp. 55-75.

A. Atkinson and J. E. Stiglitz (1980), Lectures on Public Economics, New York and London: McGraw-Hill.

R. Arnott, 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.

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

M. Rothschild and J. E. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets," Quarterly Journal of Economics, Vol. 90, pp. 629-650.

N. Stern (1979), "Optimum Taxation with Errors in Administration," Warwick University mimeo.

J. E. Stiglitz (1976), "Utilitarianism and Horizontal Equity: The Case for Random Taxation," IMSSS Technical Report No. 214, Stanford University.

______________ (1977), "Monopoly, Non-linear Pricing and Imperfect Information: The Insurance Market," Review of Economic Studies, Vol XLIV, October, pp. 407-430.

______________ (1981), Information and Economic Analysis, Oxford University Press.