NBER WORKING PAPER SERIES

PARETO EFFICIENT TAX STRUCTURES

Dagobert L. Brito

Jonathan H. Hamilton

Steven M. Slutsky

Joseph E. Stiglitz

Working Paper No. 3288

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1990

We would like to thank the referees for detecting errors in the original draft and for suggesting generalizations and better lines of proof for some of the results. Joseph Stiglitz thanks the National Science Foundation, the Olin Foundation, and the Hoover Institute for support. This paper is part of NBER's research program in Taxation. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3288
March 1990

PARETO EFFICIENT TAX STRUCTURES

ABSTRACT

Most analyses of optimal income taxation make restrictive technical assumptions on preferences (such as single-crossing) and only derive properties of welfare-maximizing tax schedules. Here, for an economy with any finite numbers of groups and commodities, Pareto efficient tax structures are described assuming only continuity and monotonicity of preferences. Most results follow directly from a property of self-selection: at an optimum, one group will never envy the bundle of another group which pays a larger total tax. The bundle of a group paying the largest total tax is undistorted. Assuming normality, undistorted outcomes for a group form a connected segment on the constrained utility possibility frontier. The tax structure at distorted outcomes is also described.

Dagobert L. Brito
Department of Economics
Rice University
Houston, TX 77251

Jonathan H. Hamilton
Department of Economics
University of Florida
Gainesville, FL 32611

Steven M. Slutsky
Department of Economics
University of Florida
Gainesville, FL 32611

Joseph E. Stiglitz
Department of Economics
Stanford University
Stanford, CA 94305

I. Introduction

The important and influential literature growing out of Mirrlees' [1971] seminal paper on optimal income taxation has stressed the trade-offs between incentive and distributional considerations in the design of income tax schedules. There were, however, two limitations. First, by solving for the tax schedule which maximized a utilitarian (or other) social welfare function, it failed to make the distinction at the heart of the New Welfare Economics between the economist who was to present a policy maker with the feasible set and the policy maker who was to choose, according to her values, from among the feasible allocations. Second, it imposed the technical assumption that for individuals of different abilities, the derived indifference curves between before tax income and consumption could only cross once (the "single crossing" property). This condition, analogous to the condition on no factor reversals which had long played a central role in trade theory, is equally objectionable in this context; it rules out, for instance, the possibility that more able individuals might have an elasticity of substitution between leisure and consumption different from that of less able individuals. Additionally, it may be difficult to define natural single crossing restrictions when there are more than two commodities.

This paper attempts to provide a general description of Pareto efficient tax structures, without imposing single crossing. In the context of a finite set of different groups, Pareto efficient tax structures maximize the utility of one individual (group) given the utility of others and given the budget balance and informational constraints on the government. We consider both the utility possibility frontier and the tax schedules which are associated with each point on that frontier.

The approach taken in this paper has one further advantage: the elementary proofs employed in establishing the basic characterization results show that these results follow from properties of the self-selection constraints and do not depend on properties of the utility functions or on the distribution of ability types.[1] Not only are single crossing restrictions irrelevant, but for most results, so are other commonly employed restrictions such as differentiability, quasiconcavity, and normality. We further do not restrict ourselves to a model with only two commodities as in most of the optimal income taxation literature.

The paper is in the spirit of Guesnerie and Seade [1982] who also use simple direct arguments to consider Pareto efficient taxation in an economy with a finite number of types of consumers and only two commodities but without single crossing.[2] Some of our results reproduce theirs. The contribution of our paper is threefold. Besides imposing less restrictive assumptions than are used in their paper,[3] we show that most of the results follow directly as simple corollaries of a property arising from the imposition of self-selection constraints independent of the nature of preferences. This simple but powerful property is that, <u>at any efficient allocation</u>, one group will always view as strictly inferior the bundle of another group which has a larger total tax.[4] We also use this property to derive new results, including specifying central properties of the constrained utility possibility frontier.

II. The Model

The society we consider is composed of m different classes of individuals. Individuals within each class are identical but the classes may differ in tastes and abilities. The government knows the number of individuals in each class, denoted N_i, $i=1,\ldots m$, but initially does not know to which class any individual belongs. Individuals consume a vector $x \varepsilon R^n$, where $x_i > 0$ denotes a net purchase for consumption and $x_j < 0$ denotes a sale or factor supply. The components of the vector x are defined in units observable to the government, for example, labor income instead of labor hours in the standard optimal income tax model or net trades instead of final consumption of any commodity of which the consumer has an initial endourment. These commodities are produced at constant producer prices, $p \varepsilon R^n_+$. Individuals in each class have a closed consumption set X^i specifying possible consumption bundles. In addition to nonnegativity constraints for goods and nonpositivity constraints for supplies, X^i is bounded from below for factor supplies or net trades. For example, given limited time, there is a maximum labor income members of each class can earn.

Individuals in each class have a utility function $U^i(x^i)$ which is continuous and strictly monotonic on X^i. The utility of any bundle not in X^i can be assigned an arbitrarily low utility number guaranteeing that such bundles would never be chosen. Note that neither quasiconcavity of each individual's utility nor single crossing of preferences across types is assumed.

A special case of this structure is the standard optimal income taxation model with only two commodities, consumption and labor income. Additionally, in that model, it is often assumed that all classes have a common underlying utility function over consumption and labor hours but differ in ability, leading to different preferences over consumption and income. As shown by Sadka (1976) for that model, if consumption is not inferior, then at any consumption-income bundle, a higher ability individual would always need less additional consumption as compensation for working enough to earn an extra dollar than would a lower ability person. Thus, indifference curves of two ability classes could cross only once. In our model, even when restricted to the two commodity world, preferences of two classes can differ arbitrarily and hence single crossing is not imposed.

In this structure, the government imposes taxes to redistribute among classes and to raise an exogenously given amount of revenue G, which may be zero. Since the government cannot distinguish individuals of different classes, it must offer the same tax possibilities to everyone consuming the same bundle. In addition, either for administrative convenience or to satisfy horizontal equity, only one tax can be charged on each bundle. Thus, randomization, which otherwise might be optimal, is ruled out by assumption.[5]

Let $T(x)$ denote the tax charged on bundle x. Each individual then maximizes utility subject to the budget constraint $p \cdot x + T(x) \leq 0$, taking p and the function $T(\cdot)$ as given. A tax function is <u>allowable</u> if, for each i, there exists an $x^i \in X^i$ with $p \cdot x^i + T(x^i) \leq 0$. For each allowable tax schedule, there is an optimal consumption bundle for individuals in class i, denoted $x^i(T)$ and a utility level $U^i(x^i(T))$. Since all classes have the same budget set and since $x^i(T)$ maximizes for class i, then $U^i(x^i(T)) \geq U^i(x^j(T))$, all i and $j \neq i$. A tax function is <u>feasible</u> if it is allowable and if the revenue raised given the optimizing choices by individuals in each class is at least as great as the government's exogenous revenue requirement, $\Sigma_i N_i T(x^i) \geq G$. Taxes are specified as an arbitrary function of the entire bundle. Given the budget constraint for each individual, taxes actually could be defined only over bundles with n-1 components. For example, in the standard model with consumption and income, taxes need only be specified as functions of income.

Associated with each feasible tax function is the distribution of utilities $U^i(X^i(T))$, all i. One tax function T would Pareto dominate another $\hat{T}$ if $U^i(x^i(T)) \geq U^i(x^i(\hat{T}))$, all i, with strict inequality for some j. We seek to specify properties of the Pareto efficient tax functions and of the utility distributions associated with these functions. Since searching over tax functions directly is difficult, we consider the equivalent problem of choosing a set of allowable bundles, x^i, i=1,..., m, for which the implicit taxes $T^i(x) = - p\cdot x^i$ raise sufficient revenue, $\Sigma_i N_i p\cdot x^i + G \leq 0$, and which satisfy self-selection constraints that no individual can strictly prefer the bundle assigned to another class to that assigned to his own, $U^i(x^i) \geq U^i(x^j)$, all i and j. As argued above, budget balance and self-selection constraints are necessary for the optimizing choices from a feasible tax schedule. They are also sufficient. To see this, note that no restrictions, such as continuity or monotonicity, are placed on the tax schedule. Therefore, given m bundles, it is possible to specify a tax schedule for which these are the only bundles that individuals can afford. Then, if self-selection constraints are satisfied, each class will choose the bundle intended for it.

The same result would still hold for some restricted classes of tax schedules. In the standard two commodity special case, it suffices to require that tax functions lead to after tax budget constraints in consumption-income space that are step functions. Individuals could be optimizing only at points where consumption has jumped up and not elsewhere. For any set of bundles satisfying self-selection constraints, there exists a tax function of this form that would allow only those bundles to be chosen by the different groups.

Given this transformation of the problem, a constrained Pareto efficient allocation is a vector of utilities, $w = (w_1,\ldots,w_m)$, and a set of consumption vectors, $\{x^i\} = (x^1,\ldots, x^m)$, such that $U^i(x^i) = w_i$, all i, $U^i(x^j) \leq w_i$, all i, j, and $p\cdot\Sigma_i N_i x^i + G \leq 0$, and so that no other allocation vector $(\hat{w}, \{\hat{x}^i\})$ satisfying these constraints has $\hat{w}_j \geq w_j$, all j, and $\hat{w}_i > w_i$, some i. The set of vectors w which are part of a constrained Pareto efficient allocation is called the constrained utility possibility frontier. Restricting G to that which can feasibly be produced given the resource limits incorporated in the consumption sets X^i, Pareto efficient allocations always exist since all preferences are continuous, all constraints are closed, and resource limits in X^i bound attainable allocations.

If the government spends all revenue raised, then, given a vector w on the constrained utility possibility frontier, it is necessary for the Pareto efficient consumption bundles associated with w to solve the following expenditure minimization problem:[6]

$$\min_{\{x^i\}} p\cdot\Sigma_i N_i x^i, \text{ s.t. } U^i(x^i) = w_i, \text{ all } i, \text{ and } U^i(x^j) \leq w_i, \text{ all } i, j \qquad (1)$$

This problem has the government maximizing its tax revenue given that target utility levels must be achieved and self-selection must hold. Note that constraints that x^i must belong to X^i are not imposed since, on the Pareto frontier, each w_i is restricted to being greater than the arbitrarily low numbers assigned to x^i not in X^i.

Since the objective function in (1) is additive in the x^i and since each constraint involves only the bundle assigned to a single class, that problem can be separated into the following m separate problems:

$$\min_{x}\ p\cdot x, \quad \text{s.t. } U^i(x) = w_i \text{ and } U^j(x) \le w_j, \text{ all } j\neq i \qquad (2)$$

Given a vector w from the utility possibility frontier, (2) can be used to find the associated efficient consumption bundles. Note that all bundles x for which $p\cdot x$ equals a constant raise the same tax revenue from group i and are called constant revenue hyperplanes.

Although (1) is necessary, it is not sufficient for a Pareto efficient allocation. Let $-T(w)$ be the minimized value in (1) given an arbitrary w. Assume budget balance is imposed so that for some $\hat{w}$, $T(\hat{w}) = G$. Such a $\hat{w}$ need not be on the utility possibility frontier since the self-selection constraints in (1) are written as $U^i(x^j) \le w_i$, where w_i is a parameter, and not as $U^i(x^j) \le U^i(x^i)$ and thus do not incorporate the direct interaction between the bundles x^i and x^j of two groups. Assume $U^j(x^i) = w_j$ and w_j is increased. Less revenue is raised from group j, so more must be raised from i which by itself could require that w_i be reduced. However, the self-selection constraint is also weakened which by itself would allow an increase in w_i. The two effects go in opposite directions. In some cases, weakening the self-selection constraint dominates so a rise in w_j leads to a rise in w_i even though the solution to (1) at the initial values just balances the government's budget. Figure 1 presents an example in which this occurs.

III. The Results

A. Relative Total Taxes and Self-Selection Constraints

The first property is the simple but powerful result that at a Pareto efficient allocation, a group must strictly prefer its own bundle to that of a group which pays a larger total tax.

Proposition 1: If $\{\hat{x}_i\}$ is constrained Pareto efficient and $U^i(\hat{x}^i) = U^i(\hat{x}^j)$, some i and j, then $-T^i(\hat{x}^i) = p\cdot\hat{x}^i \leq p\cdot\hat{x}^j = -T^j(\hat{x}^j)$.

Proof: Since the bundle $\hat{x}^j$ satisfies all the relevant constraints in (1) and since $U^i(\hat{x}^j) = w_i$ by assumption, then $\hat{x}^j$ is feasible in (2). The result is then immediate since $\hat{x}^i$ minimizes $p\cdot x$ over all such feasible bundles.

Q.E.D.

This result is illustrated in Figure 2 in a two commodity world. We consider a putative equilibrium with different groups choosing different consumption-income bundles $(-x_1, x_2)$, focusing only on groups i and j. The fact that group i is supposed to choose bundle $(-x_1^i, x_2^i)$ means that all other bundles are on or below i's indifference curve. For simplicity, the bundles intended for the kth and mth group are specified as strictly below this indifference curve. The fact that the self-selection constraint is binding for the jth bundle means that the jth bundle lies on the same indifference curve. Government revenue is measured by the distance from the bundle to the 45 degree line. If government revenue were higher at j than at i, then eliminating bundle i and assigning group i to x^j would increase government revenue without affecting any other constraint so such a situation could not be a solution to (2).

Proposition 1 does not rule out $\hat{x}^i \neq \hat{x}^j$, $U^i(\hat{x}^i) = U^i(\hat{x}^j)$, and $p\cdot\hat{x}^i = p\cdot\hat{x}^j$ and, in fact, there can exist Pareto efficient allocations with such situations, especially when single crossing does not hold. See Figure 3. However, the same utility vector $\hat{w}$ could be achieved by pooling groups i and j at the bundle $\hat{x}^j$. It would thus always be possible without loss to limit consideration to Pareto efficient allocations in which having $\hat{x}^i \neq \hat{x}^j$ and $U^i(\hat{x}^i) = U^i(\hat{x}^j)$ implies $p \cdot \hat{x}^i < p \cdot \hat{x}^j$.

B. Absence of Self-Selection Cycles

In the standard two commodity optimal income taxation model when single crossing is assumed, it is structurally impossible to have a cycle of binding self-selection constraints among groups receiving different allocations since the indifference curves of two classes can never cross more than once. Under the weaker assumptions of this paper, multiple crossings of indifference curves are possible. Although we cannot rule out the existence of Pareto efficient allocations at which such cycles exist, in Proposition 2, we show that if an efficient allocation exists which has such a cycle then there exists another allocation yielding the same utilities for every group in which, due to pooling, no cycle arises among groups receiving distinct bundles. Thus, no loss would occur from considering only allocations without such cycles.

Proposition 2. Assume $\{x^i\}$ is Pareto efficient and there exists a group K of k types that are separated ($x^i \neq x^j$, all i, j in K) and that can be ordered so that $U^i(x^i) = U^i(x^{i+1})$, $i=1,\ldots,k-1$, and $U^k(x^k) = U^k(x^1)$. Then, by pooling some of the types, there exists another efficient allocation $\{\bar{x}^i\}$ with $U^i(\bar{x}^i) = U^i(x^i)$, $i=1,\ldots,m$, and which has no cycle of binding self-selection constraints among separated groups.

Proof: From Proposition 1, $p \cdot x^i \leq p \cdot x^{i+1}$, $i=1,\ldots,k-1$, and $p \cdot x^k \leq p \cdot x^1$. Hence, $T^i(x^i)$ is the same for all i in k. Assigning group k to x^1 instead of x^k leaves all the constraints in (1) unaffected and does not change the objective function. If the self-selection constraints still cycle, then assign group k-1 to x^1 as well. This pooling can continue until no cycle exists or all k groups are pooled. Q.E.D.

Our result that allocations with cycles could be replaced by ones without was shown for many commodities and did not require quasiconcavity. Guesnerie and Seade (1982), in their Proposition 3, completely rule out an optimum having cycles of binding self-selection constraints involving <u>all</u> groups. In their two commodity model, since they assumed strict concavity of utility functions, a stronger result is available.

<u>Corollary 1</u>: If there are only two commodities and if $U^i(x^i)$ is strictly quasiconcave for all i, then at a Pareto efficient allocation some group is not part of any cycle of binding self-selection constraints and no cycle can involve more than two groups who are assigned distinct bundles.

<u>Proof</u>. Assume there exists a cycle among k groups, k > 2, each of which is assigned a different bundle. From Proposition 1, all these bundles must lie on the same constant revenue line and can be ordered in sequence. For a cycle to exist at least one group must be indifferent between two of these bundles (one of which is its own) which are not adjacent to each other. From the self-selection constraints, this group must view any bundles inbetween as no better than its own bundle. This violates strict quasiconcavity, showing the second part of the result.

From Proposition 1, the bundles of any groups on the highest tax revenue line are viewed as strictly inferior by any groups paying lower taxes. If these groups do not cycle, the result is shown. If they do cycle, then as in the proof of Proposition 2, alternative allocations yielding the same Pareto efficient utilities can be created by pooling at each of the original highest taxed bundles all groups who viewed that bundle as indifferent to their own. After the pooling, each of these bundles is viewed as strictly inferior by any group not assigned there. From Proposition 3 below, each group assigned there after pooling must have a tangency at that bundle between its indifference curve and the constant revenue line. Groups must have multiple tangencies, a contradiction of strict quasiconcavity. Thus, the highest taxed bundle or bundles cannot be in a cycle. Q.E.D.

In the two commodity world, strict quasiconcavity does not rule out a two cycle as can be seen by modifying Figure 3. Redraw $U^j(x^j) = w_j$ to run through x^i and a two cycle exists between i and j. Single crossing but not strict quasiconcavity is violated. With more than two commodities strict quasiconcavity does not seem to rule out any size cycle. Assume there are three commodities and a cycle among four bundles which are uniformly spaced around a circle lying on a constant revenue plane. The cycle can have groups indifferent only to an adjacent bundle and thus strict quasiconcavity need not be violated.

Proposition 2 does not assert that the groups could be ordered by a chain of binding self-selection constraints. Many other possibilities exist. First, there can exist situations in which no self-selection constraint is binding. For example, in the neighborhood of the no tax situation no constraint need bind. Second, several different groups can all have binding self-selection constraints with respect to the same group. That is, for some group T, $U^i(x^i) = U^i(x^j)$, some j and all $i \epsilon T$. Third, order the optimal bundles by the size of the tax raised. Assume $T^i > T^{i+1}$, all i. Then it is possible that $U^i(x^i) = U^i(x^{i+2})$ and $U^i(x^i) > U^i(x^{i+1})$. That is, the binding self-selection constraints are for group 1 with respect to group 3's bundle and for group 2 with respect to group 4's bundle. Note in this case, the bundles of groups 1 and 3 are viewed as strictly inferior by all other groups. To show that this can occur, assume that G=0 and begin with two groups whose preferences are quite different. The competitive equilibrium (both groups paying zero tax) is Pareto efficient, and in it neither self-selection constraint binds. Now introduce two additional groups, one close to each of the original groups. There exist Pareto efficient allocations in which the self-selection constraint is binding among the groups near each other, but not among those which are far apart. The net resource pattern can then be ordered as specified. Finally, there can exist chains of binding self-selection constraints in subsets of the set of groups. Group 1 is indifferent to group 2's bundle, 2 is indifferent to 3, and so on. Such a set of groups which can be ordered in this manner will be referred to as a connected set of groups.

C. <u>Existence of Undistorted Bundles</u>

One of the most important questions in the optimal income taxation literature is when should some individuals not be subject to distortionary taxation, that is, when should they face a zero marginal tax rate. The standard result (see, for example, Sadka (1976)) in a two commodity model with single crossing and a utilitarian government is that the top earner should face a zero marginal tax rate. To consider this question in our more general model, let $\hat{x}^i$ be the bundle assigned to group i in a Pareto efficient allocation and let $T(\hat{x}^i)$ denote the tax collected from this group. Group i is undistorted at this bundle if and only if $\hat{x}^i$ would still be chosen were the group instead charged a lump sum tax $\bar{T}$ equal to $T(\hat{x}^i)$, that is, if $\hat{x}^i$ maximizes $U^i(x)$ subject to $p \cdot x + \bar{T} \leq 0$. Clearly, then, at $\hat{x}^i$, there must be a tangency between i's indifference curve and the constant revenue line and the indifference curve must be locally quasiconcave there.

Proposition 3: At a Pareto efficient allocation $(\hat{w},\{\hat{x}^i\})$, if a bundle $\bar{x}$ is assigned to one or more groups and if all other groups find $\bar{x}$ inferior $(U^k(\bar{x}) < U^k(\hat{x}^k)$, all k with $\bar{x} \neq \hat{x}^k)$, then all groups assigned $\bar{x}$ are undistorted.

Proof. $\bar{x}$ must solve (2) for any group i with $\hat{x}^i = \bar{x}$. Since, by assumption $U^k(\hat{x}^k) < \hat{w}_k$ for all k not assigned $\bar{x}$, then in (2), these self-selection constraints are strictly nonbinding and can be deleted. In addition, any groups other than i assigned $\bar{x}$ can be disregarded. This follows since if a change in $\hat{x}^i$ from $\bar{x}$ is preferred by such groups then they should also be moved and if such a move would be worse for them, then their self-selection constraints with respect to the new $\hat{x}^i$ would still hold. Therefore, (2) reduces to $\bar{x}$ solving Min $p \cdot x$ subject to $U^i(x) = w_i$. For $\tilde{T} \equiv T(\bar{x})$, $p \cdot \bar{x} = -\tilde{T}$. From duality theory, $\bar{x}$ must also solve Max $U^i(x)$ subject to $p \cdot x + \tilde{T} \leq 0$, making group i undistorted at $\bar{x}$. Q.E.D.

Proposition 3 gives a condition ensuring that a group not face distortionary taxation. Proposition 4 shows that in a Pareto efficient allocation, some group always satisfies this condition.

Proposition 4: At a Pareto efficient allocation $(\hat{w},\{\hat{x}^i\})$, let $T^{max} \equiv \max_k[-p \cdot \hat{x}^k]$ be the largest tax payment by any group. Then any group j with $-p \cdot \hat{x}^j = T^{max}$ is undistorted.

Proof: From Proposition 1, if group j pays a tax equal to T^{max} and group i pays a smaller tax $(p \cdot \hat{x}^i > p \cdot \hat{x}^j)$, then $U^i(\hat{x}^i) > U^i(\hat{x}^j)$. Further, if some group k has $U^k(\hat{x}^k) = U^k(\hat{x}^j)$, from Proposition 1, $p\hat{x}^k \leq p\cdot\hat{x}^j$. Since looking across types, j is the group for which $p\cdot x^j$ is smallest, then $p\cdot\hat{x}^k = p\cdot\hat{x}^j$. Pooling group k with group j, maintains budget balance and achieves the same utilities for all groups. In this new allocation, every group k separated from j has $U^k(\hat{x}^k) > U^k(\hat{x}^j)$. From Proposition 2, all groups at $\hat{x}^j$ are undistorted. Groups pooled with j were originally on the same budget line and indifference curve as at $\hat{x}^j$. Hence, they were initially undistorted as well.

Q.E.D.

This result is the analogue of the standard optimal income taxation result that the top earner is undistorted. Here, the largest tax payer, who need not have the largest income, is undistorted.[7] Note that the highest taxed groups need not be the only ones that are undistorted. Lower taxed groups can also be undistorted as in some of the examples in the discussion following Proposition 2.

D. Description of Pareto Efficient Tax Schedules

Another question of interest is the nature of the tax structure at distorted bundles. Clearly, a bundle assigned to group k is distorted only if some other group i has a binding self-selection constraint with k's bundle, $U^i(x^k) = U^i(x^i)$. In order to characterize the set of preferences of all groups at the distorted bundle of some groups, it is convenient to add the assumption that preferences are differentiable. By using subgradients, differentiability could be relaxed at a cost of greater notational complexity. Let α be any n-1 dimensional vector of unit length and let x(t) denote the commodity vector with the t^{th} component eliminated. For any group j, starting from $\bar{x}$, let $M^j(\bar{x}, t,\alpha)$ denote the rate of change of x_t needed to hold U^j constant at $U^j(\bar{x})$ when the remaining components change in direction α, that is, along the line

$x(t) = \bar{x}(t) + b\alpha$, for b a scaler. Then $M^j(\bar{x}, t, \alpha) = -dx_t^{u=c}/db =$ $(\sum_{h\neq t} \alpha_h U_h^j(\bar{x}))/U_t^j(\bar{x})$. Let $p_\alpha = \alpha\cdot p(t)$ denote the price of a unit move in direction α. Since p is the vector of production costs, p_α is the cost of producing a unit of the mixture of goods given by the vector α.

Proposition 5: Let $(w,\{x^i\})$ be a Pareto efficient allocation and consider any group j, commodity t, and unit vector $\alpha\epsilon R^{n-1}$. If $M^j(x^j, t,\alpha) > p_\alpha/p_t$ then there exists a group k with $U^k(x^j) = w_k$ and $M^j(x^j, t, \alpha) < M^k(x^j, t, \alpha)$ and if $M^j(x^j, t, \alpha) < p_\alpha/p_t$ then there exists a group k with $U^k(x^j) = w_k$ and $M^j(x^j, t, \alpha) > M^k(x^j, t, \alpha)$.

Proof: Consider $M^j(x^j, t, \alpha) > p_\alpha/p_t$ and assume the result is false. Given the line $x(t) = x^j(t) + b\alpha$, the indifference curves and the constant tax lines can be graphed in (b, x_t) space. See Figure 4. For small $b > 0$, the line $\alpha p_\alpha\cdot x(t) + p_t x_t = p\cdot x^j$ lies above the indifference curve of group j through x^j. If $M^j(x^j, t, \alpha) > \text{Max}[M^k(x^j, t, \alpha)$, all k with $U^k(x^j) = w^k]$, then for small $b > 0$, the indifference curves through x^j for all k with $U^k(x^j) = w_k$ lie above that for group j. Therefore, a small change in x^j along the indifference curve $U^j(x) = w_j$ in the direction of $b > 0$, raises taxes and satisfies self-selection constraints for all such groups k. Thus, x^j would not have solved (2), a contradiction of it being Pareto efficient. If $M^j(x^j, t, \alpha) = \text{Max}[M^k(x^j, t, \alpha)$, all k with $U^k(x^j) = w_k]$, then moving x^j along j's indifference curve may cause a violation of the self-selection constraint for a group k with $U^k(x^k) = w_k$ and $M^k(x^j, t, \alpha) = M^j(x^j, t, \alpha)$ if the indifference curve for k through x^j is below that of j. However, for small b, the gain in utility to group k would be a second order effect given the equal slopes of the indifference curves. The gain in tax revenue is a first order effect. The extra revenue could be used to raise x^k and weaken the self-selection constraint. If x^k were constrained by some other group's self-selection constraint, then the extra tax revenue can raise the utilities of all groups tied to group k in a chain of binding self-selection constraints. This leads to a new allocation in which all self-selection constraints and resource balance are satisfied and which Pareto dominates the original. This contradiction means that $M^j(x^j, t, \alpha) < \text{Max } [M^k(x^j, t, \alpha)$, all k with $U^k(x^j) = w_k]$ as required. If $M^j(x^j, t, \alpha) < p_\alpha/p_t$, a similar proof follows using $b < 0$.

Q.E.D.

This result is easiest to understand in the two commodity special case of consumption and income where the utility function is $U(x,y)$, $U_y < 0$ and the marginal rate of substitution is $-U_y/U_c$.[8] Proposition 5 asserts that if a group's choice is distorted then its marginal rate of substitution must lie between the marginal rate of transformation (the price ratio) and the marginal rate of substitution of some other group which envies its bundle. Thus, depending upon the preferences of these other groups, the distorted marginal rate of substitution can be greater or less than the price ratio. To interpret this in terms of marginal taxes, the tax structure generally has kinks at those bundles that groups are intended to consume. Marginal tax rates are therefore not well defined at such bundles. However, if the marginal rate of substitution for a distorted group is less than the price ratio, then income is implicitly taxed on the margin and if the marginal rate of substitution is greater than the price ratio, income is implicitly subsidized.

Depending upon the circumstances, income could be implicitly taxed or subsidized. In fact, if several self-selection constraints bind at a bundle and if several groups are pooled at that bundle, then some of the pooled groups could be implicitly taxed, others could be implicitly subsidized, and some could even be undistorted. If only one group has a binding self-selection constraint at a bundle, then all groups consuming there must be implicitly taxed or implicitly subsidized.

It is precisely at this point that stronger results can be obtained by assuming single crossing. With that property, Pareto efficient taxation requires that every group within a connected set of groups (that is, which are tied together by a chain of binding self-selection constraints) either must face an implicit non-negative marginal tax rate at its own bundle or must have an implicit non-positive marginal tax rate. Only the first member of the chain faces a zero marginal tax rate. In the absence of single crossing, one group in the chain may be implicitly taxed at its bundle and another subsidized at its bundle. Furthermore, if, in the chain, redistribution is toward lower ability classes, then, given single crossing, distorted low ability income is implicitly taxed, while if redistribution is in favor of higher ability classes, then distorted high ability income is implicitly subsidized.

The results for n commodities are essentially the same. A composite commodity is created from an arbitrary mixture of n-1 of the commodities and then the composite commodity and the remaining n^{th} commodity form a two commodity problem. The composite commodity is implicitly taxed if the marginal rate of substitution is less than the price ratio and implicitly subsidized if the reverse holds. Note that it may not be possible to state that overall a group is implicitly taxed or subsidized. Some specifications of the composite commodity may be taxed and others subsidized.

The condition in Proposition 5 is necessary but not sufficient for Pareto efficiency at a distorted bundle. These conditions may be satisfied at some x^j

but, as the example in Figure 1 showed, by making simultaneous moves in x^j and some x^k with $U^k(x^k) = U^k(x^j)$, a Pareto improvement may occur. Whether this can happen requires quantitive not just qualitative considerations. The relative magnitudes of the effects of relaxing self-selection constraints versus tightening the budget constraint must be compared. Few additional general insights seem to arise from these considerations.

E. The Constrained Utility Possibility Frontier

The final two propositions describe the constrained utility possibility frontier and compare it to the utility possibility frontier arising when the government has full information about each individual's type so that self-selection constraints are not imposed. Obviously, when none of the self-selection constraints in the constrained problem bind, the two frontiers coincide and when some self-selection constraints strictly bind, the constrained frontier lies below the full information one. In addition, as shown in the next proposition, the constrained frontier is truncated relative to the full information frontier. That is, at any Pareto efficient allocation self-selection constraints can impose a minimum level of utility for some groups greater than what they could be limited to in a full information problem. Let w_i^{min} and $\hat{w}_i^{min}$ be the inf of the w_i on the full information and the constrained utility possibility frontiers respectively.

Proposition 6: In some economies, for some i, $\hat{w}_i^{min} > w_i^{min}$.

Proof: Consider some group i and a point on the full information utility possibility frontier which has this group receiving a level of utility at or near w_i^{min}. In the constrained problem, to achieve this level, group i would be heavily taxed and thus other groups would find its bundle strictly inferior to their own. From Proposition 3, i's bundle must be undistorted and, therefore, the maximum tax consistent with the given utility level is raised from i. The binding constraints are that all other groups must have bundles which are on or below i's indifference curve and that revenue at least equal to $G-T^i$ must be raised from them. If i's utility is low enough, it is possible that the resource constraint does not bind. An increase in i's utility would then allow increases in other groups utilities by relaxing the self-selection constraints. The original utility assigned to i could then not be on the constrained utility possibility schedule. This result is illustrated in Figure 5 for the two class-two commodity case with single crossing. The able group is at a low utility level. The highest utility available for the unable group on i's indifference curve is at zero income. The resource constraint does not bind.

Q.E.D.

This result means that if the Pareto problem is defined by maximizing one group's utility subject to target constraints on the utilities of all other groups in addition to the resource and self-selection constraints, then the target constraints sometimes will not bind.

The final result shows that, provided all goods are normal, if group i is undistorted at some point on the constrained utility possibility frontier, then group i is undistorted at all points on the frontier where i gets lower utility and the utilities of all other groups are not decreased.

<u>Proposition 7</u>: There always exists a region on the constrained utility possibility frontier on which no group is distorted and which thus coincides with the full information utility possibility frontier. In addition to the basic assumptions, assume that all goods are normal. Let w be a point on the

constrained utility possibility frontier at which group k is assigned an undistorted bundle. Then, at any Pareto efficient utility vector $\hat{w}$ at which $\hat{w}_k < w_k$ and $\hat{w}_i \geq w_i$, $j \neq k$, group k is still assigned an undistorted bundle.

<u>Proof</u>. The maximizing choices by each group on the no redistribution budget hyperplane $p \cdot x + G/\Sigma_i N_i = 0$ form a Pareto efficient allocation which satisfies self-selection and hence is both constrained and full information Pareto efficient. If all groups have different maximizing choices, then, at least for small redistributions, self-selection will continue to hold and the two frontiers will still coincide. Let $\{x^i\}$ be the allocation associated with w. By assumption, x^k is undistorted so that x^k minimizes $p \cdot x$ subject to $U^k(x) = w_k$. Consider $\hat{w}_k < w_k$. The undistorted bundle achieving utility $\hat{w}_k$ is the y^k which minimizes $p \cdot x$ subject to $U^k(x) = \hat{w}_k$. By normality, $y^k \leq x^k$. Therefore, by monotonicity $U^i(y^k) \leq U^i(x^k)$, all $i \neq k$. Since self-selection held initially, $U^i(x^k) \leq w_i$, $i \neq k$. Given $\hat{w}_i \geq w_i$, it then follows that $U^i(y^k) \leq \hat{w}_i$, all $i \neq k$. Hence, the undistorted bundle y^k is feasible in (2) for utility vector $\hat{w}$. Note that any other feasible bundle y in (2) for which $p \cdot y = p \cdot y^k$ must also be undistorted. Assume utility vector $\hat{w}$ were sustained by group k being assigned bundle $\hat{x}^k$ with $p \cdot \hat{x}^k > p \cdot y^k$. Then $\hat{x}^k$ could not solve (2) and $\hat{w}$ would not be Pareto efficient, a contradiction. Hence, $p \cdot \hat{x}^k = p \cdot y^k$ must hold as required. Q.E.D.

In a two class economy, as shown in Figure 6, this result implies that the constrained utility possibility frontier can be divided into three connected segments: one segment where one of the self-selection constraints strictly binds, another segment where neither self-selection constraint binds and a third segment where the other self-selection constraint strictly binds. The middle region includes the point of no redistribution ($T^1 = T^2 = G/(N_1 + N_2)$). When a self-selection constraint binds it is that of a high taxed group envying the bundle of a low taxed (subsidized) group. Note, in the model with different ability classes, there can be Pareto efficient allocations in which the low ability group desires the high ability bundle provided redistribution is favorable to the high ability group.

IV. Conclusions

This paper has established that several of the well-known qualitative properties of optimal tax schedules are really characteristics of Pareto efficient tax structures, and that these properties can be derived using quite elementary arguments under quite general conditions (not requiring special conditions on the number of commodities, on the preference maps, or on the relationships among preferences of the different types such as the single crossing property).

In the process of deriving these results, we have established some further properties, both of the Pareto efficient tax schedules and of the constrained utility possibility frontier. Besides the greater generality, our approach allows the restoration of the dichotomy which was central to the New Welfare Economics: the separation between identifying feasible (efficient) allocations and choosing among those feasible allocations, which must rest explicitly on values.

The techniques we have developed in this paper are not specific to the optimal taxation problem but can be applied to any self-selection problem such as those arising in the analyses of labor markets, insurance markets, regulated monopolies, and auction design.

FOOTNOTES

1. Though we limit ourselves here to discrete distributions, the results would be true so long as there is a bounded support to the distribution of types. See Mirrlees [1971] for how results for unbounded distributions may differ from those reported here.

2. Deb and Ramachandran[1986] in a finite class model also relax convexity and differentiability assumptions. They also seek to impose a self-selection constraint which holds with strict inequality arguing that if equality holds, the government may not be able to assign a group to the desired bundle if it is also indifferent to another. However, an alternate view is that the solution when the self-selection holds with weak inequality is really an ε-equilibrium. Although it cannot itself be achieved, a bundle arbitrarily close to that solution can be found which satisfies resource balance and which has the self-selection constraint hold with strict inequality. If the self-selection constraints must hold with strict inequality, then there may exist no solution to the maximization problem.

3. Guesnerie and Seade (1982) impose strict concavity of preferences, Inada type conditions on the indifference map, and an assumption that, even if single crossing does not hold globally, at the optimum, all groups who are pooled at the same choice bundle have locally different preferences (their assumption A). This latter assumption rules out pooling at an undistorted bundle. We do not require any of these conditions.

4. Guesnerie and Seade recognize this property in the proof of their Proposition 3 (analogous to our Proposition 4) but they also utilize concavity and their assumption A which are not required. They do not emphasize the central role of this property in deriving other results.

5. See Brito et. al. [1989] for an almost complete characterization of the conditions under which randomization is desirable in a model with two types of individuals.

6. Production efficiency thus must hold for (1) to be necessary for Pareto efficient bundles. As argued by Guesnerie and Seade (1982), that this holds is not obvious given self-selection constraints since "the type of argument used by Diamond and Mirrlees (1971) in the optimal commodity uniform-pricing case..... does not apply." Production efficiency can be shown by proving that, at a Pareto efficient allocation, there is a group whose bundle is strictly inferior to all groups not consuming that bundle. Any unused resources can then be given this group. In Proposition 4 below we show such a group exists. However, since to prove that Proposition we use (1), we cannot then use Proposition 4 to show production efficiency. A valid proof under the weak assumptions of our paper proceeds by assuming production efficiency does not hold. Then resources could be discarded by society until the same point on the utility possibility frontier is reached and production efficiency holds in the modified economy. Now (1) is necessary and Proposition 4 holds in the modified economy. But then any extra resources could be given to the unenvied group or groups and a Pareto superior outcome would be reached. This contradicts the assumption that resources could be discarded from the original society at no cost.

Note a simpler proof would be possible if G were not exogenously given but were a desirable commodity entering utility functions in a separable fashion. Then G would always be increased to use any slack resources and production efficiency would hold at any Pareto optimum.

7. Guesnerie and Seade (1982) show a similar result in a two commodity world. They, however, rule out by assumption pooling on the T^{max} budget line. We allow such pooling which can occur only if two groups have the same optimal bundle on that budget line.

8. The result in this special case generalizes that in Stiglitz [1982] where it was derived for a two class model with quasiconcavity and single crossing.

References

Brito, D., J. Hamilton, S. Slutsky and J. Stiglitz, "Randomization in Optimal Income Tax Schedules," mimeo, University of Florida (1989).

Deb, R. and R.V. Ramachandran, "Efficiency, Non-Distortion and Non-Linear Taxes in a Finite Economy," mimeo, Southern Methodist University, (1986).

Diamond, P.A. and J.A. Mirrlees, "Optimal Taxation and Public Production: I-Production Efficiency," American Economic Review 61: 8-27 (1971).

Guesnerie, R. and J. Seade, "Nonlinear Pricing in a "Finite Economy," Journal of public Economics 17: 157-179 (1982).

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

Sadka, E., "On Income Distribution, Incentive Effects and Optimal Income Taxation," Review of Economic Studies 43: 261-268 (1976).

Stiglitz, J.E., "Self-Selection and Pareto Efficient Taxation," Journal of Public Economics 17: 213-240 (1982).

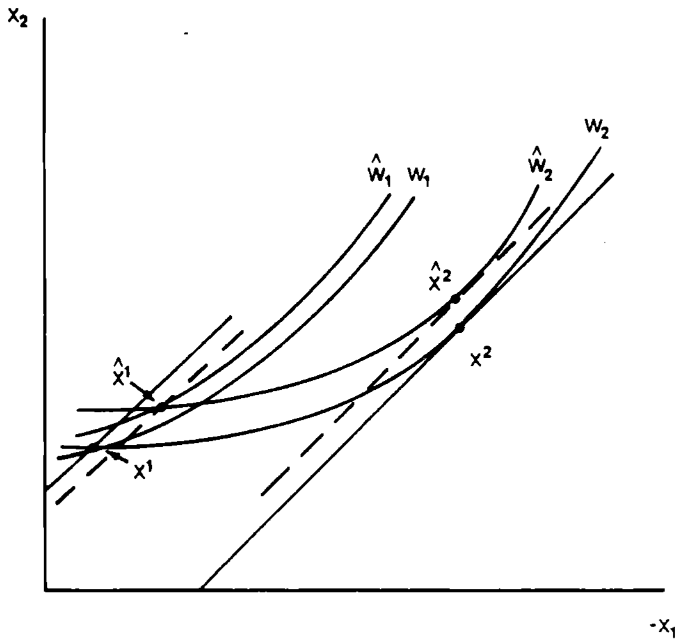

FIGURE 1

The bundles x^1 and x^2 solve (1) given (w_1, w_2), but $(\hat{w}_1, \hat{w}_2)$ satisfies all constraints and is Pareto superior.

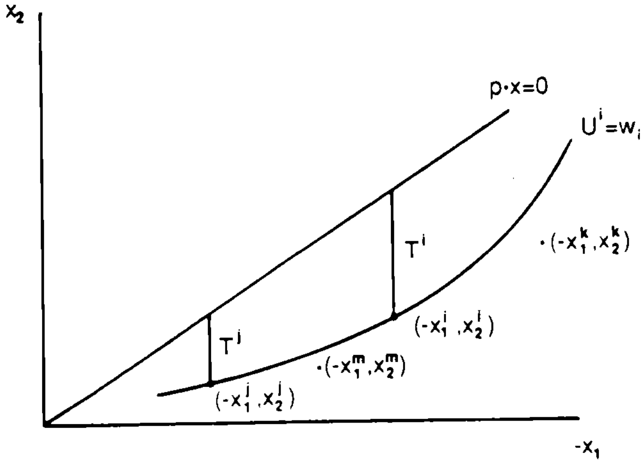

FIGURE 2

Normalizing all prices to 1, the vertical distance from a bundle to the 45° line through the origin is the total tax charged to someone consuming that bundle. If T^i were less than T^j, then the government would raise total revenue by having group i consume x^j.

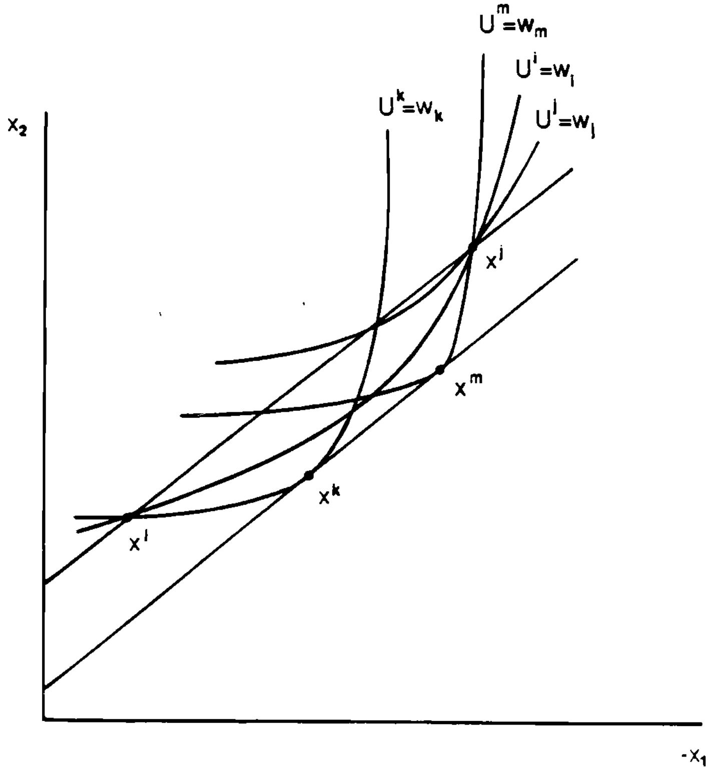

FIGURE 3

Since $U^i(x^i) = U^i(x^j)$ and $U^k(x^i) = U^k(x^k) > U^k(x^j)$, multiple crossings of i's and k's indifference curves must occur.

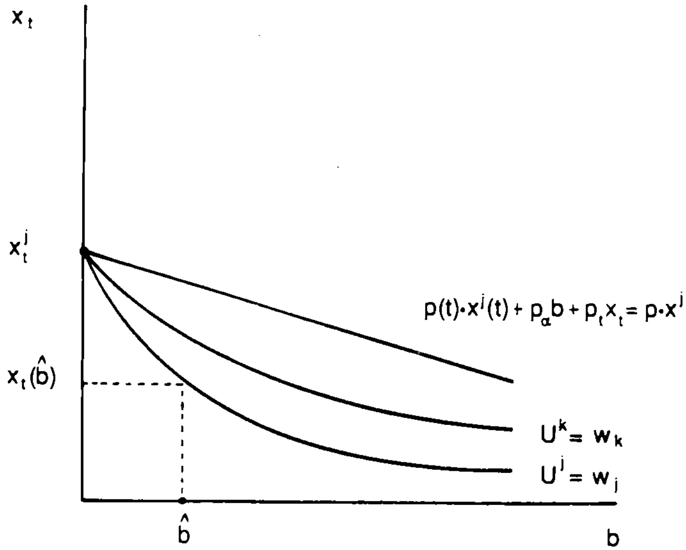

FIGURE 4

The bundle $(x_t(\hat{b}),\ x_t^j + \hat{b}\alpha)$ satisfies the self-selection constraint for group k and raises extra revenue since it lies below the constant revenue line through x^j.

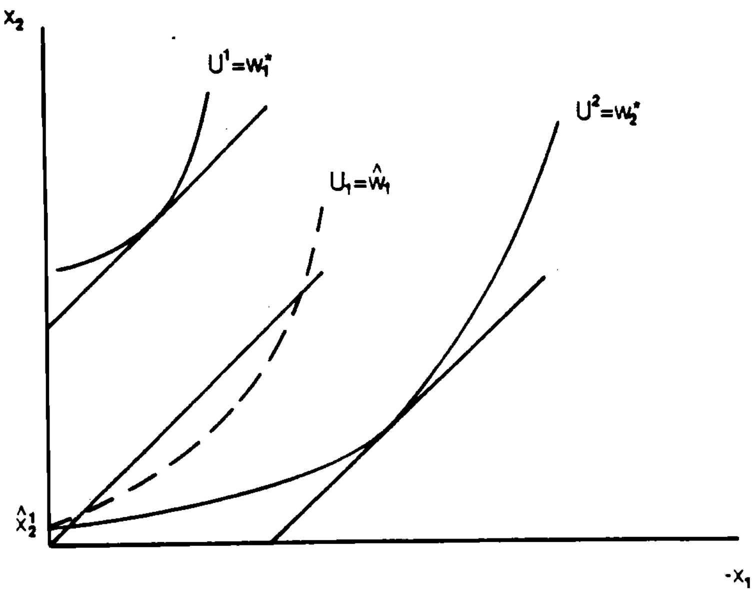

FIGURE 5

(w_1^*, w_2^*) is on the full information utility possibility frontier. Given self-selection constraints, for $U^2 = w_2^*$, the best 1 can achieve is $U^1 = \hat{w}_1$ at $(0, \hat{x}_2^1)$. The government budget is in surplus. An increase in U^2 would allow U^1 to increase without causing a deficit.

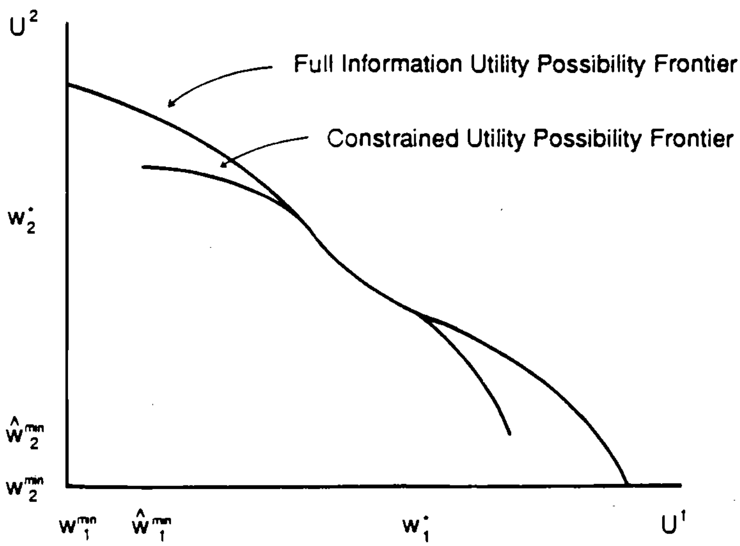

FIGURE 6

For all $w_1 \leq w_1^*$, group 1 is undistorted, and for all $w_2 \leq w_2^*$, group 2 is undistorted.