NBER WORKING PAPERS SERIES

THE OPTIMAL TWO-BRACKET LINEAR INCOME TAX

Joel Slemrod

Shlomo Yitzhaki

Joram Mayshar

Working Paper No. 3847

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1991

May 1991. This research has been supported by Grant No. 88-00230-1 from the U.S.-Israel Binational Science Foundation. We are grateful to Marianne Page for invaluable research assistance, and to Eytan Sheshinski for constructive discussions. 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 #3847
September 1991

THE OPTIMAL TWO-BRACKET LINEAR INCOME TAX

ABSTRACT

We investigate the optimal rate structure of an income tax system that is constrained to have only two brackets, plus a demogrant. We find that, in a two-class economy, Pareto-efficient tax schedules feature at least one marginal tax rate equal to zero, and that the marginal tax rate may be increasing or declining. We next use numerical optimization techniques to study the optimal structure of such a tax system in a multi-person model that is a stylized version of an actual economy. We discover that in all cases the tax rate in the second (higher) bracket is less than the tax rate that applies to the first bracket but that progressivity, in the sense of a uniformly rising average tax rate, generally obtains. Compared to the optimal one-bracket (linear) tax system, both the highest and lowest income individuals are better off, while a middle range of taxpayers is worse off.

Joel Slemrod
Department of Economics
University of Michigan
Ann Arbor, MI 48109-1220
and NBER

Shlomo Yitzhaki
Department of Economics
Hebrew University
P. Naphtali Building
- Mt. Scopus
Jerusalem 91905 Israel

Joram Mayshar
Department of Economics
Hebrew University
P. Naphtali Building - Mt. Scopus
Jerusalem 91905 Israel

THE OPTIMAL TWO-BRACKET LINEAR INCOME TAX

Joel Slemrod, Shlomo Yitzhaki and Joram Mayshar

1. Introduction

In 1980 a married couple filing jointly in the U.S. faced a graduated income tax system consisting of 14 brackets with marginal tax rates ranging from 14% to 70%. As of 1991 there are three explicit brackets, of 15%, 28% and 31%. In the 1980's many other countries carried out similar, yet usually less extreme, changes in the income tax structure designed to compress the rate structure and reduce the number of brackets.

The widespread reliance on income tax structures with a small number of brackets suggests that it is appropriate to revisit the optimal income tax literature. Following the seminal analysis of general income tax structure by Mirrlees (1971), most work focused on the optimal linear income tax. This shift in attention was no doubt due largely to the computational complexity of analysis in more general cases. An additional reason was the discovery by Mirrlees that, at least in the cases he considered, the optimal nonlinear income tax structure was approximately linear! If that result is generally true, then the computational advantages of investigating only linear tax structures can be achieved without sacrificing important insights about optimal tax structure.

In this paper we investigate the implications of adding to the linear income tax one additional level of generality--a two-bracket linear structure. We find that, in a two-class economy, Pareto-efficient tax schedules of this type feature at least one marginal tax rate equal to zero, and that the marginal tax rate may be increasing or declining. We then investigate the optimal structure of taxation when the social welfare function, utility function and distribution of abilities are characterized as in the standard optimal linear income tax problem, as in Stern (1976). We discover that in all cases the second marginal tax rate is less than the first rate but that progressivity, in the sense of uniformly rising average tax rate, generally obtains. The additional instrument of a second tax bracket allows the lower marginal tax rate on high-wage people to coax out enough additional labor supply so that the optimal demogrant is increased. Thus, compared to

the simple linear income tax system, both the highest and lowest wage individuals are better off, while a middle range of taxpayers are worse off.

We proceed as follows. Section 2 briefly reviews the conceptual issues that are important. In Section 3 we characterize Pareto-efficient tax schedules in a two-class economy. In Section 4 we discuss the methodology underlying the numerical simulations, and in Section 5 present the results of these exercises. Section 6 concludes.

2. Review of Conceptual Issues

Mirrlees (1971) asked what general income tax structure that raised a prespecified amount of revenue maximized a utilitarian social welfare function, when individuals were free to choose how much labor to supply. Imposing no restrictions on either the utility function, social welfare function, or distribution of wages, Mirrlees could show only that (i) the optimal marginal income tax lay, at all levels of income, between zero and one, inclusively and that (ii) under the optimal income tax structure some low-wage individuals may choose not to work at all. In order to obtain more specific conclusions, Mirrlees then proceeded to perform several numerical optimizations assuming a utilitarian social welfare function, an identical Cobb-Douglas utility function of goods and leisure for each individual, and a lognormal distribution of wages. With these assumptions he found that the optimal tax structure is approximately linear (i.e., it features an almost constant marginal tax rate and an exemption level of income below which tax liability is negative). Furthermore, the marginal tax rates were quite low by the standards of 1971, usually between twenty and thirty percent and always lower than forty percent. Note that although the marginal tax rate was approximately constant, the average tax rate (tax liability divided by income) increased with income due to the presence of the positive lump-sum transfer.

For the reasons mentioned in the introduction, most work subsequent to Mirrlees focused on the optimal linear income tax where all individuals receive a demogrant, and all income is taxed at a constant rate. Stern (1976) argued that the degree of labor supply responsiveness implied by the Cobb-Douglas utility function is excessive and thus overstates the costs of increasing tax

progressivity. He claimed that when a more reasonable estimate of labor supply responsiveness is used (with an elasticity of substitution of 0.4 rather than the 1.0 implicit in the Cobb-Douglas formulation) the value of the optimal tax rate is substantially higher than those found by Mirrlees. In Stern's central case, the optimal marginal tax rate a of linear tax system is fifty-four percent.[1] This compares to a marginal tax rate of thirty-five percent and a correspondingly lower zero-tax income level if the elasticity of substitution is set at one. More egalitarian social welfare functions and less substitutability imply even higher marginal tax rates.

In a separate development Phelps (1973), Sadka (1976), and Seade (1977) showed the following striking result: if there is a finite and known top to the distribution of wages, then the marginal tax rate applied to the highest individual's income should be zero. To see the intuition behind this result, consider an income tax system in which the marginal tax rate applicable to the highest observed income is positive. Now consider a second tax schedule which is identical to the first except that it allows the highest-earning household to pay no taxes on any excess of income it had previously decided not to make the effort to earn, partly because of the tax due on the additional income. When faced with the second tax schedule, this household is certainly better off, works more hours, and pays no less tax than under the first tax schedule. Thus, the highest-earning household is better off under the zero marginal tax rate scheme, and all other households are also at least as well off (they may be strictly better off if the increased tax revenue from the highest-earning household allows a reduction in tax rates in the lower brackets). A simple explanation is that raising the marginal tax at the top from zero distorts the labor supply decision of the highest earner but raises no revenue.

The practical significance of this result is unclear. It does not imply that marginal taxes should be zero or very low *near* the top, only precisely *at* the top. In fact, numerical calculations

---

[1] In this example, the government must raise about twenty percent of national income through the income tax and the social marginal valuation of income decreases with the square of income.

done by Mirrlees (1976) suggest that zero "is a bad approximation to the [optimal] marginal tax rate even within most of the top... percentiles."

What do these results imply about the optimal two-bracket linear income tax? An eyeballing of Mirrlees' numerical simulation results and the insights of the Phelps-Sadka-Seade work suggest that there is likely to be a two-bracket linear income tax with the second marginal tax rate lying below the first ($t_2 < t_1$) that dominates the simple linear tax. Whether this is the global optimum is, however, another question.

This intuition is apparently contradicted by Sheshinski (1989), who presents a proof that a declining rate structure ($t_2 < t_1$) can never be optimal. However, as we demonstrate in Appendix I of this paper, Sheshinski's proof is not correct, leaving as an open question the optimal structure of a two-bracket tax system. In what follows we pursue this issue, first by investigating what characterizes efficient tax structures in a two-class economy, and then by calculating the social welfare maximizing tax structure in stylized economies.

3. Efficient Two-Bracket Linear Income Tax Structure in a Two-Class Economy

A more recent development in the theory of optimal income taxation is Stiglitz' (1982, 1987) general characterization of Pareto-efficient tax schedules, which impose only self-selection constraints and do not require a particular social welfare function. Under that approach the tax schedule is designed so that each individual taxpayer (weakly) prefers exactly the consumption-labor choice that the tax planner intends him to select. In the case of two (classes of) taxpayers, the redistributive tax system features a zero marginal tax rate for the high-ability individual, while the low-income individual generally locates at a corner point in the tax schedule with an implicit positive marginal tax rate. For the case of N taxpayers, such a Pareto-efficient schedule requires in general N-1 tax brackets, and these have to be custom-tailored exactly to taxpayers' preferences and skills. Thus any minor variation in preferences or in the skill distribution requires a redesign of the tax schedule. This extreme sensitivity and the large number of required tax parameters put in question the practical usefulness of that approach.

In what follows we adopt an alternative approach to the study of Pareto-efficient income tax schedules. We posit that exogenous administrative considerations limit the tax schedule to have an extremely simple form--an intercept plus only two tax brackets, within which the tax function is linear. The resulting efficient restricted tax schedule should be expected to be rather robust to minor variations in the underlying information. On the other hand, the structural constraints that result from the parsimony in the number of tax brackets have to hold in addition to the self-selection constraints. If one ignores administrative considerations, the resultant tax schedule will thus clearly be Pareto-dominated by efficient tax schedules that are not hampered by this structural constraint. The distinction between the two approaches will become evident in the illustrative two-class case examined below, but is clearly much more pronounced in the more policy-relevant case that features an entire distribution of taxpayers and large costs to fine-tuning the tax schedule.

Assume that each individual i selects labor effort L_i and consumption C_i in order to maximize a standard common utility function $U(C_i, 1 - L_i)$. Individual i's before-tax income is $Y_i = w_i L_i$, where w_i is the wage rate. Consumption is $C_i = Y_i - T(Y_i, \Theta)$, where the tax function is continuous and piece-wise linear such that

$$T(Y, \Theta) = \begin{cases} -G + t_1 Y & Y \le \bar{Y} \\ -G + t_1 \bar{Y} + t_2 (Y - \bar{Y}) & Y \ge \bar{Y}. \end{cases}$$

The tax parameters are: $\Theta = (G, t_1, t_2, \bar{Y})$. A star superscript will identify the preferred levels of the choice variables by each individual. In particular, $U_i^*(\Theta)$ is the indirect utility function of individual i, and $T_i^*(\Theta)$ is his tax payment. We seek to characterize Pareto-efficient and socially optimal tax schedules that raise a given revenue R.

Assume now that the economy consists of only two types of individuals: n_L of type L and n_H of type H, where $w_H > w_L$. The following proposition provides a rather general characterization of the socially optimal tax schedule for this case.

Proposition: The socially optimal Θ can be restricted to be one of the following five cases: Progressive, that is, redistributing towards the less able:

(1) $0 = t_1 < t_2$, with $Y_L^* = \bar{Y}$

(2) $t_1 > t_2 = 0$, with U_H^* touching both branches of the budget set.

Regressive, that is, redistributing towards the more able:

(3) $t_1 < t_2 = 0$, with $Y_H^* = \bar{Y}$

(4) $0 = t_1 > t_2$, with U_L^* touching both branches of the budget set.

Nondistributive:

(5) $t_1 = t_2 = 0$.

The proof of this proposition applies the standard techniques employed by Sadka (1976), Seade (1978) and Stiglitz (1987), and is therefore only sketched in Appendix II. The linear tax schedule, which gives up a degree of freedom in selecting a second tax bracket, is seen to be efficient only when it is tantamount to equal lump-sum taxation of both individuals. Otherwise, one can always improve on it by setting to zero either the higher or the lower income tax rate. On the other hand, if we permit a third bracket, it is clear that we would have enough degrees of freedom to obtain Seade's result that one could restrict attention to schedules where the marginal tax rate is zero at both the highest and the lowest brackets.

As an illustration, consider the utility function examined in Sheshinski (1989): $U = C - 0.5L^2$. In this case labor supply has a zero income elasticity. Consider further the special case of $n_L/n_H = 1$, $R = 0$, $w_L = 2^{1/2}$ and $w_H = 2$. The utility possibility frontier (UPF) that results in this example from alternative tax structures is shown in Figure 1. Given the ability to use discriminatory lump sum taxes the UPF is AOB, where point O represents the case of no taxes. When discrimination is not possible, so that both individuals have to face the same tax schedule, but the tax schedule is unrestricted (the case considered by Sadka, Seade and Stiglitz), the UPF is

confined to be above the 45° line and is marked as A'A°OB°B'. In the case of a linear tax schedule the UPF becomes A_0OB_0.

The two-bracket piece-wise linear tax schedule, our concern here, can be calculated with the aid of the proposition above. The UPF in this case lies between the UPF for the linear tax schedule and the UPF for the case where the tax schedule is unrestricted but common to both individuals. The no-tax point, O, is common to all the UPF's. If attention is restricted to the cases of decreasing marginal tax rates, the UPF is A_2OB_4; reflecting the tax schedule of type (2) in the progressive range and the tax of type (4) in the regressive range. Restricting attention to increasing marginal tax rates, the UPF is A_1OB_3; employing the tax schedule of type (1) to the right of point O and the tax schedule of type (3) to the left.[2]

The efficient UPF for the two-bracket piece-wise linear schedule in this example is thus A_2OB_3, consisting of type (2) with diminishing rates at the progressive segment of the UPF, and of type (3) with increasing marginal (yet negative) rates at the regressive range of the UPF.

This illustrative calculation provides a counterexample of Sheshinski's claim to have shown that the optimal two-rate schedule will always exhibit increasing marginal rates. This example also suggests that the case of decreasing marginal rates may have to be taken more seriously than it has been until now. Figure 1 demonstrates, however, that even when decreasing marginal rates are Pareto-dominant, the linear tax system is Pareto-inferior even to the increasing marginal rates schedule. Increasing the complexity of the tax schedule enables society to increase the welfare of both classes. Given a specific social welfare function on the one hand, and given a measure of the

---

[2] In range (2), where $t_1 > t_2 = 0$, the problem is : Max $U_L^*(\Theta_1) = G_1 + (1 - t_1)^2$, given $U_H^*(G_2, 1) = G_2 + 2 = U_H$; $U_H^*(G_1, t_1) = G_1 + 2(1 - t_1)^2 = U_H$; $T_L^*(\Theta_1) + T_H^*(\Theta_2) = [2(1 - t_1)t_1 - G_1] - G_2 = 0$; $0 \le t_1 \le 1$. In range (1), where $t_1 = 0 < t_2$, the problem is: Max $U_L = G_1 + \bar{Y} - (\bar{Y})^2/4$, subject to $\bar{Y}t_2 = G_2 - G_1$; $U_H^*(G_2, t_2) = G_2 + 2(1 - t_2)^2 = U_H$; $T_L^*(\Theta_1) + T_H^*(\Theta_2) = - G_1 + [4(1 - t_2)t_2 - G_2] = 0$.

social costs associated with complexity on the other, one can then find the optimal degree of complexity of the income tax schedule.

Yet how general is this example? A partial answer to this question can be gleaned by considering the extreme case where for at least one of the two types of individuals, the local elasticity of substitution between consumption and leisure is zero. Figure 2 presents some indifference curves of the two types of individuals in the Y-C space (see Appendix II). In Figure 2-a, both individuals have a zero elasticity of substitution. Assuming for simplicity that R = 0 and the groups are of equal size, denote by $U_j^{(k)}[U_i]$ the highest level of U_j that can be achieved by a tax schedule of type (k) for any given level of U_i. Given U_H^* at point B, it is clear that $U_L^{(2)}[U_H^*] = U_L^*$ is reached at point A with the type (2) tax schedule OABE. On the other hand, given U_L^*, the efficient tax schedule of type (1) is O'ABE', affording the original point B for the more able individual. Thus, in this case, $U_H^{(1)}[U_L^{(2)}[U_H^*]] = U_H^*$. This demonstrates that in the progressive rate case, the UPF's for the taxes of type (1) and (2) coincide.

In Figure 2-b only the less able individual exhibits no substitutability.[3] In this case, given U_H^* as in point B, the supporting efficient tax schedule of type (2) is OADBE implying that $U_L^{(2)}[U_H^*] = U_L^*$ is reached at point A. On the other hand, given U_L^* at A, the efficient tax schedule of type 1 is O'AB'E', implying that point B' and $U_H' < U_H^*$ for the more able individual. Thus, in this case, $U_H^{(2)}[U_L^{(1)}[U_H^*]] < U_H^*$. This demonstrates that in the progressive range, the tax of type (1) is Pareto-dominated by that of type (2). The reverse happens in Figure 2-c, where the high ability person has a zero elasticity of substitution. Here, with the same notation, $U_H^{(2)}[U_L^{(1)}[U_H^*]] > U_H^*$, demonstrating that in the progressive range, the increasing marginal tax rates case of type (1) is Pareto-superior.

These examples demonstrate that it is impossible to establish in general that either the case of increasing or decreasing marginal tax rates is superior. These examples further suggest that if

---

[3] Even though we refer to one individual displaying substitutability with the other displaying none, this need not violate our assumption of common preferences; the elasticity of substitution is a local property of preferences and the two individuals end up in different C, L combinations.

the elasticity of substitution is lower at higher levels of utility, other things equal the increasing tax rates system is more likely to be optimal, with the reverse happening when the elasticity of substitution is higher at higher levels of utility. Given our limited knowledge on how the elasticity of substitution varies, it seems that one should then confine attention to the case where all taxpayers share a constant elasticity of substitution utility function.

4. Social Welfare Maxima with Many Individuals - Methodology

The investigation of efficient tax structures in a two-class economy reveals that the shape of the tax structure will depend on the particulars of the problem studied. A logical next step is to investigate what characterizes tax structures that maximize social welfare, when the form and parameterization of the social welfare function, utility function and distribution of wages are specified. This approach, which requires numerical optimization methods, has been followed by Stern (1976) and others in the study of the optimal linear tax system. In what follows we adapt that methodology to the study of the optimal two-bracket linear income tax system.

We approximate a lognormal distribution of wages with an economy of 1,000 individuals, whose wages are drawn from a lognormal distribution with $\mu = -1$ and $\sigma = 0.39$. Thus the lowest wage is at the cumulative frequency of 0.0005, the second lowest is at 0.0015, and so on until the highest wage is at 0.9995 of the cumulative distribution. We compute the optimal decision of a taxpayer that maximizes a CES utility function with elasticity of substitution ε while facing a two-bracket linear income tax, i.e.,

$$\underset{C_i,L_i}{\text{Max}}\ U_i = \left(\alpha\ C_i^{\frac{\varepsilon-1}{\varepsilon}} + (1-\alpha)(1-L_i)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}$$

subject to $C_i = w_i L_i - T_i$

where $T_i = -G + t_1 w_i L_i \qquad w_i L_i \le \bar{Y}$

$= -G + t_1 \bar{Y} + t_2(w_i L_i - \bar{Y}) \qquad w_i L_i > \bar{Y}$

Here C_i is consumption, L_i is labor, w_i is the wage rate, T is total tax payments, G is the demogrant, t_1 and t_2 are the marginal tax rates applied to the first and second brackets, respectively, and $\bar{Y}$ is the cutoff level of income between the two brackets. Solution of this maximization problem requires checking of the possible corner solutions that can arise.

The next step is to choose the values of t_1, t_2, $\bar{Y}$, and G that maximize the social welfare function, which takes the standard form of

$$W = \sum_{i=1}^{1000} \frac{1}{v}\left[U_i\left(C_i^*(w_i, G, t_1, t_2, \bar{Y}), L_i^*(w_i, G, t_1, t_2, \bar{Y})\right)\right]^v$$

subject to $\sum_{i=1}^{1000} T_i = R$

where C_i^* and L_i^* are the solutions to the problem of expression (1) and R is the required revenue of the government. We execute this maximization by conducting a grid search, with grid size of 0.001, over values of t_1, t_2, $\bar{Y}$ and the value of G which assures that the revenue constraint is met.[4]

5. Social Welfare Maxima with Many Individuals - Results

Table 1 presents the results of these calculations. To facilitate comparison with earlier results, we focus on certain parameterizations investigated by Stern (1976). We present the eighteen permutations of v = 1, -2, of R = 0, 0.05, and 0.10 and of ε = 0.2, 0.4, and 1.0.[5] The parameter settings are shown in the first three columns of Table 1. The fourth and fifth columns

---

[4]The value of G is solved using the tax programming model described in Yitzhaki (1982); the revenue constraint is satisfied within a tolerance of 10^{-10} of national income. The fact that when $t_2 < t_1$ tax revenue is not continuous requires the combination of a grid search and standard numerical optimization techniques.

[5]Following Stern (1976) we select the value of α so that, in the absence of taxes, as we vary ε, L_i for the taxpayer with mean skills equals 0.67. This procedure yields α = 0.33 when ε = 1, α = 0.41 when ε = 0.4, and α = 0.56 when ε = 0.2.

present the demogrant and tax rate of the simple linear case[6] and the next four columns present the parameters of the optimal two-bracket linear income tax.

The most striking aspect of these results is that in all cases the marginal tax rate that applies to the second bracket (t_2) lies below the marginal tax rate that applies to the first bracket (t_1). Note, however, that in all cases where the optimal G is positive, the average tax rate is increasing in income. This latter result can be checked by verifying that in all cases $t_2 > (t_1 \bar{Y} - G)/\bar{Y}$. Thus, using the terminology of Musgrave and Thin (1948), the optimal two-bracket linear income tax generally exhibits average rate progressivity but not marginal rate progressivity.

A lower value of v, which corresponds to a greater social preference for equality, increases the optimal demogrant as well as both t_1 and t_2. The gap between t_1 and t_2 also widens. Figure 3 shows the effect of moving from a v of one to a v of -2 in the case where R = 0.05 and ε = 0.4. Higher values of the revenue requirement reduce the optimal demogrant and increase both t_1 and t_2, while the gap between t_1 and t_2 widens. Figure 4 shows the effect of increasing R from 0 to 0.05 or 0.10, when v is -2 and ε = 0.4. Higher values of ε reduce the demogrant and both marginal tax rates, but do not have a consistent effect on the gap between t_1 and t_2. Figure 5 shows the optimal tax schedule for the three values of ε, when R = 0.05 and v = -2.

Figure 6 illustrates the difference between the optimal one-bracket and two-bracket linear income tax schedules for the case where v = -2, ε = 0.4, and R = 0.05. Compared to the optimal linear income tax, the optimal two-bracket linear income tax leaves both the highest and lowest wage taxpayers better off, at the expense of the middle-income earners.

The optimal cutoff point between the two brackets, $\bar{Y}$, lies for these simulations between 0.233 and 0.371. Its value is not highly sensitive to the value of v or R, but does decline

---

[6]Because our optimization techniques are not identical, our solution to the optimal linear income tax generally is slightly different from what is presented by Stern (1976). The optimal values of G and t generally differ by no more than 5 percent.

One other methodological difference is worth mentioning. We allow the demogrant to be negative while Stern does not allow such a solution. Thus, in the three cases where the optimal G is negative, our results diverge from Stern's more than usual.

significantly when ε is 1.0 compared to when it is either 0.2 or 0.4. The fraction of taxpayers in the first bracket, denoted f in Table 1, varies much more widely than $\bar{Y}$, from a low of 0.459 to a high of 0.832. Note that, in comparing two tax structures, a higher $\bar{Y}$ does not necessarily correspond to a higher fraction of taxpayers in the first bracket. This is because not only $\bar{Y}$, but also G, t_1, and t_2, influence where on the budget set a taxpayer chooses to locate. The column of Table 1 headed W/L/W shows the number of taxpayers who, respectively, gain, lose, and gain in moving from the optimal one-bracket to the optimal two-bracket linear income tax. These figures vary within a fairly narrow range, with the poorest 22 to 34% gaining, the middle 54 to 58% losing, and the top 11 to 24% gaining. The number of upper-income gainers is positively related to the elasticity of substitution between leisure and consumption.

Figures 7 and 8 plot, by the cumulative distribution of the wage rate, for two different elasticities of substitution but the same value of v and R, the difference in taxes paid and utility under the optimal linear and two-bracket systems.[7] Note the discrete increase in taxes paid at the point where it is optimal for a taxpayer to "jump" to the lower marginal rate second bracket; utility is, however, continuous at this point. Note also that an individual could have higher welfare when facing a tax system under which he pays more tax, reflecting either higher consumption, higher leisure, or both. Until the cutoff income is reached , the higher the income, the greater the loss, because the two-bracket tax always has a higher demogrant and a higher t_1. At incomes greater than $\bar{Y}$, the utility difference is made up because t_2 under the two-bracket tax is below the optimal linear rate. Note finally that the number of high-income gainers is greater in Figure 7 ($\varepsilon = 1$) than in Figure 8 ($\varepsilon = 0.2$) because the bracket cutoff comes much sooner in the wage distribution. Figure 9 shows in C - Y space the indifference curves of the critical taxpayers for the parameterization of Figure 6, and the budget sets for the optimal linear and optimal two-bracket tax schedules. Individuals 288 and 852 are both indifferent between the two schedules. Individual

[7] To facilitate the comparison, these figures show the change in utility from moving to a two-bracket tax, and minus the change in tax liability.

771 is the taxpayer who, under the two-bracket schedule, jumps between the lower and upper branches. Individuals with a wage rate higher than that of individual 771 will have flatter indifference curves in C - Y space, and will therefore locate on the upper portion of the schedule.

How valuable, from a social point of view, is it to have the extra tax instrument of a second bracket? Figure 6, which shows that the optimal two-bracket tax does not differ markedly from the optimal linear tax, suggests that the value is limited. The column of Table 1 headed B provides some information to evaluate that apparent conclusion. B is defined as the additional amount of revenue that could be raised by a two-bracket tax, compared to the optimal linear tax, while holding social welfare constant at the value achieved by the two-bracket tax. While at its greatest value of 0.0005040, the gain is equal to 0.199% of national income (where national income is defined as $\sum_i w_i L_i$), at its lowest value of 0.0000005 it is equal to just 0.0002% of national income. Thus, for the range of parameters we have investigated, the gain varies by a factor of one thousand.

As a standard comparison, consider that in the United States the budget of the Internal Revenue Service is currently 0.145% of national income. If, for example, the incremental cost of administering and enforcing a constant-quality two-bracket tax is one-tenth of the current total administrative cost, then moving to a two-bracket system increases national income if the gain is greater than 0.0145% of national income. This standard is met in many, but clearly not all, of the examples considered here.

The value of B is, holding the other parameters constant, positively related to R and negatively related to v. The partial relationship of B to ε is not, however, monotonic. The intuition for this result is as follows. The value of having two brackets lies in the ability to impose a lower marginal rate on higher-income people, thus inducing them to supply more labor. An increase in ε increases the gain from a lower tax rate, but at the same time reduces the attractiveness of all non-lump-sum taxes. When v is -2, as ε rises the latter effect comes to dominate the former, so that the optimal difference between t_1 and t_2 first rises as ε goes from 0.2 to 0.4, but falls when ε goes to 1.0 (in the cases where R = 0 or 0.05). The social value of being able to set t_2 lower than t_1 is lower when a lower general level of all marginal tax rates is called for. This also explains

the effect of R and v on B. A higher R calls for higher marginal taxes at an optimum, so that differentiating t_2 from t_1 is more valuable. The same reasoning applies to declines in v.

6. Conclusions

A linear income tax, with demogrant, offers considerable administrative advantages over more complex graduated income tax systems. Yet historically countries have accepted the higher administrative costs in order to achieve a more progressive distribution of the tax burden than that offered by a linear income tax system.

The results presented in this paper challenge the wisdom of this pervasive policy. We show that the benefits of allowing two brackets rather than one are very sensitive to the parameterization of the problem, and thus may or may not be sufficient to justify the additional administrative cost. Most strikingly, we show that if a second bracket is to be employed, income in the higher bracket should be taxed at a lower rate than income in the first bracket, not at a higher rate as characterizes the statutory tax schedules of most countries. The second bracket should be utilized to induce greater labor supply from the most productive segment of society, with the increased tax revenue used to lower the tax burden of the least productive segment. Although the calculated optimal tax system features declining marginal tax rates, it still generally features increasing average tax rates, so that it is progressive but not graduated, in the standard sense of these terms.

All of the foregoing conclusions are derived from a highly stylized model with only labor income and using a small set of parameter choices, and therefore at this stage ought to be treated with some caution as to their generality. However, the model and parameterizations are standard in the optimal income tax progressivity literature, so the conclusions drawn from them are important extensions to previous knowledge. Furthermore, we have demonstrated in a much more general setting that efficient income tax structures may exhibit either declining or increasing marginal tax rates, contrary to earlier work which purports to demonstrate the sub-optimality of declining marginal rate structures. Which kind of tax structure is optimal and what its quantitative

characteristics are depend on obtaining more precise information about the structure of individual preferences, social tradeoffs between equity and efficiency, the distribution of skills sand the administrative costs of tax structures of varying complexity.

TABLE 1

Comparison of Optimal One-Bracket and Optimal Two-Bracket Linear Income Taxes

Parameters			Linear Tax		Two-Bracket Tax					Comparison of Two Tax Structures	
v	R	ε	G	t	G	t_1	t_2	$\bar{Y}$	f	W/L/W	B (x 1,000)
1	0	0.2	0.0897	0.356	0.0933	0.373	0.323	0.306	0.768	340/550/110	0.0132
1	0	0.4	0.0561	0.223	0.0584	0.234	0.202	0.300	0.748	316/561/124	0.0074
1	0	1.0	0.0316	0.125	0.0323	0.129	0.123	0.233	0.498	252/580/168	0.0005
1	0.05	0.2	0.0576	0.401	0.0623	0.422	0.362	0.322	0.756	335/550/115	0.0215
1	0.05	0.4	0.0178	0.253	0.0224	0.274	0.229	0.298	0.678	316/555/129	0.0162
1	0.05	1.0	-0.0123	0.140	-0.0110	0.149	0.133	0.263	0.551	246/571/183	0.0034
1	0.10	0.2	0.0291	0.453	0.0356	0.481	0.408	0.328	0.712	313/553/134	0.0341
1	0.10	0.4	-0.0175	0.290	-0.0100	0.326	0.257	0.295	0.597	279/559/162	0.0324
1	0.10	1.0	-0.0545	0.159	-0.0500	0.181	0.149	0.257	0.459	234/579/187	0.0120
-2	0	0.2	0.1591	0.656	0.1614	0.668	0.613	0.312	0.800	305/569/126	0.0687
-2	0	0.4	0.1227	0.522	0.1258	0.539	0.469	0.295	0.772	293/566/141	0.1300
-2	0	1.0	0.0795	0.355	0.0813	0.367	0.320	0.291	0.773	276/586/138	0.0698
-2	0.05	0.2	0.1337	0.709	0.1357	0.719	0.662	0.348	0.832	314/573/113	0.0824
-2	0.05	0.4	0.0963	0.584	0.0996	0.603	0.518	0.314	0.771	288/564/148	0.2033
-2	0.05	1.0	0.0480	0.412	0.0517	0.435	0.368	0.265	0.661	242/563/195	0.2012
-2	0.10	0.2	0.1098	0.760	0.1120	0.770	0.717	0.371	0.832	322/568/110	0.0921
-2	0.10	0.4	0.0727	0.648	0.0773	0.671	0.585	0.316	0.724	277/551/172	0.2981
-2	0.10	1.0	0.0210	0.481	0.0272	0.518	0.423	0.259	0.590	224/540/236	0.5040

APPENDIX I

In this appendix we refute Sheshinski's (1989) proof that, in a two-bracket linear income tax, at the optimum t_2 cannot be less than t_1. In what follows we use Sheshinski's notation, in which $\bar{y} = \bar{Y}$, $\beta_1 = 1 - t_1$, $\beta_2 = 1 - t_2$, $\alpha_1 = -G$, and $\alpha_2 = -G + t_1 \bar{Y}$.

Sheshinski's proof is flawed because it fails to account for the fact that, in the case where $\beta_2 > \beta_1$, as the tax instruments (α_1, α_2, β_1, β_2, or $\bar{y}$ change, $\tilde{w}$ will change, where $\tilde{w}$ is the wage rate at which an individual will be indifferent between which segment of the budget set he locates. In fact, as (say) $\bar{y}$ increases, some individuals will "jump" from the higher segment of their budget constraint to the lower segment. Although utility is continuous with respect to w at $\tilde{w}$, tax revenue is discontinuous. Thus each of the first-order conditions (27) through (30) in Sheshinski (1989) needs to have an extra term appended to it for the revenue cost of "jumpers." In each case the extra term has the form $\lambda(\Delta R)\, f(\tilde{w}) \frac{d\tilde{w}}{dx}$, where, λ is the LaGrange multiplier for the revenue constraint, ΔR is the change in revenue from a jump from the highest segment to the lower segment of the budget constraint and x is a tax instrument.

Using the definition of $\tilde{w}$ (expression (25) in Sheshinski), we can derive $\frac{d\tilde{w}}{d\bar{y}}$ as follows:

$$u^*\left(\beta_1, \alpha_1, \tilde{w}\right) = u^*\left(\beta_2, \alpha_2, \tilde{w}\right)$$

$$u^*\left(\beta_1, \alpha_1, \tilde{w}\right) = u^*\left(\beta_2, \alpha_1 + \left(\beta_1 - \beta_2\right)\bar{y}, \tilde{w}\right)$$

Let $\bar{y}$ change, and calculate the change in $\tilde{w}$ that must occur to maintain the above equality

$$\frac{\partial u^*}{\partial w}\left(\beta_1, \alpha_1, \tilde{w}\right) d\tilde{w} = \frac{\partial u^*}{\partial \alpha_2}\left(\beta_2, \alpha_2, \tilde{w}\right)\left(\beta_1 - \beta_2\right) d\bar{y} + \frac{\partial u^*}{\partial w}\left(\beta_2, \alpha_2, \tilde{w}\right) d\tilde{w}$$

But we know (see Sheshinski's footnote 3) $\frac{\partial u^*}{\partial w}$, yielding

$$\frac{\partial u^*}{\partial c}\beta_1 \ell^*\left(\beta_1 \tilde{w}\right) d\tilde{w} = \frac{\partial u^*}{\partial \alpha_2}\left(\beta_2, \alpha_2, \tilde{w}\right)\left(\beta_1 - \beta_2\right) d\bar{y} + \frac{\partial u^*}{\partial c}\beta_2 \ell^*\left(\beta_2 \tilde{w}\right) d\tilde{w}$$

Rearranging, we can write

$$\frac{d\tilde{w}}{d\bar{y}} = \frac{\frac{\partial u^*}{\partial \alpha_2}\left(\beta_2, \alpha_2, \tilde{w}\right)\left(\beta_1 - \beta_2\right)}{\frac{\partial u^*}{\partial c}\left[\beta_1 \ell^*\left(\beta_1 \tilde{w}\right) - \beta_2 \ell^*\left(\beta_2 \tilde{w}\right)\right]},$$

which is greater than zero if $\beta_2 > \beta_1$. (Note from Sheshinski's footnote 3 that $\beta_2 \ell^*\left(\beta_2 \tilde{w}\right) > \beta_1 \ell^*\left(\beta_1 \tilde{w}\right)$.)

The change in tax revenue upon moving from the higher to the lower segment, ΔR, is equal to

$$\left(1 - \beta_2\right)\left(w\ell^*\left(\beta_2 \tilde{w}\right) - \bar{y}\right) - \left(1 - \beta_1\right)\left(\bar{y} - w\ell^*\left(\beta_1 \tilde{w}\right)\right),$$

which is strictly negative when $\beta_2 > \beta_1$.

Now focus on expression (31), on which Sheshinski's proof of the non-optimality of concave budget sets rests. To this term must be added

$$\lambda\left[\left(1 - \beta_2\right)\left(w\ell^*\left(\beta_2 \tilde{w}\right) - \bar{y}\right) - \left(1 - \beta_1\right)\left(\bar{y} - w\ell^*\left(\beta_1 \tilde{w}\right)\right)\right] \cdot$$

$$f(\tilde{w})\left[\frac{\frac{\partial u^*}{\partial \alpha_2}\left(\beta_2, \alpha_2, \tilde{w}\right)\left(\beta_1 - \beta_2\right)}{\frac{\partial u^*}{\partial c}\left[\beta_1 \ell^*\left(\beta_1 \tilde{w}\right) - \beta_2 \ell^*\left(\beta_2 \tilde{w}\right)\right]}\right]$$

which is strictly positive if $\beta_2 > \beta_1$.

Sheshinski's proof rests on the negativity of (31), which means that there is no interior solution to $\bar{y}$. But now that this additional positive term has been added, it is possible that an interior $\bar{y}$ will be optimal.

APPENDIX II

Following Sadka (1976), it is convenient to transform the choice space and to consider each individual as selecting C_i and Y_i to maximize $U^i(C_i, Y_i) = (C_i, 1 - Y_i/ w_i)$, given the common budget constraint: $C_i = Y_i - T(Y_i, \Theta)$. An efficient tax system Θ, can be considered to maximize $\delta U^*_L(\Theta) + U^*_H(\Theta)$, where $\delta = n\delta'$, with $\delta' \geq 0$ representing the social relative weight for the utility of the less able individual and $n = n_L/n_H$. The revenue constraint on Θ can be presented as: $nT^*_L(\Theta) + T^*_H(\Theta) \geq R^*$, where $R^* = R/n_H$. Given the single-crossing property of indifference curves it is possible to establish that at the socially optimal Θ: (1) The revenue constraint is met with equality, (2) $t_1 \leq 1$ and $t_2 \leq 1$, (3) if $t_1 \neq t_2$, then $Y^*_H(\Theta) \geq \bar{Y} \geq Y^*_L(\Theta)$. Sadka's argument establishes that when $t_1 \geq t_2 > 0$ (and also when $t_1 \leq t_2 < 0$) one can increase $\bar{Y}$ and change t_2 to zero, so as to keep the lower ability individuals unperturbed while improving the welfare of the more able individuals without reducing their tax payments. Seade's mirror argument implies that if, $t_2 \geq t_1 > 0$ (or if $t_2 \leq t_1 < 0$) one can lower $\bar{Y}$ and set t_1 to zero, keeping the high ability individuals unperturbed, while improving the welfare of the low ability individuals, without altering their tax payments. The cases where $t_1 < 0 < t_2$ or where $t_1 > 0 > t_2$ can be eliminated by the same reasoning.

The more novel parts of the proposition relate to the tangency conditions. Presenting only the case of redistribution towards the less able, suppose that $t_1 > t_2 = 0$, and yet, as illustrated in Figure 10-a, U^*_H does not touch the lower branch of the tax schedule OAD. By changing the schedule from OADBE to O'AFD'BE one is able to increase U^*_L while collecting more (net) tax revenue and leaving U^*_H unchanged. Similarly, suppose that $t_2 > t_1 = 0$ and yet $Y^*_L < \bar{Y}$, as illustrated in Figure 10-b. By changing the schedule from OADBE to OABE', U^*_H increases while increasing the tax revenue. The intuition behind these tangency conditions is simply that, holding constant the zero tax rate for either the top or bottom bracket, the schedule should be adjusted to make the other marginal tax rate as low as possible.

REFERENCES

Atkinson, A. "How Progressive Should Income Tax Be?" In M. Parkin and A. Nobay (eds.), Essays in Modern Economics (London: Longman, 1973).

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

Mirrlees, J. "Optimal Tax Theory: A Synthesis," Journal of Public Economics, 6, 1976, pp. 327-358.

Musgrave, J. and T. Thin, "Income Tax Progression, 1929-1948," Journal of Political Economy, 56, 1948, pp. 498-514.

Phelps, E. "The Taxation of Wage Income for Economic Justice," Quarterly Journal of Economics, 3, 1973, pp. 331-354.

Sadka, E. "On Income Distribution, Incentive Effects, and Optimal Income Taxation," Review of Economic Studies, 9, 1976, pp. 261-268.

Seade, J. "On the Shape of Optimal Tax Schedules," Journal of Public Economics, 7, 1977, pp. 203-236.

Sheshinski, E. "The Optimal Linear Income Tax," Review of Economic Studies, 39, 1972, pp. 297-302.

Sheshinski, E. "Note on the Shape of the Optimum Income Tax Schedule," Journal of Public Economics, 40, 1989, 201-215.

Slemrod, J. "Do We Know How Progressive the Income Tax System Should Be?" National Tax Journal, 36, 1983, pp. 361-369.

Stern, N. "On the Specification of Models of Optimum Income Taxation," Journal of Public Economics, 6, 1976, pp. 123-162.

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

Stiglitz, J. "Pareto Efficient and Optimal Taxation and the New New Welfare Economics," In A. Auerbach and M. Feldstein (eds.), Handbook of Public Economics, Vol. 2 (Amsterdam: North-Holland, 1987).

Tuomola, M. Optimal Income Tax and Redistribution (Oxford: Clarendon, 1990).

Yitzhaki, S. "A Tax Programming Model," Journal of Public Economics, 19, 1982, pp. 107-120.

FIGURE 1
Utility Possibility Frontiers for Different Tax Systems

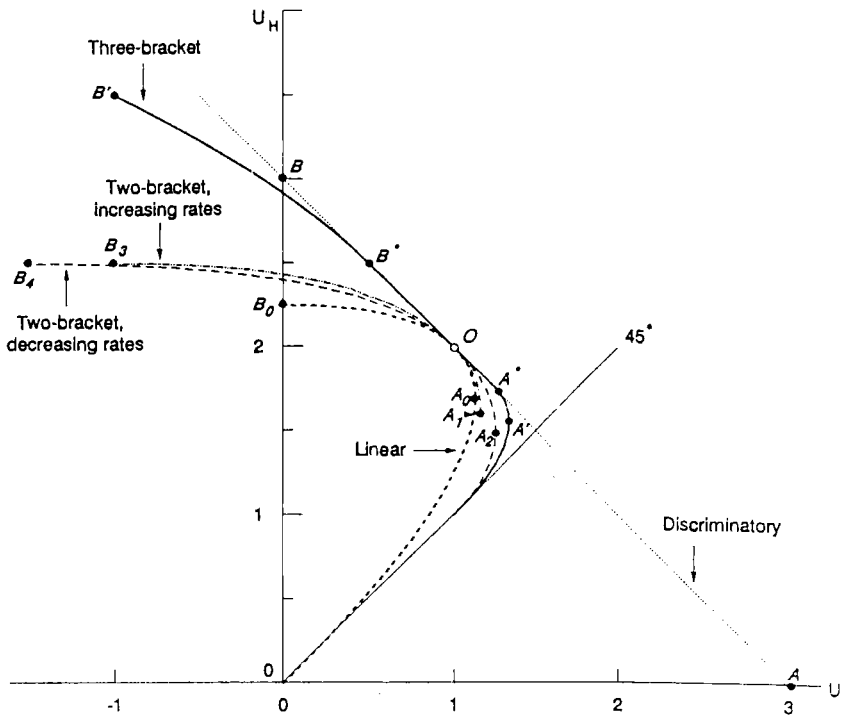

FIGURE 2a

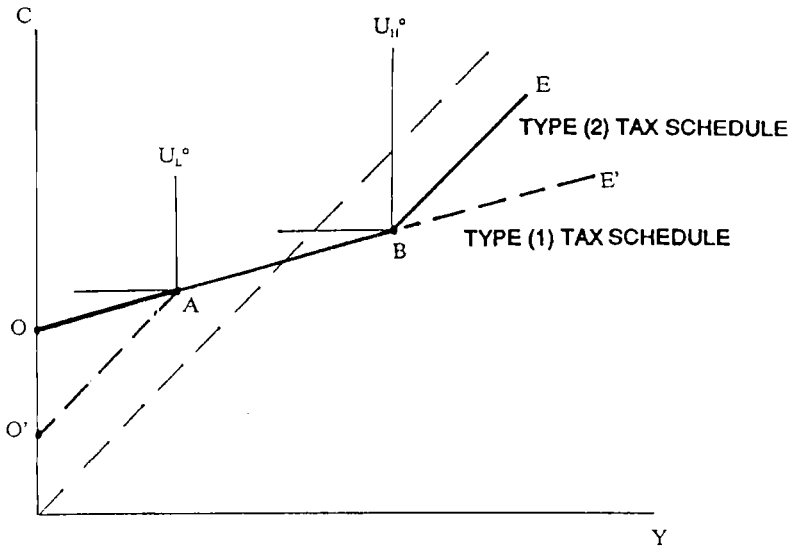

FIGURE 2b

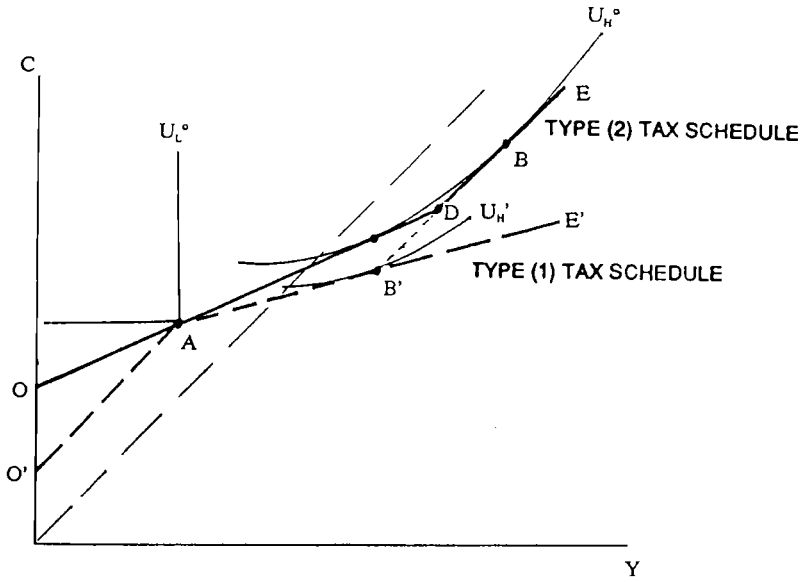

FIGURE 2c

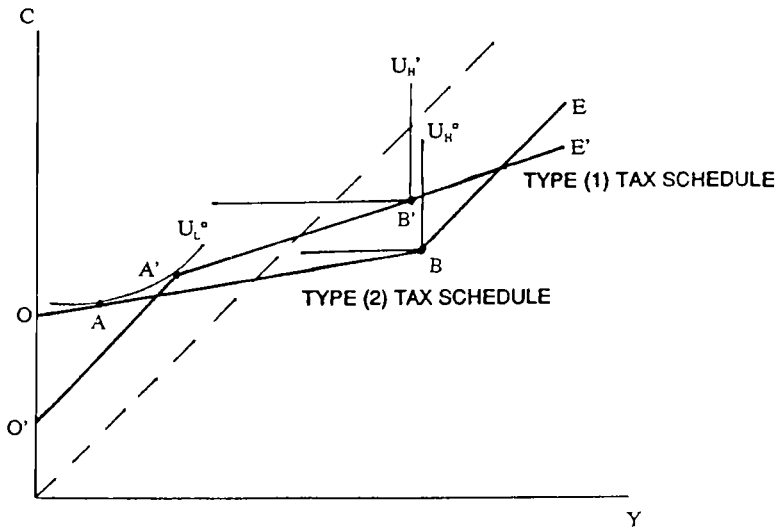

FIGURE 3
Effect of Changing v
(e=0.4, R=0.05)

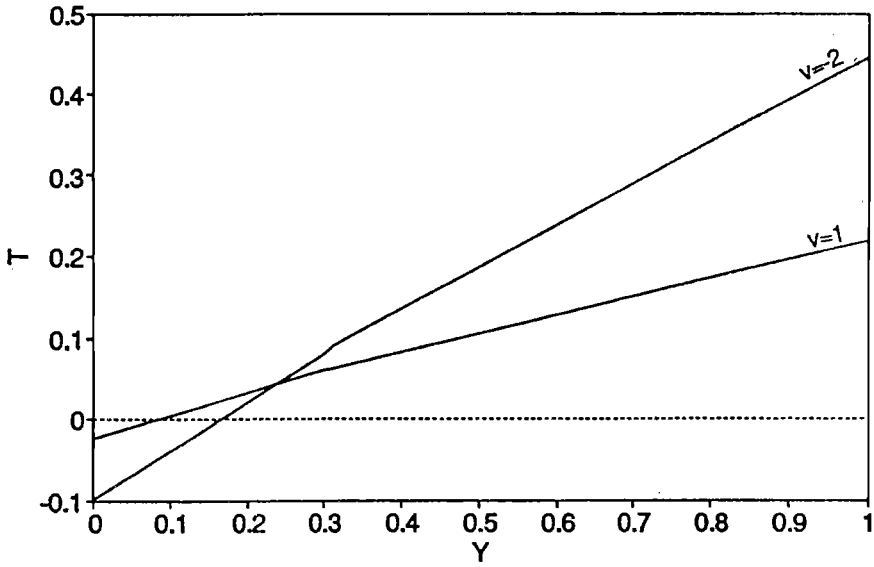

FIGURE 4
Effect of Changing R
(v=-2, e=0.4)

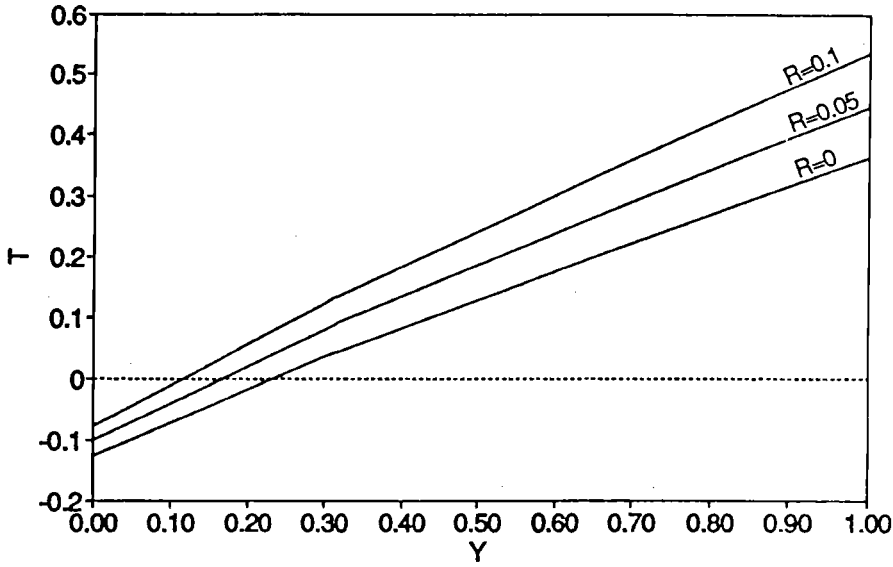

FIGURE 5
Effect of Changing e
(v=-2, R=0.05)

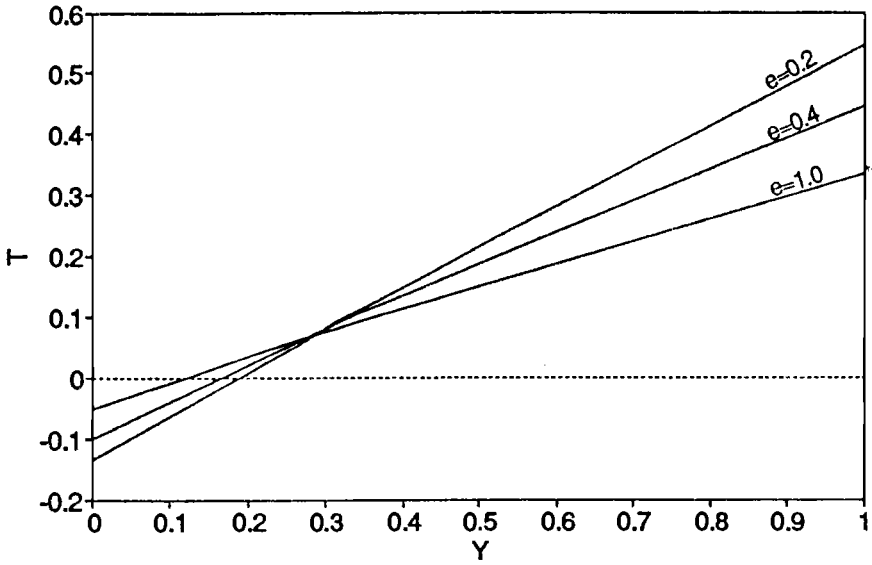

FIGURE 6
Comparison of Optimal Linear and
Two-Bracket Tax Schedules
(v=-2, e=0.4, R=0.05)

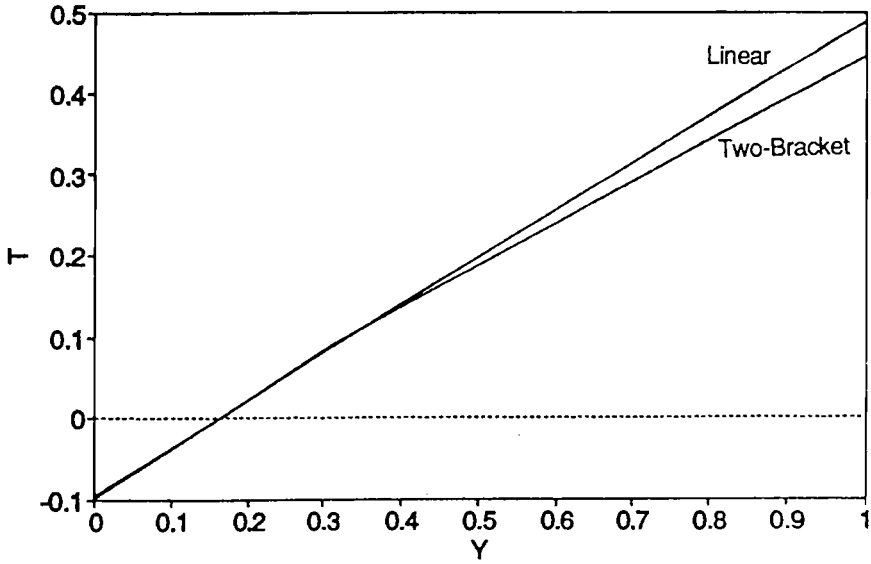

FIGURE 7
Comparison of Utility and Tax Paid Under Linear and Two-Bracket Systems
(v=-2, e=0.2, R=0.05)

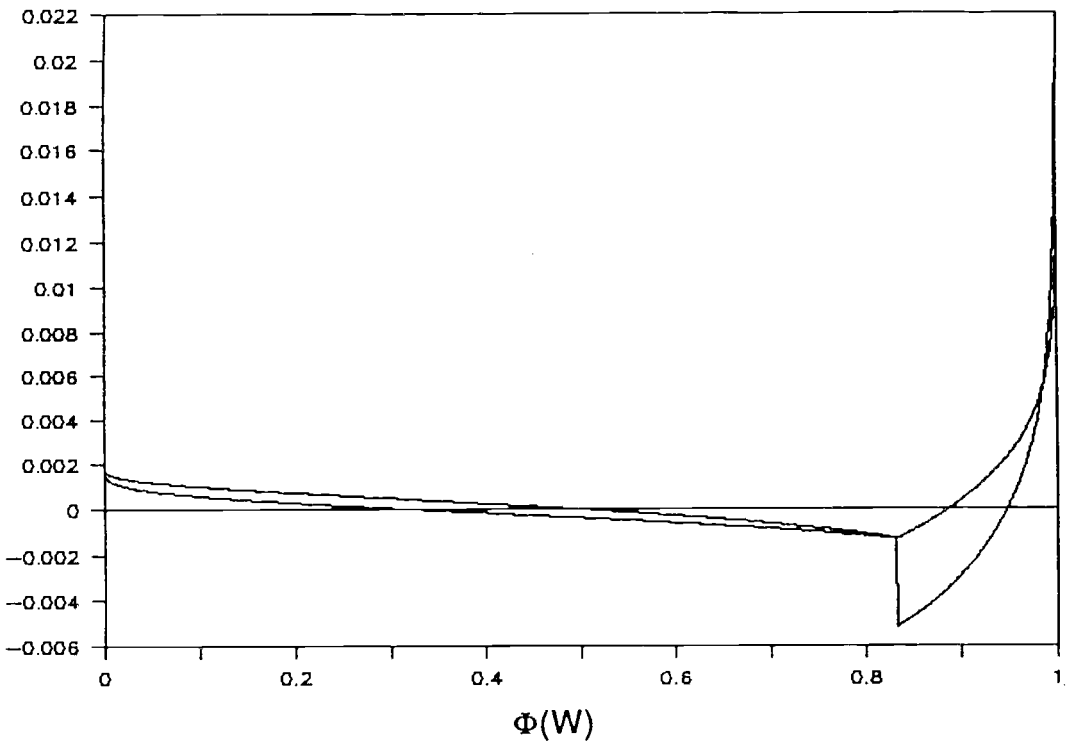

FIGURE 8
Comparison of Utility and Tax Paid Under Linear and Two-Bracket Systems
(v=-2, e=1.0, R=0.05)

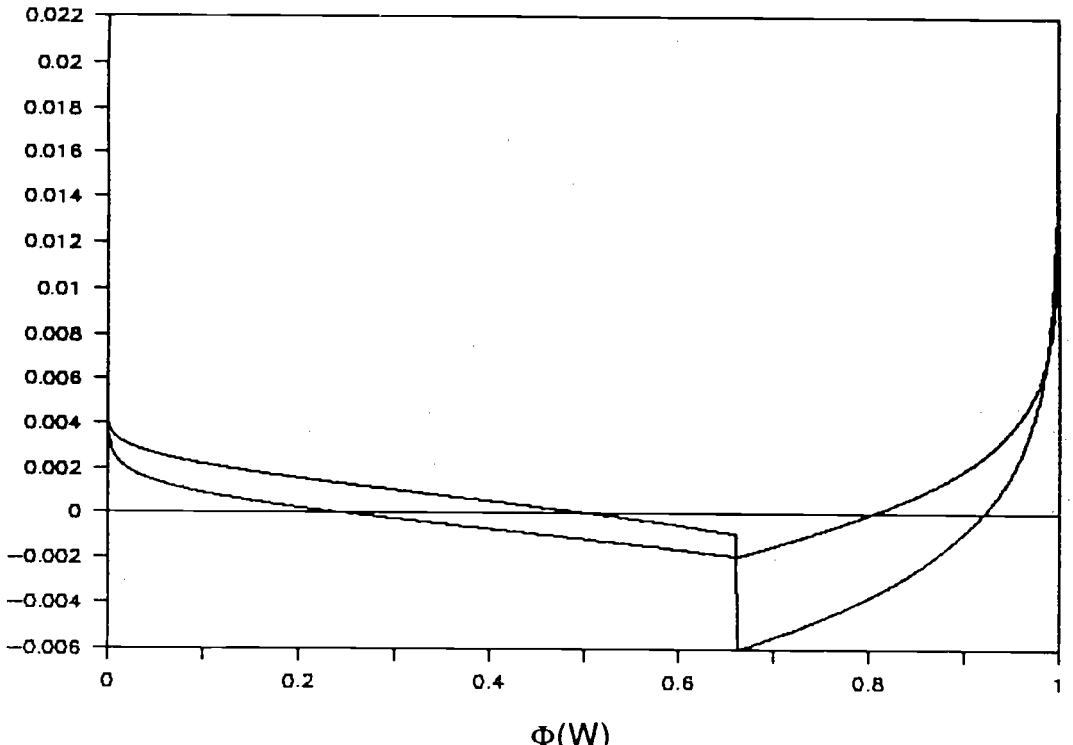

FIGURE 9
Optimal Tax Systems and Indifference Curves of Critical Taxpayers
(v=-2, e=0.4, R=0.05)

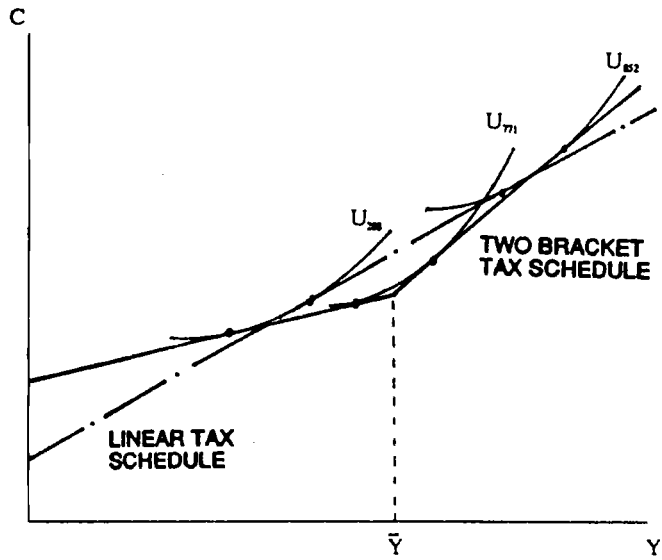

FIGURE 10a

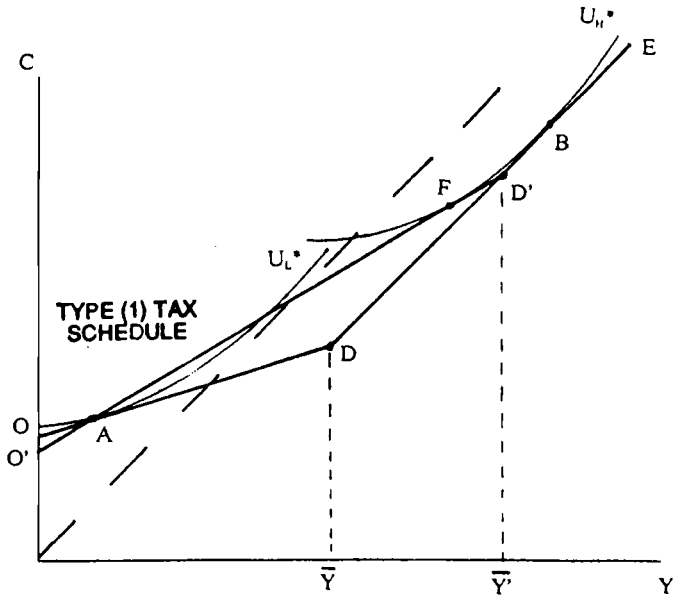

FIGURE 10b