Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates

Peter A. Diamond

The American Economic Review, Vol. 88, No. 1. (Mar., 1998), pp. 83-95.

Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28199803%2988%3A1%3C83%3AOITAEW%3E2.0.CO%3B2-8

The American Economic Review is currently published by American Economic Association.

Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates

By PETER A. DIAMOND *

Using the Mirrlees optimal income tax model with quasi-linear preferences, the paper examines conditions for marginal tax rates to be rising at high income levels and declining in an interval containing the modal skill. It examines conditions for the marginal tax rate to be higher at a low skill level than at the high skill level with the same density—an argument only holding for skill levels above a cutoff where resources of a worker are marginally of the same value as resources of the government. Data on earnings rates are presented. (*JEL* H21)

The trade-off between efficiency and income distribution plays a central role in analyzing tax policy.[1] The modern framework for analyzing this trade-off using nonlinear income taxes was created in James A. Mirrlees (1971). While this formulation crystallized a presentation of the income tax problem and derived some of the properties of optimal income taxation, the implications for policy have been somewhat limited. For example, the *Financial Times* (September 11, 1995 p. 24) has summarized the policy impact of the optimal income tax literature as: "A few general principles none the less gained the status of received wisdom, for example that marginal tax rates should be constant and modest over most of the income range, but zero at the top and bottom." The public finance community has recognized that the results deriving zero marginal tax rates at the top and the bottom of the income distribution are of little or no relevance for policy. This paper argues that the case for nonconstant and high marginal tax rates in the Mirrlees model is considerably stronger than has been realized. The technical contribution of this paper is very modest, being primarily a rearrangement of terms in the standard first-order condition for optimal income taxation. This rearrangement leads to a different way of approaching the combinations of assumptions that will sign the change in marginal tax rates with income level. In addition, by concentrating on the case where there is a zero income derivative of labor supply, the intuition behind the first-order condition becomes clearer.

Section I reviews some of the previous literature. Section II presents the optimal income tax problem. Section III examines conditions for marginal tax rates to be rising at income levels above the modal skill level. Section IV examines the level of marginal tax rates on very high incomes. Section V examines conditions for the marginal tax rate to be declining and to be higher at a skill level below the modal skill than at the skill level above the mode with the same density of skills. These arguments apply for skill levels above a cutoff level, where resources are of the same value in the hands of the government and in the hands of a worker with the cutoff skill level. Section VI considers another example. Section VII looks at data on earnings rates to suggest the relevance of alternative empirical assumptions on the distribution of skills. Some closing remarks are in Section VIII.

* Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139. I am grateful to Jon Gruber for suggesting the inclusion of data in this paper and providing me with the figures presented below, and to Marcus Berliant, Svetlana Danilkina, Jim Mirrlees, Eytan Sheshinski, Jay Wilson, and anonymous reviewers for helpful suggestions. I am also grateful to the National Science Foundation for research support under Grant SBR-9307876.

[1] If part of the population has potential income below per capita government expenditures, then it is impossible to finance government spending without some distorting taxes. This paper assumes that concern for the income of the poor is large enough that there is a positive transfer to those with zero income.

I. Review of the Literature

Formal results about the optimal income tax are fairly limited. (For recent expositions, see Matti Tuomala, 1990 Ch. 6 or Gareth D. Myles, 1995 Ch. 5.) Assuming that labor supply can be continuously adjusted, there is no gain from having marginal tax rates above 100 percent since no one will have such a tax at the margin. That is, the same outcome can be achieved with taxes no greater than 100 percent. It is usually assumed that preferences are such that consumption is an increasing function of the wage. Then, earnings will be nondecreasing in skill. It then follows that the optimal tax structure has nonnegative marginal rates (Mirrlees, 1971) and positive rates in the interior of the income distribution (Jesus K. Seade, 1982).

Assuming that there is a finite maximum to the skill distribution, the marginal tax rate should be zero at the income level of the top skill (Efraim Sadka, 1976; Seade, 1977). The argument for this result is quite intuitive. Assume this were not the case, then, extending the tax function to higher incomes with a zero tax rate would lead the top earner to work more, raising social welfare without losing any tax revenue. However, this condition need not convey information about optimal taxes over any significant region of incomes—the optimal rates need not approach zero until very close to the top. This point has been made by the numerical calculations in Tuomala (1984).

At the bottom of the skill distribution, in the presence of optimal taxes, there may or may not be an atom of individuals doing no work. If everyone works, then the argument for a zero marginal tax rate carries over (Seade, 1977). However, if there is an atom of nonworkers, the optimal tax has a positive marginal tax rate at the level where earnings begin (Udo Ebert, 1992). This latter case seems empirically more relevant.

In addition to these analytical results, presentation of the first-order condition for optimal taxes has generally been accompanied by observations on the factors leading to high or low rates, ceteris paribus. Considerable effort has gone into simulations, starting with that by Mirrlees. In his simulation Mirrlees assumed a utility function $u = \log[x] + \log[1 - y]$, where x is consumption and y is labor supply (in percentage terms), a social evaluation $G(u) = -\exp(-bu)/b$ $(b > 0)$, and a lognormal distribution of skills. He concluded (p. 206) that "Perhaps the most striking feature of the results is the closeness to linearity of the tax schedules." As seen in the survey in Tuomala (1990), similar results followed with some other simulations, but some simulations have shown other patterns, including a significantly inverse-U shaped pattern (e.g., Ravi Kanbur and Tuomala, 1994). As will be clarified below, simulation results are sensitive to both the utility function and the family of distributions of skills assumed, opening up the possibility of different conclusions.

II. Optimal Income Tax Problem

The Mirrlees optimal income tax problem is the maximization of the integral over the population of a concave function of individual utilities, subject to an aggregate budget constraint and subject to the constraint that individuals optimize in their choice of labor supply given the relationship between work and after-tax income. The only difference across individuals in the model is a difference in skills, with an individual of skill n having a marginal product equal to n. The model is a one-period model with only labor income. It is assumed that the government can observe income received but not hours worked or skill. Denoting consumption of someone with skill n by $x(n)$, labor (in percentage terms) by $y(n)$, and the concave utility function by $u(x, y)$, the social objective function can be stated as

$$(1) \qquad \int_{n_0}^{n_1} G\{u[x(n), y(n)]\} f(n)\, dn,$$

where $G(u)$ is an increasing and strictly concave function of utility, with G independent of n, and the distribution of skills is written as $F(n)$, with density $f(n)$. It is assumed that the distribution of skills is single-peaked, with a mode at n_m. The density is assumed to be positive and continuous between the bottom and the top skill levels, n_0 and n_1.

The resource constraint on this maximization can be stated in terms of output—that aggregate consumption be less than aggregate production minus government expenditures, E:

$$(2) \quad \int_{n_0}^{n_1} x(n)f(n)\,dn \leqq \int_{n_0}^{n_1} ny(n)f(n)\,dn - E.$$

This constraint can be stated alternatively in terms of taxes. Denoting taxes as a function of earnings as $T[ny(n)]$, consumption equals the difference between earnings and taxes, $x(n) = ny(n) - T[ny(n)]$. In this case, the government budget constraint is that taxes cover government expenditures:

$$(3) \quad \int_{n_0}^{n_1} T[ny(n)]f(n)\,dn \geqq E.$$

That the resource constraint can be stated equivalently in terms of government budget balance or in terms of aggregate supply and demand is a consequence of Walras Law.

In addition to the resource constraint, there is an incentive compatibility constraint. The government observes earnings, not hours worked or skill. Thus the government is restricted to setting taxes as a function only of earnings. The incentive compatibility constraint is that the selected labor supply, $y(n)$, maximizes utility, given the tax function, $u\{ny(n) - T[ny(n)], y(n)\}$. The relevant part of the tax function is just the part that is selected by someone—taxes can be set arbitrarily high at earnings levels that no one chooses with the optimal tax structure. Thus the incentive compatibility constraint can be stated in the familiar form that a worker with skill n does not prefer to imitate the earnings of a worker with a different skill level:

$$(4) \quad u\{ny(n) - T[ny(n)], y(n)\} \geqq u\{n'y(n') - T[n'y(n')], n'y(n')/n\} \quad \text{for all } n \text{ and } n'.$$

That is, someone of skill n would have to work n'/n times as much as someone with skill n' in order to have the same earnings level.

This paper will concentrate on the special case where there are no income effects on labor supply.[2] That is, it is assumed that utility is linear in consumption (referred to as quasi-linear):

$$(5) \quad u(x, y) = x + v(1 - y) = ny - T(ny) + v(1 - y),$$

where v is assumed to be strictly concave.[3] This assumption seems appropriate at very high income levels, since people at the top of the income distribution are likely to leave large estates—with a linear utility of bequests, neither consumption nor earnings vary with the exact level of estate. (The receipt of such bequests is not part of the model.) In addition, this assumption removes a source of considerable complication in tax analysis. In the presence of distorting taxes, income effects imply that lump-sum taxes have efficiency effects since they change distorted labor supply decisions.

This problem has some complexity in the derivation of the first-order condition for an optimal tax function, but is familiar from a number of mechanism design problems.[4] The simplest way to proceed is to replace the incentive compatibility conditions, (4), with the first-order condition for individual choice, which, from (5), can be written:

$$(6) \quad v'[1 - y(n)] = n\{1 - T'[ny(n)]\},$$

where T' is the marginal tax rate. For later use, it is convenient to note that for the quasi-linear utility function the elasticity of labor supply

[2] For a discussion of this case with a constant elasticity of labor supply, see Anthony B. Atkinson (1990). The complementary case where utility is linear in leisure has been studied; see Stefan Lollivier and Jean-Charles Rochet (1983) and John A. Weymark (1987).

[3] I assume that $G'[v(1)]$ is infinite, so that someone doing no work is given positive consumption and the nonnegativity constraint on consumption can be ignored.

[4] For an exposition of the mechanism design problem, see Drew Fudenberg and Jean Tirole, 1991 Ch. 7.

evaluated at the chosen labor supply of a worker of skill n, $e(n)$, satisfies

$$(7) \quad e(n) = -v'[1 - y(n)] \div \{ y(n)v''[1 - y(n)] \}.$$

Since the wage equals the skill level, this is the elasticity with respect to the wage, evaluated at the labor supply level that is chosen by someone with skill n.

More complicated than deriving the first-order condition is the problem of checking when the first-order condition does indeed characterize an optimum. The complication comes from the need to check that individual labor supplies satisfying the first-order conditions are globally optimal choices, and not just the solution to a first-order condition. This problem arises since the budget set is not convex when marginal tax rates are declining over some income levels. For any particular economy, one can check whether individual labor supplies are optimal. This issue raises the possibility that with the optimal tax, the distribution of skills results in a distribution of incomes that either has bunching at some income level (an atom of workers choosing the same income level) or a gap in the distribution of incomes. (Bunching at zero income or a gap between zero and the lowest positive income are not issues for the interpretation of the optimal tax structure below.) I do not explore this issue for this particular class of preferences, but proceed with analysis of the first-order condition; the analysis holds where the equilibrium distribution of incomes has no bunching and no gap, since generically the equation is a necessary condition for the optimal tax where this is true.

The first-order condition for the optimal tax can be calculated by specializing the condition in Mirrlees (1971) for quasi-linear preferences or deriving it directly, as is done in the Appendix. As usually written, the condition is:

$$(8) \quad p(n - v')f = [(v' - yv'')/n] \times \left[\int_n^{n_1} (p - G')\, dF \right],$$

where p is the Lagrange multiplier on the government's budget constraint and the functions v and G are evaluated at the appropriate labor supplies and consumption levels.

It is convenient to use the elasticity of labor supply and the marginal condition for individual choice, (6), to rewrite (8) as

$$(9) \quad T'/(1 - T') = [(e^{-1} + 1)/n] \times \left[\int_n^{n_1} (p - G')\, dF \right] \Big/ [pf].$$

Multiplying and dividing (9) by $(1 - F)$ to turn the integral into an average term, (9) can be rewritten as:

$$(10) \quad T'/(1 - T') = A(n)B(n)C(n)$$

$$\text{where } A(n) = e^{-1}(n) + 1;$$

$$B(n) = \int_n^{n_1} (p - G')\, dF / \{ p[1 - F(n)] \};$$

$$C(n) = [1 - F(n)]/[nf(n)].$$

The analysis below examines these three functions, $A(n)$, $B(n)$, and $C(n)$, under alternative assumptions on the functions $e(n)$, $f(n)$ and $G(u)$.

The absence of income effects allows an intuitive grasp of the factors that determine the optimal tax structure. Increasing the marginal tax rate affecting some skill level involves an increase in the deadweight burden for people at this skill level. Thus, the optimal marginal tax rate at some income level depends on the elasticity of labor supply at that income level, since this is important for marginal distortions. Increasing the marginal tax rate also transfers income from all individuals with higher skills to the government, without changing the distortions of their labor supplies. The weights on these two elements depend on the ratio of individuals with skills above this level to individuals with skills at this level and on the level of skill which links the tax on hours to the tax on income. This intuition is displayed in equation (10), where the first-order condition for the optimal income tax is written as a product of these three terms.

The same approach to signing the change in marginal tax rates can be used with the assumption of additive preferences, $u_{xy} = 0$, without the further assumption of quasi-linear preferences. The mathematical conditions for signing the change in the marginal tax rate are similar, although the economic interpretation of the conditions is more complex. $A(n)$ no longer depends only on the compensated elasticity of labor supply, but also has a term involving the second derivative of the utility-of-consumption, which is no longer assumed to be zero. Thus different economic assumptions are needed to sign the mathematical expressions.

III. Increasing Marginal Tax Rates

I turn now to analysis of (10), the first-order condition for the optimal income tax in the presence of quasi-linear preferences. In general, the variation in the elasticity of labor supply with skill will depend on the tax function, since taxes will affect the level of labor supplied and the elasticity varies with the quantity of labor supplied. One obvious exception, making for simpler analysis, is that of a constant elasticity of labor supply. In this case the utility of leisure satisfies $v(1 - y) = c\{1 - [1 - (1 - y)]^k\} = c(1 - y^k)$ for some constants c and k.

LEMMA A: *If* $v(1 - y) = c\{1 - [1 - (1 - y)]^k\} = c(1 - y^k)$, *then* $A(n)$ *is a constant.*

With quasi-linear preferences, a uniform transfer from the government to all workers has no effect on labor supply, and so no extra impact on the government budget. The welfare impact of such a transfer is the average of G' over the entire population. Thus one can conclude that the Lagrangian on the government budget constraint, p, is equal to the average of G':

$$(11) \qquad p = \int_{n_0}^{n_1} G'(n) f(n)\, dn.$$

Thus, $B(n_0)$ is equal to zero.

Given the incentive compatibility constraint, utility must be nondecreasing in skill and increasing where earnings are positive, since a worker can always have the same consumption as a worker with lower skill while doing less work, provided the level of work is positive. That is, above the skills at which there is no work, utility is increasing in n. With G a concave function, G' is then decreasing in n. Since $B(n)$ is the average of $[p - G']$ from the level n to the top of the skill distribution, $B(n)$ is increasing in n.[5]

Since p is equal to the average of G' and G' is nonincreasing, there is a critical value of n, denoted n_C, at which G' is equal to p:

$$(12) \qquad G'[u(n_C)] = p.$$

If n_C occurs at a level of skill where there is positive work, then n_C is unique; otherwise n_C is set equal to the highest skill at which there is no work. The level of n_C is endogenous, varying with both the structure of the economy and the nature of the social welfare function. To simplify the statement of results, analysis is restricted to economies where this critical level is below the modal level of skill:

$$(13) \qquad n_C < n_m.$$

This seems like the more interesting case, assuming that the mode of skills is near the median and the government would like to redistribute toward a fraction of the labor force well below one-half.

I note that $[1 - F(n)]B(n)$ is increasing in n up to n_C and then decreasing in n. These results are summarized as:

LEMMA B: $B(n)$ *is increasing in* n. $[1 - F(n)]B(n)$ *decreases in* n *for* $n > n_C$.

I turn now to the shape of the distribution of skills. Given the assumption of a single-

[5] Formally, differentiating $B(n)$, the derivative has the same sign as the average of $(p - G')$ from n to n_1 minus the value $[p - G'(n)]$. Since G' is decreasing in n, this difference is positive.

peaked density of skills, $nC(n)$ is decreasing in n for n below the modal level, n_m. For values of n above the modal level, the shape of $C(n)$ depends on the family of distributions assumed for skills. With a Pareto distribution above the modal skill level, (i.e., the density is proportional to $1/n^{1+a}$ for $a > 0$), then $C(n)$ is a constant above the modal skill level.

LEMMA C: *For $n < n_m$, $nC(n)$ is decreasing in n. For $n > n_m$, $C(n)$ is constant if $F(n)$ is the Pareto distribution above n_m.*

One can now put these lemmas together to identify sufficient conditions for marginal tax rates to be increasing with income for incomes above the modal level. Where all three of $A(n)$, $B(n)$, and $C(n)$ are nondecreasing and at least one is increasing, then marginal tax rates are increasing.

PROPOSITION 1: *Marginal tax rates are increasing above the modal skill if, above this skill, the elasticity of labor supply is constant and the distribution of skills is Pareto.*

With the conditions in Proposition 1, $A(n)$ and $C(n)$ are constants, so that $T'/(1-T')$ varies with n as $B(n)$ varies with n. With $B(n)$ increasing, so too is T'. The result carries over if the elasticity of labor supply falls with skill at the equilibrium labor supplies. Similarly, it is sufficient to have a distribution of skills such that $[1-F(n)]/[nf(n)]$ is increasing. Moreover, the result of rising tax rates will hold for part of the skill distribution (above the mode) if the conditions are met for that part; one does not need conditions on the entire distribution.

IV. Asymptotic Marginal Tax Rates

With a known finite top to the distribution of skills, the optimal marginal tax rate is zero at the top of the income distribution. As noted in the review of the literature and is clear from the argument behind Proposition 1, this need not imply that rates approach zero until very close to the top. Thus it is natural to consider the case of an unbounded distribution of skills and to consider the behavior of the optimal marginal tax rate as skills rise without limit.

In addition to assumptions on the distribution of skills and the elasticity of labor supply, the shape of the social welfare of individual utility, $G(u)$, needs to be examined. One possibility is that the marginal welfare weight of consumption of those at the top tends to zero as skill rises without limit[6]. For example, this is the case in the example in Mirrlees (1971) where $G = -\exp(-bu)/b$ $(b > 0)$ and $u = a\log(x) + \log(1-y)$. Similarly, it is the case in Martin Feldstein's (1985) study of social security, where $G = u$ and $u = \log(x)$. If G' goes to zero as n rises without limit, then $B(n)$ goes to 1. Alternatively, one might assume that G' has a positive lower bound which is approached as n rises without limit. For example, Atkinson (1990) considers the case of a "charitable Conservative" position, where the marginal welfare weight of consumption takes on two values—a high one for "poor" people and a low one for "nonpoor" people. I denote by g the ratio of the lower bound on G' to the Lagrangian on the government budget constraint, which is equal to the average of G' in the entire population. Thus, $B(n)$ converges to $1-g$ as skill rises without limit.

Assuming a constant elasticity of labor supply, e, and a Pareto distribution for skills above the mode with coefficient a, so that $C(n)$ equals $1/a$, (10) becomes

$$(14) \quad T'/(1-T') = (e^{-1}+1)B(n)/a.$$

Solving for T' and taking the limit as n rises, one has:

PROPOSITION 2: *Assuming a Pareto distribution of skills above the modal skill and a constant elasticity of labor supply, as skill rises without limit the optimal tax rate converges to*

$$(15) \quad T' = (e^{-1}+1)(1-g) \div [a + (e^{-1}+1)(1-g)].$$

[6] In this case, the tax rate tends to the revenue-maximizing rate, since, in the limit, the only effect of taxes on welfare is through the budget constraint.

TABLE 1—ASYMPTOTIC MARGINAL TAX RATES

	$g = 0$			$g = 0.25$			$g = 0.5$		
$a =$	0.5	1.5	5.0	0.5	1.5	5.0	0.5	1.5	5.0
e									
0.2	92	80	55	90	75	47	86	67	38
0.5	86	67	38	82	60	31	75	50	23

Notes: Asymptotic marginal tax rates, in percent, with a constant elasticity of labor supply, e, a Pareto distribution of skills with parameter a, and a ratio of social marginal utility with infinite income to average social marginal utility of g.

To examine the implications for the taxation of very high earners, values need to be selected for a, e, and g. Identifying skill with the wage yields an elasticity based on adjusting hours of labor supply; identifying skill with an underlying ability suggests a larger elasticity since education is also variable. With a zero income effect, compensated and ordinary labor supply elasticities are the same. Presumably it is the compensated elasticity that one would want to use for illustrative purposes. Recognizing that I am seeking an elasticity for high earners, looking at the elasticity for prime-age males provides an approximation. Based on the survey by John Pencavel (1986), calculations are done for a range of elasticities from 0.2 to 0.5. A range of g from 0 to 0.5 seems very wide. For the coefficient of the Pareto distribution, using tax data Daniel R. Feenberg and James M. Poterba (1993) find a varying between 0.5 and 1.5 over the years 1951–1990 for the incomes of the top 0.5 percent of the population. The calculations reported below suggest the possibility of a considerably higher value for the distribution of skills, perhaps as large as 5. Values of the asymptotic marginal tax rate [from (15)] are shown in Table 1. Thus I conclude that there is a case for high marginal tax rates in the quasi-linear Mirrlees model with plausible empirical parameters.

V. Decreasing Marginal Tax Rates

In considering decreasing marginal tax rates, I consider only the levels of skills above n_C, the level at which G' equals p. At skill levels above n_C, it would be desirable to transfer resources away from this skill level (if it could be done costlessly). It is now convenient to work with equation (9), which I rewrite

$$(9) \quad T'/(1 - T') = [e^{-1} + 1]\left[\int_n^{n_1} (p - G')\, dF\right] \div [pnf(n)].$$

As noted in Lemma B, at skill levels above n_C, the integral in (9) is decreasing with skill. Below the mode, the density is rising and so $1/[nf(n)]$ is falling with skill. Thus with a constant or rising elasticity of labor supply, the marginal tax rate is declining with skill. This argument also goes through above the mode where $nf(n)$ is rising with skill. This is summarized in:

PROPOSITION 3: *Above the critical skill level, n_C, marginal tax rates are decreasing where the elasticity of labor supply is constant and the distribution of skills has $nf(n)$ rising with skill.*

While one would expect $nf(n)$ to be increasing in n just above the modal skill, empirically, this seems unlikely at high skills, as is indicated in the data discussed below.

One can also use (9) to compare tax rates at two income levels above n_C, on either side of the modal skill and such that the density is equal at the two points. With G' less than p at the lower of the two skill levels being compared, the marginal tax rate would be higher at the lower income level with a constant or rising elasticity of labor supply. This is summarized in:

PROPOSITION 4: *Above the critical skill level, n_C, marginal tax rates are higher at the lower of two skill levels that have the same density and the same the elasticity of labor supply at the chosen labor supplies.*

Combining results, one can see the pattern of tax rates when the density of skills is single-peaked (and such that the workers with the modal skill work and have G' less than p in equilibrium). With a constant elasticity of labor supply and the Pareto distribution of skills where the density is falling, the pattern of marginal tax rates is U-shaped above n_C, with the minimum of marginal rates occurring at the modal skill. Moreover, marginal rates are higher at the lower income levels. Plausibly, the density of skills does not have a kink at the mode, but changes smoothly from rising to declining as a Pareto density. Then, with the conditions in Proposition 3 the minimum of the tax rate (over the range above n_C) occurs above the modal skill. It is worth reiterating that the range with declining marginal rates need not begin at zero earned income.

VI. Another Example

The assumption of a constant elasticity of labor supply relates the optimal tax to a familiar concept in the analysis of deadweight burdens. By moving the term "n" from $C(n)$ to $A(n)$, one finds another example with similar conclusions. Consider the logarithmic case, $v(1-y) = \log(1-y)$. In this case, the elasticity of labor supply is equal to $(1-y)/y$. Thus one has:

LEMMA A′: *If $v(1-y) = \log(1-y)$, then $A(n) = n(1-T')$.*

If, above the modal skill level, the distribution is the exponential distribution, then $nC(n)$ is a constant.

LEMMA C′: *For $n < n_m$, $nC(n)$ is decreasing in n. For $n > n_m$, $nC(n)$ is constant if $F(n)$ is the exponential distribution above n_m.*

I can now put together Lemmas A′, B, and C′.

PROPOSITION 1′: *Marginal tax rates are increasing above the modal skill if, above this skill, the utility-of-leisure is logarithmic and the distribution of skills is exponential.*

For Proposition 1′, it is noted that $T'/(1-T')^2$ varies with n as $B(n)$ does. With $B(n)$ increasing, so too is T'. As above, from the arguments that led to Proposition 1′, one can see that the result carries over if, at the equilibrium labor supplies, the elasticity of labor supply falls with skill more than in the stated condition. Similarly, it is sufficient to have a distribution of skills such that $[1-F(n)]/f(n)$ rises. Moreover, the result of rising tax rates will hold for part of the skill distribution (above the mode) if the conditions are met for that part; one does not need conditions on the entire distribution.

For the case just analyzed, the asymptotic marginal rate is calculated. With a logarithmic utility-of-leisure function and an exponential distribution of skills (above the mode) with coefficient b, (10) becomes:

$$(16)\quad T'/(1-T')^2 = B(n)(1-F)/f = B(n)/b.$$

Solving for T' and taking the limit, one has:

PROPOSITION 2′: *Assuming an exponential distribution of skills above the modal skill with parameter b and logarithmic utility-of-leisure, as skill rises without limit the optimal tax rate converges to*

$$(17)\quad T' = 1 - [(b'^2 + 4b')^{1/2} - b']/2,$$

where $b' = b/(1-g)$.

For g equal to 0 and $(1-F)/f$, that is, $1/b$, of 5, 10, and 15 (see Figures 1 and 2), the optimal marginal tax rate tends to 64, 73, and 77 percent. For g equal to 0.5 and the same values of $1/b$, the optimal marginal tax rate tends to 54, 64, and 70 percent.

Similarly, one can examine conditions for declining marginal tax rates with the log-

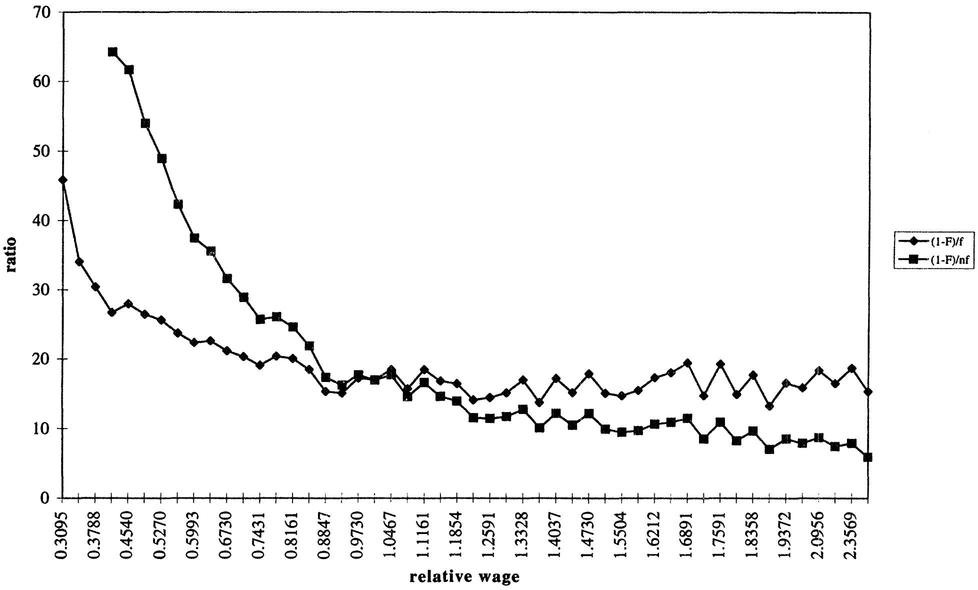

FIGURE 1. RATIOS $[1 - F(n)]/f(n)$ AND $[1 - F(n)]/[nf(n)]$ CALCULATED FROM RELATIVE WAGES

arithmic utility-of-leisure. In this case, (9) becomes:

$$(18) \quad T'/(1 - T')^2 = \left[\int_n^{n_1} (p - G')\, dF\right] \Big/ [pf].$$

From Lemma B, it can be concluded that the tax rate is declining above n_C and below the mode.

PROPOSITION 3′: *Between the critical skill level, n_C, and the mode, n_m, marginal tax rates are decreasing if the utility-of-leisure is logarithmic.*

Similarly, from (18) one can conclude:

PROPOSITION 4′: *Above the critical skill level, n_c, marginal tax rates are higher at the lower of two skill levels that have the same density if the utility-of-leisure is logarithmic.*

VII. Data on the Distribution of Skills

While a careful attempt to fit this model to available data is beyond the scope of this paper, it does seem interesting to examine the distribution of wages. For this purpose, calculations have been done using the March 1992 CPS. This survey asked individuals for annual earnings in 1991, as well as weeks worked and typical hours per week. From these numbers one can calculate an implied average wage.[7] Using these wages, calculations were made of the mean wage per cell; the number of observations per cell, adjusted by interval width in order to be proportional to the density; and the number of observations with higher wages. Approximately 17 percent of the sample report wages below $1 or no work and are omitted. In order to have reasonable cell sizes, the wage intervals are first $0.50, but are expanded above a wage of $26. As expected, a smoothing of the data would show a single-peaked distribution, as assumed in the analysis above. In Figure 1 is shown the ratios $(1 - F)/f$ and $(1 - F)/(nf)$, where n is measured as the wage relative to the mean wage. Because the series are very noisy, the graph is a centered three-cell moving average.

[7] No attempt was made to consider both earners in a two-earner family or wages of single females.

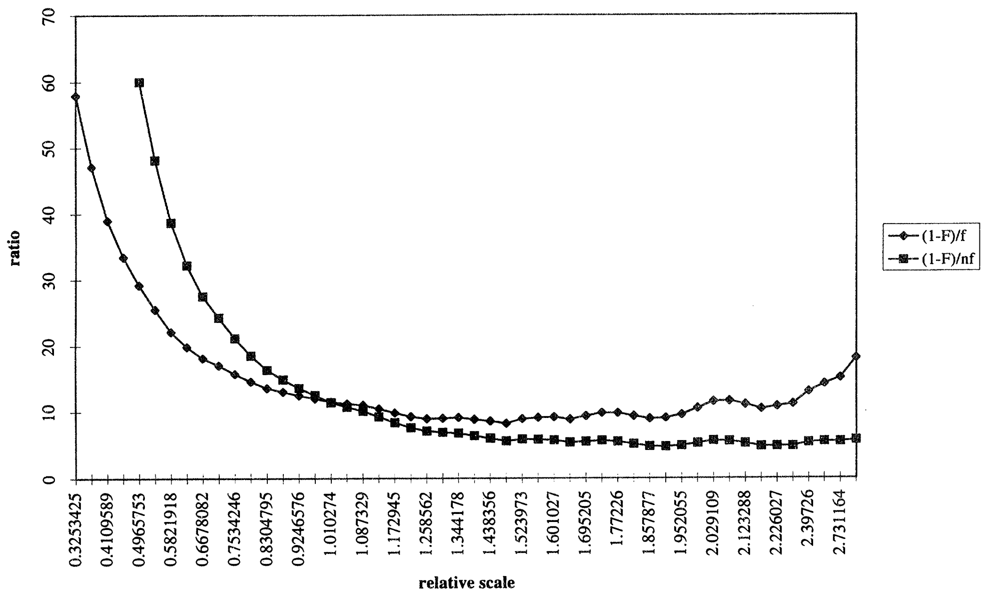

FIGURE 2. RATIOS $[1 - F(n)]/f(n)$ AND $[1 - F(n)]/[nf(n)]$ CALCULATED FROM RELATIVE SKILLS

For readability, the lowest wages are dropped from the graph since the ratios are very large. The figure shows sharply falling values of $(1 - F)/f$ through the range where the density is first rising and then roughly flat, that is, up to a wage of roughly \$13, a little below the mean of \$13.70. Beyond this point $(1 - F)/f$ is roughly constant at a value around 15. This implies a downward trend in $(1 - F)/(nf)$. A constant value of $(1 - F)/f$ is consistent with an exponential distribution over this range of values.

With a longer time horizon than one year, one would consider education to be an endogenous variable somewhat responsive to tax incentives. One might also be interested in the distribution of skills within a cohort. Thus a further calculation was done by regressing the log of the wage on education, age, and age squared, and plotting the exponentiated residuals. In Figure 2 are the same curves for this distribution as shown in Figure 1 for the distribution of wages (except that a moving average was not used). This distribution shows a fatter tail than the distribution of wages, with $(1 - F)/f$ rising and $(1 - F)/(nf)$ roughly constant for the top 15 percent of the skill distribution. A constant value of $(1 - F)/(nf)$ is consistent with a Pareto distribution over this range of values.

VIII. Concluding Remarks

The absence of income effects allows an intuitive grasp of the factors that determine the optimal tax structure. Increasing the marginal tax rate affecting some skill level involves an increase in the deadweight burden for people at this skill level. Thus, the optimal marginal tax rate at some income level depends on the elasticity of labor supply at that income level, since this is important for marginal distortions. Increasing the marginal tax rate also transfers income from all individuals with higher skills to the government, without changing the distortions of their labor supplies. The weights on these two elements depend on the ratio of individuals with skills above this level to individuals with skills at this level and on the level of skill which links the tax on hours to the tax on income. This intuition is displayed in the equations above, where the first-order condition for the optimal income tax is written with a product of these three terms.

The rewriting of the first-order condition also highlights the critical role of the assumed family of distributions for the upper tail of skills, as opposed to just the value of the parameters. With the Pareto distribution, $(1 - F)/(nf)$ is a constant and the change in the marginal tax rate reflects the rate of decline in social marginal utility of income as well as the change in labor supply elasticity with skill. With the exponential distribution, $(1 - F)/(nf)$ declines at the rate $1/n$. Thus either the elasticity of labor supply or the social marginal utility of income needs to be falling sufficiently rapidly to have constant or rising marginal tax rates. In the simulations in Mirrlees (1971), it was assumed that the distribution of skills was lognormal, so that $(1 - F)/(nf)$ declines at the rate $1/\log(n)$. Presumably the relatively constant marginal tax rate in the those simulations would have had a different shape with a different assumed family of distributions. Exploration of the shape of this distribution is clearly important for the normative case for different degrees of income tax progressivity.

There is not a simple route between the Mirrlees model and policy implications for annual income taxes levied repeatedly on families and covering both capital and labor incomes. The assumption of a zero income elasticity of labor supply and the limited information on both the shape of the skill distribution and the pattern of elasticities of labor supply by skill level would limit inferences even if there were a simple route. Nevertheless there are some lessons from the analysis. The sharp fall in $(1 - F)/f$ as skills approach the mode of the skill distribution from below seems highly relevant, especially if one wants to redistribute from people near the mode, rather than to them. This finding on the shape of an optimal (negative) income tax seems relevant in thinking about the phaseout of the earned income tax credit, and, possibly, welfare reform. That is, labor supply depends on the net return to earnings, which, in turn, depends on both the income tax and the phaseout of income-tested benefits. The presence of high marginal tax rates in the region of phaseout of benefits is not necessarily a basis for criticism of the programs—the optimal program may well have such a shape because of the advantage of higher marginal rates over a shorter range of skills where the skill density is large and rising. In other words, a sizable implicit marginal tax rate where benefits are being phased out is consistent with the U-shaped pattern of marginal rates and may well be optimal.

Second, this model confirms the implication of Mirrlees' calculations that the optimality of a zero tax rate at the highest income level is not a finding that sheds much light on optimal taxes, especially in the absence of knowledge of exactly where the top is. That is, if one replaced an unbounded distribution of skills by a bounded one with the same distribution up to some level and a concentration of skills at the highest levels, the result of rising marginal tax rates continues to hold until the concentration at the top is reached. There is no need for tax rates to decline slowly toward zero as one approaches the absolute top of the skill distribution.

Third, the sensitivity of the pattern of marginal rates to the measure of skill seems relevant, although different formulations of "skill" will be associated with different estimates of the elasticities of labor supply as well as different estimates of the shape of the distribution of skills. This analysis emphasizes the importance of the shape of the distribution of skills for optimal tax rates.

Appendix: Heuristic Derivation of the Optimal Tax First-Order Condition

Using just the first-order condition for labor supply for the quasi-linear utility function as a constraint, the optimal tax problem is

$$\text{(A1)} \quad \operatorname{Max} \int_{n_0}^{n_1} G\{u[x(n), y(n)]\} f(n)\,dn,$$

$$\text{subject to:} \int_{n_0}^{n_1} x(n) f(n)\,dn$$

$$\leqq \int_{n_0}^{n_1} n y(n) f(n)\,dn - E;$$

$$v'[1 - y(n)] = n(1 - T'[ny(n)]).$$

With $x(n) = ny(n) - T[ny(n)]$, and using the first-order condition for labor supply, the change in consumption with skill satisfies:

$$\text{(A2)} \quad x'(n) = y(n)(1 - T') + n(1 - T')y'(n) = [y(n) + ny'(n)]v'/n.$$

With the quasi-linear utility function, one can calculate the derivative of u with respect to n:

$$\text{(A3)} \quad u'(n) = x'(n) - v'y'(n) = y(n)v'/n.$$

Treating $u(n)$ as a state variable and $y(n)$ as a control variable, the optimal tax problem can be rewritten as

$$\text{(A4)} \quad \text{Max} \int_{n_0}^{n_1} G[u(n)]f(n)\,dn,$$

subject to:

$$\int_{n_0}^{n_1} \{u(n) - v[1 - y(n)]\}f(n)\,dn \leqq \int_{n_0}^{n_1} ny(n)f(n)\,dn - E;$$

$$u'(n) = y(n)v'[1 - y(n)]/n.$$

Forming a Hamiltonian for this expression,

$$\text{(A5)} \quad H = \{G[u(n)] - p[u(n) - v[1 - y(n)] - ny(n)]\} \times f(n) + h(n)y(n) \times v'[1 - y(n)]/n,$$

where p and $h(n)$ are multipliers. The derivative of h is equal to minus the partial derivative of the Hamiltonian with respect to u:

$$\text{(A6)} \quad h'(n) = -\{G'[u(n)] - p\}f(n).$$

Maximizing the Hamiltonian with respect to $y(n)$,

$$\text{(A7)} \quad -p\{n - v'[1 - y(n)]\}f(n) = h(n)\{v'[1 - y(n)] - y(n)v''[1 - y(n)]\}/n.$$

Recognizing that $h(n_1)$ is equal to zero, (A6) can be integrated from n to n_1 to have an expression for $h(n)$. Substituting in (A7), one then has the first-order condition in the text, (8).

REFERENCES

Atkinson, Anthony B. "Public Economics and the Economic Public." *European Economic Review*, May 1990, *34*(2–3), pp. 225–48.

Ebert, Udo. "A Reexamination of the Optimal Nonlinear Income Tax." *Journal of Public Economics*, October 1992, *49*(1), pp. 47–73.

Feenberg, Daniel R. and Poterba, James M. "Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns," in James Poterba, ed., *Tax policy and the economy*. Cambridge, MA: MIT Press, 1993, pp. 145–77.

Feldstein, Martin. "The Optimal Level of Social Security Benefits." *Quarterly Journal of Economics*, May 1985, *100*(2), pp. 303–20.

***Financial Times*.** "The Markets: Still Waiting for Answers on Tax—Economics Notebook." September 11, 1995, p. 24.

Fudenberg, Drew and Tirole, Jean. *Game theory*. Cambridge, MA: MIT Press, 1991.

Kanbur, Ravi and Tuomala, Matti. "Inherent Inequality and the Optimal Graduation of Marginal Tax Rates." *Scandinavian Journal of Economics*, June 1994, *96*(2), pp. 275–82.

Lollivier, Stefan and Rochet, Jean-Charles. "Bunching and Second-Order Conditions: A Note on Optimal Tax Theory." *Journal of Economic Theory*, December 1983, *31*(2), pp. 392–400.

Mirrlees, James A. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, April 1971, *38*(114), pp. 175–208.

Myles, Gareth D. *Public economics*. Cambridge: Cambridge University Press, 1995.

Pencavel, John. "Labor Supply of Men: A Survey," in Orley Ashenfelter and Richard Layard, eds., *Handbook of labor economics*. Amsterdam: North-Holland, 1986, pp. 3–102.

Sadka, Efraim. "On Income Distribution, Incentive Effects and Optimal Income Taxation." *Review of Economic Studies*, June 1976, *43*(2), pp. 261–67.

Seade, Jesus K. "On the Shape of Optimal Tax Schedules." *Journal of Public Economics*, April 1977, *7*(2), pp. 203–35.

———. "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies*, October 1982, *49*(4), pp. 637–43.

Tuomala, Matti. "On the Optimal Income Taxation: Some Further Numerical Results." *Journal of Public Economics*, April 1984, *23*(3), pp. 351–66.

———. *Optimal income tax and redistribution*. Oxford: Clarendon Press, 1990.

Weymark, John A. "Comparative Static Properties of Optimal Nonlinear Income Taxes." *Econometrica*, September 1987, *55*(5), 1165–85.