Queen's Economics Department Working Paper No. 984

OPTIMAL INCOME TAXATION WITH QUASI-LINEAR PREFERENCES REVISITED

Robin Boadway
Queen's University

Katherine Cuff
Queen's University

Maurice Marchand
Universite Catholique de Louvain

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

4-1999

OPTIMAL INCOME TAXATION WITH QUASI-LINEAR PREFERENCES REVISITED

by

Robin Boadway†, Katherine Cuff†† and Maurice Marchand†††

April 1999

ABSTRACT

We study the properties of the optimal non-linear income tax when preferences are quasi-linear in leisure. When the objective function is utilitarian or maxi-min, closed-form solutions for the marginal tax rates and the marginal utility of consumption are obtained which depend only on the skill distribution. Bunching occurs over intervals in the distribution when the second-order incentive condition is binding, and this depends solely on the skill distribution. The patterns of consumption and tax rates in the non-bunched range are however independent of whether the second-order incentive constraints are binding. Bunching at the bottom can also occur if a non-negative constraint on incomes is binding for some households. The pattern of incomes can be determined as a function of the skill distribution and the revenue requirement. In the absence of bunching, an increase in required revenue causes an equal increase in all household incomes, which leads to a simple characterization of the shadow price of government funds. Examples for specific skill distributions are given, and it is shown that the pattern of marginal tax rates depend critically on whether the skill distribution is truncated at the upper end.

JEL Classification: H21, H23

Keywords: optimal income tax, quasi-linear preferences

Acknowledgements: Financial support of the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

† Queen's University, Canada; boadwayr@qed.econ.queensu.ca

†† Queen's University, Canada; cuffk@qed.econ.queensu.ca

††† Université Catholique de Louvain, Belgium; mmarchand@core.ucl.ac.be

OPTIMAL INCOME TAXATION WITH QUASI-LINEAR PREFERENCES REVISITED

by

Robin Boadway, Katherine Cuff and Maurice Marchand

I. INTRODUCTION

The properties of resource allocations under optimal non-linear income taxes are notoriously difficult to characterize explicitly. As a result, authors commonly rely on numerical calculations to give some idea about the structure of the tax schedule.[1] The difficulty with this approach is that these calculations leave somewhat obscure the features of the problem which give rise to the tax structures computed. For example, the relative roles played by the form of the policy-maker's objective function, the substitutability of leisure for consumption in household preferences, the distribution of wages, and the government revenue requirement are difficult to infer from numerical calculations. As well, the intuition behind some of the results is not altogether clear.

It turns out that there is a special case in which a closed-form solution can be obtained for the optimal structure of tax rates, and that is when consumer preferences are quasi-linear in leisure and the objective function is utilitarian. The possibility of obtaining an explicit solution in this case was first pointed out by Lollivier and Rochet (1983), and some of the properties for this case were derived by Weymark (1986a, 1986b, 1987) for the discrete case (finite number of skill levels) and by Ebert (1992) for the continuous case. It turns out that the maxi-min case leads to an even simpler solution. The purpose of this paper is to provide a full characterization of the optimal non-linear income tax for the quasi-linear preference case under utilitarian and maxi-min objectives. In particular, we can determine the structure of marginal and average tax rates as well as the patterns of consumption and income. We are also able to provide an intuitive explanation of the shadow price of public funds and its relevance for the structure of taxes, as well as for the effect of the government's revenue requirement on the shadow price and therefore on the progressivity of the tax.

[1] For a summary of the literature, see Tuomala (1990).

Our approach allows us to show explicitly the conditions under which bunching can occur at the bottom or elsewhere in the wage distribution, and to explain the intuition behind such bunching. The explanation builds on Ebert's insight in the continuous case that bunching can be caused by a failure of the second-order conditions characterizing the households' self-selection constraints. But we also demonstrate that bunching can occur for another reason, and that is when the non-negativity constraint that must be imposed on household income is binding. Although the need for such a constraint has been recognized in the above literature (Weymark, 1986a), its consequences have not been studied. We show that if such a constraint is active, it drastically affects the optimal tax structure.

The assumption that preferences are quasi-linear in leisure is not unreasonable. But, it is also rather restrictive. The benefit of using this case is that it does allow us to gain some new insight into the optimal tax problem. Its ease of solution is also of obvious pedagogical value. Moreover, it provides an interesting contrast with the recent results of Diamond (1998), who characterized the pattern of optimal marginal income taxes when preferences are quasi-linear in consumption rather than leisure. Some of our results on the structure of marginal tax rates parallel those he obtained. We show however that his main finding of a U–shaped pattern of marginal tax rates depends critically on the skill distribution being unbounded at the top. If the maximum skill level is finite, his results no longer apply.

We begin in the next section with an outline of the basic features of the model. Section III considers the properties of the optimal tax function for the utilitarian case, paying particular attention to the validity of the first-order approach and the consequences for the tax structure of the second-order incentive constraint being binding over various ranges of the skill distribution. In section IV, the tax structure under the maxi-min objective function is studied. Qualitative results for some specific skill distributions are considered in the next section. Sections VI and VII then take up the determination of the pattern of before-tax incomes in the optimum, and analyze the consequences for the tax structure of the non-negative income constraint being binding at the bottom. The final section provides some concluding comments.

II. THE BASIC MODEL

Our setting is similar to the standard approach of Mirrlees (1971), except for the assumption of quasi-linear preferences. Household utility functions are given by $u(x) - \ell$, where x is consumption, ℓ is labour, and $u(x)$ is increasing and strictly concave. Households differ in their abilities (or skills), which correspond with their wage rates. Households are distributed by wage rates w according to the distribution function $F(w)$, for $w \in \mathcal{W} = [\underline{w}, \overline{w}]$, where $\underline{w} > 0$ and $\overline{w}$ can be either finite or infinite. It turns out that the value of the upper bound on skills is of some importance in determining the shape of the marginal tax structure. In particular, truncating the distribution at a finite upper bound can cause the marginal tax rates at the upper end to decline more rapidly than in the untruncated case, and this can have a qualitatively important effect on the tax structure. The corresponding density function, $f(w)$, is assumed for simplicity to be differentiable and strictly positive for all wages in $\mathcal{W}$. Population size is normalized to unity, so $F(\overline{w}) = 1$.

As is standard in the optimal income tax literature, the government can observe incomes but not wage rates. Households obtain all their income from wages, so before-tax income is given by $y \equiv w\ell$. Therefore, we can rewrite household utility as $u(x) - y/w$. The difference between y and x is simply the tax paid to the government, $T = y - x$, which can be either positive or negative (i.e. a transfer to the household). Given some tax function, the government can effectively observe both y and x. In (x, y)–space with x and y on the horizontal and vertical axes respectively, the indifference curves of households earning a given wage w are horizontally parallel, and at any given point in this space the higher is the wage rate of a household the flatter is its indifference curve. The first of these properties is equivalent to the absence of income effects on the household's choice of consumption, and the second implies that the monotonicity or single-crossing property is verified.

The government chooses the tax imposed on each household or, equivalently, the consumption-income bundle intended for each household $\{(x(w), y(w)), w \in \mathcal{W}\}$, subject to three sorts of constraints. The first is the government budget constraint, which takes the form:

$$\int_{\underline{w}}^{\overline{w}} [y(w) - x(w)] f(w) dw \geq R \tag{1}$$

where $y(w)$ and $x(w)$ denote the income and consumption of a type-w person, and R is an exogenous revenue requirement.

The second is the set of incentive-compatibility, or self-selection, constraints. These require that a household of type w choose the consumption-income bundles intended for it, that is, $u(x(w)) - y(w)/w \geq u(x(w')) - y(w')/w$ for any w and $w' \in \mathcal{W}$. Following Lollivier and Rochet (1983), it is analytically convenient to transform a household's utility function by multiplying it by its wage rate and defining $V(w)$ as follows:

$$V(w) \equiv wu(x(w)) - y(w)$$

This transformation allows us to substitute $y(w)$ out of the problem readily, thereby exploiting the quasi-linearity property to simplify the incentive-compatibility conditions.[2]

The latter can indeed be written as $wu(x(w)) - y(w) \geq wu(x(w')) - y(w')$, or:

$$V(w) \geq V(w') + (w - w')u(x(w')) \tag{IC}$$

Since w' can be either lower or higher than w, this requires that for w' approaching w, we have in the limit $\dot{V}(w) = u(x(w))$, for all $w \in \mathcal{W}$, where we use a 'dot' to refer to derivatives with respect to w. This is the so-called *first-order incentive-compatibility* (FOIC) condition. And, given that $u(x(w))$ is the slope of $V(w)$ at w, the inequality in condition (IC) requires that $V(w)$ be strictly convex, or $d\dot{V}(w)/dw = u'(x)\dot{x}(w) > 0$. Since $u'(x) > 0$, $x(w)$ must then be increasing in w, which is called the *second-order incentive-compatibility* (SOIC) condition. It should also be noted for future purposes that the first-order conditions for the household maximization problem imply that $\dot{y}(w)$ takes on the same sign as $\dot{x}(w)$.[3]

The third constraint requires that labour supply and therefore before-tax income be non-negative ($y(w) \geq 0$, $w \in \mathcal{W}$).[4] For now, *we ignore the non-negativity constraint on*

[2] The usual approach in the literature is to substitute x out of the problem and treat y as a control variable (Mirrlees, 1971; Tuomala, 1990; Diamond, 1998). With quasi-linear in leisure preferences, it is more convenient to do the reverse.

[3] In response to the pair of functions imposed by government policy $\{x(w), y(w)\}$, $w \in \mathcal{W}$, households of skill w maximize $wu(x(\hat{w})) - y(\hat{w})$ with respect to $\hat{w}$. This yields the first-order condition $wu'\dot{x}(\hat{w}) - \dot{y}(\hat{w}) = 0$, which under incentive compatibility will be satisfied at $\hat{w} = w$. Therefore, $\dot{x}(w)$ and $\dot{y}(w)$ take the same sign.

[4] It seems reasonable also to assume an upper bound on labour supply and therefore income. The implications of a binding maximum income constraint are considered briefly below, but until then the constraint is assumed to be non-binding.

income and assume it to be satisfied. In a later section, we consider the circumstances in which it might be violated.

III. UTILITARIAN OBJECTIVE FUNCTION

We can now formulate the government's problem. In this section, we assume that the objective function is utilitarian: the government maximizes the sum of individuals' utilities. In a later section, a maxi-min social welfare function is considered. Using the definition of $V(w)$ to eliminate $y(w)$ from the government budget constraint, the problem is:

$$\max \quad \int_{\underline{w}}^{\overline{w}} \frac{V(w)}{w} f(w) dw \tag{P}$$

subject to:

$$\int_{\underline{w}}^{\overline{w}} [wu(x(w)) - V(w) - x(w)] f(w) dw \geq R$$

$$\dot{V}(w) = u(x(w))$$

$$\dot{x}(w) \geq 0$$

If we let $z(w) \equiv \dot{x}(w)$, the non-negativity constraint on $\dot{x}(w)$ (the SOIC condition) can be written as $z(w) \geq 0$. We can treat $z(w)$ as the control variable, and $V(w)$ and $x(w)$ as state variables in this dynamic optimization problem. The Hamiltonian function is:

$$H(w) = V(w)\frac{f(w)}{w} + \lambda[wu(x(w)) - V(w) - x(w)]f(w) + \pi(w)u(x(w)) + \mu(w)z(w) + \kappa(w)z(w)$$

where λ can be interpreted as the shadow price of government funds, $\pi(w)$ is the co-state variable associated with $\dot{V}(w) = u(x(w))$, $\mu(w)$ is the co-state variable associated with $\dot{x}(w) = z(w)$, and $\kappa(w)$ is the shadow value of the non-negativity constraint on $z(w)$. The optimal solution must satisfy the following necessary conditions:

$$\frac{\partial H}{\partial z(w)} = 0 = \mu(w) + \kappa(w) \tag{2}$$

$$\frac{\partial H}{\partial x(w)} = -\dot{\mu}(w) = \lambda[wu'(x(w)) - 1]f(w) + \pi(w)u'(x(w)) \tag{3}$$

$$\frac{\partial H}{\partial V(w)} = -\dot{\pi}(w) = \frac{f(w)}{w} - \lambda f(w) \tag{4}$$

$$\pi(\underline{w}) = \pi(\overline{w}) = \mu(\underline{w}) = \mu(\overline{w}) = 0 \tag{5}$$

$$\kappa(w) > 0 \to \dot{x}(w) = 0; \quad \dot{x}(w) > 0 \to \kappa(w) = 0 \tag{6}$$

The first three conditions are the standard necessary conditions on the control and state variables. Relations (5) are the transversality conditions. Conditions (6) are required to ensure that the SOIC condition is satisfied: if they are not binding for a given w, then $\kappa(w) = 0$. The rest of this section is concerned with interpreting these necessary conditions.[5]

It is useful to begin by solving for λ and $\pi(w)$, both of which turn out to depend solely on the skill distribution and can be given intuitive interpretations. Using the fact that $\pi(\underline{w}) = 0$ by the transversality conditions (5), we can solve (4) for $\pi(w)$:

$$\pi(w) = \int_{\underline{w}}^{w} \left(\lambda - \frac{1}{m}\right) f(m)dm = \lambda F(w) - G(w) \tag{7}$$

where, following Ebert (1992), we have defined $G(w) \equiv \int_{\underline{w}}^{w}(f(m)/m)dm$. The function $G(w)$ is simply the expected value of $1/m$ over the interval $\underline{w} \leq m \leq w$ multiplied by the proportion of individuals on the interval, or $G(w) = E(1/m|m \leq w)F(w)$. Using the transversality condition $\pi(\overline{w}) = 0$ we obtain from (7):

$$\lambda = G(\overline{w}) = E\left(\frac{1}{w}\right) \tag{8}$$

Thus, *the shadow price of government revenue depends only upon the distribution of the population.* The intuition for this is as follows. With quasi-linear preferences, an increase in the revenue required by the government (dR) will come entirely from increased labour income: the consumption of any household will not change because it is not affected by income. It turns out, as shown below, that if more tax revenue is needed, all households will be required to increase their labour incomes by the same amount. That will impose a utility cost of $1/w$ on a type-w household. The aggregate effect of this utility cost on the utilitarian objective is simply its expected value over the entire distribution $E(1/w)$, or $G(\overline{w})$.

[5] We assume throughout that the second-order conditions for the government's problem (P) are satisfied. These, of course, should not be confused with the SOIC condition, which refers to the household's incentive problem.

Turning back to $\pi(w)$, we can derive an alternative expression for $\pi(w)$ from (4) using the transversality condition $\pi(\overline{w}) = 0$:

$$\pi(w) = -\int_{w}^{\overline{w}} \left(\lambda - \frac{1}{m}\right) f(m)dm$$

The co-state variable $\pi(w)$ is the shadow price of increasing $V(w) = wu(x(w)) - y(w)$ by one unit. To do this, the government reduces the income that type-w households must earn by one unit. To ensure the incentive-compatibility constraints continue to be satisfied, the government must reduce by the same amount the income required from all households with skill greater than w. This income reduction lowers tax revenue and at the same time increases the utility of all households with skill above w. The total welfare cost from the lost revenue and the total welfare benefit from the increases in utility are given by the first and second term in the above expression, respectively. Substituting (8) into (7), we obtain:

$$\pi(w) = G(\overline{w})F(w) - G(w) \tag{9}$$

Except at $w = \overline{w}$, the shadow price of increasing $V(w)$ is negative (that is, an increase will result in a net social loss)[6] and, as already mentioned, depends only on the distribution of skills.

Equation (3) can now be used to determine the pattern of consumption and marginal tax rates. We begin by considering the case where the SOIC condition is satisfied over the entire population distribution, so the 'first-order approach' is adequate. Then we consider the possible violations of the SOIC condition and its consequences for the optimal consumption path.

First-Order Approach

So let us first assume that $\dot{x}(w) > 0$ for all w, so $\kappa(w) = 0$ by (6), and $\mu(w) = 0$ by (2). Then, using (3) the path of consumption is determined by:

$$\lambda[wu'(x(w)) - 1]f(w) + \pi(w)u'(x(w)) = 0 \tag{10}$$

[6] The sign of (9) follows from the fact that the expected value of the reciprocal of the wage rate for $w < \overline{w}$, normalized by the size of the population at w, $G(w)/F(w))$, will be less than the expected value of the reciprocal of the wage rate over the entire distribution, $G(\overline{w})$.

As both λ and $\pi(w)$ depend only on the distribution of skills, expression (10) indicates that the pattern of $u'(x(w))$ depends only on the distribution of skills, which simplifies the characterization of the optimal tax allocation considerably. Using (8) and (9), we can rewrite (10) as follows:[7]

$$u'(x(w)) = \frac{f(w)}{wf(w) + F(w) - G(w)/G(\overline{w})} \tag{11}$$

From the negativity of $\pi(w)$ in (9) and the transversality conditions (5), we obtain that $u'(x(w)) < 1/w$ except at the end points of the skill distribution. Also, when preferences are quasi-linear in leisure, the optimal path of consumption $x(w)$ depends only on the skill distribution and the functional form for $u(x)$.

Next, we can obtain an expression for the pattern of marginal tax rates. Let $T(y(w))$ be the tax function. Then, from the household optimization problem, the marginal tax rate for a type-w household is given by:[8]

$$T'(y(w)) = 1 - \frac{1}{wu'(x(w))}$$

Using (11), the marginal tax rate for a person of type w can be rewritten as:

$$T'(y(w)) = \frac{G(w)/G(\overline{w}) - F(w)}{wf(w)} \tag{12}$$

Unlike the pattern of consumption which also depends on the form of $u(x)$, *the pattern of tax rates depends only upon the distribution of skills.* The above yields the standard result (Seade, 1977) that (in the absence of bunching) the marginal tax rates at the top and bottom of the distribution are zero, $T'(y(\underline{w})) = T'(y(\overline{w})) = 0$. Moreover, it is straightforward to show that $0 \le T'(y(w)) \le 1$ for all w in the interior of $\mathcal{W}$.[9] Therefore, the marginal tax

[7] Equation (10) could alternatively be written $u'(x(w)) = 1/\widehat{w}(w)$, where $\widehat{w}(w)$ is analogous to the *adjusted wage* introduced by Weymark (1986a) and is defined here as $\widehat{w}(w) = w + [F(w) - G(w)/G(\overline{w})]/f(w)$.

[8] The household chooses y to maximize $u(y - T(y)) - y/w$. The first-order condition immediately yields the expression for $T'(y)$.

[9] To see this, recall from footnote 6 that $G(w) > F(w)G(\overline{w})$ for all $w \in (\underline{w}, \overline{w})$, so $T'(y(w)) > 0$ for all w in the interior. From footnote 3, $\text{sign}\{\dot{x}(w)\} = \text{sign}\{\dot{y}(w)\}$. Since $\dot{x}(w) > 0$ and $x(w) = y(w) - T(y(w))$, so $\dot{x}(w) = \dot{y}(w)(1 - T'(y(w))$, it follows that $1 - T'(y(w)) > 0$ and thus, $T'(y(w)) < 1$.

rate must rise on an interval starting at the lower bound $y(\underline{w})$ and decrease on an interval close to the upper bound $y(\overline{w})$. Between these two extreme income intervals it is difficult in general to infer the optimal pattern of the marginal tax rates from (12).

To get a better idea of the shape of the marginal tax rate schedule, let us multiply and divide (12) by $1 - F(w)$ to obtain:

$$T'(y(w)) = B(w) \cdot C(w) \tag{13}$$

where

$$B(w) = \frac{G(w)/G(\overline{w}) - F(w)}{1 - F(w)} \quad \text{and} \quad C(w) = \frac{1 - F(w)}{wf(w)}$$

The factors $B(w)$ and $C(w)$ are the analogues of those in Diamond (1998) (his equation (10)): $C(w)$ is identical, while $B(w)$ has the same interpretation in our setting.[10] To determine the shape of T', we can look at either the patterns of $B(w)$ and $C(w)$ or the patterns of $B(w)(1-F(w))$ and $C(w)/(1-F(w))$, whichever is more instructive for the case at hand. To investigate these different patterns, let w_m denote the modal wage rate for a single peaked distribution ($\dot{f}(w_m) = 0$) and w_c denote the wage rate such that $w_c = 1/G(\overline{w})$ (referred to as the 'critical wage rate' by Diamond, reflecting the fact that at this wage rate the marginal disutility of labour income in terms of forgone leisure just equals the shadow price of government revenue). For the unimodal distributions we consider below, we suppose (as in Diamond, 1998) that $w_m > w_c$.[11] Given this, the following properties of $B(w)$, $C(w)$, $B(w)(1 - F(w))$, and $C(w)/(1 - F(w))$ are proved in the Appendix:

Properties:

i) $B(w)$ is positive and increasing for all $w \in (\underline{w}, \overline{w}]$ and zero at $\underline{w}$;

ii) $B(w)(1 - F(w))$ is increasing in w for $w < w_c$, decreasing in w for $w > w_c$, and zero at $\underline{w}$ and $\overline{w}$;

---

[10] The factor $B(w)$ represents the value of the additional income the government would receive by increasing $T'(y(w))$ from all individuals of skill level greater than w given that their labour supply distortions are unchanged (Diamond, 1998), or in this case where preferences are linear in leisure, given that their consumption is unaffected. Diamond also has a third factor, $A(w) = 1 + e^{-1}(w)$, where $e(w)$ is the elasticity of labor supply and takes into account the marginal labour supply distortions. In our context, $A(w) = 1$.

[11] This follows whenever the mode of a distribution is greater than its mean. Since $1/w$ is a strictly convex function, the mean of $1/w$ is always greater than one over the mean of w by Jensen's inequality.

iii) $C(w)$ is decreasing in w for $w < w_m$ and depends on the specific skill distribution for $w > w_m$;

iv) $C(w)/(1-F(w))$ is decreasing in w for $w < w_m$ and depends on the specific distribution for $w > w_m$.

As these results indicate, except in special cases, $B(w)$ and $C(w)$ (and $B(w)(1-F(w))$ and $C(w)/(1-F(w))$) can have conflicting effects on T' as w rises. Only in the range $w_c < w < w_m$ is there no ambiguity: by Properties ii) and iv), T' is decreasing in this range. For $w < w_c$, the two factors $B(w)$ and $C(w)$ (as well as $B(w)(1-F(w))$ and $C(w)/(1-F(w))$) change in opposite directions, so it is not clear on this basis alone whether T' is rising or falling (except, of course, near $w = \underline{w}$). For $w > w_m$, the pattern of T' is also ambiguous and depends critically on the distributional assumptions (except, again, near $w = \overline{w}$). In the next section, we consider various skill distributions and characterize the shape of the marginal tax rate schedule under each distribution.

It is worth stressing that we are investigating how the marginal tax rate (the change in total tax liability in response to a change in income) changes as the wage rate w — and not income y — rises. This is the typical approach in the literature. To determine how the marginal tax rate varies with income, note that

$$\frac{dT'(y(w))}{dw} = T''(y(w))\dot{y}(w) \tag{14}$$

where $T''(y(w))$ is the second derivative of the total tax function $T(y)$ with respect to y. This allows us to make some inferences about the shape of the total tax function in (T, y)–space. We know from (12) that the marginal tax rate is always non-negative, $T'(y(w)) \geq 0$, so total tax liabilities rise with y. Given the first-order approach is valid, it must be $\dot{y}(w) > 0$. Therefore, the second derivative of the tax function with respect to income will take on the same sign as $dT'(y(w))/dw$.[12]

So far our analysis has been conducted as if the first-order approach is valid. It will be as long as (11) yields a solution for $x(w)$ which is everywhere increasing in w. We now consider the possibility that $x(w)$ is not everywhere increasing in w.

[12] If it is necessary to take into account the SOIC condition, then $\dot{y}(w) \geq 0$, such that $\dot{y}(w) > 0$ over any non-bunched region and $\dot{y}(w) = 0$ over any bunched interval. All individuals in the bunched region earn the same income and have the same tax liabilities.

Second-Order Approach

To check for a violation of the SOIC condition under the first-order approach, we need determine whether $x(w)$ is increasing with w. To verify this, rewrite (11) using the definition of the marginal tax rate:

$$u'(x(w)) = \frac{1}{w(1 - B(w)C(w))} \tag{15}$$

Then, $x(w)$ will be increasing with w if $u'(x(w))$ is decreasing with w. Therefore, the SOIC condition will be satisfied if $w(1 - B(w)C(w))$ is increasing in w. A sufficient, though not necessary, condition for this to be true is that $B(w)C(w)$ be decreasing, that is, that T' be decreasing. As we have seen above, there is no guarantee that this will be the case. Note that whether the SOIC condition is satisfied depends solely on the form of the skill distribution function.

That leaves open the possibility that the SOIC condition will be violated for some w under the first-order approach. If so, there must be some range over which $\kappa(w) > 0$, so $\dot{x}(w) = 0$. In this range, $\dot{y}(w) = 0$ as well, so there is bunching. This illustrates the insight in Ebert (1992) that bunching will be induced by a violation of the SOIC condition. We shall see later that bunching can occur for another reason as well — the violation of a non-negativity constraint on income. For now, we concentrate on bunching induced by a violation of the SOIC condition. In principle, bunching can occur anywhere in the distribution of skills. It is useful to consider separately the consequences for the tax structure of bunching at the bottom and bunching in the interior.[13]

Bunching at the Bottom

Bunching occurs at the bottom of the skill distribution if the derivative of the denominator of (15) is negative at the lowest skill level. This can be shown to be equivalent to having

[13] As pointed out by Ebert (1992), there cannot be bunching at the highest income $y(\overline{w})$. He shows that if the second-order conditions are satisfied, then the marginal tax rate at the highest income level is zero and the marginal tax rates at all other income levels are positive (unless there is no bunching at $y(\underline{w})$, in which case $T'(\underline{w}) = 0$). From the households' optimization problem, the marginal tax rate is increasing over the bunched interval, $dT'(y(w))/dw = 1/(w^2 u'(x)) > 0$, where $u'(x)$ is fixed over the interval. This implies that if there is bunching at the highest income level those bunched individuals with $w < \overline{w}$ will have a negative marginal tax rate which violates the above result. Weymark (1986b) shows this result also holds in the discrete case when the non-negativity constraint on consumption is satisfied.

$\underline{w} < 1/(2G(\overline{w}))$. So there will be bunching at the bottom if $\underline{w}$ is small enough, and this can occur for any skill distribution.[14]

Suppose there is a wage rate w_2 such that for all $w > w_2$, $\dot{x}(w) > 0$ so the SOIC condition is satisfied ($\kappa(w) = 0$), while for $\underline{w} \leq w \leq w_2$, it is not ($\kappa(w) \neq 0$). There will therefore be a range of skills, say $[\underline{w}, w_0]$, such that all households obtain the bundle $(x(w_0), y(w_0))$. In this range, $\mu(w) \neq 0$, and $\mu(w)$ is determined by integrating (3) and using (5), (8), and (9) to give:

$$\mu(w) = \int_{\underline{w}}^{w} [G(\overline{w})(1 - mu'(x(w_0)))f(m) - (G(\overline{w})F(m) - G(m))u'(x(w_0))]dm$$

Since $\mu(w_0) = 0$, we obtain using (11) and after some simplification:

$$\frac{F(w_0)}{u'(x(w_0))} = \int_{\underline{w}}^{w_0} \frac{f(w)}{u'(x(w))} dw \tag{16}$$

where $u'(x(w)$ in the integral is given by (11) simply ignoring the SOIC condition. This equation, along with the requirement that $u'(x(w_0))$ satisfies (11) for $w = w_0$, determines w_0 and $x(w_0)$.

Figure 1 illustrates the case of bunching at the bottom. The J-shaped consumption profile depicts the pattern of consumption which would be generated by condition (11). The portion with the negative slope violates the SOIC condition. The minimum point on the consumption curve occurs at the skill level we have called w_2. The point w_0 is determined where (16) is satisfied. The value of $1/u'(x(w_0))$ just equals the average value of $1/u'(x(w))$ over the range $\underline{w} \leq w \leq w_0$. This, of course, implies that $w_0 > w_2$. Outside the bunching range, the path of consumption and the marginal tax rates are precisely the same as under the first-best approach. This implies that the tax rate faced by households with skill w_0 will be positive.

— FIGURE 1 NEAR HERE —

[14] Mirrlees (1971), in his simulations, assumed that $\underline{w} = 0$, implying that there is necessarily bunching at the bottom.

Bunching in the Interior

The same argument can be extended with relatively little amendment to bunching ranges in the interior.[15] Suppose, as in Figure 2, that there is a range in the interior of the consumption profile from w_{22} to w_2 that would be generated by the first-order approach (11) such that consumption is declining. This implies that there will be a range over which the non-negativity constraint on $\dot{x}(w)$ is binding, or $\kappa(w) > 0$. Suppose the bounds of that range are denoted by w_{00} and w_0. Over this range $x(w)$ is constant at $x(w_0)$ as shown in the diagram. To determine the values of w_0, w_{00} and $x(w_0)$, note that $\mu(w_0) = \mu(w_{00}) = 0$, but $\mu(w) \neq 0$ for $w_{00} < w < w_0$.

Proceeding as above, integrating (3) now yields for the range $w_{00} < w < w_0$:

$$\mu(w) = \int_{w_{00}}^{w} [G(\overline{w})(1 - mu'(x(w_0)))f(m) - (G(\overline{w})F(m) - G(m))u'(x(w_0))]dm$$

Using $\mu(w_0) = 0$, this simplifies to:

$$\frac{F(w_0) - F(w_{00})}{u'(x(w_0))} = \int_{w_{00}}^{w_0} \frac{f(w)}{u'(x(w))} dw \tag{17}$$

In this case, w_0, w_{00} and x_0 are jointly determined by (17) and by the requirement that $u'(x(w_0))$ satisfy (11) for both $w = w_0$ and $w = w_{00}$. As in the case of bunching at the bottom, (17) indicates that the bunching range is determined by the condition that the reciprocal of $u'(x(w_0))$ equals the average value of the reciprocal of $u'(x(w))$ over the bunching interval. Figure 2 depicts the consumption pattern when the SOIC condition fails over the range $[w_{22}, w_2]$ so that bunching occurs in the interior.

— FIGURE 2 NEAR HERE —

The upshot of the above discussion is that with preferences that are quasi-linear in leisure, *bunching induced by a violation of the* SOIC *condition affects neither the consumption path nor the marginal tax rates in the non-bunched interval(s).* It will, however, affect average tax rates, as will become clear later.

[15] The possibility of this case was considered by Ebert (1992) as an example of the relationship between bunching and violation of the SOIC condition, though not explicitly solved.

IV. MAXI-MIN OBJECTIVE FUNCTION

The above analysis is based on the government adopting a utilitarian objective function in which all households' utilities are given equal weight.[16] In this section, we consider the case of a maxi-min objective function where the concern of the government is solely with the least able persons (those with $w = \underline{w}$). This case, though extreme, enables us to obtain additional results on the pattern of the tax structure, which turns out to take a very simple form. For the sake of simplicity, the SOIC condition is assumed to be satisfied everywhere as well as the non-negativity condition on income. Therefore, our analysis will not consider the possibility of bunching although we shall spell out the conditions for the first-order approach to be valid. It is straightforward to extend our analysis to account for bunching, but no new insights are gained.

In this case, the government's problem can be written:

$$\max \quad \frac{V(\underline{w})}{\underline{w}}$$

subject to:

$$\int_{\underline{w}}^{w} [wu(x(w)) - V(w) - x(w)]f(w)dw \geq R$$

and

$$V(w) = V(\underline{w}) + \int_{\underline{w}}^{\overline{w}} u(x(m))dm$$

where the first constraint is the budget constraint and the second constraint is simply the integral of the FOIC condition, $\dot{V}(w) = u(x(m))$.

Substituting the second constraint into the first allows us to eliminate $V(w)$ for $w > \underline{w}$

[16] It is straightforward to extend our analysis to a general quasi-concave additive social welfare function with differing weights on individuals' utilities. The results on the optimal path of consumption and marginal tax rates are analogous to those obtained in (11) and (13) except the term $B(w)$ now depends on the social welfare weights given to each individual. If the individual weights take a very simple linear weighting scheme, it is possible to characterize the amount of aversion to inequality the government's objective function has with a single parameter. It can then be shown that the marginal cost of public funds and the optimal path of consumption in the interior of the skill distribution is decreasing in the government's aversion to inequality.

from the problem. The revenue constraint becomes:[17]

$$\int_{\underline{w}}^{\overline{w}} \left[(wf(w) - 1 + F(w))u(x(w)) - x(w)f(w)\right] dw - V(\underline{w}) \geq R$$

The government problem is then simply to choose $V(\underline{w})$ and $x(w)$ to maximize $V(\underline{w})$ subject to this revenue constraint. The Lagrangian expression is:

$$\mathcal{L} = \frac{V(\underline{w})}{\underline{w}} + \lambda \left(\int_{\underline{w}}^{\overline{w}} \left[(wf(w) - 1 + F(w))u(x(w)) - x(w)f(w)\right] dw - V(\underline{w}) - R \right)$$

The first-order conditions with respect to $V(\underline{w})$ and $x(w)$ may be written:

$$\lambda = \frac{1}{\underline{w}}$$

and

$$u'(x(w)) = \frac{f(w)}{wf(w) + F(w) - 1} = \frac{1}{w(1 - C(w))} \tag{18}$$

where $C(w)$ is defined as before.

The first condition indicates that the shadow price of public funds is simply $1/\underline{w}$: a unit increase in revenue will cause all households, including the least skilled to supply one more unit of income.[18] Only the latter is accounted for as far as social welfare is concerned. From the second equation, we conclude that $u'(x(w)) < 1/w$ except at the top of the distribution where it is equal to the wage rate. And, since $T'(y(w)) = 1 - 1/(wu'(x(w)))$,

$$T'(y(w)) = C(w) \tag{19}$$

which is useful to compare with (13) in the utilitarian case.[19] As usual, the marginal tax rate is zero at the top ($T'(y(\overline{w})) = 0$), while in the interior $1 > T'(y(w)) > 0$. But, at the bottom, $T'(y(\underline{w})) = 1/(\underline{w}f(\underline{w})) > 0$ contrary to the utilitarian case in which the

[17] We have again used the rule for iterative integration referred to earlier.

[18] As will be shown later, income changes on a one-for-one basis with the government's required revenue. That is, $dy(w)/dR = 1$.

[19] Solving the maxi-min problem for the case where preferences are quasi-linear in consumption (the Diamond (1998) case), it is straightforward to show that $T'(y(w)) = C(w)A(w)$ where $A(w)$ is defined as in footnote 10. As in Diamond, when the elasticity of labour supply is constant, so is $A(w)$. Then, qualitatively the same pattern of marginal tax rates emerges.

marginal tax rate is zero in the absence of bunching at the bottom of the skill distribution.[20] From Property iii), we infer that with a unimodal distribution T' is declining in the range $(\underline{w}, w_m)$.

To determine whether or not this first-order approach is valid, i.e. the SOIC condition is satisfied, we first note from (18) that $\dot{x}(w) > 0$ if $w(1 - C(w))$ is increasing in w which is equivalent to:

$$2 + C(w)\frac{w\dot{f}(w)}{f(w)} > 0 \tag{20}$$

For $\dot{f}(w) \geq 0$, (20) will be positive and the SOIC condition will be satisfied.

V. RESULTS FOR SPECIFIC DISTRIBUTIONS

In the above analysis, we showed that both the marginal tax rates and violations of the SOIC condition depend solely on the skill distribution. This is true whether the government has a utilitarian or maxi-min objective function. In this section, we adopt specific skill distributions and use our general analytical results to characterize the marginal tax rates and the possible violations of the SOIC condition under the two different objective functions. The distributions we consider include the uniform, the log-normal, the truncated and untruncated Pareto, and the exponential. For the last two distributions, we consider that they only apply after the modal skill level.

Uniform Distribution (Lollivier and Rochet (1983))

As shown by Lollivier and Rochet (1983), the uniform distribution enables one to obtain clear-cut results. This is the reason why we consider it here despite its unrealistic form.

Marginal Tax Function

The density function of the uniform distribution is given by $f(w) = 1/(\overline{w} - \underline{w})$ for $w \in [\underline{w}, \overline{w}]$. Using (13), we obtain the following expression for the marginal tax rate under the utilitarian objective:

$$T'(y(w)) = \left[\frac{\log w - \log \underline{w}}{\log \overline{w} - \log \underline{w}} - \frac{w - \underline{w}}{\overline{w} - \underline{w}}\right] \frac{\overline{w} - w}{w}$$

[20] For a discussion of the possibility of a non-zero marginal tax rate at the bottom for the maxi-min case, see Seade (1977), page 231, footnote 29.

As expected, the marginal tax rate is nil at both the lower and upper bounds of the skill support. Differentiating the above expression yields:

$$\frac{dT'(y(w))}{dw} = \frac{\overline{w} - \underline{w}}{w^2}\left[\frac{1 + \log \underline{w} - \log w}{\log \overline{w} - \log \underline{w}} - \frac{w}{\overline{w} - \underline{w}}\right]$$

which is positive at $w = \underline{w}$ and negative at $w = \overline{w}$, also as expected. Since both factors on the right-hand side of the above expression decrease with w, the marginal tax function is concave with respect to w and so has an inverted-U shape under the utilitarian objective.

These results may be contrasted to those obtained under the maxi-min objective. In this case, we infer from (18):

$$T'(y(w)) = \frac{\overline{w}}{w} - 1$$

implying that starting from a positive value at $w = \underline{w}$ the marginal tax rate is monotonically decreasing in w and is nil at $w = \overline{w}$. Furthermore, the marginal tax rate is convex with respect to w.

Violations of the SOIC Condition

From (13) and (15) (or (18) and (19) for the maxi-min case), the SOIC condition is satisfied if $w(1 - T'(y(w)))$ increases with w. Under the utilitarian objective, this can be shown to be equivalent to:

$$w > \frac{\overline{w} - \underline{w}}{2(\log \overline{w} - \log \underline{w})}$$

implying that there may only be bunching on an interval of skills starting at the bottom of the distribution. Moreover, for bunching to occur $\underline{w}$ must be small enough.[21] On the contrary, under the maxi-min objective there is never any bunching since $w(1 - T'(y(w)))$ always rises with w.

Log-Normal Distribution (Mirrlees (1971))

A more realistic skill distribution is the log-normal, according to which $\log w$ is normally distributed for $w \in \mathcal{W}$. We can show that $C(w)/(1 - F(w))$ is increasing in w for $w > w_m$.

Unfortunately, other than the general results given above, no further results can be derived analytically under either the utilitarian or maxi-min objectives. That is, T' declines

[21] Lollivier and Rochet (1983) show that under a quasi-concave additive social welfare function with an arbitrary linear weighting scheme for individuals utilities, the SOIC condition can only be violated on an interval at the bottom of the uniform distribution.

in w over the range $w \in [w_c, w_m]$ under a utilitarian objective and declines in w for all $w \leq w_m$ under a maxi-min objective. This implies that the SOIC condition is satisfied below the modal skill level, but may be violated at some $w > w_m$.

As in the more general case where preferences are not quasi-linear, we would have to rely on simulations to get further results. Such simulations have proven to be somewhat inconclusive. For example, Tuomala (1990), following Mirrlees (1971), takes the mean of $\log w$ to be -1 and allows the standard deviation to vary from 0.39 to 1.0. Then for a variety of assumptions about the elasticity of substitution between consumption and leisure, he finds that, except close to the bottom of the income distribution, the schedule of marginal tax rates is monotonically declining regardless of whether the objective function utilitarian or maxi-min. Other authors have adopted more realistic assumptions on the variance of the skill distribution and have shown that the optimal marginal tax rate schedule increases over most skill levels and declines near the top of the distribution.[22]

Pareto Distribution (Diamond (1998))

Suppose instead a Pareto skill distribution. By assumption, this distribution only applies after the modal skill level. That is, over the range $(\widehat{w}, \overline{w})$ where $\widehat{w} \geq w_m$. As a standard, we adopt the realistic assumption that $\overline{w} < \infty$. However, we also consider the case when $\overline{w} = \infty$. This is the assumption used in Diamond (1998). As we shall see, his qualitative results depend critically on the Pareto distribution being untruncated, that is, on there being some households of infinitely large productivity.

Marginal Tax Function

Over the range $(\widehat{w}, \overline{w})$ the density of the Pareto distribution is $f(w) = Aw^{-(1+a)}$ where $a > 0$ and $A = a(\widehat{w}^a - \overline{w}^a)^{-1}[1 - F(\widehat{w})]$. Solving for $C(w)$, it can be shown that $C(w) = (1/a)\left[1 - (w/\overline{w})^a\right]$, which implies that with an untruncated distribution ($\overline{w} = \infty$), $C(w) = 1/a$. Differentiating $C(w)$ in the truncated case ($\overline{w} < \infty$) yields $\dot{C}(w) = -(1/\overline{w})^a aw^{a-1}$ which is negative. Therefore, under a truncated Pareto distribution $C(w)$ is decreasing, while with an untruncated Pareto distribution $C(w)$ is constant.

[22] See Kanbur and Tuomala (1990). In their paper, they assume a CES utility function and set the standard deviation of $\log w$ to 0.7 and 1.0. In both cases, the optimal marginal tax rate schedule exhibits an inverse U–shape when the government maximizes a utilitarian social welfare function.

From Property i), this implies that when the Pareto distribution is untruncated, T' is increasing in w under a utilitarian objective function. This indicates that Diamond's result (his Proposition 1), which generates the U–shaped pattern of T' above w_m, applies not only to the case where preferences are quasi-linear in consumption, but also to the case when they are quasi-linear in leisure. But, this is clearly a very special result. As soon as the distribution is truncated, the Diamond result no longer applies: $C(w)$ is now decreasing rather than constant.

To get a better idea of the shape of the marginal tax rate under the utilitarian objective, we use expression (13) to solve for T' under the Pareto distribution:

$$T'(y(w)) = \frac{1}{a}\left[1 - \left(\frac{w}{\overline{w}}\right)^a\right] + \frac{1}{G(\overline{w})(1+a)}\left[\frac{1}{\overline{w}}\left(\frac{w}{\overline{w}}\right)^a - \frac{1}{w}\right]$$

Differentiating $T'(y(w))$ with respect to w, we obtain after some simplification

$$\frac{dT'(y(w))}{dw} = w^{a-1}\left(\frac{1}{\overline{w}}\right)^a\left[\frac{a}{1+a}\frac{1}{G(\overline{w})\overline{w}} - 1\right] + \frac{1}{w^2 G(\overline{w})(1+a)}$$

The first term in this expression is negative, and therefore, whether the marginal tax rate is increasing or decreasing when the distribution is truncated depends on the relative magnitudes of the two terms. When $a \geq 1$, both terms are decreasing in w, which means that the marginal tax rate is concave in w; and, there is a range of skills starting with $\widehat{w}$ over which T' increases with w if and only if the above derivative is positive at $w = \widehat{w}$. When $a < 1$, it is impossible to reach unambiguous conclusions in the utilitarian case. By contrast, if the distribution is untruncated ($\overline{w} = \infty$), then the marginal tax schedule is increasing and strictly concave in w, as in Diamond.

Turning to the maxi-min objective function, clear-cut results are obtained for the Pareto distribution. Using (19), the marginal tax rate in this case is given by

$$T'(y(w)) = \frac{1}{a}\left[1 - \left(\frac{w}{\overline{w}}\right)^a\right]$$

which implies that with a truncated Pareto distribution, the marginal tax rate is decreasing in w and is zero at $w = \overline{w} < \infty$. In addition, it will be strictly concave when $a > 1$ and strictly convex when $a < 1$. If the distribution is untruncated, then the marginal tax rate is constant and equal to $1/a$.

Violations of the SOIC Condition

The SOIC condition is satisfied if $w(1 - T'(y(w)))$ increases with w. Under the utilitarian objective, this can be shown to be equivalent to:

$$1 - \frac{1}{a} + \left(\frac{w}{\overline{w}}\right)^{a} \left[\frac{1+a}{a} - \frac{1}{G(\overline{w})\overline{w}}\right] > 0$$

where the expression in the square brackets is positive. As the left-hand side is monotonically increasing in w, bunching may only occur on an interval of skills starting at $\widehat{w}$ when the distribution is truncated. For very small values of $a > 0$, it is possible that the SOIC condition is violated even at w close to $\overline{w}$. If the distribution is untruncated, then the above expression reduces to $1 - 1/a > 0$ which is satisfied for all $a > 1$. If $a \leq 1$, then the SOIC condition is always violated when the Pareto distribution is untruncated.

Under a maxi-min objective function, expression (20) must be positive for the SOIC condition to be satisfied. If the distribution is truncated, expression (20) is equivalent to $(w/\overline{w})^{a} > (1-a)/(1+a)$, which is everywhere satisfied when $a \geq 1$. If $a < 1$, then it is possible that the SOIC condition is violated. However, since the left-hand side of the expression is monotonically increasing in w, if there is bunching it must only occur on an interval $(\widehat{w}, \widehat{\widehat{w}})$ where $\widehat{\widehat{w}} < \overline{w}$. If instead the distribution is untruncated, then the SOIC condition under a maxi-min objective is equivalent to $a > 1$. Thus, if the distribution is untruncated the SOIC condition can nowhere be satisfied when $a \leq 1$.

Exponential Distribution

An alternative distribution which we also assume to apply after the modal skill level ($\widehat{w} > w_m$) is the exponential distribution. Although this distribution is not used in the literature, we adopt it here to contrast the results obtained with it to those obtained under a Pareto distribution.

Marginal Tax Function

The density of an exponential distribution is $f(w) = Ae^{-w/b}$, where $A > 0$, the exponential rate of decline $(1/b)$ is positive, and $\hat{w} > w_m$. This distribution yields $C(w) = (b/w)\left(1 - e^{w/b}/e^{\overline{w}/b}\right)$. Differentiating this expression, we find

$$\dot{C}(w) = -(b/w^2)\left(1 - e^{w/b}/e^{\overline{w}/b}\right) - e^{w/b}/(we^{\overline{w}/b})$$

which is everywhere negative whether $\overline{w}$ is finite or not. On the other hand, differentiating $C(w)/(1-F(w))$ gives $d(1/(wf(w)))/dw = [1/(w^2 f(w))]\,[w/b-1]$ which is negative when $w < b$, and positive otherwise. Therefore, from Property ii), T' is decreasing in w under the utilitarian objective function in the range $[\widehat{w}, b)$. On the contrary, if the objective is maxi-min then T' is necessarily decreasing over the entire exponential distribution.

Violations of the SOIC Condition

Under a utilitarian objective, the SOIC condition is satisfied in the range $[\widehat{w}, b)$ since the denominator of (15), i.e. $w(1-T')$, is increasing in this range. Using the same argument, under a maxi-min objective, the SOIC condition is always satisfied above $\widehat{w}$ since T' is decreasing.

VI. THE PATTERN OF INCOME AND AVERAGE TAX RATES

So far we have studied the optimal path of consumption and the tax structure that it implies, but we have said little about the path of income. In the above problem, y has been substituted out of the revenue constraint (1) using the definition of $V(w)$. We can therefore determine $y(w)$ by working back through the revenue constraint. As we shall see, there is no guarantee that this yields a pattern of income, and therefore of labour supply, which is non-negative throughout the income range. This implies that we need to introduce explicitly a non-negativity constraint on income; following the standard literature we so far have implicitly assumed it to be satisfied. We show in this section that such a constraint can affect the patterns of both tax payments and consumption across households.

We begin by showing how $y(w)$ can be recovered from the above model. *Let us for the moment continue to assume that the non-negativity constraint on $y(w)$ is not binding.* From the definition of $V(w)$, we have $y(w) = wu(x(w)) - V(w)$. To obtain an expression for $V(w)$, we integrate the first-order IC condition, $\dot{V}(w) = u(x(w))$, to give $V(w) = V(\underline{w}) + \int_{\underline{w}}^{w} u(x(m))dm$, which we substitute in the above relation to obtain:

$$y(w) = wu(x(w)) - V(\underline{w}) - \int_{\underline{w}}^{w} u(x(m))dm \tag{21}$$

Then, integrating this over the entire population and rearranging, we obtain $V(\underline{w}) =$

$\int_{\underline{w}}^{\overline{w}}[wu(x(w))-y(w)]f(w)dw-\int_{\underline{w}}^{\overline{w}}u(x(w))(1-F(w))dw$.[23] Eliminating from this expression $\int_{\underline{w}}^{\overline{w}}y(w)f(w)dw$ by using the government budget constraint yields:

$$V(\underline{w}) = \int_{\underline{w}}^{\overline{w}} [wu(x(w)) - x(w)]f(w)dw - \int_{\underline{w}}^{\overline{w}} u(x(w))(1 - F(w))dw - R$$

Finally, using this expression in (21), we obtain:

$$y(w) = K(w) + R \tag{21$'$}$$

where

$$K(w) = wu(x(w)) - \int_{\underline{w}}^{\overline{w}} \{u(x(w))[wf(w) + 1 - F(w)] - x(w)f(w)\}\, dw - \int_{\underline{w}}^{w} u(x(m))dm$$

Note that $K(w)$ depends only on the distribution of skills and the functional form of $u(x)$, since $x(w)$ depends only on these two elements.[24] This implies that:

$$\frac{dy(w)}{dR} = 1 \qquad \forall w \in \mathcal{W} \tag{22}$$

Thus, an increase in revenue requirement causes all persons to increase their income by an equal amount. This has a number of implications. First, the effect on the path of $y(w)$ of a change in revenue requirements confirms our earlier intuition for the interpretation of λ as the shadow price of public funds in (8). Since all household incomes change by the same amount when revenue requirements change, a unit increment of tax revenue involves a household of type w increasing its labour supply by $1/w$ with no change in consumption. The utility cost of these labour supply changes is simply $1/w$ summed over the entire population, which is $E(1/w)$, or $G(\overline{w})$.

---

[23] We have made use of Fubini's Theorem to evaluate the iterated integral as follows:

$$\int_{\underline{w}}^{\overline{w}} \left(\int_{\underline{w}}^{w} u(x(m))dm \right) f(w)dw = \int_{\underline{w}}^{\overline{w}} u(x(w))(1 - F(w))dw$$

See, for example, Stewart (1991).

[24] Since R is independent of w, it must be that $\dot{y}(w) = \dot{K}(w)$. Using the above expression for $K(w)$, $\dot{K}(w) = wu'(x(w))\dot{x}(w) + u(x(w)) - u(x(w)) = wu'(x(w))\dot{x}(w)$ (as in footnote 3) which equals $\dot{x}(w)/(1 - T'(y(w)))$, implying $\dot{y}(w) > \dot{x}(w)$.

Following Musgrave and Thin (1948), we define the tax system to be progressive if the average tax rate increases with income, where the average tax rate is given by $A(w) \equiv T(y(w))/y(w)$. The slope of a ray from the origin to the tax function $T(y(w))$ in (y, T)-space is equal to the average tax rate $A(w)$. How the slope of this ray changes as y increases determines the progressivity of the tax system and will depend on both the distributional assumptions and the amount of required revenue to be raised. The distributional assumptions determine the shape of the total tax function as shown above. To see how the degree of progressivity depends on the amount of revenue to be raised, rewrite $A(w)$ as follows:

$$A(w) = \frac{y(w) - x(w)}{y(w)} = 1 - \frac{x(w)}{y(w)}$$

Using (22), differentiating yields:

$$\frac{dA(w)}{dR} = \frac{x(w)}{y^2(w)} > 0$$

Thus, not surprisingly, average tax rates increase for all households if the revenue requirement increases.

To illustrate how changes in R affect the progressivity of the tax system consider two persons in some non-bunched range with wages $w_2 > w_1$. Suppose initially that the tax is progressive, so $A(w_2) > A(w_1)$. This implies that $c(w_1)/y(w_1) > c(w_2)/y(w_2)$. Therefore, since $y(w_2) > y(w_1)$, $dA(w_1)/dR > dA(w_2)/dR$: the average tax rate increases more for the low-wage person than for the high-wage one. Consider the effect of an increase in R on the ratio $A(w_1)/A(w_2)$. If this ratio increases, the tax system becomes less progressive. Differentiating $A(w_1)/A(w_2)$ with respect to R yields:

$$\frac{d}{dR}\left(\frac{A(w_1)}{A(w_2)}\right) = \frac{1}{A^2(w_2)}\left[A(w_2)\frac{dA(w_1)}{dR} - A(w_1)\frac{dA(w_2)}{dR}\right] > 0$$

In words, if the tax system is progressive, increasing the revenue requirement makes it less so. On the other hand, if the tax system is regressive to begin with ($A(w_1) > A(w_2)$), we cannot predict whether an increase in required revenue causes it to become more or less progressive.

Finally, the fact that incomes for all households change by one-to-one with the revenue requirement, as shown by (22), implies that we need to take account of the possibility that a non-negativity constraint on income may be required. Suppose, for example, that the SOIC condition is satisfied, so that bunching of the sort discussed earlier does not occur. In this case, consumption is everywhere increasing in w, and incomes will be determined by (21′), at least assuming they are all positive, and will also be increasing. Imagine now decreasing the revenue requirement. As R decreases, so too will $y(w)$. Eventually a point will be reached where incomes for the lowest wage households fall to zero. Obviously, incomes cannot fall below zero. As R is reduced further, an increasing range of low-wage persons find their incomes reduced to zero. Since they must all obtain the same consumption, we have a situation in which bunching at the bottom occurs even though the SOIC condition is not violated. Let us turn to a more formal analysis of this case.

VII. THE NON-NEGATIVE INCOME CONSTRAINT

To analyze what happens to the stream of consumption and tax rates when a non-negativity constraint on incomes is binding but the SOIC condition $\dot{x}(w) \geq 0$ is not active, let us amend the first-order approach to our optimal income tax problem to incorporate the non-negativity constraint on income. The latter can be written $y(w) = wu(x(w)) - V(w) \geq 0$. Associating the Lagrange multiplier $\delta(w)$ with this non-negativity constraint, the Hamiltonian becomes:

$$H = V(w)\frac{f(w)}{w} + \lambda[wu(x(w)) - V(w) - x(w)]f(w) + \pi(w)u(x(w)) + \delta(w)(wu(x(w)) - V(w))$$

where $x(w)$ now becomes a control variable (since we have assumed the constraint $\dot{x}(w) \geq 0$ is not binding). The necessary conditions are now:

$$\frac{\partial H}{\partial x(w)} = 0 = \lambda[wu'(x(w)) - 1]f(w) + \pi(w)u'(x(w)) + \delta(w)wu'(x(w)) \tag{23}$$

$$\frac{\partial H}{\partial V(w)} = -\dot{\pi}(w) = \frac{f(w)}{w} - \lambda f(w) - \delta(w) \tag{24}$$

$$\pi(\underline{w}) = \pi(\overline{w}) = 0 \tag{25}$$

$$\delta(w) > 0 \rightarrow y(w) = 0; \quad y(w) > 0 \rightarrow \delta(w) = 0 \tag{26}$$

By assumption, R is low enough for there being at the optimum a range of skills $\underline{w} \leq w \leq w_y$ such that the non-negative income constraint is binding. Thus, $\delta(w) > 0$ for $w < w_y$, and zero elsewhere. Denote the level of consumption for those earning zero income by x_0.

Using the transversality condition $\pi(\overline{w}) = 0$, we can integrate (24) to obtain $\pi(w)$ for $w \geq w_y$:

$$\pi(w) = \int_w^{\overline{w}} \left(\lambda - \frac{1}{m}\right) f(m)dm, \quad w \geq w_y \tag{27}$$

since $\delta(w) = 0$ for $w \geq w_y$. Then, substituting $\delta(w)$ from (24) into (23) yields for $\underline{w} \leq w \leq w_y$:

$$(u'(x_0) - \lambda)f(w) + u'(x_0)\frac{\partial w\pi(w)}{\partial w} = 0, \quad w \leq w_y$$

Integrating this, and using the transversality condition $\pi(\underline{w}) = 0$, we obtain:

$$(u'(x_0) - \lambda)F(w_y) + u'(x_0)w_y\pi(w_y) = 0 \tag{28}$$

Finally, substituting $\pi(w_y)$ from (27) into this expression provides us with the following formula of the shadow price of public funds:

$$\lambda = \frac{\frac{F(w_y)}{w_y} + \int_{w_y}^{\overline{w}} \frac{1}{m} f(m)dm}{\frac{F(w_y)}{w_y u'(x_0)} + (1 - F(w_y))} \tag{29}$$

where λ now depends on the value of x_0 and w_y. This must be compared with (8), which holds when the non-negativity constraint $y(w) \geq 0$ is not binding at any skill level. The difference between the two formulas is easily understood. To see this, let us increase the labour income of households with skills $w \geq w_y$ by one unit. This increase yields $1 - F(w_y)$ in tax revenue and causes a loss in social welfare of $\int_{w_y}^{\overline{w}} (1/m)f(m)dm$. To continue satisfying the incentive compatibility constraint of households with skill w_y, $x_0 = x(w_y)$ needs to be adjusted by $dx_0 = -(1/(u'(x_0)w_y))dy < 0$. As any household with $w < w_y$ has its utility decreased by the same amount as households with skill w_y, this causes a loss in social welfare equal to $F(w_y)/w_y$. It also decreases public expenses by an amount given by the first term in the denominator in (29).

Substituting $\pi(w)$ from (27) into (23), the path of consumption for households with skill above w_y can be inferred from the following condition:

$$u'(x(w)) = \frac{f(w)}{wf(w) + \frac{1}{\lambda}\int_{w_y}^{\overline{w}} \frac{1}{m} f(m)dm - 1 + F(w)}, \quad w \geq w_y \tag{30}$$

where, $u'(x(w)) < 1/w$ for $w_y \leq w \leq \overline{w}$ and λ is given by (29). As earlier, we can also find an expression for the marginal tax rate faced by households with skill above w_y:

$$T'(w) = \frac{\lambda(1 - F(w)) - \int_{w_y}^{\overline{w}} \frac{1}{m} f(m)dm}{\lambda w f(w)} \tag{31}$$

where λ is given by (29). The expression for T' is positive except for $w = \overline{w}$ where it is nil, and individuals at the end of the bunching interval (type-w_y households) have a positive marginal tax.[25]

It is worth emphasizing that once the non-negativity constraint $y \geq 0$ is binding on some range of skill, the property that the pattern of marginal tax rates depends only upon the skill distribution no longer holds. The level of revenue requirements R also matters because it affects the values of w_y and x_0 and so that of λ (see (29)), which in turn influences the consumption path (see (30)). The larger is the level of R the higher will be the value of λ, and therefore for a given skill $w \geq w_y$, the lower consumption will be.

So far, our discussion of income constraints has focused on the case of a minimum income constraint. This is a natural case to consider since it generates bunching at the bottom of the skill distribution, a phenomenon which has been widely emphasized in the optimal income tax literature. But, a maximum income constraint is equally plausible, especially in our model with quasi-linear preferences. To see this, consider the full-information analogue of our model, where the government can observe ability. With no incentive constraint to contend with, the government simply chooses $x(w)$ and $y(w)$ to maximize $\int_{\underline{w}}^{\overline{w}}[u(x(w)) - y(w)/w]f(w)dw$ subject to the revenue constraint $\int_{\underline{w}}^{\overline{w}}[y(w) - x(w)]f(w)dw \geq R$, the non-negative income constraint $y(w) \geq 0$, and a maximum income constraint of the form $y(w)/w \leq \bar{\ell}$, where $\bar{\ell}$ is the maximum amount of labour that can be supplied to the market. With λ denoting the shadow price of public funds, the solution to this problem is such that $u'(x(w)) = \lambda \ \forall w \in \mathcal{W}$, implying that all individuals receive the same consumption allocation in the first-best outcome. Income allocations, on the other hand, will be such that only individuals with skill above some level $\tilde{w}$ provide labour, and this at the maximum amount of labour ($\bar{\ell}$). This level is such that $u'\left(\bar{\ell}\int_{\tilde{w}}^{\overline{w}} mf(m)dm\right) = 1/\tilde{w}(=\lambda)$. For all $w < \tilde{w}$, income is zero and the non-negative income constraint binds, while for all $w > \tilde{w}$, income is equal to $w\bar{\ell}$.

When the government cannot observe skill levels, this stark result will not occur. Nonetheless, it is certainly conceivable that, as well as there being some low skill levels for which the non-negative income constraint is binding, there will be some individuals supplying the maximum amount of labour, $\bar{\ell}$. Unlike the non-negativity constraint, the

[25] This follows from evaluating expression (31) at $w = w_y$.

maximum income constraint does not cause people to be bunched at a given (x, y)–bundle as they have differing skill levels, and therefore, different maximum incomes. In addition, although $\dot{y}(w) > 0$ by the SOIC condition, labour supply does not have to rise monotonically with ability. Therefore, the maximum income constraint may bind only in the interior and/or at the bottom of the skill distribution in the second-best outcome. Simulations using a log-normal skill distribution illustrate this result, i.e. $\dot{\ell}(w) < 0$ near and at the top of the skill distribution (Tuomala, 1990). However, if the highest skilled individuals are constrained to earn their maxi-min income then they will have a positive marginal tax rate.[26]

VIII. CONCLUDING REMARKS

Our purpose in this paper has been to provide as full a characterization as possible of the solution to the optimal income tax problem when preferences are quasi-linear in leisure. This case has the unique feature that under utilitarian and maxi-min objective functions, closed-form solutions can be obtained for the path of the marginal utility of consumption and the pattern of income tax rates in terms of the skill distribution function. This is obviously of some pedagogical value. But in addition, in the process of examining this case, we have uncovered a number of interesting features of the optimal income tax problem and its solution. These include the following.

First, bunching at the bottom of the skill distribution can occur for two reasons. Either the SOIC condition is violated, as emphasized by Ebert (1992), or a non-negativity constraint on incomes is binding. In the former case, bunching will generally involve a positive income for those bunched, while in the latter it necessarily involves zero income. Bunching can also occur in the interior if the SOIC condition is violated there. In the case where bunching arises because of a failure of the SOIC condition, the solutions for the marginal utility of consumption and the marginal tax rates in the non-bunched regions are identical to those obtained using the first-order approach.

Second, a maximum income constraint reflecting the maximum available labour supply

[26] This can be shown be adding the constraint, $y(w)/w \geq \bar{\ell}$ to the government's optimization problem in the non-negativity section and noting that increases in $\bar{\ell}$ increase the value of the Hamiltonian. Deriving an expression for the optimal marginal tax rate, and evaluating it at $w = \overline{w}$ given the maximum income constraint is binding at $w = \overline{w}$, we obtain $T'(y(\overline{w})) > 0$.

must also be specified for the optimal income tax problem. It may be binding at the top of the income distribution, but will not lead to bunching since the maximum labour supply reflects different maximum incomes for different skilled persons.

Third, with preferences that are quasi-linear in leisure, as long as the minimum and maximum income constraint are not binding, changes in government revenue requirement induce equal per person changes in income, with no change in consumption. This obviously reflects the zero income elasticity of labour supply, and results in a very simple representation of the shadow price of public funds under the two objective functions. A consequence of this is that as government revenue is reduced, eventually some bunching will be induced at the bottom as the non-negative income constraint becomes binding. Once that happens, further reductions in the revenue requirement will affect the entire path of consumption.

Fourth, in the absence of bunching, interesting patterns for the tax structure emerge. Under fairly plausible assumptions about the skill distribution, the pattern of marginal tax rates takes an inverted U–shape for the utilitarian case. This implies that the tax function itself is concave at lower skill levels and convex at higher ones. The implication is that while the tax structure can be progressive at the lower end (average tax rates can be rising with skills), it will be regressive for higher skill levels. The skill level at which regressivity sets in depends upon the amount of revenue to be raised, it being lower the greater the revenue requirement. For the other case of the maxi-min objective function, the results are even more stark. In this case, the tax function is concave throughout under fairly mild restrictions on the distribution function. Thus, even in this case of extreme inequality aversion, the tax can be regressive over a significant range of the skill distribution. These general qualitative results will continue to hold when there is bunching at the bottom.

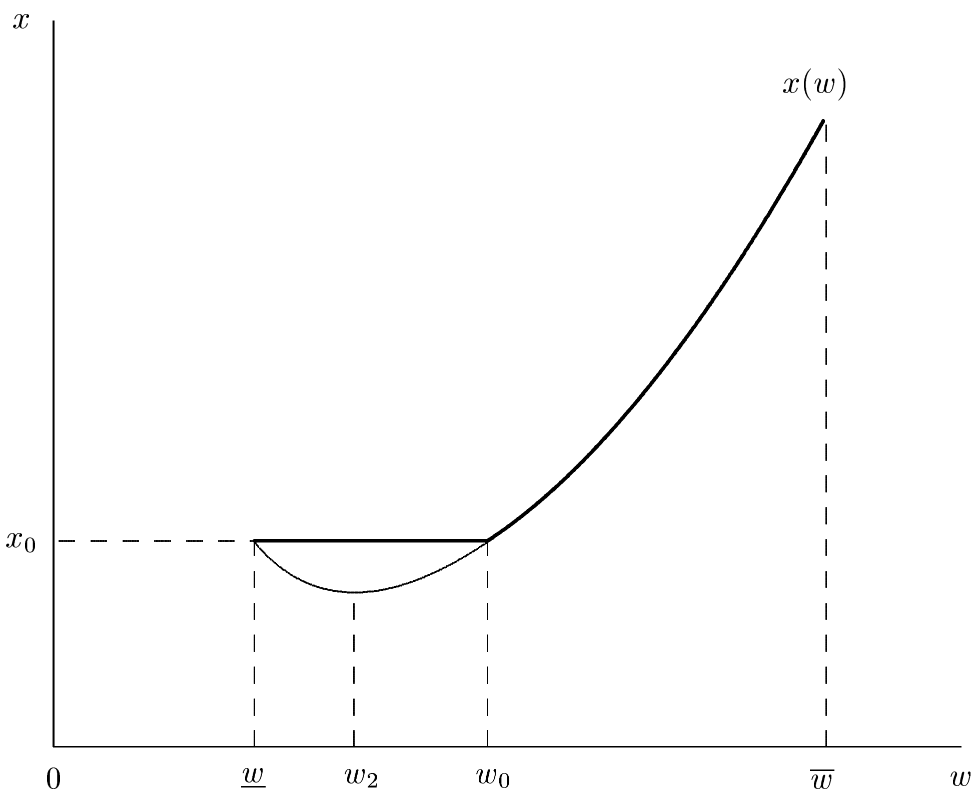

Figure 1

Bunching at the Bottom

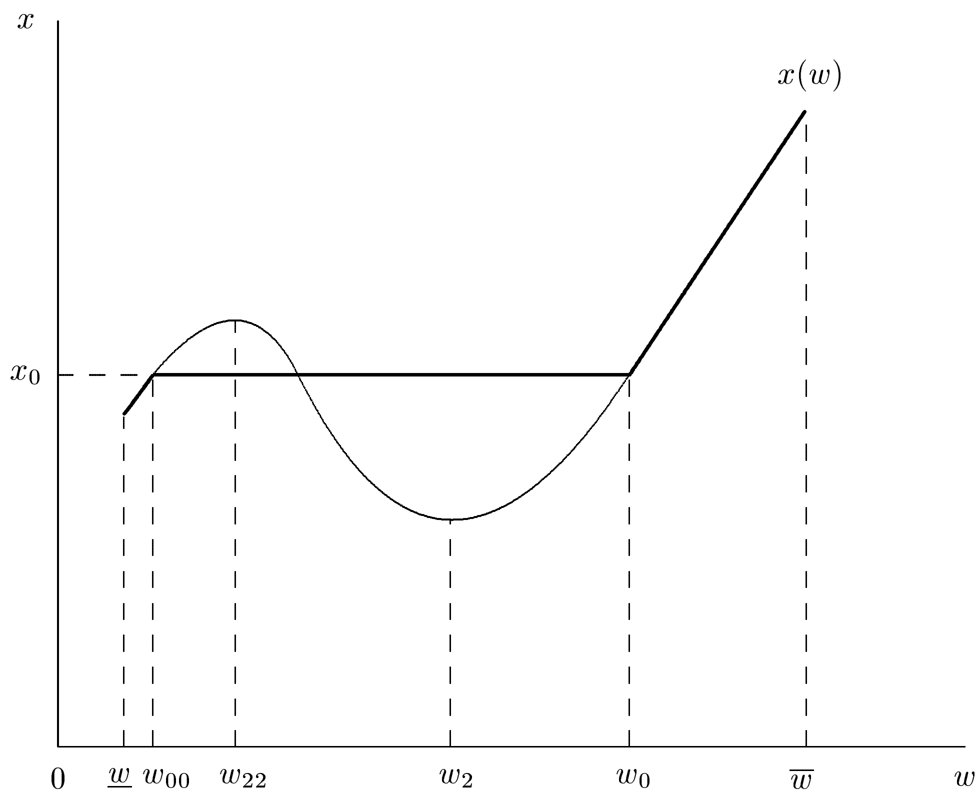

Figure 2

Bunching in the Interior

REFERENCES

Diamond, Peter A. (1998), 'Optimal Income Taxation: An Example with a U–Shaped Pattern of Optimal Marginal Tax Rates', *American Economic Review* **88**, 83–95.

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

Kanbur, Ravi and Matti Tuomala (1994), 'Inherent Inequality and the Optimal Graduation of Marginal Tax Rates', *Scandinavian Journal of Economics* **96**, 275–282.

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

Mirrlees, J.A. (1971), 'An Exploration in the Theory of Optimum Income Taxation', *Review of Economic Studies* **38**, 175–208.

Musgrave, R.A. and T. Thin (1948), 'Income Tax Progression 1929-1948', *Journal of Political Economy* **56**, 498–514.

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

Stewart, James, *Calculus, 2nd edition* (California: Brooks/Cole Publishing Company, 1991).

Tuomala, Matti (1990), *Optimal Income Tax and Redistribution* (Oxford: Clarendon Press).

Weymark, John A. (1986a), 'A Reduced-Form Optimal Nonlinear Income Tax Problem', *Journal of Public Economics* **30**, 199–217.

Weymark, John A. (1986b), 'Bunching Properties of Optimal Nonlinear Income Taxes', *Social Choice and Welfare* **3**, 213–32.

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

APPENDIX

Proof of Property i): By (13), $B(w) = [G(w)/G(\overline{w}) - F(w)]/[1 - F(w)]$. From (9) and the transversality conditions (5), $B(w) > 0$ for $w \in (\underline{w}, \overline{w})$ and zero at $w = \underline{w}$. By L'Hôpital's rule, $B(\overline{w}) = 1 - 1/(\overline{w}G(\overline{w})) > 0$. Differentiating $B(w)$:

$$\dot{B}(w) = \frac{f(w)\left[\frac{1-F(w)}{w} - G(\overline{w}) - G(w)\right]}{G(\overline{w})(1-F(w))^2} = \frac{f(w)[\int_w^{\overline{w}} f(x)dx - w\int_w^{\overline{w}} \frac{1}{x}f(x)dx]}{G(\overline{w})(1-F(w))^2} \qquad (A1)$$

which is positive for all $w < \overline{w}$.

Proof of Property ii): Differentiate $B(w)(1 - F(w))$ to obtain:

$$\frac{\partial B(w)(1-F(w))}{\partial w} = f(w)\left[\frac{1}{wG(\overline{w})} - 1\right] \qquad (A2)$$

which is positive for all $w < 1/G(\overline{w}) = w_c$, negative for all $w > w_c$, and zero for $w = w_c$.

Proof of Property iii): By (13), $C(w) = (1 - F(w))/(wf(w))$, which is positive for all $w \in [\underline{w}, \overline{w})$ and zero for $w = \overline{w}$. Differentiating $C(w)$:

$$\dot{C}(w) = -\frac{1}{w}\left[1 + \frac{1-F(w)}{wf}\left(1 + \frac{w\dot{f}(w)}{f(w)}\right)\right] \qquad (A4)$$

For any unimodal skill distribution up to the modal skill level, $\dot{f}(w) \geq 0$ and $(A4)$ is negative.

Proof of Property iv): Differentiate $C(w)/(1 - F(w)) = 1/(wf(w))$ to obtain:

$$\frac{d(1/(wf(w)))}{dw} = -\frac{1}{(wf(w))^2}\left[w\dot{f}(w) + f(w)\right] \qquad (A5)$$

For any unimodal skill distribution up to the modal skill level, $\dot{f}(w) \geq 0$ and $(A5)$ is negative.