Queen's Economics Department Working Paper No. 1073

Optimal Marginal and Average Income Taxation under Maxi-min

Robin Boadway
Queen's University

Laurence Jacquet
Universite Catholique de Louvain

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

5-2006

Optimal Marginal and Average Income Taxation under Maxi-min*

Robin Boadway[†]and Laurence Jacquet[‡]

May 19, 2006

Abstract

Using the Mirrlees optimal income tax model with a maxi-min social welfare function, we derive conditions for a decreasing marginal tax rate throughout the skill distribution, a strictly concave tax function in income and a single-peaked average tax schedule. With additive preferences and a constant labor supply elasticity, marginal tax rates are decreasing below the modal skill level, and will also decrease above the mode if aggregate skills are non-decreasing with the skill level. In this case and with a bounded skill distribution or with a constant hazard rate, the tax function is strictly concave in income and the average tax rate single-peaked. When quasilinear utility functions apply in either consumption or leisure, under fairly mild restrictions on the truncated or untruncated distribution function, marginal tax rates are decreasing in skill and the average tax profile is single-peaked. The distribution of skills has the same qualitative influence for either case of quasilinearity. These results continue to hold when there is bunching at the bottom due to a binding non-negativity constraint. We also illustrate how relaxing the assumption of constant elasticity of labor supply, generally used in the literature, modifies the results.

Key Words: maxi-min, optimal income taxation

JEL Classification: H21

1 Introduction

The theory of the optimal income tax structure is a technically complex one that yields few clear-cut analytical results. The most commonly cited ones involve properties of the marginal tax rate at the top or bottom of the skill distribution, which are of limited interest. Even using the relatively simple model devised by Mirrlees (1971), one has to rely on simulations to obtain typical tax schedules (Tuomala, 1990; Saez, 2001). That is because the optimal tax structure is the result of several sorts of influences interacting with

*We would like to thank Jan Boone, Mathias Hungerbühler, Etienne Lehmann and Benteng Zou for their extremely valuable comments and suggestions.

[†]Department of Economics, Queen's University, Kingston, Canada. Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

[‡]Belgian National Scientific Fund (FNRS), Department of economics, Université Catholique de Louvain. Financial Support from FNRS and Fonds Spécial de la Recherche which has made it possible for Laurence Jacquet to visit Queen's University is gratefully acknowledged. This research is part of a programme supported by the Belgian government (pôles d'attraction interuniversitaires PAI P5/21).

one another. One is the responsiveness of labor supply to redistributive taxes. Another is the social welfare function that determines the equity effects of redistribution. A third is the shape of the skill distribution (Diamond, 1998; Saez, 2001). And, the manner in which these influences interact is affected by the incentive constraints that limit the feasibility of redistribution. This prevents one from analytically deriving the entire optimal tax schedule.

In this paper, we exploit the fact that one of these influences can be virtually suppressed if we take an extreme form of the social welfare function, the maxi-min social welfare function emphasized by Rawls (1971).[1] In the context of optimal income tax theory, the maxi-min social welfare function becomes simply the utility of the least-skilled individual, since the incentive constraints ensure that utility is non-decreasing in skills. This effectively shuts down one of the three influences mentioned above, and we are left with two main sources of influence on the optimal tax structure: the variability of labor and the shape of the skill distribution. This turns out to simplify the optimal tax problem considerably, and in many cases to yield explicit solutions and characterizations of the optimal tax structure. The structure of the optimal income tax function can take interesting forms that contrast both with those observed in the real world, and those obtained for social welfare functions with any non-negative degree of aversion to inequality, especially Diamond (1998).

The analysis of the optimal income tax under a maxi-min objective function is not new (Atkinson, 1975; Phelps, 1973; Kanbur and Tuomala, 1994; Piketty, 1997; d'Autume, 2001; Salanié, 2003). Our approach can be seen as synthesizing the maxi-min approach and extending it in some interesting directions. When preferences are additive in consumption and leisure, marginal tax rates are decreasing over the entire distribution of skills under reasonable assumptions, unlike in the case with general social welfare functions. Moreover, the tax function itself is strictly concave with income and average tax rates are generally single-peaked. We also show in particular that when quasilinear utility functions apply in either consumption or leisure, so that income effects are suppressed, the optimal tax analysis and the results obtained from it turn out to be quite simple. In these cases, the distribution of skills is a key determinant of the tax schedule, and has the same qualitative influence for either case of quasilinearity. The shape of the tax structure can depend upon whether the skill distribution is bounded at the top.

[1] Following Okun's (1975) interpretation of Rawls, when all the weight is placed on the most needy, the optimal tax literature uses the term 'Rawlsian objective'. Obviously, this interpretation does not adequately reflect the richness of the *Theory of Justice* of Rawls (1971), but it is the common practice in economics.

In the following section, we outline the basic features of the model. Section 3 considers the properties of the optimal tax function with general additive preferences. We present the optimal tax formula in terms of the elasticity of labor supply, the pattern of marginal utilities of consumption, and the shape of the skill distribution. We derive sufficient conditions for decreasing marginal tax rates in skill and a single-peaked average tax profile in income. We pay particular attention to the validity of the so-called first-order approach and derive sufficient conditions for this to be valid with general additive preferences. In Section 4, we then turn to quasilinear-in-leisure preferences, which have an obvious pedagogical value since these preferences ease the solution considerably. Qualitative results for some specific skill distributions are derived. Section 5 studies the tax structure with quasilinear-in-consumption preferences. When the elasticity of labor supply is constant, the optimal tax patterns are similar for both types of quasilinear preferences. We also allow for a variable elasticity of labor supply using a logarithmic utility of leisure, as in Diamond (1998). In our analysis, bunching can only occur at the bottom when the non-negativity constraint that must be imposed on the household income is binding. Section 6 then analyzes the consequences for the tax structure of this binding constraint. The final section summarizes.

2 The model

Assume that all households have the same additively separable utility function:

$$u(x, \ell) = v(x) - h(\ell) \tag{1}$$

where x is consumption and ℓ is labor (so $1 - \ell$ is leisure), with $v' > 0 \geq v''$, $h' > 0$ and $h'' \geq 0$, with either $v'' < 0$ or $h'' > 0$. Such preferences satisfy the single-crossing or monotonicity property that is important for optimal income tax analysis. Households differ only in skills, which correspond with their wage rates given that aggregate production is linear in labor. These wage rates w are distributed according to the distribution function $F(w)$ for $w \in W = [\underline{w}, \overline{w}]$, where $0 < \underline{w} < \overline{w} \leq \infty$. The corresponding density function, $f(w) = F'(w)$, is assumed to be differentiable and strictly positive for all $w \in W$. It is assumed that the distribution is single-peaked, with a mode at w_m.

Households obtain all their income from wages, and before-tax income is denoted by $y \equiv w\ell$. Let $x(w)$, $\ell(w)$ and $y\,(w)$ be consumption, labor supply and income for a household with skill w. The government can observe incomes but not wage rates or amounts of labor supplied, so it bases its tax scheme on total income. Then, the budget constraint for this household is:

$$x(w) = y(w) - T(y(w)) \tag{2}$$

where $T(y(w))$ symbolizes the income tax imposed on type-w households. The household maximizes (1) subject to (2), yielding the first-order condition:

$$\frac{h'(\ell(w))}{wv'(x(w))} = 1 - T'(y(w)) \tag{3}$$

If we use the definition of income to rewrite the utility function as $v(x) - h(y/w)$, the lefthand side can be interpreted as the marginal rate of substitution between income y and consumption x. This is chosen to equal one minus the marginal income tax rate. For later use, it is convenient to note that, from (3), the compensated elasticity of labor supply, $e^c(w_n)$, satisfies

$$e^c(w_n) = \frac{w_n}{\ell}\frac{\partial \ell}{\partial w_n}\bigg|_{\overline{u}} = \frac{h'(\ell)}{(h''(\ell) - w_n^2 v''(x))\,\ell} > 0 \tag{4}$$

and the uncompensated elasticity of labor supply, $e^u(w_n)$ satisfies

$$e^u(w_n) = \frac{w_n}{\ell}\frac{\partial \ell}{\partial w_n} = \frac{h'(\ell) + v''(x)w_n^2\ell}{(h''(\ell) - w_n^2 v''(x))\ell} \tag{5}$$

where $w_n \equiv w(1 - T'(y(w)))$ is the after-tax wage rate.

Throughout the paper, we assume that the objective function of the government is maxi-min, so the government is only concerned about the welfare of the least well-off households. Given the information assumptions we are making, the worst-off will be those at the bottom of the skill distribution whose wage is $\underline{w}$. The determination of the optimal income tax structure can be formulated as a mechanism design problem. The government chooses the tax schedule $T(y(w))$ or, equivalently, the consumption-income bundle intended for each household $\{(x(w), y(w), w \in W\}$, to maximize the welfare of the least well-off households, subject to three sorts of constraints.

The first is the government budget constraint, which takes the form:

$$\int_{\underline{w}}^{\overline{w}} [y(w) - x(w)]f(w)dw \geq R \tag{6}$$

where R is an exogenous revenue requirement. This constraint must be binding at the optimum since u is increasing in x.

The second is the set of incentive-compatibility constraints. These require that a household of type w choose the consumption-income bundle intended for it, that is,

$$v(x(w)) - h\left(\frac{y(w)}{w}\right) \geq v(x(w')) - h\left(\frac{y(w')}{w}\right) \quad \forall w, w' \in W \tag{7}$$

Ensuring these self-selection constraints is equivalent to reformulating (7) as a minimization problem as follows. Define the value function of utility for a type-w household by:

$$u(w) \equiv v\left(x\left(w\right)\right) - h\left(\frac{y(w)}{w}\right) \tag{8}$$

For a type w' household, (7) and (8) give:

$$u(w') = v\left(x\left(w'\right)\right) - h\left(\frac{y(w')}{w'}\right) \geq v\left(x\left(w\right)\right) - h\left(\frac{y(w)}{w'}\right)$$

Therefore, combining the last two expressions for w and w':

$$0 = u(w) - v\left(x\left(w\right)\right) + h\left(\frac{y(w)}{w}\right) \leq u(w') - v\left(x\left(w\right)\right) + h\left(\frac{y(w)}{w'}\right)$$

This implies that w' minimizes $u(w') - v\left(x\left(w\right)\right) + h\left(y(w)/w'\right)$ at $w = w'$. Evaluating the first-order condition at $w = w'$, we obtain

$$\dot{u}(w) = h'(.)\frac{y(w)}{w^2} = h'(.)\frac{\ell(w)}{w} \qquad \forall w \tag{9}$$

This is the set of first-order incentive-compatibility (FOIC) conditions. The first-order approach uses only these FOIC conditions. Since these are only necessary conditions, their solution may not indicate a minimum. It may be necessary to adopt a second-order approach and include the second-order conditions in the government's problem.

The second-order conditions for the incentive compatibility (SOIC) constraint to be satisfied is found using the second derivative of $u(w') - u(x(w), y(w)/w')$ with respect to w' evaluated at $w = w'$:

$$\ddot{u}(w) + \frac{2h'(.)y(w)}{w^3} \geq 0 \qquad \forall w \tag{10}$$

These are the SOIC conditions, and can be rewritten as follows. Differentiate the FOIC conditions (9) with respect to w.

$$\ddot{u}(w) + \frac{2h'(.)y(w)}{w^3} - \frac{h'(.)\dot{y}(w)}{w^2} - \frac{h''(.)y(w)}{w^3}\dot{y}(w) = 0$$

which gives, by using (10)

$$\left(\frac{h'(.)}{w^2} + \frac{h''(.)y(w)}{w^3}\right)\dot{y}(w) \geq 0$$

which implies $\dot{y}(w) \geq 0$. Moreover differentiating (8) with respect to w and using (9), we obtain:

$$v'(.)\dot{x}(w) - h'(.)\frac{\dot{y}(w)}{w} = 0$$

Therefore, $\dot{y}(w) \geq 0$ or $\dot{x}(w) \geq 0$ are equivalent ways to write the SOIC conditions. Note that if the SOIC constraints are slack ($\dot{y}(w) > 0$), the first-order approach is appropriate. Where they are binding, we have $\dot{x}(w) = \dot{y}(w) = 0$, so there is bunching of households of different skills.

The third constraint (highlighted in Boadway, Cuff and Marchand, 2000) requires that labor supply and therefore before-tax income be non-negative ($y(w) \geq 0, w \in W$).[2] For now, we neglect the non-negativity constraint on income and assume it to be satisfied. We investigate the consequences of this constraint being binding in Section 6.

The solution to the government's maxi-min problem will give the highest level of $u(\underline{w})$ that can be achieved given R and the incentive constraints, where $u(\underline{w}) = v(x(\underline{w})) - h(\ell(\underline{w}))$. Let the solution to this problem be denoted $\underline{u}$. This solution can also be obtained from an equivalent revenue-maximizing problem which can be formulated as follows. Take $\underline{u}$ as given and consider tax profiles that will generate this level of utility for the worst-off households, given the incentive conditions. It is apparent that $\underline{u}$ can be supported by a large number of tax profiles such that tax revenues are no greater than R, that is,

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

As long as the incentive constraints are satisfied for all w, we know from the above problem that generating the utility level $\underline{u}$ requires that the tax revenue generated cannot exceed R, so the above inequality must be satisfied. In fact, if we maximize the amount of tax revenue that will yield utility $\underline{u}$ for the worst-off households, that level of revenue will be precisely R. Therefore, maximizing tax revenue subject to $u(\underline{w}) \geqslant \underline{u}$ and the incentive conditions is an equivalent problem to the one of maximizing $u(\underline{w})$ subject to tax revenues being at least equal to R and the incentive conditions. The purpose of drawing attention to these two equivalent approaches to solving the maxi-min optimal income tax problem is to relate our approach to the revenue-maximizing approach used by other authors (e.g., d'Autume, 2001; Piketty, 1997).

3 The general additive case

We proceed by first studying the optimal income tax problem for the case where household utilities take the general additive form. Unambiguous qualitative results in this

[2] A maximum labor supply constraint could also be required, although we ignore this requirement in this paper. By the same token, we should impose the requirement that consumption be non-negative ($x(w) \geq 0$). However, given our maxi-min objective and the fact that $x(w)$ is increasing in w, this constraint will not be binding in our analysis.

case are somewhat limited, though the sources of influences at work can be identified. In the next section, we turn to two quasilinear forms of (1) that have been prominent in the literature—quasilinear-in-leisure and quasilinear-in-consumption—where somewhat sharper results are obtained.

For expositional purposes, we follow the first-order approach and ignore the SOIC conditions. Later we can verify whether the latter are satisfied for the maxi-min case. In the government's maxi-min problem, to deal with the fact that the objective function is simply $u(\underline{w})$, we transform the objective function to make it amenable to the control theory approach. Let $I(w)$ be an indicator function such that $I(w) = 1$ if $w = \underline{w}$, and $I(w) = 0$ otherwise. Furthermore, from (1), we have that $u(w) = v(x(w)) - h(\ell(w))$, which implies that $u(w)$, $x(w)$, and $\ell(w)$ are not independent. We can therefore treat $x(w)$ as a function of the other two and write it as $x(w) = X(u(w), \ell(w))$, where $X(.)$ is increasing in both arguments. In particular, by differentiating the utility function, we obtain:

$$\frac{\partial X(u(w),\ell(w))}{\partial \ell} = \frac{h'(\ell(w))}{v'(x(w))}, \qquad \frac{\partial X(u(w),\ell(w))}{\partial u} = \frac{1}{v'(x(w))}$$

This allows us to suppress $x(w)$ from the government problem.

The government solves for the feasible tax structure that maximizes the utility of the least well-off households subject to the budget constraint (6) and the FOIC conditions (9):

$$\text{Max}_{u(w),\ell(w)} \quad \int_{\underline{w}}^{\overline{w}} u(w) I(w)\, dw \tag{11}$$

$$\text{s.t.} \quad \int_{\underline{w}}^{\overline{w}} [w\ell(w) - X(u(w),\ell(w))] f(w) dw = R \quad \text{and} \quad \dot{u}(w) = \frac{-\ell(w)h'(\ell(w))}{w}$$

The equivalent problem consists in maximizing tax revenue subject to a boundary condition on the utility level of the least-well off and to the FOIC constraints (9):

$$\text{Max}_{u(w),\ell(w)} \quad \int_{\underline{w}}^{\overline{w}} [w\ell(w) - X(u(w),\ell(w))] f(w) dw \tag{12}$$

$$\text{s.t.} \quad u(\underline{w}) = \underline{u} \quad \text{and} \quad \dot{u}(w) = \frac{\ell(w)h'(\ell(w))}{w}$$

For both problems, Appendix 1 describes the variational techniques used and derives the necessary conditions for a maximum.

The first-order conditions characterizing the optimum nonlinear marginal tax in each case can be written as:

$$\frac{T'(y(w))}{1-T'(y(w))} = \left[1 + \frac{h''(\ell(w))\ell(w)}{h'(\ell(w))}\right] \frac{1}{wf(w)} v'(x(w)) \int_{w}^{\overline{w}} \frac{f(t)}{v'(x(t))} dt \quad \forall w \in W \tag{13}$$

Note that this formula for optimal marginal tax rates does not depend on either $\underline{u}$ or R explicitly, although of course the levels of tax liabilities do. Formula (13) allows us to obtain some simple properties with a maxi-min criterion and additive preferences.

First, the marginal tax rate at the top is zero ($T'(y(\overline{w})) = 0$) with a bounded skill distribution.[3] The zero marginal tax rate at the top is a standard result with a bounded skill distribution (see Sadka, 1976; Seade, 1977): raising the marginal tax rate at the top above zero is suboptimal because it would distort the labor supply decision of the highest earner but would raise no revenue. With an unbounded distribution, this intuition may not apply since in this case, there is always a household with a higher wage level than the one at which we calculate the optimal marginal tax rate. Even so, the asymptotic tax rate tends to zero if $[1 + \ell h''(\ell)/h'(\ell)]$ and the hazard rate $(1 - F(w))/f(w)$ are constant.[4]

Second, at the bottom of the skill distribution where $w = \underline{w}$, the optimal marginal tax rate is strictly positive.[5] This maxi-min result is in sharp contrast with the standard result of zero marginal tax rate at the bottom when using as an objective a social welfare function with non-negative aversion to inequality and assuming no bunching at the bottom (Seade, 1977; Myles, 1995). To understand this contrasting result, recall the equity-efficiency tradeoff from which the optimal tax schedule results. A rise in $T'(.)$ at any skill level $\widetilde{w}$ distorts the labor supply of those with skill $\widetilde{w}$, implying an efficiency loss. However, an increase of $T'(y(\widetilde{w}))$ increases the tax receipts from those with $w > \widetilde{w}$, implying more redistribution towards those with skills $w \leq \widetilde{w}$. As long as the former contribute less to the welfare criterion than the latter, such transfers are positively valued hence, an equity

[3]Proof: Multiplying and dividing (13) by $(1 - F(w))$, we obtain:

$$\frac{T'(y(w))}{1 - T'(y(w))} = \left[1 + \frac{h''(\ell(w))\ell}{h'(\ell(w))}\right] \frac{1 - F(w)}{wf(w)} \frac{v'(x(w))}{1 - F(w)} \int_w^{\overline{w}} \frac{f(t)}{v'(x(t))} dt \quad \forall w \in W$$

The term $\left[v'(x(w)) \int_w^{\overline{w}} f(t)/v'(x(t))\, dt\right] / [1 - F(w)]$ is asymptotically finite, as can be shown using l'Hôpital's rule:

$$\lim_{w \to \overline{w}} \frac{v'(x(w))}{1 - F(w)} \int_w^{\overline{w}} \frac{f(t)}{v'(x(t))} dt = \lim_{w \to \overline{w}} v'(x(w)) \lim_{w \to \overline{w}} \frac{\int_w^{\overline{w}} f(t)/v'(x(t))\, dt}{1 - F(w)} \to 1$$

At the top, we have $[1 - F(\overline{w})]/[\overline{w} f(\overline{w})] = 0$ (which is obvious if $f(\overline{w}) \neq 0$ and remains valid if $f(\overline{w}) = 0$, as can be shown using l'Hôpital's rule). Therefore, since $[1 + \ell h''(\ell)/h'(\ell)]$ is always positive, the asymptotic marginal tax rate is zero with bounded skill distributions.

[4]To see this, let us denote H the constant hazard rate $(1 - F(w))/f(w)$. Therefore $\lim_{\overline{w} \to \infty} \frac{1 - F(w)}{wf(w)} = \lim_{\overline{w} \to \infty} H/\overline{w} \to 0$. From the first equation in the previous footnote and the asymptotically finite term $\left[v'(x(w)) \int_w^{\overline{w}} f(t)/v'(x(t))\, dt\right] / [1 - F(w)]$, T' tends to zero when $[1 + \ell h''(\ell)/h'(\ell)]$ and the hazard rate are constant.

[5]There will still be a positive tax rate at the bottom if there is bunching, either due to an interval of low-skilled persons who are not working (Seade, 1977) or due to the violation of SOIC constraints (Guesnerie and Laffont, 1984 and Ebert, 1992).

gain appears. Suppose $\widetilde{w} = \underline{w}$. With a Bergson-Samuelson social welfare function, the equity effect vanishes since there is an infinitesimally small number of persons at $\underline{w}$ (and no one below).[6] However, with a maxi-min criterion, everyone in the objective function is at $w = \underline{w}$, so the equity effect is positive. Therefore, $T'(y(\underline{w})) > 0$ at the optimum.[7]

Finally, the righthand side is always nonnegative, so $T'(y(w)) \geq 0$. To characterize the pattern of marginal tax rates for skills in the interior, it is useful to understand the intuition behind the multiplicative components of (13).

The first component, $[1 + \ell h''(\ell)/h'(\ell)]$, is a measure of the elasticity of labor supply and as such reflects an *efficiency effect.* Using (4) and (5), this factor is equivalent to $[1 + e^u(w_n)]/e^c(w_n)$. A priori, it is not clear how this term varies with skills. Following Diamond (1998), we shall often assume that this term is constant.

The second term, $1/[wf(w)]$, can be called the *density effect.* It indicates that the optimal marginal tax rate is lower the higher are aggregate skills $wf(w)$ at skill level w. The loss of revenue from increasing the marginal tax rate is higher, the higher is $wf(w)$. With single-peaked skill distributions, this term always decreases below the mode of the distribution. Above the mode, it will either increase or decrease depending on how rapidly $f(w)$ falls with w.

The final term in (13), $v'(x(w)) \int_w^{\overline{w}} f(t)/v'(x(t))\, dt$, is the product of the marginal utility of consumption and the gain in tax revenue from decreasing the utility of everyone above w by one unit. Let us call it the *global revenue effect.* The intuition of this effect is as follows. Suppose we reduce the utility of everyone above w by a marginal unit (so that the FOIC constraints are still satisfied in that range). The gain in increased revenue is $1/v'(x(w))$ per person, while there is no loss of social welfare since the utility of those with $w > \underline{w}$ does not count with a maxi-min criterion. Thus, the integral represents the net effect of a marginal reduction in u above w, and depends on the number of people above w. The entire term $v'(x(w)) \int_w^{\overline{w}} f(t)/v'(x(t))\, dt$ is declining with w since we assume that the $v(x)$ function is concave and since the integral term decreases in w.

From the above, we can infer that when $h(\ell)$ takes the constant elasticity form so the first term in (13) is constant, the marginal tax rate always decreases in w below the

[6] If there is bunching at the bottom, there will be a mass of people there and the equity effect will be positive.

[7] Note that as the aversion to inequality approaches ∞, the social welfare function approaches maxi-min (e.g. $\lim_{\nu\to\infty} \int_{\underline{w}}^{\overline{w}} \frac{u(w)^{1-\nu}}{1-\nu} f(w)dw \to \min_{w} \{u(w)\} f(w)$ pointwise). Taking the limit of the solution to the standard optimization problem as $\nu \to \infty$ results in $T'(y(\underline{w})) = 0$ in the limit. Thus, the maxi-min solution is not the same as the limit of the solution to the problem with a Bergson-Samuelson social welfare function.

mode ($\dot{T}' < 0$) with single-peaked skill distributions. After the mode of w, the shape of the optimal tax profile will depend on the relative influence of the density effect, which is ambiguous, and the global revenue effect, which is decreasing. A sufficient, though not necessary, condition for decreasing T' over the entire skill distribution is that aggregate skills $wf(w)$ are non-decreasing beyond the mode.

So far, our analysis has been conducted as if the first-order approach were valid. This will be the case as long as (13) yields a solution for $T'(y(w))$ such that $x(w)$ (or $y(w)$) is everywhere increasing in w. To derive a sufficient condition for the optimal marginal tax profile to satisfy the SOIC conditions, rewrite (3) as

$$\frac{v'(x(w))}{h'\left(y(w)/w\right)} = \frac{1}{w(1 - T'(y(w)))} \tag{14}$$

Given that $v' < 0$, $h'' \geq 0$, the lefthand side of (14) will be decreasing if and only if $\dot{x}(w) > 0$ (and therefore, $\dot{y}(w) > 0$). Moreover, the righthand side will be decreasing in w if and only if $w(1 - T'(y(w))$ is increasing in w. Therefore, the latter is a necessary and sufficient condition for the SOIC conditions to be satisfied ($\dot{x}(w) > 0$). Thus, a non-increasing marginal tax rate ($\dot{T}'(y(w)) \leq 0$) is sufficient for the SOIC conditions to be satisfied. An implication of this is that since T' decreases in w below the mode when $h(\ell)$ is constant elasticity, the SOIC conditions will be satisfied there and there will be no bunching at the bottom on that account.

So far, and as is typical in the literature, we have focused on how the marginal tax rate changes with the wage rate w, and not with income y. Since the marginal tax rate is always non-negative, $T'(y(w)) \geq 0$, total tax liabilities rise with y. To determine how the marginal tax rate varies with income, note that

$$\frac{dT'(y(w))}{dw} = T''(y(w))\dot{y}(w) \tag{15}$$

where $T''(y(w))$ is the second derivative of the total tax function $T(y)$ with respect to y. As long as the SOIC condition $\dot{y}(w) > 0$ is satisfied for any w, $T''\left(y\left(w\right)\right)$ takes the same sign as $dT'(y(w))/dw$. Therefore, where $T'\left(y\left(w\right)\right)$ declines in w, the optimal tax function $T(y(w))$ is increasing and strictly concave in y.

This last result has implications for the pattern of average tax rates, a dimension that has drawn little attention. Assume that in the maxi-min optimum, the least well-off receive a transfer, that is, $T(y(\underline{w})) < 0$. (This requires the government revenue requirement not too large.) Therefore, the average tax rate $T(y)/y$ will be single-peaked in income if $T'(y)$ is decreasing and $T'(\overline{y}) \to 0$. This is illustrated in Figure 1, where $T(y)$ is strictly concave and $T'(\overline{y}) \to 0$. The average tax rate (that is, the slope of dashed lines from the origin) first increases up to income y_2, and then decreases. A formal proof is given in Appendix 1.

INSERT FIGURE 1 HERE

To summarize this section:

Proposition 1: Assuming a maxi-min criterion and general additive preferences, the optimal marginal tax profile has the following characteristics:

(i) At the bottom of the skill distribution, the optimal marginal tax rate is strictly positive (even without bunching).

(ii) The optimal marginal tax rate at the top is zero with a bounded distribution. With unbounded distributions, when the hazard rate and the elasticity of labor are constant, the marginal asymptotic tax rate equals zero.

(iii) If the elasticity of labor is constant, the optimal marginal tax rate is decreasing below the modal skill with a single-peaked skill distribution.

(iv) If the elasticity of labor is constant, aggregate skills $wf(w)$ non-decreasing with the productivity level is a sufficient condition for a decreasing $T'(.)$ over the entire skill distribution.

(v) A non-increasing $T'(.)$ is sufficient for satisfying the SOIC conditions. Below the mode, the SOIC conditions are always satisfied with a constant elasticity of labor supply.

(vi) $T'(.)$ decreasing in skill is necessary and sufficient for $T(y)$ to be increasing and strictly concave in income. The average tax rate will then be single-peaked in income if $T(\underline{y}) < 0$ and $T'(\overline{y}) \to 0$.

Note for future reference that (13) can be written in a way that is directly analogous to Diamond (1998)'s formula for optimal marginal tax rates under a general social welfare function:

$$\frac{T'(y(w))}{1 - T'(y(w))} = A(w)B(w)C(w) \qquad \forall w \in W \tag{16}$$

where

$$A(w) = \left[1 + \frac{h''(\ell(w))\ell}{h'(\ell(w))}\right], \quad B(w) = \frac{v'(x(w))}{1 - F(w)} \int_w^{\overline{w}} \frac{f(t)}{v'(x(t))} dt \quad \text{and} \quad C(w) = \frac{1 - F(w)}{wf(w)}$$

This way of writing the factors will be useful in the next sections.

4 Quasilinear-in-leisure preferences

In this section, we assume that preferences are quasilinear in leisure, so $h(\ell) = \ell$ and $v(x)$ is increasing and strictly concave. This setting is the same as Lollivier and Rochet (1983), Weymark (1986a, 1986b, 1987), Ebert (1992) and Boadway, Cuff and Marchand (2000). Utility can then be written $u(x, y/w) = v(x) - y/w$. Such preferences are characterized by the absence of income effects in the choice of consumption. Furthermore, $A(w) = 1 + \ell h''(\ell)/h'(\ell) = 1$ in this case, so another of the three influences on the optimal tax structure is shut down. We begins by deriving general properties for the optimal tax structure, and then consider the tax structure for specific skill distributions.

4.1 The optimal tax structure

The optimal tax structure can be obtained as a special case of the general additive one. The optimal tax formula (13) becomes:

$$\frac{T'(y(w))}{1 - T'(y(w))} = \frac{v'(x(w))}{wf(w)} \int_w^{\overline{w}} \frac{f(t)}{v'(x(t))} dt \tag{17}$$

This can be simplified considerably as follows. As shown in Appendix 2, from the first-order conditions for the optimal tax problem, we obtain:

$$\int_w^{\overline{w}} \frac{f(t)}{v'(x(t))} dt = w(1 - F(w)) \tag{18}$$

Then substituting (18), the first-order condition for the household (3) and $h'(\ell) = 1$ into (17), and using the above definition of $C(w)$, we get:[8]

$$T'(y(w)) = \frac{1 - F(w)}{wf(w)} = C(w) \tag{19}$$

Thus, the structure of optimal marginal tax rates depends only on the properties of the distribution function for skills as in Boadway, Cuff and Marchand (2000).

Since, as we showed above, a non-increasing $T'(y(w))$ with w is a sufficient condition for satisfying the SOIC conditions, $\dot{C}(w) \leq 0$ implies that the SOIC conditions are satisfied. Therefore, the properties of the skill distribution may be sufficient to determine if the SOIC conditions are satisfied.

With quasilinear-in-leisure preferences, it is useful for pedagogical purposes to note that the optimal tax schedule can be derived from a simple Lagrangian optimization problem.

[8] With quasilinear-in-leisure preferences, $C(w)$ is assumed to be below 1 so there is no bunching at the top due to a violation of the SOIC conditions.

This is shown in Appendix 2 where the following first-order conditions are derived:

$$\lambda = \frac{1}{\underline{w}} \tag{20}$$

and

$$v'(x(w)) = \frac{f(w)}{wf(w) + F(w) - 1} \tag{21}$$

where λ is the multiplier associated with the government budget constraint. The first condition indicates that the marginal cost of public funds depends only on the level of skill of the lowest skilled person. Intuitively, with quasilinear-in-leisure preferences, an increment in government revenue requirements involves all households supplying an additional unit of labor, that is, $dy(w)/dR = 1 \ \forall w \in W$. This is shown in Appendix 2. The utility cost to this person—who is the only one whose utility counts—from supplying an additional unit of income is simply $1/\underline{w}$. The second condition, (21), yields the optimal tax rate, (19), after substituting in the first-order conditions for the household, (3).

4.2 Results for specific distribution functions

Most simulations use a log-normal distribution which matches roughly the single-moded empirical income distribution (Aitchison and Brown, 1957) but has also an unrealistic thin upper tail. It has been argued that the Pareto distribution fits the empirical income distribution at high income levels reasonably well (Feenberg and Poterba, 1993 and Saez, 2001[9]). However, these simulations rely on the assumption that the skill distribution takes the same shape as the income distribution which is a strong an assumption. Therefore, we use skill distributions which encompass the traditionally used ones and the Weibull distribution which is flexible enough to match a wide variety of decreasing and single-peaked skill distributions and which also allows for a thicker upper tail than the log-normal. Furthermore, it seems realistic to assume these distributions have an upper bound. The distributions we then consider include the Pareto, the Weibull and the log-normal, both in their truncated and untruncated forms.

Since $C(w)$ is always declining below the modal skill level w_m, so are marginal tax rates. We therefore focus especially on how the various distributions apply above w_m.

[9] In Saez (2001), the skill distribution used is calibrated such that given the chosen utility functions and a flat tax (reproducing approximately the real US tax schedule), the resulting distribution replicates the empirical earnings distribution in the U.S. This empirical earnings distribution seems remarkably well approximated at the upper end by a Pareto earnings distribution. (Pareto already noted this almost one century ago.) As the tax profile is roughly linear in the U.S., a Pareto distribution is then also a good approximation for skills at the upper end.

For those distributions whose densities are decreasing everywhere (e.g. the Pareto), we suppose the distribution applies above w_m.

Pareto distribution

Some well-known results depend on the distribution being Pareto and untruncated (Diamond, 1998 and Salanié, 2003) above the modal skill level. The density $f(w)$ of an untruncated Pareto is proportional to $1/w^{1+a}$ for $a > 0$. Therefore, from (19), $C(w) = 1/a$,[10] so the marginal tax rate is *constant* (and strictly positive) where the Pareto distribution applies. Thus, the optimal income tax over the entire skill distribution has a so-called *hockey-stick profile* in skills: it has a constant marginal tax rate beyond the mode (analogous to a linear progressive tax) where the Pareto distribution applies, and a monotonically declining marginal tax rate before that.[11]

Intuitively, this result is due to the wide tail of the Pareto distribution which gives an incentive to have high marginal tax rates at high levels of productivity since these raise a lot of additional revenue. The top rate depends negatively of the thinness of the top tail distribution. (With Pareto distributions, the higher the Pareto parameter a, the thinner is the tail of the distribution.)

Since $T'(.)$ decreases below w_m and $T'(.)$ is constant where the Pareto distribution applies, we infer from (15) that the tax function is concave in income. It is strictly concave below the income corresponding to w_m and linear above that. From Proposition 1(v) and $T'(.)$ non-increasing, the SOIC conditions are satisfied. Therefore, there is no bunching on this account.

Since the marginal tax rate as $w \to \overline{w}$ is larger than zero, the sufficient conditions of Proposition 1(vi) are not satisfied, so we cannot be sure that the average tax rate is single-peaked: it may be increasing throughout. However, in the (unrealistic) case where the whole distribution of skill is an untruncated Pareto, the tax profile is linear over the whole distribution of y. In Figure 1, if the tax function were increasing and linear in y, it can easily be seen that the average tax rate is everywhere increasing if $T(\underline{y}) < 0$ (cf. Hindriks, Lehmann and Parmentier, 2006).

A truncated Pareto distribution ($\overline{w} < \infty$) applying beyond the mode seems more

[10] With quasilinear-in-leisure preferences, $a \geq 1$ is assumed to ensure that marginal tax rates lie below 1.

[11] These global results assume that the Pareto distribution applies for all skill levels beyond the mode. It may be more realistic to suppose that it applies only begining at some skill level $\widehat{w}$ beyond the mode, as in Diamond (1998). It is possible that in the range between the mode and $\widehat{w}$ the marginal tax rate not be declining.

realistic. The density is then given by $f(w) = a\underline{w}^a w^{-a-1}[1-(\underline{w}/\overline{w})^a]^{-1}$, so:

$$C(w) = \frac{1}{a}\left[1 - \left(\frac{w}{\overline{w}}\right)^a\right] \tag{22}$$

where $\dot{C}(w) = -w^{a-1}/\overline{w}^a < 0$. The marginal tax rate therefore now (monotonically) decreases with w both before and beyond the mode. In addition, $T'(y(w))$ is strictly concave in w when $a > 1$ and strictly convex when $a < 1$. The SOIC conditions are again always satisfied. Moreover, the marginal tax rate is zero at $w = \overline{w} < \infty$, which confirms the standard result that the optimal marginal tax rate is zero at the top with a bounded skill distribution. Therefore, since the tax function is strictly concave in income, (15) implies that the optimal average tax rate has a single-peaked pattern if $T(\underline{y}) < 0$ (Proposition 1(vi)).

Weibull and log-normal distributions

With the untruncated Weibull distribution, the density is

$$f(w) = \frac{\beta}{\eta}\left(\frac{w}{\eta}\right)^{\beta-1} e^{-(w/\eta)^\beta}$$

where $\beta > 0$ is the shape parameter and $\eta > 0$ the scale parameter.[12]With an untruncated log-normal distribution, the density is

$$f(w) = e^{-(\ln w-\mu)^2/(2\sigma^2)}/(\sigma w\sqrt{2\pi})$$

where μ is the mean and σ^2 the variance. With an untruncated Weibull distribution, $\dot{C}(w) < 0$, while with a log-normal distribution, $\dot{C}(w) < 0$ if:[13]

$$e^{(\ln w-\mu)^2/(2\sigma^2)}\left(1 - erf\left(\frac{\ln w - \mu}{\sigma\sqrt{2}}\right)\right)(\ln w - \mu) < 0.8\sigma \tag{23}$$

Therefore, $T'(y(w))$ is necessarily decreasing over the entire Weibull and log-normal distributions assuming (23) is satisfied. From Proposition 1(v), this also implies that the SOIC

[12]Varying the shape β of a Weibull density drastically modifies the density profile. For $0 < \beta < 1$, $f(w)$ decreases and is convex as w increases. For $\beta = 1$, it becomes the exponential distribution, as a special case. For $\beta > 1$, the density assumes wear-out type shapes and $f(w)$ decreases after the mode. For $\beta = 2$, it becomes the Rayleigh distribution as a special case. For $\beta < 2.6$, $f(w)$ is positively skewed (has a right tail). For $2.6 < \beta < 3.7$, its coefficient of skewness approaches zero (no tail), consequently, it may approximate the normal density. And for $\beta > 3.7$, it is negatively skewed (left tail).

[13]With a Weibull distribution, we have $C(w) = [1/\beta(w/\eta)^\beta]$, $\dot{C}(w) = -w^{-\beta-1}\eta^\beta < 0$ and $\ddot{C}(w) = (\beta+1)w^{-\beta-2}\eta^\beta > 0$. Therefore, $C(w)$ is decreasing and convex in w. With a log-normal, we have

$$C(w) = (\sigma\sqrt{2\pi}/2)\left[1 - erf(E(w))\right] e^{(\ln w-\mu)^2/(2\sigma^2)}$$

where $erf(.)$ is the error function (encountered in integrating the normal distribution and defined by, $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$) with $derf(E(w))/dw > 0$ and $erf(.) \in [-1, 1]$. In this case, the sufficient condition (23) for $\dot{C}(w) < 0$ is derived.

conditions are always satisfied so there can be no bunching on that account. As well, $T'(y(w))$ tends to zero as w increases, so the income tax function is strictly concave in income. The sufficient conditions of Proposition 1 (vi) being satisfied, the optimal average tax rate has a single-peaked pattern.

Suppose now that the Weibull and log-normal distributions are truncated at $\overline{w} < \infty$. We derive $\dot{C}(w) < 0$ with a truncated Weibull distribution (with a mode larger than the scale parameter which is usually standardized to 1) and with a truncated log-normal distribution if:[14]

$$e^{(\ln w - \mu)^2/(2\sigma^2)} \left(g\left(\overline{w}\right) - g(w)\right)(\ln w - \mu) < 0.8\sigma \tag{24}$$

The marginal tax rate is then decreasing, and tends to zero at the top of the skill distribution. Therefore, the SOIC conditions are satisfied, the tax function is strictly concave in income and the average tax rate single-peaked (if $T(\underline{y}) < 0$).

We can summarize our results for quasilinear-in-leisure preferences and distinct skill distributions as follows.

Proposition 2: With a maxi-min criterion and quasilinear-in-leisure preferences:

(i) The tax pattern depends exclusively on the distribution of skills through $C(w)$.

(ii) The optimal marginal tax profile is decreasing in skill with truncated Pareto, with untruncated and, if w_m is larger than the scale density parameter, with truncated Weibull, and if conditions (23) and (24) apply, with untruncated and truncated log-normal skill distributions.

[14]With a truncated Weibull, the density becomes $f(w) = (\beta/\eta)(w/\eta)^{\beta-1}e^{-(w/\eta)^\beta}/(1-e^{-\overline{w}/\beta})$. So

$$C(w) = \frac{-e^{-(\overline{w}/\eta)^\beta} + e^{(-w/\eta)^\beta}}{\beta(\frac{w}{\eta})^\beta e^{(-w/\eta)^\beta}}$$

Differentiating with respect to w, we obtain:

$$\dot{C}(w) = -\frac{1}{w}\left[1 + \left(\left(\frac{w}{\eta}\right)^\beta - 1\right)(1-e)^{(w/\eta)^\beta - (\overline{w}/\eta)^\beta}\right]$$

which is < 0 when $w \geq \eta$. Therefore, a modal w larger than η is a sufficient condition for $\dot{C}(w) < 0$.

When the log-normal distribution is bounded, the density becomes

$$f(w) = \frac{1}{w\sigma\sqrt{2\pi}}e^{-(\ln w - \mu)^2/(2\sigma^2)}/(1 + \mathrm{erf}(g(\overline{w})))$$

where $g(w) = erf\left((\ln w - \mu)/(\sigma\sqrt{2})\right)$. Hence,

$$C(w) = \frac{g\left(\overline{w}\right) - g(w)}{\left(2/(\sigma\sqrt{2\pi})\right)e^{-(\ln w - \mu)^2/(2\sigma^2)}}$$

In this case, we find that $\dot{C}(w) < 0$ if and only if inequality (24) is satisfied.

(iii) The asymptotic tax rate is zero for all distributions considered except with an untruncated Pareto distribution where it is finite. With this distribution applying beyond w_m, the marginal tax profile has a hockey-stick profile in skill and income.

(iv) The tax profile is increasing and strictly concave in income and the average tax rate is single-peaked in income (if $T(\underline{y}) < 0$) with all the above distributions except the untruncated Pareto. When the whole skill distribution is untruncated Pareto, the average tax rate is increasing.

5 Quasilinear-in-consumption preferences

In this section, we consider preferences that are quasilinear-in-consumption as in Diamond (1998): $v(x) = x$ and $h''(\ell) > 0$.[15] Hence there are no income effects on labor supply, and both the compensated or uncompensated elasticities may be written:

$$e(w) = \frac{h'(\ell)}{\ell h''(\ell)}$$

Since $v'(x(w)) = 1$, the expression for marginal tax rates, (16), becomes:[16]

$$\frac{T'(y(w))}{1 - T'(y(w))} = \left(1 + \frac{\ell h''(\ell)}{h'(\ell)}\right)\left(\frac{1 - F(w)}{wf(w)}\right) = A(w)C(w) \tag{25}$$

Since only a substitution effect now prevails, the larger the labor supply elasticity, the lower $A(w)$, and the lower the optimal marginal tax rates.

Suppose further that the elasticity of labor supply $e(w)$ is constant (e.g., $h(\ell) = \ell^\alpha$), so that $A(w) = \overline{A}$. In this case, $T'(y(w))$ depends only on $C(w)$, that is, on the properties of the distribution function $F(w)$. With quasilinear-in-leisure preferences, we also derived in (19) that $T'(y(w))$ depends only on $C(w)$. The results of the previous section then apply. All the properties we highlighted with distinct distribution functions for the optimal tax profile with quasilinear-in-leisure preferences are also valid with quasilinear-in-consumption preferences and $A(w)$ constant.

Our previous results with Pareto distributions may then be contrasted with Diamond (1998) who uses a Pareto distribution and quasilinear-in-consumption preferences. He finds that with a standard social welfare function that is increasing and concave in all

[15] Actually, Diamond defines $h(1-\ell)$ as the utility of leisure, with $h'(1-\ell) > 0 > h''(1-\ell)$. Therefore, the expressions we derive are not exactly identical to his.

[16] The factor $B(w)$ disappears from (16) because, as it can be readily verified,

$$B(w) = \frac{v'(x(w))}{1 - F(w)} \int_w^{\overline{w}} \frac{f(t)}{v'(x(t))} dt = 1 \text{ when } v' = 1$$

its arguments and assuming an untruncated Pareto skill distribution beyond the mode, the marginal tax rate is U-shaped. The use of a maxi-min objective eliminates the rising part of his U-shaped marginal tax rate structure. As soon as the Pareto distribution is truncated, the difference with Diamond's result becomes even sharper. In this case, the marginal tax rate decreases monotonically with w both before and after the mode.

With quasilinear-in-consumption preferences, it is straightforward to derive the optimal tax schedule in a simple Lagrangian problem. Appendix 3 presents this derivation and shows that the shadow price of public funds is simply unity with quasilinear-in-consumption preferences. A unit increase of revenue requirements R will cause all households, including the least-skilled, to decrease their consumption by one unit. This has a value of unity to the least well-off household.

5.1 An example with non-constant labor supply elasticity

When the elasticity of labor supply—and therefore $A(w)$—is not constant, the problem becomes rather more complicated unless we assume special functional forms. A particularly useful case is where the utility of leisure is logarithmic, a case considered by Diamond (1998) and Dahan and Strawczynski (2000). The maxi-min criterion allows us to characterize the entire optimal tax profile for various skill distributions, rather than only beyond the mode as in Diamond (1998) where a general social welfare function is used.

In this case, the disutility of labor becomes $h(\ell) = -\ln(1-\ell)$ and the utility function is $u(x,\ell) = x + \ln(1-\ell)$. The compensated (and uncompensated) elasticity (4) can be rewritten as:

$$e(w_n) = \frac{1-\ell(w)}{\ell(w)} \tag{26}$$

Moreover, the first-order condition at the household level (3) becomes:

$$1 - T'(y(w)) = \frac{1}{(1-\ell(w))w} \tag{27}$$

In the absence of taxation, labor supply increases with productivity and approaches unity as w approaches $+\infty$. Hence the individual is induced to supply more labor up to a maximum level, $\ell = 1$ in our case. Then, from (26), the elasticity is decreasing in w and tends to zero: $\lim_{w\to\infty} e(w_n) \to 0$. [17]

[17] Empirically, Gruber and Saez (2002) estimates a labor supply elasticity higher at the top of the income distribution. It is plausible that people at the top of the observed income distribution are also the most productive. This then questions the relevance of assuming a logarithmic utility of leisure. However, it allows to contrast and extend analytical results of Diamond (1998) and Dahan and Strawczynski (2000) in interesting directions.

The expression for the pattern of marginal tax rates is obtained using (25):

$$\frac{T'(y(w))}{1-T'(y(w))} = \frac{1}{1-\ell(w)} \frac{1-F(w)}{wf(w)} = A(w)C(w) \tag{28}$$

An alternative expression can be obtained by substituting (27) into (28) to give:

$$\frac{T'(y(w))}{(1-T'(y(w))^2} = \frac{1-F(w)}{f(w)} = wC(w) \tag{29}$$

where the middle term is the inverse of the hazard rate. Using this expression, we shall investigate the pattern of marginal tax rates for a series of different types of skill distributions as before.

Before turning to the characterization of optimal tax structures, recall that a sufficient condition for satisfying the SOIC conditions is a non-increasing marginal tax rate. From (28), $\dot{C}(w) < 0$ is no longer sufficient for $T'(y(w))$ to be decreasing in w. As long as $\ell(w)$ increases with w, $T'(y(w))$ can increase with w. In this case, the SOIC conditions can then be violated and our computations invalidated.

In what follows, we consider the pattern of optimal tax rates when the distribution takes distinct shapes. Before w_m, $wC(w) = (1 - F(w))/f(w)$ will be decreasing as before. Therefore, by (29), $T'(y(w))$ will be declining before w_m for any single-peaked distribution. The SOIC conditions are then satisfied below w_m with this non-constant labor supply elasticity so there will be no bunching at the bottom on that account. And from Proposition 1(vi), the tax function is increasing and strictly concave in y below the mode. Again, we suppose that those distributions with decreasing densities (i.e., the Pareto and Weibull with $\beta \leq 1$) apply beyond w_m, while those with single-peaked densities apply everywhere.

Pareto distribution

Suppose the skill distribution is Pareto above w_m. For an untruncated Pareto distributions, we have shown that $C(w) = 1/a$, so

$$wC(w) = \frac{w}{a}$$

This implies that the marginal tax rate is monotonically *increasing* with w where the Pareto distribution applies. Moreover, as $\overline{w} \to \infty$, we have from (29):

$$\frac{T'(y(\overline{w}))}{(1-T'(y(\overline{w}))^2} \to \infty$$

which means that $T'(y(w))$ goes to unity at the top. This result occurs because with an untruncated Pareto distribution, $C(w)$ is constant $(1/a)$. From (28), the only factor that varies along the entire skill range is $A(w) \equiv 1/(1-\ell(w))$. Therefore, $A(w)$ goes to infinity as the wage goes to infinity, so the optimal asymptotic tax rate goes to unity.

From (15), as long as the SOIC conditions are satisfied, the tax function is then increasing and strictly convex in y where the Pareto distribution applies. Therefore, we cannot draw the same conclusions about average tax rates for a Pareto distribution applying beyond the mode since the sufficient conditions of Proposition 1(vi) are not satisfied.[18]

If the Pareto distribution is truncated at $\overline{w} < \infty$, we obtain from (22):

$$wC(w) = \frac{w}{a}\left[1 - \left(\frac{w}{\overline{w}}\right)^a\right]$$

Therefore, again, for a bounded distribution, we have a zero marginal tax rate at the top. Moreover, $w\dot{C}(w) = 1/a - (w/\overline{w})^a(1/a+1) \gtrless 0$ if $w \lessgtr \overline{w}/\left(\sqrt[a]{1+a}\right)$. Hence, the optimal marginal tax profile has an inverted U-shaped pattern beyond w_m with a maximum at $w = \overline{w}^a/\left(\sqrt[a]{1+a}\right)$. Overall, the marginal tax rate declines in skill until w_m, and then takes an inverted U-shape once the Pareto distribution sets in.[19] Moreover, assuming that the SOIC conditions are satisfied where $T'(.)$ increases with w (which is not guaranteed), (15) would imply that, beyond the mode, $T'(y)$ first increases with y and becomes decreasing for higher y. This implies that the marginal tax rates cannot be U-shaped in income. The optimal average tax rate is increasing at the bottom of the income distribution (if $T(\underline{y}) < 0$) and decreasing (at least) close to the top. (We cannot conclude about the average tax profile over the rest of the distribution.)

Weibull distribution

Suppose first that the Weibull distribution is untruncated. For $\beta \leq 1$, the density is decreasing, while $\beta > 1$ gives a single-peaked density. For $\beta < 1$, we have that $wC(w) = w^{1-\beta}/\left(\beta(1/\eta)^\beta\right)$ increases, is strictly convex and tends to ∞ when $w \to \overline{w} = \infty$. (Note that $wC(w)$ is unbounded at $\underline{w}$.) Then, the marginal tax rate increases and is strictly convex in w. Moreover, at $\overline{w} = \infty$, from (29), $T'(y(w))$ goes to unity. From (15), the increasing (in y) tax function is then strictly convex once the Weibull distribution sets in, as long as the SOIC conditions are satisfied. Therefore, we cannot conclude about average tax rates for a Weibull distribution applying beyond the mode since sufficient conditions

[18] In the specific case where the whole skill distribution is Pareto, the tax function is increasing and strictly convex in y everywhere hence, as soon as we assume $T(\underline{y}) < 0$, the average tax rate is increasing in y. This last result can easily be verified through a similar graphical exercise to the one of Figure 1, here, with a strictly convex tax function.

[19] As before, if there is a range between the mode and the skill level at which the Pareto distribution applies, the marginal tax rate could be rising or falling there.

of Proposition 1(vi) are not satisfied. However, in the specific case where the whole skill distribution is Weibull with $\beta < 1$, the tax function is increasing and strictly convex in y everywhere. If we assume $T(y(\underline{w})) < 0$, the average tax rate is increasing in y.

For $\beta = 1$, which is the special case of an exponential distribution, $wC(w) = \eta$. Therefore, the marginal tax rate in the maxi-min case is constant (and strictly positive) in the range where the exponential distribution applies. Thus, the optimal marginal income tax has a hockey-stick profile: the marginal tax rate is constant for $w > w_m$ where the exponential distribution applies, and declines before that.

From (29), it is interesting to note that a (strictly positive) constant $wC(w)$ (equivalently, a constant hazard rate) implies that a 100 percent marginal tax rate at the top is never optimal. The optimal tax function itself is increasing and concave in y. It is increasing in y below $y(w_m)$ and constant above it. Therefore, our sufficient conditions of Proposition 1(vi) are not satisfied and we can draw no conclusions on the average tax rate except at the bottom where it is increasing in y (if $T(\underline{y}) < 0$). However, if the exponential distribution applies over the whole distribution, the optimal tax profile is then linear in y hence, the optimal average tax rate is increasing in y (if $T(\underline{y}) < 0$).

For $\beta > 1$, $wC(w)$ increases, is strictly concave and tends to ∞ at the top. Then, $T'(y(w))$ is increasing in w and strictly concave above w_m where this distribution applies. The marginal tax rate tends to unity at the top. Therefore, when a Weibull with $\beta > 1$ applies and when the SOIC conditions are satisfied, the tax function is increasing and strictly convex in y, with $T'(\overline{y}) \to 1$. Average tax rates are then increasing in y, if $T(\underline{y}) < 0$ when the Weibull with $\beta > 1$ applies.

Suppose now that the Weibull distribution is truncated. In this case, $wC(w)$ is always decreasing when, as previously, we assume that $w_m > \eta$ (usually standardized to 1). Therefore, $T'(y(w))$ is necessarily decreasing over the entire Weibull distribution range, and the SOIC conditions are satisfied. Again, the optimal asymptotic tax rate is equal to zero. From Proposition 1(vi), the optimal tax function is then strictly concave and the optimal average tax rate is single-peaked (if $T(\underline{y}) < 0$).

Log-normal distribution

Again, we begin with the untruncated case. Recall that the righthand side of (29) is the inverse of the hazard function. We know that the hazard function of the log-normal distribution first increases from zero and then ultimately falls back to zero. Hence, the marginal tax scheme is U-shaped. The hazard function of the log-normal distribution does in fact have a single maximum at a value of $x = (\log w - \mu)/\sigma$ satisfying $h(x) = \sigma + \mu$ (Lancaster, 1990). At the median skill ($w = e^{\mu}$), $x = 0$. The value of σ such that $h(0) = \sigma$ is 0.7978 (see Table 3.1., Lancaster, 1990). Therefore, if $\sigma \simeq 0.8$, the minimum marginal tax rate is at the median. Moreover if σ is lower than 0.8, then the minimum point is to the left of the median, and vice versa (Dahan and Strawczynski, 2004).

We can show that if $\sigma \simeq 0.4$, marginal tax rates decline until the mean ($e^{\mu+0.5\sigma^2}$ with $\sigma \simeq 0.4$) multiplied by $e^{0.3}$, and then they rise, so the minimum marginal tax rate is at the mean multiplied by $e^{0.3}$. Moreover, if $\sigma > 0.4$, then the minimum point is at the right of the mean multiplied by $e^{0.3}$, and vice versa.[20]

Again, the SOIC conditions may be violated in the range of w where $T'(.)$ increases. Moreover, as $\overline{w} \to \infty$, from (29) and the fact that $\lim_{w \to \overline{w}} \frac{f(w)}{1-F(w)} = \lim_{w \to \overline{w}} \left(1/\frac{w\sigma^2}{\log(w)-\mu}\right)$ (e.g. Lancaster, 1990; Dahan and Strawczynski, 2004), we obtain as with the untruncated Pareto distribution:

$$\frac{T'(y(\overline{w}))}{(1-T'(y(\overline{w}))^2} \to \infty$$

which means that $T'(y(w))$ goes to unity at the top. Again this is because the elasticity of labor supply goes to zero at the top when utility is linear in consumption. From (15), the tax profile is then increasing and concave in y below and beyond the mode up to a certain y where it becomes convex. Therefore, the average tax rate is increasing at the bottom of the income distribution and we cannot conclude for the rest of the distribution.

Finally, suppose that the log-normal distribution is truncated at $\overline{w} < \infty$. In this case:

$$\dot{wC}(w) < 0 \quad \text{as} \quad e^{(\ln w - \mu)^2/(2\sigma^2)} \left(g\left(\overline{w}\right) - g(w)\right) \left(\sigma^2 + \ln w - \mu\right) < 0.8\sigma \qquad (30)$$

and again we have the standard result for a bounded distribution, $T'(y(\overline{w})) = 0$. $T'(y(w))$ is then decreasing in w and the SOIC conditions are then satisfied. The tax function is strictly concave in y and the average tax rate is single-peaked in y (if $T(\underline{y}) < 0$).

The results of this section are summarized as follows.

Proposition 3: Assuming a maxi-min criterion and quasilinear-in-consumption preferences, if the elasticity of labor is constant, Proposition 2 applies. If the utility of leisure is logarithmic:

[20] Using Table 3.1. in Lancaster (1990), if $\sigma \simeq 0.26$, marginal tax rates decline until the mean multiplied by $e^{0.23}$ and then they rise (i.e., the minimum marginal tax rate is at the mean multiplied by $e^{0.23}$). And if $\sigma > 0.26$, then the minimum point is at the right of the mean multiplied by $e^{0.23}$, and vice versa. If $\sigma \simeq 0.5$, marginal tax rates decline until the mean multiplied by $e^{0.4}$ and then they rise, that is, the minimum marginal tax rate is at the mean multiplied by $e^{0.4}$. If $\sigma > 0.5$, then the minimum point is at the right of the mean multiplied by $e^{0.4}$, and vice versa.

(i) $T'(.)$ is decreasing for $w < w_m$.

(ii) $T'(.)$ is decreasing in w, $T(y)$ is concave and the average tax rate is single-peaked in y (if $T(\underline{y}) < 0$) with a truncated Weibull (with $w_m > \eta$) and a truncated log-normal distribution (if condition (30) is met in the latter case).

(iii) For $w > w_m$, $T'(.)$ is constant with an untruncated Weibull distribution with $\beta = 1$ (i.e., an exponential distribution), $T'(.)$ is increasing with an untruncated Pareto and an untruncated Weibull with $\beta \neq 1$, and U-shaped with an untruncated log-normal distribution. (The SOIC conditions may be violated and these last results invalidated over any range of skills where $T'(.)$ increases.)

(iv) The asymptotic tax rate converges to unity for untruncated Pareto, untruncated Weibull when $\beta \neq 1$ and untruncated log-normal skill distributions. The asymptotic tax rate is positive but lower than unity with an untruncated exponential distribution.

(v) If $T(\underline{y}) < 0$, the average tax rate is increasing in y with untruncated Pareto and untruncated Weibull distributions, both applying over the entire skill distribution.

6 The non-negative income constraint

Throughout this paper, our analysis (especially in the case where utility is linear in leisure) requires that labor supply and therefore before-tax income be non-negative. However, as government revenue requirements R decrease, so too will gross labor income. Since the SOIC conditions are not binding below the mode, $y(w)$ will be increasing. Eventually a point will be reached where incomes from the lowest-wage households fall to zero. As R is reduced further, an increasing range of low-wage people 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 conditions are not violated. This bunching illustrates voluntarily unemployment where the low-skilled choose to work zero hours so their gross incomes are zero over the bunching interval. From (9), this implies that utility is also constant over the bunching interval. Moreover, the SOIC constraint $\dot{y}(w) \geq 0$ implies that this type of bunching can occur only at the bottom of the skill distribution. Thus, if R is low enough, at the optimum there is a range of skills $\underline{w} \leq w \leq w_y$, where the non-negativity constraint on income is binding. Over this range, $x(w)$ is constant at $x(w_y) \equiv x_0$.

Consider the case of quasilinear-in-leisure preferences. This is the case where non-negative income constraint is most likely to be relevant, and is the most transparent case to analyze. In the presence of this type of bunching, the first-order conditions continue to apply outside of the bunching interval. This implies that our qualitative results about the optimal tax rates are maintained for all $w \in [w_y, \overline{w}]$. Associating λ and $\zeta(w)$ respectively with the government's budget constraint and the FOIC constraint, the Hamiltonian becomes:

$$H_B(w) = u(w)I_{[w=\underline{w}]} + \int_{w_y}^{\overline{w}} \left\{ \lambda[w\ell(w) - x(w, u(w), \ell(w))]f(w) + \zeta(w)\frac{\ell(w)}{w} \right\} dw$$
$$+\lambda(F(w_y)x_0 - R) \tag{31}$$

with $\ell(w_y) \geq 0$ where w_y is a new control variable and $x_0 = x(w_y)$. Equation (19), which gives the optimal tax schedule, is still valid for any $w \geq w_y$. In particular, individuals at the top end of the bunching interval (type-w_y households) have a positive marginal tax rate. Therefore, if a solution to the optimization problem (11) implies $y(w_y) < 0$, then the solution to the optimization problem (31) is characterized by a non-empty range of skills $[\underline{w}, w_y]$ such that $y(w) = 0$ and $x(w) = x_0$ for all $w \in [\underline{w}, w_y]$, and the path of consumption for households with skills above w_y is characterized by (21).

Recall that if the non-negative income constraint is not binding, neither λ nor $x(w)$ is affected by changes in R: $y(w)$ increases uniformly with R. But, with the non-negative income constraint binding, λ is affected by changes in w_y as shown by the following formula (derived in Appendix 4) for the shadow price of public funds:

$$\lambda = \frac{1/w_y}{F(w_y)/[w_y v'(x_0)] + (1 - F(w_y))} \tag{32}$$

It is evident that λ depends now on R, unlike the case when the non-negativity constraint $y(w) \geq 0$ is not binding at any skill level (see (20)). The difference between the two formula is readily understood. Consider a unit increase in the required labor income of households with skills $w \geq w_y$, this increase being triggered by a rise in R. This increase yields $1 - F(w_y)$ in tax revenue. The only loss in social welfare due to the increased in required labor income comes from people with $u(\underline{w})$ as utility level (hence, x_0 as consumption). To continue satisfying the incentive compatibility constraints of households with skill w_y, x_0 needs to be adjusted by $dx_0 = -(1/v'(x_0)w_y)dy < 0$. This causes a loss in the social welfare $u(\underline{w})$ equal to $1/w_y$. It also decreases public expenses by an amount given by the first term in the denominator in (32).

Appendix 4 undertakes a comparative static exercise, similar to Boadway, Cuff and Marchand (2000) and Boone and Bovenberg (2006) but with a maxi-min criterion, to determine the effect of a change in R on w_y and λ. We derive the following proposition.

Proposition 4: With a maxi-min criterion and bunching due to a binding non-negativity income constraint, $d\lambda/dR > 0$, $dw_y/dR < 0$, $dx_0/dR < 0$ and $dx(w)/dR = 0$ for $w > w_y$. The results of Proposition 2 apply for $w > w_y$.

Contrary to the case without bunching, a higher level of R raises the marginal cost of public funds. The required additional resources come from additional work effort of the employed persons and new entrants into the labor market (as the bunching interval decreases), and a lower consumption level of the unemployed.

7 Conclusions

The purpose of this paper has been to provide as full a characterization as possible of the solution of the optimal income tax problem when preferences are additive and the normative criterion is maxi-min. Focusing on the maxi-min objective is fruitful for a number of reasons. First, it is straightforward to obtain a clear understanding of the economic effects underlying the optimal tax profile since we are left with only two main sources of influence on the optimal tax structure: the variability of labor and the shape of the skill distribution. Second, one can derive analytical results on the shape of the tax profile in both skills and income without resort to numerical simulations. For example, we can derive sufficient conditions for a decreasing marginal tax profile over the entire range of skills or incomes, in contrast to the U-shape profile obtained when social welfare has any non-negative degree of aversion to inequality (e.g. Diamond 1998, Saez 2001). Further, we can derive sufficient conditions for a single-peaked average tax rate. Third, considering the specific and often used quasilinear preferences, and assuming the elasticity of labor supply constant when preferences are quasilinear-in-consumption, allows us to show how the distribution of skills alone determines the optimal tax profile, and generally results in a decreasing marginal tax profile in skills, a concave tax function in income and a single-peaked average tax rate. With quasilinear-in-leisure preferences, as long as the minimum income constraint is not binding, changes in government revenues induce equal per person changes in income, with no change in consumption. This reflects the zero income elasticity of labor supply and results in a very simple representation of the shadow price of public funds. A consequence of this is that as government revenue is reduced, eventually some bunching will be induced at the bottom as the non-negativity income constraint becomes binding. Once that happens, further reductions in the government public spending will affect both the size of the bunching interval and the structure of the marginal tax rates.

However, our previous general qualitative results continue to hold outside of the bunching interval.

References

[1] Aitchison, J. and Brown, J.A.C. (1957), "The lognormal distribution", Cambridge, Cambridge University Press, London.

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

[3] Atkinson, A.B. (1975), "La 'Maxi-Min' et l'imposition optimale des revenus", Cahiers du Seminaire d'Econometrie, 74-86.

[4] Atkinson, A.B. (1990), "Public economics and the economic public", European Economic Review, 34(2): 225-248.

[5] Boadway, R., Cuff, K. and Marchand, M. (2000), "Optimal income taxation with quasi-linear preferences revisited", Journal of Public Economic Theory, 2(4): 435-60.

[6] Boone, J. and Bovenberg, L. (2006), "The simple economics of bunching", Journal of Public Economic Theory, forthcoming.

[7] Chander, P. and Wilde, L.L. (1998), "A general characterization of optimal income tax enforcement", Review of Economic Studies, 65(1): 165-183.

[8] d'Autume, A. (2001), "L'imposition optimale du revenu : une application au cas français", Revue française d'économie, 15(3): 3-63.

[9] Dahan, M. and Strawczynski, M. (2000), "Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates: Comment", American Economic Review, 90(3): 681-86.

[10] Dahan, M. and Strawczynski, M. (2004), "The optimal asymptotic income tax rate", Bank of Israel Discussion Paper, 15.

[11] Diamond, P.A. (1998), "Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates", American Economic Review, 88(1): 83-95.

[12] Ebert, U. (1992), "A reexamination of the optimal nonlinear income tax", Journal of Public Economics, 49(1): 47-73.

[13] Feenberg, D. and Poterba, J.M. (1993), "Income inequality and the income of the very-high income taxpayers: Evidence from tax returns.", in James Poterba, ed., Tax Policy and the Economy, Cambridge, MA: MIT Press, 145-77.

[14] Guesnerie, R. and Laffont, J.-J. (1984), "A complete solution to a class of principal-agent problems with an application to the control of self-managed firm", Journal of Public Economics, 25(3): 329-369.

[15] Gruber, J. and Saez, E. (2002), "The elasticity of taxable income: evidence and implications", Journal of Public Economics, 84(1): 1-32.

[16] Hindriks, J. Lehmann, E. and Parmentier, A. (2006), "Optimal income taxation and the shape of average tax rates", Economics Bulletin, 8(2), 1-6.

[17] Kanbur, R. and Tuomala, M. (1994), "Inherent inequality and the optimal graduation of marginal tax rates", Scandinavian Journal of Economics, 96(2), 275-282.

[18] Lancaster (1990), The Econometric Analysis of Transition Data, Cambridge University Press.

[19] Lollivier, S. and Rochet, J.-C. (1983), "Bunching and second-order conditions: A note on optimal tax theory", Journal of Economic Theory, 31(2): 392-400.

[20] Mirrlees, J.A. (1971), "An exploration in the theory of optimum income taxation", Review of Economic Studies 38(2): 175-208.

[21] Myles, G. (1995), Public Economics, Cambridge University Press.

[22] Okun, A.M. (1975), Equality and Efficiency: The Big Tradeoff, Washington D.C.: The Brookings Institution.

[23] Phelps, E.S. (1973), "Taxation of wage income for economic justice", The Quarterly Journal of Economics, 87(3): 331-354.

[24] Piketty, T. (1997), "La redistribution fiscale face au chômage", Revue française d'économie, 12(1): 175-208.

[25] Rawls, J. (1971), A Theory of justice, Oxford: Oxford University Press.

[26] Salanié, B. (2003), The Economics of Taxation, MIT Press: Cambridge.

[27] Sadka, E. (1976), "On income distribution, incentive effects and and optimal income taxation", Review of Economic Studies, 43(2): 261-267.

[28] Saez, E. (2001), "Using elasticities to derive optimal income tax rates", Review of Economic Studies, 68(1): 205-229.

[29] Seade, J.K. (1977), "On the shape of optimal tax schedules", Journal of Public Economics, 7(2): 203-235.

[30] Tuomala, A.M. (1990), Optimal Income Tax and Redistribution, Oxford: Clarendon Press.

[31] Weymark, J.A. (1986a), "A reduced-form optimal nonlinear income tax problem", Journal of Public Economics, 30(2): 199-217.

[32] Weymark, J.A. (1986b), "Bunching properties of optimal nonlinear income taxes", Social Choice and Welfare, 3(3): 213-232.

[33] Weymark, J.A. (1987), "Comparative static properties of optimal nonlinear income taxes", Econometrica, 55(5): 1165-1185.

Appendix 1. General additive preferences

Derivation of optimal marginal tax formula (13)

The Hamiltonian function for the maxi-min problem (11) is:

$$H_u(w) = u(w)I(w) + \lambda[w\ell(w) - X(u(w), \ell(w))]f(w) + \zeta(w)\frac{\ell(w)h'(\ell(w))}{w}$$

where $\ell(w)$ is the control variable, $u(w)$ the state variable (given by FOIC conditions (9)), λ is the multiplier associated with the budget constraint (6) and $\zeta(w)$ is the multiplier of the FOIC constraints (the shadow value of the state variable $u(w)$). The first-order and transversality conditions are:

$$\frac{\partial H_u}{\partial \ell} = \lambda\left[w - \frac{h'(\ell)}{v'(x)}\right]f(w) + \frac{\zeta(w)h'(\ell)}{w}\left(1 + \frac{\ell(w)h''(\ell)}{h'(\ell)}\right) = 0 \qquad (33)$$

$$\frac{\partial H_u}{\partial u} = 1 - \frac{\lambda f(w)}{v'(x(w))} = -\frac{\partial \zeta(w)}{\partial w} \quad \text{for } w = \underline{w} \qquad (34)$$

$$\frac{\partial H_u}{\partial u} = -\frac{\lambda f(w)}{v'(x(w))} = -\frac{\partial \zeta(w)}{\partial w} \quad \forall w > \underline{w} \qquad (35)$$

$$\zeta(\overline{w}) = 0 \qquad (36)$$

Integrating of both sides of (35) we have:

$$-\int_w^{\overline{w}} \frac{f(t)}{v'(x(t))}dt = -\frac{\zeta(\overline{w}) - \zeta(w)}{\lambda} \quad \forall w > \underline{w}$$

and using the transversality conditions (36) we obtain:

$$-\frac{\zeta(w)}{\lambda} = \int_w^{\overline{w}} \frac{f(t)}{v'(x(t))}dt > 0 \qquad (37)$$

This equation is still true at $\underline{w}$ since there is no mass point at $\underline{w}$. The shadow price $\zeta(w)$ is the effect of an increase of $u(w)$ on the value of the objective function, $u(\underline{w})$. Using (3), (33) may be rewritten as:

$$\frac{T'(y(w))}{1 - T'(y(w))} = -\frac{\zeta(w)v'(x(w))}{\lambda w f(w)}\left(1 + \frac{\ell h''(\ell)}{h'(\ell)}\right) \quad \forall w \in W \qquad (38)$$

Consider now the equivalent problem (12) where tax revenue is maximized. The corresponding Hamiltonian is:

$$H_R(w) = [w\ell(w) - X(u(w), \ell(w))]f(w) + \pi(w)\frac{\ell(w)h'(\ell(w))}{w}$$

where the control and state variables are the same as in the primal problem and the multipliers of the FOIC conditions are now $\pi(w)$. The necessary conditions are:

$$\frac{\partial H_R}{\partial \ell} = \left[w - \frac{h'(\ell)}{v'(x)}\right] f(w) + \frac{\pi(w)h'(\ell)}{w}\left(1 + \frac{\ell(w)h''(\ell)}{h'(\ell)}\right) = 0 \tag{39}$$

$$\frac{\partial H_R}{\partial u} = -\frac{f(w)}{v'(x(w))} = -\frac{\partial \pi(w)}{\partial w} \tag{40}$$

and $u(\underline{w}) = \underline{u}$. The transversality condition is:

$$\pi(\overline{w}) = 0 \tag{41}$$

The solution to this problem is equivalent to the solution to problem (11). By integrating of both sides of (40) and using the transversality conditions (41) we obtain:

$$-\pi(w) = \int_w^{\overline{w}} \frac{f(t)}{v'(x(t))} dt \tag{42}$$

This is similar to (37) except that the lefthand side of the latter was the ratio of the shadow price associated with the FOIC conditions to the marginal cost of public funds. The shadow price $\pi(w)$ measures the effect of an increase of $u(w)$ on the value of tax revenue, and is essentially equivalent to the ratio of the shadow prices in the problem (11). Moreover, using (3), (39) may be rewritten as:

$$\frac{T'(y(w))}{1 - T'(y(w))} = -\frac{\pi(w)v'(x(w))}{wf(w)}\left(1 + \frac{\ell h''(\ell)}{h'(\ell)}\right) \quad \forall w \in W \tag{43}$$

This is exactly analogous to (38) for the problem (11), and confirms that the solutions to the two problems are the same.

Equivalently, (37) and (38) or (42) and (43) may be combined to yield (13).

Derivation of the optimal average tax profile

From the binding budget constraint and the strict concavity of $T(y)$, as soon as R is not too large, $T(y(\underline{w})) \equiv T(\underline{y}) < 0$. From the average tax rate $t(y) = T(y)/y$, we derive:

$$t'(y) = \frac{T'(y) - t(y)}{y}$$

$$t''(y) = \frac{T''(y) - t'(y)}{y} - \frac{T'(y) - t(y)}{y}$$

From $T'(y) > 0$, $T''(y) < 0$ and $T'(\overline{y}) \to 0$, $T(y)$ has an inflexion point at $y = \widehat{y} < \infty$. For $y \geq (<)\widehat{y}$: $T'(\widehat{y}) \geq (<)t(\widehat{y})$. From the two previous equations, we then have: $t'(y) > 0$ and $t''(y) < 0$ below $\widehat{y}$. Beyond $\widehat{y}$: $t'(y) < 0$ and note the ambiguous sign of $t''(y)$. Therefore, $t(y)$ is single-peaked when $T(\underline{y}) < 0$ and $T'(\overline{y}) \to 0$.

Appendix 2. Quasilinear-in-leisure preferences

Derivation of (18)

Using $h'(\ell) = 1$, $h''(\ell) = 0$ and substituting (33) into (35) we have

$$\dot{\zeta}(w) = \frac{\zeta(w)}{w} + \lambda f(w)w \quad \forall w > \underline{w} \tag{44}$$

This is a linear differential equation that can be solved using the method of the varying constant. Considering only the homogeneous part, we can write $\dot{\zeta}(w)/\zeta(w) = dw/w$. Integrate both parts and define $\beta(w) = \ln(\zeta(w))/\ln(w) \Leftrightarrow \zeta(w) = we^{\beta(w)}$. Then, differentiating with respect to w, we get $\dot{\zeta}(w) = e^{\beta(w)} + we^{\beta(w)}\dot{\beta}(w)$. Substituting this into (44) yields:

$$e^{\beta(w)} + we^{\beta(w)}\dot{\beta}(w) = \frac{\zeta(w)}{w} + \lambda f(w)w \quad \forall w > \underline{w}$$

Substituting $\zeta(w) = we^{\beta(w)}$ obtained above into this last equation, we have: $e^{\beta(w)}\dot{\beta}(w) = \lambda f(w)$, or equivalently $de^{\beta(w)}/dw = \lambda f(w)$. Integrating up to w, we can rewrite:

$$e^{\beta(w)} = \lambda \int_{\underline{w}}^{w} f(t)dt + \beta_0$$

for some constant β_0. Substituting this into $\zeta(w) = we^{\beta(w)}$ yields::

$$\zeta(w) = w \left[\beta_0 + \lambda \int_{\underline{w}}^{w} f(t)dt\right] \quad \forall w > \underline{w}$$

With the transversality condition (36), evaluating $\zeta(w)$ at $w = \overline{w}$ implies that $\lambda = -\beta_0 > 0$. Hence, the above equation for $\zeta(w)$ can be rewritten as

$$\zeta(w) = -w\lambda(1 - F(w)) \quad \forall w > \underline{w} \tag{45}$$

Substituting (45) into (37), we obtain (18).

Derivation of (19) by the Lagrangian method

Transform the utility as in Lollivier and Rochet (1983), $V(w) \equiv wu\left(x\left(w\right), y\left(w\right)/w\right) = wv(x(w)) - y(w)$. The government's problem is to maximize $V(\underline{w})$ subject to

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

and

$$V(w) = V(\underline{w}) + \int_{\underline{w}}^{w} v(x(t)) dt$$

where the first constraint is the budget constraint (6) and the second constraint is the integral of the FOIC conditions (7) using $\dot{V}(w) = u(x(t))$. Substituting $V(w)$ from the second constraint into the first yields:

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

where we have made use of Fubini's theorem to evaluate the double integral. The government problem is to choose $V(\underline{w})$ and $x(w)$ to maximize $V(\underline{w})$ subject to this revenue constraint. The first-order conditions immediately yield (20) and (21). Substituting (21) and $h'(\ell) = 1$ into (3), we obtain (19).

Proof that $\mathbf{dy(w)/dR = 1}$ with quasilinear-in-leisure preferences

From the utility function, $\ell(w) = v(x(w)) - u(w)$. Substituting this into (9) gives:

$$\dot{u}(w) = \frac{v(x(w))}{w} - \frac{u(w)}{w} \tag{46}$$

This linear differential equation can be solved using the method of varying constant. From the homogeneous part, we know that

$$\frac{du(w)}{dw} = -\frac{u(w)}{w} \quad \text{or} \quad \frac{du(w)}{u\left(w\right)} = -\frac{dw}{w}$$

Integration yields:

$$\int \frac{du(w)}{u\left(w\right)} = -\int \frac{1}{w} dw$$

so, $\ln u(w) = -[\ln(w) + \gamma(w)]$, where $\gamma(w)$ is an arbitrary constant. Taking antilogs, we arrive at:

$$u(w) = e^{-\ln(w)} e^{-\gamma(w)} = (1/w)(1/e^{\gamma(w)}) \tag{47}$$

Differentiating with respect to w we obtain:

$$\dot{u}(w) = \frac{-1}{e^{\gamma(w)}w}\left(1/w + \frac{\partial\gamma(w)}{\partial w}\right)$$

Using (46) and (47), we can rewrite this as:

$$v(x(w)) = \frac{-1}{e^{\gamma(w)}}\dot{\gamma}(w) \Leftrightarrow v(x(w)) = \frac{d}{dw}e^{-\gamma(w)}$$

Integrating up to w, we get:

$$e^{-\gamma(w)} = \int_{\underline{w}}^{w} v(x(t))dt + I(w)$$

for some constant $I(w)$. Using (47) again:

$$u(w) = (1/w)\left(\int_{\underline{w}}^{w} v(x(t))dt + I(w)\right) \tag{49}$$

where we can redefine $I(w)$ as $J(w) - R$. From the definition of the utility, we have

$$y(w) = w(v(x(w)) - u(w)) = w(v(x(w)) - \int_{\underline{w}}^{w} v(x(t))dt - J(w) + R \tag{50}$$

Substituting (49) into (50) and then into the binding budget constraint (6) we get:

$$\int_{\underline{w}}^{\overline{w}} f(w)wv(x(w))dw - \int_{\underline{w}}^{\overline{w}}(1-F(w))v(x(w))dw - J(w) - \int_{\underline{w}}^{\overline{w}} x(w)f(w)dw = 0$$

where we have used Fubini's theorem. Substitute this into (50) to obtain:

$$y(w) = K(w) + R$$

where (using Ebert's (1992) notation)

$$K(w) = wv(x(w)) - \int_{\underline{w}}^{\overline{w}} \{v(x(w))[wv(x(w)) - 1 + F(w)] - x(w)f(w)\}\, dw - \int_{\underline{w}}^{\overline{w}} u(v(t))dt$$

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

$$\frac{dy(w)}{dR} = 1 \qquad \forall w \in W \tag{51}$$

Appendix 3. Quasilinear-in-consumption preferences

Derivation of (25) by the Lagrangian method

Integrating the FOIC constraint (9) from $\underline{w}$ to w gives:

$$u(w) = u(\underline{w}) + \int_{\underline{w}}^{w} \frac{h'(y(t)/t)}{t^2} y(t)dt \tag{52}$$

and substituting this into the revenue constraint (6) for $u(w)$, where $x(w) = u(w) + h(\ell(w))$, gives:

$$\int_{\underline{w}}^{\overline{w}} \left[y(w) - h\left(\frac{y(w)}{w}\right)\right] f(w)dw - u(\underline{w}) - \int_{\underline{w}}^{\overline{w}} \frac{h'(y(w)/w)}{w^2} y(w)(1-F(w))dw = R$$

where we have used Fubini's theorem to evaluate the double integral. The government problem is then to choose $u(\underline{w})$ and $y(w)$ to maximize $u(\underline{w})$ subject to this revenue constraint. The first-order conditions yield $\lambda = 1$, and

$$\frac{h'(\ell)}{w} = \frac{wf(w)}{wf(w) + (1-F(w))(1+e^{-1}(w))}$$

Then, substituting (3), where $v'(x) = 1$, into the latter yields (25).

Proof that $\mathbf{dx(w)/dR = -1}$ with quasilinear-in-consumption preferences

From the utility function, we can write:

$$x(w) = u(w) + h(\ell(w)) \tag{53}$$

Substituting this into the binding government budget constraint (6), we obtain:

$$\int_{\underline{w}}^{\overline{w}} [y(w) - u(w) - h(\ell(w))]f(w)dw - R = 0$$

Using the FOIC condition (52) and Fubini's Theorem, this can be written:

$$u(\underline{w}) = -R + \int_{\underline{w}}^{\overline{w}} [y(w) - h(\ell(w))]f(w)dw - \int_{\underline{w}}^{\overline{w}} \frac{h'(\ell(w))}{w} \ell(w)(1-F(w))dw$$

Substitute this into (53) using (52) for $u(w)$ and obtain:

$$x(w) = -R + L(w)$$

where

$$L(w) = \int_{\underline{w}}^{\overline{w}} [w\ell(w) - h(\ell(w))]f(w)dw - \int_{\underline{w}}^{\overline{w}} \frac{h'(\ell(w))}{w}\ell(w)(1-F(w))dw + h(\ell(w))$$

Since $L(w)$ depends only on the distribution of skills and the functional form of $h(\ell(w))$ (since $\ell(w)$ depends only on these two elements), this implies that

$$\frac{dx(w)}{dR} = -1 \quad \forall w \in W$$

Appendix 4. Derivation of Proposition 4

With bunching due to a binding non-negativity income constraint, we can derive two equations in λ and w_y. Integrating both sides of (35) from $\underline{w}$ to w with $w = \underline{w} + \varepsilon$ (with ε approaching 0), we find[21]

$$\zeta(w) - \zeta(\underline{w}) = -w\lambda(1 - F(w))$$

Since the necessary conditions for the Hamiltonian optimization problem associated with (11) and the Lagrangian optimization problem in Appendix 2 are equivalent, λ will be as well. Therefore, we can substitute (20) into (45) and obtain

$$\lim_{w \to \underline{w}} \zeta(w) = -1$$

Substituting this into the previous equation, we obtain $\zeta(w) = -w\lambda(1-F(w))$. From this and (45), setting $w = w_y$, we may rewrite

$$1 - \lambda w_y(1 - F(w_y)) = \lambda \frac{F(w_y)}{v'(x_0)}$$

which gives (32). Substituting (21) for $w = w_y$ into (32) for $u'(x_0)$ yields the following expression for λ in terms of w_y:

$$\lambda = \frac{f(w_y)}{f(w_y)w_y - F(w_y)(1 - F(w_y))} \tag{54}$$

[21] Although (34) (slightly modified such that it is satisfied $\forall\, w \in [\underline{w}, w_y]$) should hold (as all people in the range $[\underline{w}, w_y]$ should get the same outcome), integrating it would move us away from maxi-min. It would imply that we weigh the utility of type $\underline{w}$ by the number of people receiving this utility. For maxi-min, this is irrelevant.

It can be shown that (54) is downward sloping as follows. Substituting $w_y = \underline{w}$ in (54), we obtain

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

Similarly for $w_y = \overline{w}$, we can write:

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

Since $1/\underline{w} > (\overline{w} - \underline{w} + 1)/\overline{w}$, this curve must be downward sloping over parts of the range $[\underline{w}, \overline{w}]$ in (w_y, λ) space.

To obtain the second relationship between w_y and λ, rewrite the government budget constraint as

$$-F(w_y)x_0 + \int_{w_y}^{\overline{w}} [y(w) - x(w)] f(w) dw = R$$

Following the same procedure as in Appendix 2, the expression for $y(w)$ when the non-negative income constraint is binding is

$$\begin{aligned} y(w) &= wv(x(w)) - \int_{w_y}^{w} v(x(t))dt + (1 - F(w_y))^{-1} \\ &\times \left[R + F(w_y)x_0 - \int_{w_y}^{\overline{w}} \{v(x(w))[wf(w) + F(w) - 1] - x(w)f(w)\} \, dw \right] \end{aligned}$$

Evaluating this at $w = w_y$ where $y(w_y) = 0$ yields an equation determining w_y:

$$\begin{aligned} &(1 - F(w_y))w_y v(x(w_y)) + R + F(w_y)x_0 \\ &= \textstyle\int_{w_y}^{\overline{w}} \{[f(w)w - (1 - F(w))]v(x(w)) - f(w)x(w)\} \, dw \end{aligned} \tag{56}$$

which is upward sloping in (w_y, λ) space if $\dot{x}(w) \geq 0$ as shown in Boone and Bovenberg (2006). Following them, when $x'(w_y) > 0$, (w_y, λ) are determined by the intersection of the downward sloping curve (54) and the upward sloping curve (56). Clearly, (54) is not affected by a change in R. It can be shown (see Boone and Bovenberg 2006) that (56) shifts upwards (and to the left) as R increases. Hence, w_y falls and λ rises with R. By (32), x_0 varies inversely with changes in λ. However, by (21) we see that with a maximin criterion, the rise in λ does not imply a reduction of consumption and a rise in the marginal tax rate for all types $w > w_y$ as it would be the case with a utilitarian criterion.

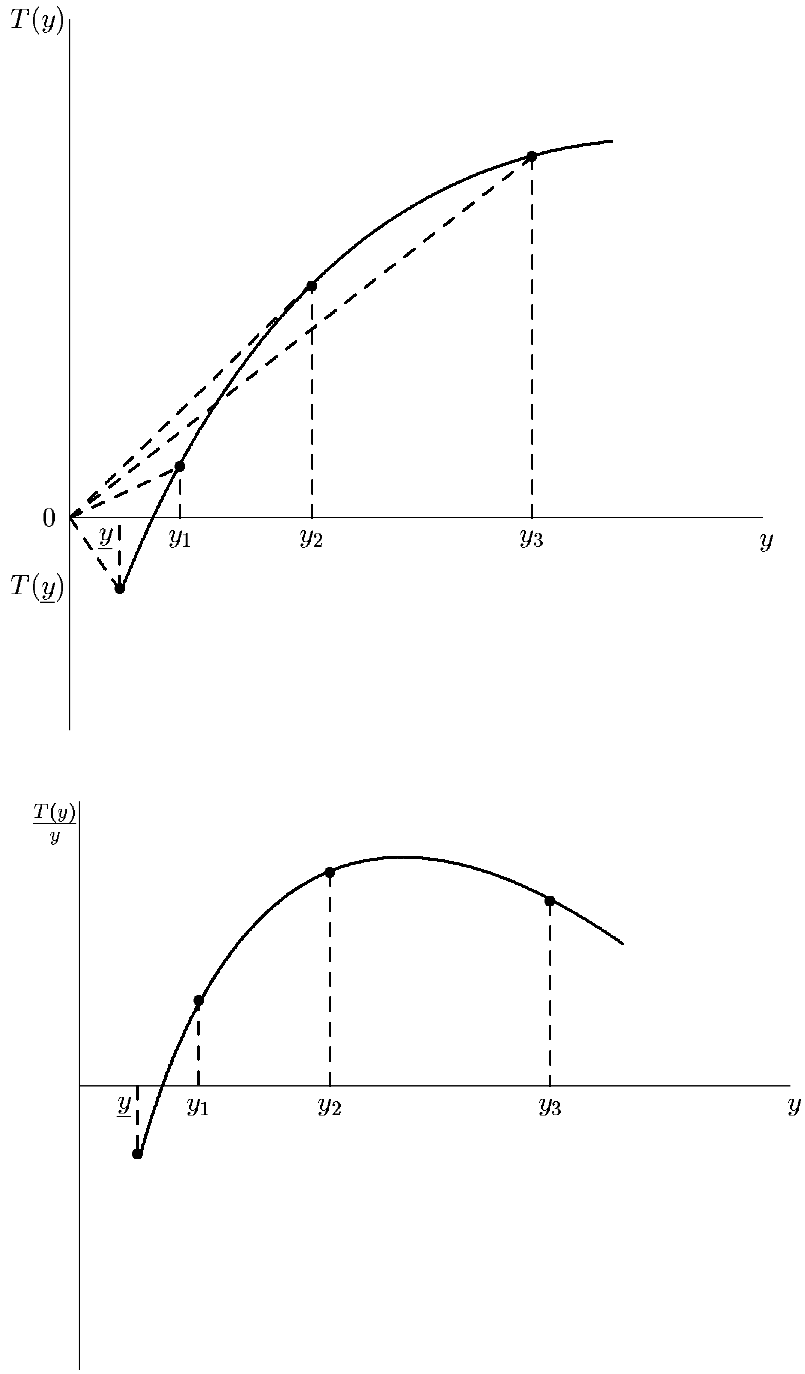

Figure 1 Single-peaked average tax rate