Soc Choice Welf
DOI 10.1007/s00355-009-0437-x

Optimal nonlinear income tax and nonlinear pricing: optimality conditions and comparative static properties

Laurent Simula

Received: 15 August 2008 / Accepted: 3 December 2009

Abstract Using the Mirrlees optimal income tax model, with no income effects on labour supply, this article shows that the discrete population approach provides new insights into the characterization of the optimal tax system, which complements the previous findings. The analysis is based on the "Spence-Mirrlees wedge" which corresponds, at each observed gross income level, to a ratio between the marginal tax rate of the individual for whom the bundle is designed and that of his nearest more productive neighbour if he chooses to mimic. Using this wedge, a necessary and sufficient condition for bunching to be optimal is obtained in terms of the primitives of the model, separating optima are characterized geometrically, and comparative statics properties derived, notably with respect to skill levels and individual social weights. It is then shown that the analysis extends to adverse-selection problems where participation constraints replace the tax revenue constraint.

Keywords Optimal tax · Income tax · Nonlinear pricing · Adverse-selection · Comparative statics

JEL Classification D82 · D86 · H21 · H31

1. Introduction

This article analyses the nonlinear income tax problem à la Mirrlees (1971) with a finite distribution of workers as in Guesnerie and Seade (1982). In this setting, the optimal tax structure is the product of different sorts of interacting influences:

L. Simula (✉)
Uppsala Center for Fiscal Studies, Department of Economics, Uppsala University, P.O. Box 513, SE-75120 Uppsala, Sweden
e-mail: laurent.simula@nek.uu.se

individual skill levels, government's aversion to income inequality and responsiveness of labour supply. Moreover, the way in which these influences interact is affected both by the incentive-compatibility constraints and the tax revenue constraint. Merely assuming that individual preferences are concave and increasing both in consumption and leisure does not yield many general results. Consequently, investigations often resort to numerical simulations (Tuomala 1990). However, if simulations are very useful to quantify the optimal tax rates, they are not ideally suited for shedding light on the economic intuitions behind the results. The present article follows an alternative path: it concentrates on the benchmark case where there are no income effects on labour supply to derive new qualitative features of the solution to the optimal income tax schedule.

Following Lollivier and Rochet (1983), many articles have focused on the polar case in which all income effects are absorbed by labour supply, i.e. on preferences over consumption and leisure which are separable and linear in leisure. In the finite population framework, it was used to characterize the optimal allocation (Weymark 1986a,b), provide comparative static results with respect to a preference parameter, individual productivities and social weights (Brett and Weymark 2008; Weymark 1987), and examine the impact of changing individual productivity in an economy with two classes of agents where the government adopts a maximin or maximax objective function (Hamilton and Pestieau 2005). It was also used in the continuous population framework to provide an example in which different types of individuals are bunched together when the tax schedule is set optimally (Ebert 1992) and to characterize the optimum allocation with a particular focus on bunching (Boadway et al. 2000; Boone and Bovenberg 2007). There are however limitations in the linear-in-leisure model. In particular, the tax schedule is then independent of the labour response because the disutility of labour is constant, a fact which seems to be incompatible with the empirical findings.

Because most of the empirical studies give credence to small income effects relative to substitution effects as regards labour supply (Blundell 1992; Blundell and MaCurdy 1999), assuming no income effects on labour supply, i.e. focusing on separable preferences in consumption and leisure, linear in consumption (referred to as quasilinear-in-consumption), provides a relevant benchmark for the policymaker. This has been acknowledged since the work by Diamond (1998) who examines conditions for marginal tax rates to be U-shaped when the population is continuous. The present article shows that the discrete population approach provides new insights into the characterization of the optimal tax system, which complements the previous findings. Most of the analysis is based on a wedge corresponding, at each observed gross income level, to a ratio between the marginal tax rate of the individual for whom the bundle is designed and that of his nearest more productive neighbour if he chooses to mimic. This wedge is referred to as the *"Spence-Mirrlees" wedge* and cannot be defined when there is a continuum of agents because the space between two productivity levels has then measure zero. Thanks to the Spence-Mirrlees wedge, a necessary and sufficient condition under which bunching is not optimal is obtained in terms of the primitives of the model. When the optimal allocation is fully separating, the optimum is characterized by the equality, at each observed income level, between the Spence-Mirrlees wedge and the cumulative social weight of the individuals with earnings up to this income level. This feature is first used to construct the optimal allocation geometrically in a series of very simple steps. It is then employed to examine the impact of changing the marginal utility of money, skill levels, and social weights. Skill levels are the fundamental source of the self-selection problem in Mirrlees model since they basically condition the effort a given individual has to undertake when he chooses to mimic everyone else. It is shown that a slight alteration in the skill level of any individual only affects the implicit marginal tax rates and pre-tax incomes of himself and his nearest less productive neighbour, which is in sharp contrast with Brett and Weymark (2008): for quasilinear-in-leisure preferences, all marginal tax rates and incomes are usually affected, except for the top individual. Changes in skill levels are then combined with adjustments in social weights since a policymaker might be less concerned by the well-being of someone who becomes more productive. The analysis is then extended to adverse-selection problems where participation constraints replace the tax revenue constraint. The Spence-Mirrlees wedge appears as a key determinant of the optimum menu of contracts and its bunching properties, while hazard rates play the part of cumulative social weights.

The article is organized as follows. Section 2 sets up the model. Section 3 derives the reduced form of the optimal nonlinear income tax problem and characterizes the optimal allocation. Section 4 is devoted to the comparative static analysis. Section 5 extends the analysis to nonlinear pricing. Section 6 concludes. Most proofs are relegated to the Appendix.

2. The model

The population consists of $I \geq 2$ individuals, indexed by $i \in \mathcal{I} := \{1, \ldots, I\}$. There are two goods, consumption and leisure. Units of the consumption goods are chosen so that one unit costs one euro. Person i's consumption and labour supply are denoted x_i and ℓ_i, respectively. The economy is competitive, with constant-returns-to-scale technology; so person i's wage rate is fixed and equal to his productivity θ_i. For convenience, only one person has a given productivity level. Individuals are thus indexed in terms of productivity. This simplification is not particularly restrictive as the distance between two productivity levels is free to vary. Without loss of generality, the vector of productivities $\theta := (\theta_1, \ldots, \theta_I)$ is taken to be monotonically increasing,

$$0 < \theta_1 < \cdots < \theta_I. \tag{1}$$

An individual with productivity θ_i working ℓ_i units of time has gross income $z_i := \theta_i \ell_i$. All individuals have the same preferences over consumption and leisure, represented by the utility function $U : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$,

$$U(x_i, \ell_i) := \gamma x_i - v(\ell_i), \qquad i \in \mathcal{I}, \tag{2}$$

where $\gamma \in \mathbb{R}_{++}$ is the marginal utility of money.[1] It is assumed that the disutility of labour $v : \mathbb{R}_+ \to \mathbb{R}_+$ is a $\mathcal{C}^3$-function which satisfies $v' > 0$, $v'' > 0$, $v''' > 0$, $v(0) = 0$ and $\lim_{\ell_i \to \infty} v'(\ell_i) \to \infty$. The assumption $v''' > 0$ means that the increase by which additional labour supply becomes unattractive becomes larger with ℓ_i.[2] Because $U(x_i, \ell_i) = U(x_i, z_i/\theta_i)$, individuals have *personalized utility* functions $u : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ in the gross-income/consumption space,

$$u(x_i, z_i; \theta_i) := \gamma x_i - v(z_i/\theta_i), \qquad i \in \mathcal{I}. \tag{3}$$

The *marginal rate of substitution* $s(z_i; \theta_i)$ of person i at the (x_i, z_i)-bundle is independent of his consumption level x_i, with

$$s(z_i; \theta_i) := -\frac{u'_{z_i}(x_i, z_i; \theta_i)}{u'_{x_i}(x_i, z_i; \theta_i)} = \frac{v'(z_i/\theta_i)}{\gamma \theta_i}, \qquad i \in \mathcal{I}. \tag{4}$$

So, in the gross-income/consumption space, i's indifference curves are parallel *vertical* displacements of each other. Moreover, the *Spence-Mirrlees condition* is met: for a given gross income level, more productive individuals have flatter indifference curves.

A social allocation specifies a consumption and gross income level for each individual. It is represented by a vector $a = (x, z) \in \mathbb{R}^I \times \mathbb{R}^I_+$, with $x = (x_1, \ldots, x_I)$ and $z = (z_1, \ldots, z_I)$. The tax policymaker knows the functional form of the utility function and the distribution of wages in the population. He is however unable to observe each individual's productivity. As a result, he is restricted to setting taxes as a function of gross income z_i. By the taxation principle, a nonlinear income tax schedule is equivalent to a mapping $\theta \longrightarrow a$ which satisfies the *incentive compatibility constraints*

$$IC_{ij} : u(x_i, z_i; \theta_i) \geq u\left(x_j, z_j; \theta_i\right), \qquad \forall (i, j) \in \mathcal{I}^2, \tag{5}$$

and the *tax revenue constraint*

$$\sum_{i=1}^{I} z_i \geq \sum_{i=1}^{I} x_i. \tag{6}$$

An allocation a is *budget-balanced* if the tax revenue constraint (6) is binding. It is a *simple monotonic chain to the left* ("SMCL" hereafter) if and only if all

[1] The assumption $x_i \in \mathbb{R}$ is introduced for convenience; it can be checked ex post that x_i is nonnegative. Similarly, it is typically assumed that $\ell_i \in \mathbb{R}$ in the quasilinear-in-leisure version of Mirrlees's model (Lollivier and Rochet 1983; Weymark 1986b, 1987).

[2] This assumption plays an important role in the comparative statics. The role of third derivatives in such exercises has been emphasized by the literature devoted to risk and uncertainty. For instance, see Kimball (1990), Laffont (1989), Menegatti (2001). It also yields uniqueness of the solution to the optimal income tax model, as established below. In this respect, it could be relaxed to "weakly decreasing consumption-specific risk aversion" introduced by Hellwig (2007), which is sufficient for uniqueness.

adjacent downward incentive-compatibility constraints in (5) are binding, i.e. iff $u(x_{i+1}, z_{i+1}; \theta_{i+1}) = u(x_i, z_i; \theta_{i+1})$ for $i = 1, \ldots, I-1$. Because of (3), these equalities are equivalent to

$$u(x_{i+1}, z_{i+1}; \theta_{i+1}) - u(x_i, z_i; \theta_i) = R(z_i; \theta_i, \theta_{i+1}), \qquad i = 1, \ldots, I-1 \quad (7)$$

where $R(z_i; \theta_i, \theta_{i+1})$ is defined, for later reference, as

$$R(z_i; \theta_i, \theta_{i+1}) := v(z_i/\theta_i) - v(z_i/\theta_{i+1}), \qquad i = 1, \ldots, I-1. \quad (8)$$

The *social welfare function* $W : \mathbb{R}^I \times \mathbb{R}^I_+ \to \mathbb{R}$ is a weighted sum of individual utilities,

$$W(a) := \sum_{i=1}^{I} \lambda_i u(x_i, z_i; \theta_i), \quad (9)$$

in which $\lambda := (\lambda_1, \ldots, \lambda_I)$ are individual social weights. The policymaker's taste for redistribution from the high to the low productive individuals is captured through the requirement that the higher the individual productivity the less the weight in the social objective, i.e.

$$0 < \lambda_I < \cdots < \lambda_1. \quad (10)$$

The social objective function (9) can thus be regarded as a linear approximation of $\sum_{i=1}^{I} \phi(u(x_i, z_i; \theta_i))$, where the function ϕ would be strictly concave.[3] If $I = 2$, assumption (10) amounts to considering the "normal" case studied by Stiglitz (1982) in which only the incentive compatibility constraint of the high type is binding.

As $W(a)$ is homogeneous of degree one in λ, the sum of the social weights can be normalized without loss of generality. It is convenient to define Λ_i as the cumulative social weight of the i less productive individuals, i.e. $\Lambda_i = \sum_{j=1}^{j=i} \lambda_i$ for $i \in \mathcal{I}$, and to set

$$\Lambda_I = I. \quad (11)$$

Consequently, *admissible parameters* $(\theta, \gamma, \lambda)$ belong to the set

$$\mathcal{P} := \{\theta | (1) \text{ is satisfied}\} \times \mathbb{R}_{++} \times \{\lambda | (10) \text{ and } (11) \text{ are satisfied}\}. \quad (12)$$

The optimal nonlinear income tax problem can thus be formulated as follows.

Problem 1 (*Optimal Nonlinear Income Tax Problem*) For $(\theta, \gamma, \lambda) \in \mathcal{P}$, choose an allocation $a \in \mathbb{R}^I \times \mathbb{R}^I_+$ to maximize $W(a)$ under the incentive-compatibility constraints (5) and the tax revenue constraint (6).

[3] So, pure utilitarianism in which all individual social weights λ_i are equal is not considered. This is justified since, in this case, the solution to the second-best optimal income tax problem is the laissez-faire, whose properties are already well-known.

3. The optimal allocation

The optimal nonlinear income tax problem involves two sets of control variables, gross income z and net income x. It can however be transformed into a reduced-form problem in which the policymaker chooses only one of these sets. The reduced-form problem makes it easier to interpret the social value function as well as the optimality conditions and to derive comparative static results. For this purpose, Problem 1 is separated into two subproblems. In the first one, gross income is arbitrarily chosen within the set of incentive-feasible gross income levels,

$$\mathcal{Z} := \left\{ z \in \mathbb{R}^I | 0 \leq z_1 \leq \cdots \leq z_I \right\}. \tag{13}$$

Subproblem 1 Given a gross income vector $z \in \mathcal{Z}$ and the parameters $(\theta, \gamma, \lambda) \in \mathcal{P}$, choose the consumption vector $x \in \mathbb{R}^I$ to maximize the social welfare function $W(a)$ subject to the incentive-compatibility constraints (5) and the tax revenue constraint (6).

Let $\mathcal{X}^*(z; \theta, \gamma, \lambda)$ be the set of maximizers. Then, if $\mathcal{X}^*(z; \theta, \gamma, \lambda)$ contains only one consumption vector $x^*(z; \theta, \gamma, \lambda)$, the reduced-form problem can be stated as follows.

Subproblem 2 Given the parameters $(\theta, \gamma, \lambda) \in \mathcal{P}$, choose $z \in \mathcal{Z}$ to maximize the social welfare function $W(x^*(z; \theta, \gamma, \lambda), z)$.

The incentive-compatibility constraints (5) place structure on the solutions to theses problems.

Proposition 1 *Given $z \in \mathcal{Z}$ and $(\theta, \gamma, \lambda) \in \mathcal{P}$,*

(i) *any allocation $a = (x^*, z)$ where $x^* \in \mathcal{X}^*(z; \theta, \gamma, \lambda)$, is a SMCL which is budget-balanced;*
(ii) *there is a unique solution to Subproblem 1, which is independent of λ, twice continuously differentiable and defined by $x^* : \mathcal{Z} \times \mathbb{R}^I_{++} \times \mathbb{R}_{++} \to \mathbb{R}^I_+$ with*

$$x_1^*(z; \theta, \gamma) = \frac{1}{I} \left\{ \sum_{j=1}^{I} z_j - \frac{1}{\gamma} \sum_{j=2}^{I} (I + 1 - j) \left[v\left(\frac{z_j}{\theta_j}\right) - v\left(\frac{z_{j-1}}{\theta_j}\right) \right] \right\}, \tag{14}$$

$$x_i^*(z; \theta, \gamma) = x_1^*(z; \theta, \gamma) + \frac{1}{\gamma} \sum_{j=2}^{i} \left[v\left(\frac{z_j}{\theta_j}\right) - v\left(\frac{z_{j-1}}{\theta_j}\right) \right], \quad i = 2, \ldots, I. \tag{15}$$

An implication of (i) is that (7) holds at any optimum allocation. Hence, $R(z_i; \theta_i, \theta_{i+1})$ indicates at which rate utility must be increased for the tax schedule to induce individual truthtelling. So, it may be regarded as the *marginal rent* the policymaker has to leave to the more productive $i+1$-individual to prevent him from mimicking the i-individual.

3.1. The reduced form

We can now take stock of the previous results to give a more compact formulation of Subproblem 2. For this purpose, it is convenient to introduce the new vector of social parameters $\beta = (\beta_1, \ldots, \beta_I)$, with

$$\beta_i := \Lambda_i - i, \qquad i \in \mathcal{I}. \tag{16}$$

Because of (10) and (11), the graph of $i \to \Lambda_i$ is hump-shaped and above the 45°-line. Hence, $\beta_i > 0$ for $i = 1, \ldots, I - 1$, whilst $\beta_I = 0$. Skill levels do not appear directly in (16), unlike in the analogous expression in Weymark (1986b).

The parameters β_i summarize in a transparent way the redistributive taste of the government. First, they would all be equal to zero if the government adopted pure utilitarianism as a social objective. They thus express the policymaker's *strict* aversion to income inequality. Second, to get further insight into β_i, it is instructive to consider the effects of the government's decision to give each of the i less productive individuals one extra euro of consumption in the absence of incentive effects. On the one hand, the utility of each of them is increased by γ, the marginal utility of money. Hence, the gross social benefit amounts to $\gamma\Lambda_i$. On the other hand, the consumption of every individual must be decreased by i/I in order to satisfy the tax revenue constraint (6). This reduces individual welfare by $\gamma i/I$ and thus social welfare by $\Lambda_I \times \gamma i/I = \gamma i$. Summing both effects, it appears that $\gamma\beta_i$ is the net social benefit of marginally increasing the consumption of the i less skilled individuals whilst ignoring informational externalities. So, β_i is this net social benefit expressed in monetary units. Third, the parameters β_i can alternatively be defined as

$$\beta_i := I - i - \sum_{j=i+1}^{I} \lambda_j, \qquad i = 1, \ldots, I - 1. \tag{17}$$

They thus also corresponds to the net social cost, expressed in euros, of marginally increasing the consumption of the $I - i$ most productive individuals. That is why they are henceforth referred to as *net cumulative social weights*. Thanks to them Subproblem 2 can be rewritten as follows.

Problem 2 (*Reduced Form*) For $(\theta, \gamma, \lambda) \in \mathcal{P}$, choose $z \in \mathcal{Z}$ so as to maximize the social objective function $\mathcal{W}^* (z; \theta, \gamma, \lambda) : \mathcal{Z} \times \mathcal{P} \to \mathbb{R}$ with[4]

$$\mathcal{W}^* (z; \theta, \gamma, \lambda) := \sum_{i=1}^{I} \left[\gamma z_i - v (z_i/\theta_i)\right] - \sum_{i=1}^{I} \beta_i R (z_i; \theta_i, \theta_{i+1}). \tag{18}$$

$\mathcal{W}^*$ is strictly concave over the convex set $\mathcal{Z}$ because v'' is positive and increasing. Hence, there is a *unique* gross income vector which maximizes $\mathcal{W}^*$ for every

4 Since $\beta_I = 0$, $R\left(z_I; \theta_I, \theta_{I+1}\right)$ is defined arbitrarily without loss of generality. Note that the expression of $\mathcal{W}^* (z; \theta, \gamma, \lambda)$ is justified in the proof of Proposition 2.

$(\theta, \gamma, \lambda) \in \mathcal{P}$. It remains to substitute it in $x_i^* \left(g^z \left(\theta, \gamma, \lambda\right); \theta, \gamma\right)$ to get the optimal allocation.

Proposition 2 *For $(\theta, \gamma, \lambda) \in \mathcal{P}$, there is a unique allocation $a = (g^x (\theta, \gamma, \lambda), g^z (\theta, \gamma, \lambda))$ solution to Problem 2, with*

$$\begin{cases} g_i^z (\theta, \gamma, \lambda) = \arg\max_{z_i \in \mathcal{Z}} \mathcal{W}^* (z; \theta, \gamma, \lambda) \\ g_i^x (\theta, \gamma, \lambda) = x_i^* \left(g^z (\theta, \gamma, \lambda); \theta, \gamma\right) \end{cases}, \qquad i \in \mathcal{I}. \tag{19}$$

A first observation is that both $g^x (\theta, \gamma, \lambda)$ and $g^z (\theta, \gamma, \lambda)$ are functions. Moreover, Proposition 2 implies that the social allocation solution to the optimal nonlinear income tax problem is a monotonic chain to the left. As a consequence, the optimal tax schedule is not differentiable at each observed gross income level z_i. It is nevertheless possible to use the differentiability of the indifference curves in order to define implicit marginal tax rates as

$$T' \left(z_i; \theta_j\right) := 1 - s \left(z_i; \theta_j\right) = 1 - \frac{v' \left(z_i/\theta_j\right)}{\gamma \theta_j}, \qquad (i, j) \in \mathcal{I}^2.\text{[5]} \tag{20}$$

Every $T' \left(z_i; \theta_j\right)$ is less than one because $v' > 0$ and decreases in z_i since $v'' > 0$. Since at the optimum only the adjacent downward incentive-compatibility constraints are binding, two implicit marginal tax rates are of particular interest at each observed gross income level z_i: the implicit marginal tax rate $T' (z_i; \theta_i)$ faced by i for whom the $\left(x_i^* (z_i; \theta, \gamma), z_i\right)$-bundle is designed and the implicit marginal tax rate $T' (z_i; \theta_{i+1})$ the nearest more productive $i+1$-individual would face if he were mimicking the i-individual.

The implicit marginal tax rates allow us to get further understanding of the reduced-form objective function $\mathcal{W}^* (z; \theta, \gamma, \lambda)$. Indeed, let z be a fixed gross income vector and consider that the gross income z_i of the i-individual is increased at the margin. As

$$\frac{\mathrm{d}\mathcal{W}^* (z; \theta, \gamma, \lambda)}{\mathrm{d}z_i} = \gamma T' (z_i; \theta_i) - \beta_i R' (z_i; \theta_i, \theta_{i+1}), \qquad i \in \mathcal{I}, \tag{21}$$

by (18) and (20), the impact on social welfare may be thought of as proceeding in two steps. In the first step, the i-individual pays $T' (z_i; \theta_i)$ additional euros in taxes, which relaxes the tax revenue constraint (6). As γ is the marginal utility of money, the positive effect on social welfare amounts to $\gamma T' (z_i; \theta_i)$. In the second step, the effect on incentives is taken into account. Person i receives $1 - T' (z_i; \theta_i)$ extra euros of consumption. As a result, the $I - i$ more productive individuals have to sacrifice less consumption when they decide to mimic i. So, cheating becomes more attractive to them. For $i + 1$, the marginal rent is $R' (z_i; \theta_i, \theta_{i+1})$. This person's binding incentive-compatibility constraint $IC_{i+1,i}$ is restored by adjusting consumption by

[5] Note that implicit marginal tax rates are also defined at bundles not chosen. This construction is useful in the subsequent analysis.

$R'(z_i; \theta_i, \theta_{i+1})/\gamma$. Increasing the consumption of everybody of higher ability by the same amount preserves the monotonic chain to the left. Because $\gamma\beta_i$ is the net social cost of marginally increasing consumption of the $I - i$ most productive individuals, social welfare is reduced by $\beta_i R'(z_i; \theta_i, \theta_{i+1})$. If the social optimum is interior, it is therefore obtained when the positive effect on social welfare due to the relaxation of the tax revenue constraint offsets the negative one stemming from private information, i.e.

$$\frac{\mathrm{d}\mathcal{W}^*(z; \theta, \gamma, \lambda)}{\mathrm{d}z_i} = 0 \Leftrightarrow T'(z_i; \theta_i) = \frac{\beta_i}{\gamma} R'(z_i; \theta_i, \theta_{i+1}), \qquad i \in \mathcal{I}. \tag{22}$$

3.2. Optimal allocations and bunching

In order to characterize the optimal allocation, it is useful to consider the *relaxed* form of Problem 2, obtained when all monotonicity conditions on z are removed (but not the non-negativity constraint $z \geqq 0$).

Problem 3 (*Relaxed Form*) For $(\theta, \gamma, \lambda) \in \mathcal{P}$, choose $z \geqq 0$ so as to maximize the social objective function $\widehat{\mathcal{W}}(z; \theta, \gamma, \lambda) : \mathbb{R}_+^I \times \mathcal{P} \to \mathbb{R}$ with

$$\widehat{\mathcal{W}}(z; \theta, \gamma, \lambda) := \sum_{i=1}^{I} \left[\gamma z_i - v(z_i/\theta_i)\right] - \sum_{i=1}^{I} \beta_i R(z_i; \theta_i, \theta_{i+1}). \tag{23}$$

This problem is closely related to Problem 2. Its first-order conditions are:

$$\partial\widehat{\mathcal{W}}(z; \theta, \gamma, \lambda)/\partial z_i \leq 0 \ (= 0 \text{ if } z_i > 0), \qquad i = 1, \ldots, I. \tag{24}$$

Note that $\widehat{\mathcal{W}}$ is strictly concave over the convex set $\mathbb{R}_+^I$ because v'' is positive and increasing. Therefore, the first-order conditions (24) are both necessary and sufficient and yield a *unique* gross income vector, solution to Problem 3. This solution is denoted $\widehat{z}(\theta, \gamma, p) = (\widehat{z}_1(\theta, \gamma, p), \ldots, \widehat{z}_I(\theta, \gamma, p))$and satisfies

$$T'(\widehat{z}_i; \theta_i) \leq \frac{\beta_i}{\gamma} R'(\widehat{z}_i; \theta_i, \theta_{i+1}) \ (= 0 \text{ if } \widehat{z}_i > 0), \qquad i \in \mathcal{I}. \tag{25}$$

If $\widehat{z}(\theta, \gamma, p)$ is non-decreasing, then Problems 2 and 3 have both the same solution; in other words, $\widehat{z}(\theta, \gamma, p) = g^z(\theta, \gamma, \lambda)$. Otherwise, the optimal allocation involves bunching. If that case, let $\mu_1, \ldots, \mu_I$ be the Kuhn-Tucker multipliers of the different constraints in $\mathcal{Z}$ and consider the Lagrangian of Problem 2,

$$\mathcal{L} = \mathcal{W}^*(z; \theta, \gamma, \lambda) + \mu_1 z_1 + \mu_2(z_2 - z_1) + \cdots + \mu_I(z_I - z_{I-1}). \tag{26}$$

The corresponding first-order conditions are:

$$\partial \mathcal{L}/\partial z_i = \partial \mathcal{W}^* (z; \theta, \gamma, \lambda) / \partial z_i + \mu_i - \mu_{i+1} = 0, \qquad i = 1, \ldots, I-1, \quad (27)$$

$$\partial \mathcal{L}/\partial z_I = \partial \mathcal{W}^* (z; \theta, \gamma, \lambda) / \partial z_I + \mu_I = 0, \quad (28)$$

$$\mu_1 \geq 0 \ (= 0 \text{ if } z_1 > 0), \quad (29)$$

$$\mu_i \geq 0 \ (= 0 \text{ if } z_i > z_{i-1}), \qquad i = 2, \ldots, I. \quad (30)$$

Now assume that person i is not bunched together with anyone else at the social optimum, in the sense that $0 \leq z_{i-1} < z_i < z_{i+1}$. Then, (30) implies that $\mu_i = \mu_{i+1} = 0$. Therefore, by (27), z_i satisfies (22). Because $z_i > 0$, it is clear that (25) yields the same solution as (22). The following Lemma is obtained.

Lemma 1 *For $(\theta, \gamma, \lambda) \in \mathcal{P}$, if person i is not bunched together with anyone else at the social optimum, then $g_i^z(\theta, \gamma, \lambda) = \widehat{z}_i$.*

Two types of bunching must be distinguished. (a) Bunching due to the violation of the non-negativity constraints, called $z = 0$ bunching by Boone and Bovenberg (2007), can only occur at the bottom, in which case the individuals bunched together have gross income equal to 0. (b) The other kind of bunching stems from the violation of the monotonicity constraints and can happen either at the bottom or in the interior of the skill distribution.

To gain further insight, let us assume that persons $n = i, \ldots, i + \kappa - 1$, are all bunched together but not with anyone else, and denote by $\overline{z}$ their gross income level. If there is $z = 0$ bunching, then: $i = 1$, $\mu_1 > 0$ and $\mu_{1+\kappa} = 0$. Summing (27) for $n = 1, \ldots, \kappa$ and using (30), one gets:

$$\frac{1}{\kappa} \sum_{n=1}^{\kappa} \partial \mathcal{W}^* (z; \theta, \gamma, \lambda) / \partial z_n \Big|_{z_n = 0} < 0. \quad (31)$$

If there is bunching due to the violation of the monotonicity constraints, then $\mu_i = \mu_{i+\kappa} = 0$. Summing again (27) for $n = i, \ldots, i + \kappa - 1$, one obtains:

$$\frac{1}{\kappa} \sum_{n=i}^{i+\kappa-1} \frac{\partial \mathcal{W}^* (z; \theta, \gamma, \lambda)}{\partial z_n} \Bigg|_{z_n = \overline{z}} = 0. \quad (32)$$

Consequently, the optimal gross income vector is solution to the relaxed Problem 2, except on a number of bunching sets where it is characterized by (31) and (32). Loosely speaking, the first-order conditions of Problem 2 are "decoupled" in the absence of bunching and only re-coupled inside an interval of "bunch". The following proposition summarizes these results.

Proposition 3 *For $(\theta, \gamma, \lambda) \in \mathcal{P}$, the optimal allocation is such that:*

(i) *$g_i^z(\theta, \gamma, \lambda) = \widehat{z}_i(\theta, \gamma, p)$, except on a number K of disjoint sets $\mathcal{B}^k := \{i^k, \ldots, j^k\}$, $k = 1, \ldots, K$, i^k increasing with k, where $g_n^z(\theta, \gamma, \lambda) = \overline{z}^k$ for every $n \in \mathcal{B}^k$;*

(ii) $0 < T'(z_i;\theta_i) < 1$ *for* $i = 1,\ldots,I-1$, $T'(z_I;\theta_I) = 0$ *and bunching at the top is ruled out;*

(iii) *if* $i^k = 1$,

$$\frac{1}{\#\mathcal{B}^k}\sum_{n\in\mathcal{B}^k}\left.\frac{\partial\mathcal{W}^*(z;\theta,\gamma,\lambda)}{\partial z_n}\right|_{z_n=\bar{z}^k} \le 0 \left(= 0 \text{ if } \bar{z}^k > 0\right); \tag{33}$$

(iv) *for each interior* $\mathcal{B}^k$,

$$\frac{1}{\#\mathcal{B}^k}\sum_{n\in\mathcal{B}^k}\left.\frac{\partial\mathcal{W}^*(z;\theta,\gamma,\lambda)}{\partial z_n}\right|_{z_n=\bar{z}^k} = 0. \tag{34}$$

The fact that there is no bunching and no distortion at the top follows from a well-known efficiency argument (cf. Seade 1977; Guesnerie and Seade 1982). Note that, when $I = 2$, Proposition 3 implies that the optimal allocation is fully separating.

To gain further insights into the optimal allocation, it is useful to define the function $\alpha_i : [0, \theta_i v'^{-1}(\gamma\theta_i)[\to \mathbb{R}_+$ as

$$\alpha_i(z_i;\theta_i,\theta_{i+1},\gamma) := \frac{R'(z_i;\theta_i,\theta_{i+1})}{\gamma T'(z_i;\theta_i)} = \frac{T'(z_i;\theta_{i+1}) - T'(z_i;\theta_i)}{T'(z_i;\theta_i)},\quad i = 1,\ldots,I-1. \tag{35}$$

For $i < I$, the domain of α_i corresponds to the gross incomes for which $T'(z_i;\theta_i)$ is strictly positive; hence, by Proposition 3, $g_i^z(\theta,\gamma,\lambda) \in [0, \theta_i v'^{-1}(\gamma\theta_i)[$. Moreover, α_i is continuous and strictly increasing over its domain.[6] Using (22) and (35), $\widehat{z}$ must satisfy

$$\alpha_i(\widehat{z}_i;\theta_i,\theta_{i+1},\gamma) \ge \frac{1}{\beta_i}\ (= \text{if } \widehat{z}_i > 0),\qquad i = 1,\ldots,I-1. \tag{36}$$

So, if $\widehat{z}_i > 0, \widehat{z}_i = \alpha_i^{-1}(1/\beta_i;\theta_i,\theta_{i+1},\gamma)$. Since the optimal gross income vector $g^z(\theta,\gamma,\lambda)$ lies in the interior of $\mathcal{Z}$ if and only if $0 < \widehat{z}_1 < \cdots < \widehat{z}_I$, the following characterization is obtained.

Proposition 4 *For* $(\theta,\gamma,\lambda) \in \mathcal{P}$, *the optimum is fully separating if and only if* $\beta_i \in \alpha_i\left([0, \theta_i v'^{-1}(\gamma\theta_i)[\right)$ *for* $i = 1,\ldots,I-1$ *with*

$$0 < \alpha_1^{-1}(1/\beta_1;\theta_1,\theta_2,\gamma) < \ldots < \alpha_{I-1}^{-1}(1/\beta_{I-1};\theta_{I-1},\theta_I,\gamma) < \theta_I v'^{-1}(\gamma\theta_I). \tag{37}$$

This proposition provides a necessary and sufficient condition under which bunching is either or not optimal in terms of the *primitives* of the model β, θ and γ. There

[6] For $i < I$, α_i is strictly increasing since $dT'(z_i;\theta_i)/dz_i < 0$ and $R''\left(z_i;\theta_i,\theta_{i+1}\right) > 0$.

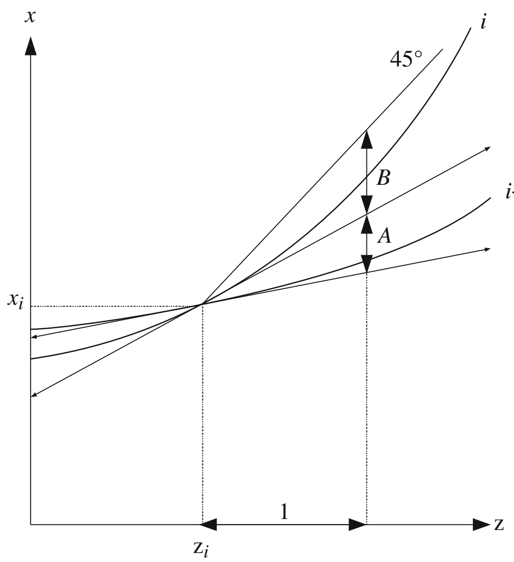

Fig. 1 $\alpha_i \left(z_i; \theta_i, \theta_{i+1}, \gamma\right) = A/B$

is no analogue condition in literature considering a continuous population. The set of parameters for which the optimum is fully separating is denoted $\mathcal{P}^0$.

3.3. Separating optimum allocations

When the primitives of the model imply that bunching is suboptimal, the optimality conditions—obtained from (36)—are strikingly simple.

Proposition 5 *For* $(\theta, \gamma, \lambda) \in \mathcal{P}^0$, z *is socially optimal if and only if*

$$\alpha_i \left(z_i; \theta_i, \theta_{i+1}, \gamma\right) = 1/\beta_i, \qquad i = 1, \ldots, I-1, \tag{38}$$

and $T' \left(z_I; \theta_I\right) = 0$.

For a gross income z_i, the wedge $\alpha_i \left(z_i; \theta_i, \theta_{i+1}, \gamma\right)$ tells us to which extent person $i+1$'s implicit marginal tax rate is higher than person i's one. Geometrically, it thus corresponds to the tangent of the angle between i's and $i+1$'s indifference curves divided by $T' \left(z_i; \theta_i\right)$, as shown in Fig. 1. It is closely related to the single-crossing condition and thus henceforth referred to as the *Spence-Mirrlees wedge*. Indeed, the single-crossing condition corresponds to a restriction on its *sign*, which must be strictly positive. The conditions for social optimality (38) introduce an additional restriction on its *magnitude*: an allocation is socially optimal only if, at each observed gross income level z_i, the Spence-Mirrlees wedge is equal to the inverse of the cumulative social weight β_i.

Proposition 5 allows a simple two-step geometric construction of the optimal allocation, as illustrated in Fig. 2. *In the first step*, the tax revenue constraint is ignored.

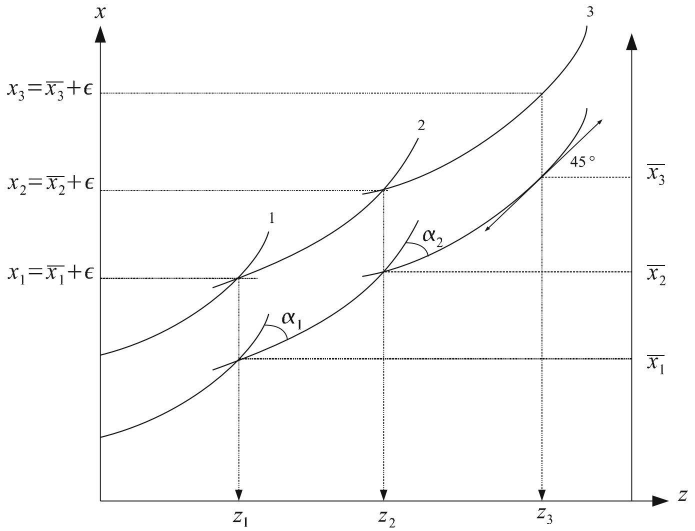

Fig. 2 Construction of the optimal allocation

Start at $(0, \widehat{x}_1)$, where $\widehat{x}_1$ is arbitrarily chosen. Move along θ_1's indifference curve until $\alpha_1(z_1; \theta_1, \theta_2, \gamma) = 1/\beta_1$. This point is $(\overline{x}_1, z_1)$. Then, move along θ_2's indifference curve through $(\overline{x}_1, z_1)$ until $\alpha_2(z_2; \theta_2, \theta_3, \gamma) = 1/\beta_2$. This point is $(\overline{x}_2, z_2)$. And so on until $i = I - 1$. The determination of z_I exploits the no-distortion-at-the-top result: starting from $(\overline{x}_{I-1}, z_{I-1})$, gross income is increased along the indifference curve of the most productive individual until the tangent to this line has slope one. By construction, the allocation $(\overline{x}, z)$ is a SMCL and is, thus, incentive compatible. However, it is not necessarily budget-balanced. So, *in the second step*, each $\overline{x}_i$ is varied by a same amount $\varepsilon = \frac{1}{I}\sum_{i=1}^{I}(z_i - \overline{x}_i)$ so as to get a binding tax revenue constraint. The resulting allocation $(x, z) = (\overline{x} + \varepsilon, z)$ is socially optimal.

4. Comparative static properties

The Spence-Mirrlees wedge allows us to examine the impact of changing the parameters of the model at the margin. For expositional reasons, the analysis will concentrate on the comparative statics of separating allocations. Yet, using Proposition 3, it can be shown that the results derived below remain valid when there is bunching at the optimum, provided changes in $(\theta, \gamma, \lambda)$ do not modify the sets $\mathcal{B}^k$ of individuals bunched together. The next property is particular useful for the following analysis.

Proposition 6 *The functions g^x and g^z are $\mathcal{C}^1$ at every $(\theta, \gamma, \lambda)$in $\mathcal{P}^0$.*

4.1. Changing the marginal utility of money

The marginal utility of money γ measures the intensity of the individual preference for private consumption. When it goes up, the marginal utility of leisure expressed in consumption good is reduced. So, the indifference curves of every individual become flatter in the (z, x)-space. Everyone is thus willing to sacrifice more leisure to obtain a certain amount of additional consumption. As a consequence, changing γ at the margin casts light on how taxes should be adjusted in a country where all individuals would like to work more to consume more.

Proposition 7 *For* $(\theta, \gamma, \lambda) \in \mathcal{P}^0$ *and* $i \in \mathcal{I}$, $\partial g_i^z(\theta, \gamma, \lambda) / \partial \gamma > 0$.

Hence, the adjustment in the tax system should not discourage more hard-working individuals to earn more. National income is increased and thus, since production efficiency is preserved by Lemma 1, total consumption goes up.

4.2. Changing individual productivities

In Mirrlees's model, individuals are born with different abilities to turn effort into output. These fixed skill levels are the sole source of heterogeneity within the population. They constitute the most basic ingredients of the model because they give rise to the adverse selection problem: indeed, the effort level z_j/θ_i a given i-individual must provide to earn the gross income of everyone else directly depends on his productivity. In practice, skill levels are subject to changes. For instance, a highly skilled individual can catch an illness which impairs his productivity while a low skilled can benefit from on-the-job training. Another argument rests on the technological side of the economy: since a person's productivity depends on the capitalistic intensity in his branch of activity, a new investment can make him more productive. The impact of changing a person's skill level on his own choices, but also on the policymaker and the whole population, is as follows.[7]

Proposition 8 *For* $(\theta, \gamma, \lambda) \in \mathcal{P}^0$ *and* $(i, j) \in \{1, \ldots, I-1\} \times \mathcal{I}$,

$$\partial z_i / \partial \theta_{i+1} < 0, \tag{39}$$

$$\partial z_{i+1} / \partial \theta_{i+1} > 0, \tag{40}$$

$$\partial z_j / \partial \theta_{i+1} = 0 \quad \textit{for} \, j \notin \{i, i+1\}, \tag{41}$$

and

$$dT'(z_i; \theta_i) / d\theta_{i+1} > 0, \tag{42}$$

$$dT'(z_{i+1}; \theta_{i+1}) / d\theta_{i+1} < 0 \quad \textit{when} \quad i \neq I-1 \; (= 0 \quad \textit{when} \quad i = I-1), \tag{43}$$

$$dT'\left(z_j; \theta_j\right) / d\theta_{i+1} = 0 \quad \textit{for} \; j \notin \{i, i+1\}, \tag{44}$$

Increasing person $i+1$'s productivity only alters his gross income and that of his nearest less productive neighbour. Indeed, by Proposition 5, only the wedges $\alpha_i(z_i; \theta_i, \theta_{i+1}, \gamma)$ and $\alpha_{i+1}(z_{i+1}; \theta_{i+1}, \theta_{i+2}, \gamma)$ depend on θ_{i+1}. Accordingly, the gross income levels of all other individuals, with $j \notin \{i, i+1\}$, remain unaltered. As regards persons i and $i+1$, the adjustment process combines *three effects*.

[7] The fact that the productivity vector θ is strictly monotonically increasing ensures that (1) remains satisfied once a given individual productivity is changed at the margin.

First, the variation in θ_{i+1} gives rise to a *local substitution effect*. The increase in person $i+1$'s productivity results in a rise in his net-of-tax wage rate, which leads him to increase his labour supply in efficiency units, z_{i+1}.

Second, changing θ_{i+1} has an *incentive effect*. As he becomes more efficient, person $i+1$ has to provide less effort if he wants to imitate person i. Consequently, his indifference curve through person i's gross-income/consumption bundle flattens. By (20), this corresponds to an increase in the implicit marginal tax rate $T'(x_i, z_i; \theta_{i+1})$ he would face if he were cheating.

Third, person i incurs an *informational externality* induced by the incentive effect. Since the cumulative social weight β_i is unaltered, the wedge $\alpha_i(z_i; \theta_i, \theta_{i+1}, \gamma)$ must stay constant (Proposition 5). Consequently, the increase in $T'(x_i, z_i; \theta_{i+1})$ must be associated with an increase in the implicit marginal tax rate $T'(x_i, z_i; \theta_i)$ and thus with a reduction in person i's net-of-tax wage rate. Finally, the substitution effect leads person i to work less.

The changes in gross income ensure that a new monotonic chain to the left a' is obtained. However, this incentive-compatible allocation is not necessarily budget-balanced. Therefore, in a second step, all consumption levels are adjusted by a same amount in order to obtain a production-efficient allocation. This corresponds to a vertical displacement of all indifference curves through the bundles of a'. However, the direction of this displacement cannot be signed in the general case, which explains why changes in consumption are ambiguous.

4.3. Changing social weights

Social and political changes often promote greater social concern for some groups of the population. We examine the impact of increasing person i's social weight while all other welfare weights λ_j, with $j \neq i$, are scaled down proportionally. This is equivalent to increasing one welfare weight, holding the others constant, in the absence of the normalization rule (11).

These changes in the individual social weights result in variations in the net cumulative social weights β: for every $j < i$, the decrease in λ_j implies a reduction in the cumulative social weight Λ_j and thus in β_j; in contrast, since the magnitude of the raise in λ_i is equivalent to the reduction in all other social weights in the population, Λ_j and β_j go up for every j such that $i \leq j < I$. By Proposition 5, (i) $g^z_j(\theta, \gamma, \lambda)$ is independent of β_i for $i \neq j$; (ii) a rise in β_i reduces $g^z_i(\theta, \gamma, \lambda)$ because $\alpha_i(.; \theta_i, \theta_{i+1}, \gamma)$is strictly increasing around the optimum as previously noted. Increasing person i's social weight thus results in the following changes.

Proposition 9 *Let $(\theta, \gamma, \lambda) \in \mathcal{P}^0$, $z \equiv g^z(\theta, \gamma, \lambda)$and $i \in \mathcal{I}$. Consider that person i's social weight is increased at the margin while all other welfare weights λ_j, with $j \neq i$, are scaled down proportionally. Then,*

(i) *z_j is increased and $T'\left(z_j; \theta_j\right)$ goes down for every $j < i$;*
(ii) *if $i \neq I$, z_j is decreased and $T'\left(z_j; \theta_j\right)$ goes up for $i \leq j \leq I-1$;*
(iii) *z_I and $T'(z_I; \theta_I)$ are unaltered.*

4.4. Productivity change accompanied by a social weight adjustment

Up to now, we have considered changes in productivity without any impact on individual social weights and changes in social weights independent of variations in skill levels. However, since the policymaker wants to redistribute incomes from the more to the less productive, he might be less concerned for the well-being of an individual who becomes more productive and reduce the social weight of the latter accordingly. For concreteness, let us consider that the policymaker responds to a marginal increase in person $i+1$'s productivity by a slight decrease in λ_{i+1}. The overall effect on the optimal gross incomes and implicit marginal tax rates follow from the combination of Propositions 8 and 9.

Proposition 10 *Let* $(\theta, \gamma, \lambda) \in \mathcal{P}^0$ *and* $(i, j) \in \{1, \ldots, I-1\} \times \mathcal{I}$. *Consider a marginal increase in* θ_{i+1} *accompanied by a slight decrease in* λ_{i+1}, *whilst all other social weights are scaled up proportionally. Then,*

(i) *for every* $j \leq i$, z_j *is reduced and* $T'\left(z_j; \theta_j\right)$ *is increased;*
(ii) *for every* $j > i$, z_j *is increased;*
(iii) *for every* $i < j < I$, $T'\left(z_j; \theta_j\right)$ *is reduced;*
(iv) $T'\left(z_I; \theta_I\right)$ *is unaltered.*

5. Extension to nonlinear pricing

The Spence-Mirrlees wedge appears as a key determinant of the properties of the optimal solution to the optimal income tax in the above quasilinear economy. This section shows how the analysis extends to generic adverse-selection problems (see, e.g. Laffont and Martimort 2002; Salanié 2005). One considers a generic situation where a 'principal' contracts with an 'agent' to produce a certain amount of some good z_i and compensates the agent with a monetary transfer x_i. The agent knows his type θ_i before signing the contract. The cost he incurs when producing z_i is given by $v\left(\theta_i \ell_i\right) \equiv v\left(z_i\right)$, where ℓ_i represents his effort level. His utility is given by $u\left(x_i, z_i; \theta_i\right) := x_i - v\left(z_i\right)$. The principal only knows that type θ_i occurs with probability λ_i / I. He designs a menu of contracts which maximizes his expected utility in the exchange with the agent.

Problem 4 (*Optimal Nonlinear Pricing*) For $(\theta, \gamma, \lambda) \in P$, choose a menu of contracts $a \in \mathbb{R}^I \times \mathbb{R}_+^I$ to maximize $\sum_{i=1}^{I} \lambda_i \left[h\left(z_i\right) - x_i\right]$, with $h' > 0$ and $h'' < 0$, subject to the incentive compatibility constraints (5) and the participation constraints

$$\gamma x_i - v\left(z_i / \theta_i\right) \geq 0, \qquad i = 1, \ldots, I. \tag{45}$$

The marginal profit $T'\left(z_i; \theta_j\right)$ is now defined as

$$T'\left(z_i; \theta_j\right) := h'\left(z_i\right) - s\left(z_i; \theta_j\right) = h'\left(z_i\right) - \frac{v'\left(z_i / \theta_j\right)}{\gamma \theta_j}. \qquad (i, j) \in \mathcal{I}^2. \tag{46}$$

Because any optimum menu of contracts is a SMCL, (15) applies. Moreover, the agent's participation constraint in state 1 is the only one to be binding; it plays the same part in determining $x_1^*(z;\theta,\gamma)$ as the tax revenue constraint, in the optimal income tax problem. So, $x_1^*(z;\theta,\gamma) = v(z_1/\theta_1)/\gamma$. Using these observations, the principal's reduced-form objective function is

$$\mathcal{W}^*(z;\theta,\gamma,\lambda) = -\frac{I}{\gamma}v\left(\frac{z_1}{\theta_1}\right) + \sum_{i=1}^{I}\lambda_i h(z_i)$$
$$-\frac{1}{\gamma}\sum_{i=2}^{I}\left(\sum_{j=i}^{I}\lambda_j\right)\left[v\left(\frac{z_i}{\theta_i}\right) - v\left(\frac{z_{i-1}}{\theta_i}\right)\right].$$

Because $\mathcal{W}^*$ is strictly concave, the necessary and sufficient conditions for an interior maximum are

$$\frac{\partial\mathcal{W}^*(z;\theta,\gamma,\lambda)}{\partial z_i} = 0 \Leftrightarrow \lambda_i\left(h'(z_i) - \frac{v'(z_i/\theta_i)}{\gamma\theta_i}\right) = \frac{1}{\gamma}\left(\sum_{j=i+1}^{I}\lambda_j\right)\left[\frac{v'(z_i/\theta_i)}{\theta_i} - \frac{v'(z_i/\theta_{i+1})}{\theta_{i+1}}\right]$$
$$\Leftrightarrow \frac{R'(z_i;\theta_i,\theta_{i+1})}{\gamma T'(z_i;\theta_i)} = \frac{\lambda_i}{I-\Lambda_i} \Leftrightarrow \alpha_i(z_i;\theta_i,\theta_{i+1},\gamma) = \frac{1}{\beta_i}, \qquad i = 1,\ldots,I-1, \tag{47}$$

where β_i is defined as

$$\beta_i := \frac{I-\Lambda_i}{\lambda_i}, \qquad i = 1,\ldots,I-1, \tag{48}$$

and $T'(z_I;\theta_I) = 0$. The analysis of the previous sections thus extends as follows.

Proposition 11 *The menu of contracts solution to Problem 4 is characterized by Propositions 3, 4 and 5 with* $T'\left(z_i;\theta_j\right)$ *and* β_i *defined by* (46) *and* (48).

Note that the inverse of the hazard rate $(I-\Lambda_i)/\lambda_i$ plays the same part, in the characterization of the optimum menu of contracts, as the net cumulative social weight in the characterization of the optimum income tax. Because the utility of the agent is not linear, the monotonicity of the hazard rate is not sufficient to prevent bunching. Thanks to the Spence-Mirrlees wedge an alternative condition is obtained: bunching is optimal if and only if the hazard rates do not satisfy (37). When the optimal menu of contracts is separating, the optimal contract is characterized by the equality between the Spence-Mirrlees wedge and the inverse of the hazard rate for all $i = 1,\ldots,I-1$.

The comparative static properties of the optimum menu of contracts are the following provided $(\theta,\gamma,\lambda) \in \mathcal{P}^0$. Proposition 11 implies that the impact of increasing in γ and θ_i is given by Propositions 7 and 8. Proposition 9 must be adapted because an increase in λ_i has not the same effects on the hazard rates as on the cumulative social weights. Let us consider a rise in the probability of state i accompanied by proportional adjustments in all other probabilities: β_j increases for all $j < i$ and decreases otherwise. By Proposition 11, z_j is independent of β_i for $i \neq j$ while a rise in β_j reduces z_j around the optimum. Therefore, in the optimum, the principal decides to ask the agent to produce more output in all states of nature with a lower β_j and less output in all other states of nature.

6. Conclusion

This article introduces the notion of Spence-Mirrlees wedge and shows its usefulness in the analysis of two important classes of incentives problems. In the optimal income tax problem, this wedge $\alpha_i\left(z_i;\theta_i,\theta_{i+1},\gamma\right)$ indicates to which extent person $i+1$'s marginal tax rate is higher than person i's one when both earn gross income z_i. In the standard adverse-selection problem, it tells us to which extent, for a given output level z_i, the marginal profit is higher in state $i+1$ than in state i. In both models, the Spence-Mirrlees wedge is a key determinant of the optimum menu of contracts. It allows the derivation of a necessary and sufficient condition, expressed in terms of the primitives of the models, under which bunching is optimal. It also allows a simple geometric construction of the separating menus of contracts as well as the derivation of their comparative static properties. These observations suggest that the Spence-Mirrlees wedge might be a useful tool to cast light on the structure of other screening problems in which incentive compatibility constraints are binding.

Acknowledgements I would like to thank two anonymous reviewers, the editor, Craig Brett, Antoine d'Autume, Julia Cagé, Guy Gilbert, Laurence Jacquet, Guy Laroque, Jacques Le Cacheux, Etienne Lehmann, Pierre Pestieau, Alain Trannoy, Gwenola Trotin and John A. Weymark for their very helpful comments and suggestions. The usual caveat applies.

Appendix

Proof of Proposition 1 (i.a) *a is a SMCL.* Assume a is not a SMCL, i.e. $u\left(x_j^*,z_j;\theta_j\right) \neq u\left(x_{j-1}^*,z_{j-1};\theta_j\right)$ for some $j \geq 2$. This is equivalent to considering that there exists $j \geq 2$ for which

$$\gamma x_j^* - v\left(z_j/\theta_j\right) > \gamma x_{j-1}^* - v\left(z_{j-1}/\theta_j\right). \tag{A.1}$$

For (A.1) to be satisfied, $\left(x_j^*,z_j^*\right) >> \left(x_{j-1},z_{j-1}\right)$. Hence,

(i) is established if $z_j \leq z_{j-1}$. If $z_j > z_{j-1}$, let $\overline{x}_i = x_i^* + \varepsilon$ for $i = 1,\ldots,j-1$, and $\overline{x}_i = x_i^* - \frac{j-1}{I-j+1}\varepsilon$ for $i = j,\ldots,I$, where $\varepsilon > 0$ is arbitrarily chosen. The new allocation $(\overline{x},z)$ satisfies all incentive compatibility constraints for sufficiently small ε and is feasible because $\sum_i \overline{x}_i = \sum_i x_i^*$. In addition,

$$W\left(\overline{x},z\right) - W\left(x^*,z\right) = \gamma \sum_{i=1}^{I} \lambda_i\left(\overline{x}_i - x_i^*\right)$$

$$= \gamma \left[\sum_{i=1}^{j-1} \lambda_i - \frac{j-1}{I-j+1} \sum_{i=j}^{I} \lambda_i \right] \varepsilon, \quad \text{(A.2)}$$

which can be minored thanks to (10) to get $W(\overline{x}, z) - W(x^*, z) \geq \gamma(j-1)\varepsilon\left[\lambda_{j-1} - \lambda_j\right] > 0$, contradicting $x^* \in \mathcal{X}^*(\overline{z}; \theta, \gamma, \lambda)$.

(i.b) *ais budget-balanced*. Fix $\overline{z}$ in $\mathcal{Z}$. The constraints (5) are satisfied, with $z_i = \overline{z}_i$ and $x_i = x_i^*$. Assume (6) is not binding. So, (6) is still satisfied if every x_i^* is increased by a sufficiently small $\varepsilon > 0$. This Pareto-improving increase is incentive compatible since a same amount $\gamma\varepsilon$ is added to both sides of $IC_{i,j}$ for every $(i, j) \in \mathcal{I}^2$, contradicting $x^* \in \mathcal{X}^*(\overline{z}; \theta, \gamma, \lambda)$.

(ii) Because a is a SMCL by (i.a),

$$x_i = x_{i-1} + \frac{1}{\gamma}\left[v(z_i/\theta_i) - v(z_{i-1}/\theta_i)\right], \qquad i = 2, \ldots, I, \quad \text{(A.3)}$$

and, thus,

$$x_i = x_1 + \frac{1}{\gamma}\sum_{j=2}^{i}\left[v\left(z_j/\theta_j\right) - v\left(z_{j-1}/\theta_j\right)\right], \qquad i = 2, \ldots, I. \quad \text{(A.4)}$$

By (i.a), the equality form of (6) can be substituted in $\sum_{i=1}^{I} x_i$ obtained from (A.4):

$$\sum_{i=1}^{I} z_i = I x_1 + \frac{1}{\gamma}\sum_{i=2}^{I}\sum_{j=2}^{i}\left[v\left(\frac{z_j}{\theta_j}\right) - v\left(\frac{z_{j-1}}{\theta_j}\right)\right] = I x_1 + \frac{1}{\gamma}\sum_{i=2}^{I}(I+1-i)\left[v\left(\frac{z_i}{\theta_i}\right) - v\left(\frac{z_{i-1}}{\theta_i}\right)\right]. \quad \text{(A.5)}$$

This equation has a unique solution in x_1, that is substituted in (A.4). □

Proof of Proposition 2 It is sufficient to establish that substitution of $x^*(z; \theta, \gamma)$ into W yields $\mathcal{W}^*$ for every $z \in \mathcal{Z}$. By (7),

$$u\left(x_i^*, z_i; \theta_i\right) = u\left(x_1^*, z_1; \theta_1\right) + \sum_{j=1}^{i-1} R\left(z_j; \theta_j, \theta_{j+1}\right), \qquad i = 2, \ldots, I, \quad \text{(A.6)}$$

from which

$$\sum_{i=1}^{I} u\left(x_i^*, z_i; \theta_i\right) = I u\left(x_1^*, z_1; \theta_1\right) + \sum_{i=1}^{I-1}(I-i)\, R\left(z_i; \theta_i, \theta_{i+1}\right). \quad \text{(A.7)}$$

In addition, summing (3) over i on $\mathcal{I}$ and employing the equality form of (6),

$$\sum_{i=1}^{I} u\left(x_i^*, z_i; \theta_i\right) = \gamma \sum_{i=1}^{I} z_i - \sum_{i=1}^{I} v\left(z_i/\theta_i\right). \tag{A.8}$$

Plugging (A.8) in (A.7) and solving for $u\left(x_1^*, z_1; \theta_1\right)$,

$$u\left(x_1^*, z_1; \theta_1\right) = \frac{1}{I}\left[\gamma \sum_{i=1}^{I} z_i - \sum_{i=1}^{I} v\left(\frac{z_i}{\theta_i}\right) - \sum_{i=1}^{I-1}(I-i)\, R\left(z_i; \theta_i, \theta_{i+1}\right)\right]. \tag{A.9}$$

Using (A.6) and (A.9),

$$\begin{aligned} W &= u\left(x_1^*, z_1; \theta_1\right)\sum_{i=1}^{I}\lambda_i + \sum_{i=2}^{I}\sum_{j=1}^{i-1}\lambda_i R\left(z_j; \theta_j, \theta_{j+1}\right) \\ &= u\left(x_1^*, z_1; \theta_1\right) I + \sum_{i=1}^{I-1}\left(\sum_{j=i+1}^{I}\lambda_j\right) R\left(z_i; \theta_i, \theta_{i+1}\right) \\ &= \sum_{i=1}^{I}\left[\gamma z_i - v\left(\frac{z_i}{\theta_i}\right) + \left(i - \sum_{j=1}^{i}\lambda_j\right) R\left(z_i; \theta_i, \theta_{i+1}\right)\right], \end{aligned} \tag{A.10}$$

in which $R\left(z_I; \theta_I, \theta_{I+1}\right)$ is an arbitrary number. □

Proof of Proposition 3 (i) If $i \notin \mathcal{B}^k \forall k$, $z_{i-1} < z_i < z_{i+1}$. Then, by (30), $\mu_i = \mu_{i+1} = 0$. Therefore, by (27), z_i satisfies $\partial \mathcal{W}^*\left(z; \theta, \gamma, \lambda\right)/\partial z_i = 0$, whose unique solution is $\widehat{z}_i$. Hence, $z_i = \widehat{z}_i$. (ii) See Guesnerie and Seade (1982, Proposition 7). (iii)–(iv) Assume $\left\{i^k, \ldots, j^k\right\}$ are bunched together at the optimum. *Case (a): bunching at the bottom due to the constraint* $\geqq 0$. $\mathcal{B}^k = \left\{1, \ldots, j^1\right\}$, $\mu_1 > 0$ and $\mu_{j^k+1} = 0$. Summing (27) for $n = 1, \ldots, j^k$,

$$\sum_{n\in\mathcal{B}^k} \partial \mathcal{W}^*\left(z; \theta, \gamma, \lambda\right)/\partial z_n\Big|_{z_n = \overline{z}^k} = -\mu_1 \leq 0, \tag{A.11}$$

that is divided by $\#\mathcal{B}^k > 0$. *Case (b): bunching due to the violation of the monotonicity constraints.* Because $\mu_{i^k} = \mu_{j^k+1} = 0$, summing (27) for $n = i^k, \ldots, j^k$, yields the equality form of (A.11). □

Proof of Proposition 6 By Proposition 5, $g_i^z\left(\theta, \gamma, \lambda\right) = \alpha_i^{-1}\left(1/\beta_i; \theta_i, \theta_{i+1}, \gamma\right)$for $i < I$ and $g_I^z\left(\theta, \gamma, \lambda\right) = \theta_I v'^{-1}\left(\gamma\theta_I\right)$. Since $\alpha_i' > 0$ and $v'' > 0$, $g^z\left(\theta, \gamma, \lambda\right) \in \mathcal{C}^1$. Thus, by Proposition 1, $g^x\left(\theta, \gamma, \lambda\right) \in \mathcal{C}^1$. □

Proof of Proposition 7 Let $z \equiv g^z(\theta, \gamma, \lambda)$. For $i = I, z_I = \theta_I v'^{-1}(\gamma\theta_I)$ because $T'(z_I; \theta_I) = 0$ by Proposition 3. Since $v'' > 0, \partial z_I/\partial\gamma > 0$. For $i \leq i - 1$, let $\phi_i := (1/\beta_i) - \alpha_i(z_i; \theta_i, \theta_{i+1}, \gamma)$ and note that $\partial\phi_i/\partial\gamma > 0$. Since $v''' > 0$ and $0 < \theta_i < \theta_{i+1}$,

$$\frac{\partial\phi_i}{\partial z_i} = \frac{\beta_i}{\theta_{i+1}^2} v''\left(\frac{z_i}{\theta_{i+1}}\right) - \frac{1+\beta_i}{\theta_i^2} v''\left(\frac{z_i}{\theta_i}\right) < 0. \tag{A.12}$$

By the implicit function theorem, $\phi_i \equiv 0$ defines z_i as a $\mathcal{C}^1$-function of γ, $z_i = \varphi_i^\gamma(\gamma)$ with

$$\frac{\partial z_i}{\partial\gamma} \equiv \frac{\mathrm{d}\varphi_i^\gamma(\gamma)}{\mathrm{d}\gamma} = -\frac{\partial\phi_i/\partial\gamma}{\partial\phi_i/\partial z_i} > 0. \tag{A.13}$$

□

Proof of Proposition 8 Let $z \equiv g^z(\theta, \gamma, \lambda)$. It is clear from Proposition 5 that z_j does not depend on θ_{i+1} except for $j = i, i+1$. Hence, $\partial z_j/\partial\theta_{i+1} = 0$ for $j \notin \{i, i+1\}$, implying (44).

It remains to examine the effect of a change in θ_{i+1} on z_i and z_{i+1}. If $\theta_{i+1} < \theta_I$, $\phi_i \equiv 0$ implicitly defines z_i and z_{i+1} as $\mathcal{C}^1$-functions of θ_{i+1}, $z_i = \varphi_i^\theta(\theta_{i+1})$ and $z_{i+1} = \varphi_{i+1}^\theta(\theta_{i+1})$, respectively, with derivatives

$$\frac{\mathrm{d}\varphi_i^\theta(\theta_{i+1})}{\mathrm{d}\theta_{i+1}} = -\frac{\partial\phi_i/\partial\theta_{i+1}}{\partial\phi_i/\partial z_i} \text{ and } \frac{\mathrm{d}\varphi_{i+1}^\theta(\theta_{i+1})}{\mathrm{d}\theta_{i+1}} = -\frac{\partial\phi_{i+1}/\partial\theta_{i+1}}{\partial\phi_{i+1}/\partial z_{i+1}}. \tag{A.14}$$

(39) and (40) hold because $\partial\phi_i/\partial z_i < 0$ and $\partial\phi_{i+1}/\partial z_{i+1} < 0$ by (A.12) while

$$\frac{\partial\phi_i}{\partial\theta_{i+1}} = -\frac{\beta_i}{\theta_{i+1}^2}\left[v'\left(\frac{z_i}{\theta_{i+1}}\right) + \frac{z_i}{\theta_{i+1}} v''\left(\frac{z_i}{\theta_{i+1}}\right)\right] < 0, \tag{A.15}$$

$$\frac{\partial\phi_{i+1}}{\partial\theta_{i+1}} = \left(\frac{1+\beta_{i+1}}{\theta_{i+1}^2}\right)\left[v'\left(\frac{z_{i+1}}{\theta_{i+1}}\right) + \frac{z_{i+1}}{\theta_{i+1}} v''\left(\frac{z_{i+1}}{\theta_{i+1}}\right)\right] > 0, \tag{A.16}$$

If $\theta_{i+1} = \theta_I$, the change in z_{I-1} is obtained as previously and that in z_I comes from the observation that $z_I = \theta_I v'^{-1}(\gamma\theta_I)$.

Two cases must be distinguished as regards marginal tax rates. (i) $i < I - 1$: as $\partial z_i/\partial\theta_{i+1} < 0$, (20) and $v'' > 0$ imply an increase in $T'(z_i; \theta_i)$; similarly, $\partial z_{i+1}/\partial\theta_{i+1} > 0$ implies a reduction in $T'(z_{i+1}; \theta_{i+2})$. (ii) $i = I-1 : T'(z_{i+1}; \theta_{i+1})$ is unaltered because of Proposition 3. $T'(z_i; \theta_i)$ changes as in (i). □

References

Blundell R (1992) Labour supply and taxation: a survey. Fiscal Stud 13:15–40

Blundell R, MaCurdy T (1999) Labor supply: a review of alternative approaches. In: Ashenfelter O, Card D (eds) Handbook of labor economics, vol 3A. North-Holland, Amsterdam, pp 1559–1695

Boadway R, Cuff K, Marchand M (2000) Optimal income taxation with quasi-linear preferences revisited. J Public Econ Theory 2(4):435–460

Boone J, Bovenberg L (2007) The simple economics of bunching: optimal taxation with quasi-linear preferences. J Public Econ Theory 9:89–105

Brett C, Weymark JA (2008) The impact of changing skill levels on optimal nonlinear income taxes. J Public Econ 92:1765–1771

Diamond P (1998) Optimal income taxation: an example with u-shaped pattern of optimal marginal tax rates. Am Econ Rev 88(1):83–95

Ebert U (1992) A reexamination of the optimal nonlinear income tax. J Public Econ 49(1):47–73

Guesnerie R, Seade J (1982) Nonlinear pricing in a finite economy. J Public Econ 57:157–179

Hamilton J, Pestieau P (2005) Optimal income taxation and the ability distribution: implications for migration equilibria. Int Tax Public Finan 12:29–45

Hellwig M (2007) The undesirability of randomized income taxation under decreasing risk aversion. J Public Econ 91:791–816

Kimball M (1990) Precautionary saving in the small and in the large. Econometrica 58:53–73

Laffont J-J (1989) The economics of uncertainty and information. MIT Press, Cambridge

Laffont J-J, Martimort D (2002) The theory of incentives. Princeton University Press, Princeton

Lollivier S, Rochet J-C (1983) Bunching and second-order conditions: a note on optimal tax theory. J Econ Theory 31:392–400

Menegatti M (2001) On the conditions for precautionary saving. J Econ Theory 98:189–193

Mirrlees JA (1971) An exploration in the theory of optimum income taxation. Rev Econ Stud 38:175–208

Salanié B (2005) The economics of contracts. MIT Press, Cambridge, MA

Seade JK (1977) On the shape of optimal tax schedule. J Public Econ 7:203–235

Stiglitz J (1982) Self-selection and pareto efficient taxation. J Public Econ 17(2):213–240

Tuomala M (1990) Optimal income tax and redistribution. Clarendon Press, Oxford

Weymark JA (1986a) Bunching properties of optimal nonlinear income taxes. Soc Choice Welf 3:213–232

Weymark JA (1986b) A reduced-form optimal nonlinear income tax problem. J Public Econ 30:199–217

Weymark JA (1987) Comparative static properties of optimal nonlinear income taxes. Econometrica 55:1165–1185