

# Negative Marginal Tax Rates and Heterogeneity

IFS Working Paper W09/12

Philippe Choné  
Guy Laroque

# Negative marginal tax rates and heterogeneity<sup>1</sup>

Philippe Choné  
INSEE-CREST

Guy Laroque  
INSEE-CREST and University College London

April 28, 2009

First version: May 2007

<sup>1</sup>We have benefited from the comments of Mark Armstrong, Richard Blundell, Martin Hellwig, Bruno Jullien, Wojciech Kopczuk, Etienne Lehmann, Pierre Pestieau, Jean-Charles Rochet, Emmanuel Saez, Gilles Saint Paul, Bernard Salanié, John Weymark and of the participants in seminars in Toulouse, University College London, Louvain la Neuve and ESEM 2006 in Vienna. We are very grateful to Robert Moffitt, the editor, and to four anonymous referees for their perceptive comments and suggestions.

## Abstract

Heterogeneity is likely to be an important determinant of the shape of optimal tax schemes. This article addresses the issue in a model à la Mirrlees with a continuum of agents. The agents differ in their productivities and opportunity costs of work, but their labor supplies depend only on a unidimensional combination of their two characteristics. Conditions are given under which the standard result that marginal tax rates are everywhere non-negative holds. This is in particular the case when work opportunity costs are distributed independently of income. But one can also get negative marginal tax rates: economies where negative tax rates are optimal at the bottom of the income distribution are studied, and a numerical illustration is given, based on UK data.

JEL classification numbers: H21, H31.

Keywords: optimal taxation, heterogeneity, welfare.

Please address correspondence to Guy Laroque, laroque@ensae.fr, INSEE-CREST J360, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France.

# 1 Introduction

The bulk of the theory of optimal taxation recommends that the marginal tax rate be everywhere positive: labor supply is distorted downwards, compared with laissez-faire. The conditions under which the result holds (continuous labor supply or intensive margin, unobserved productivity, constant opportunity cost of work, utilitarian planner with a redistributive motive towards lower incomes) have been thoroughly investigated (Seade (1977), Seade (1982), Werning (2000), Hellwig (2007)).

The purpose of the present paper is to describe how negative marginal tax rates can be optimal, keeping with the intensive setup largely studied in the literature, when there are multiple dimensions of heterogeneity. Indeed Mirrlees (1976) in its Section 4 indicates, along a line that will be pursued further here, that the sign of the marginal tax rate cannot be predicted when the agents in the economy differ along several dimensions of heterogeneity. Workers differ in both their productivities and opportunity costs of work. Under utilitarianism, heterogeneity comes into play through its impact on cardinal utilities. How social weights, or marginal utilities of income, vary with income determines the shape of the optimal tax scheme. Conditional on income, utility levels, and more importantly directly their derivatives with respect to income, may either decrease or increase with work opportunity costs. This may vary according to circumstances, depending on whether the cost is associated with poor living conditions (i.e. a handicap) or reflects a taste for leisure or opportunities outside the labor market (such as gardening at home or black market activities). For a given income, the social weight is proportional to the expected marginal utilities, where the expectation is taken over the distribution of work opportunity costs at this income level. Therefore the social weight curve both depends on the specification of the cardinal utility function and on how the distribution of work opportunity costs changes with income.

The fact that heterogeneity can blur the redistribution motive, through the possible correlation between ability and onerousness of work, has been discussed in models where agents have a finite number of types. A pioneer work is that of Stiglitz (1982) who shows that the high-skilled individual should face a negative marginal tax rate if the low-skilled individual's upward incentive constraint binds,

which is the case if social weights are increasing in type. The importance of this observation for policy was pointed out in Cuff (2000), who is the first to explicitly link social weights to the opportunity cost of work, using alternative choices of cardinal utilities. In a model with four types, combining two values for productivity and two values for the cost of work, Boadway, Marchand, Pestieau, and Racionero (2002) showed how it can be optimal to have binding upward incentive constraints when high opportunity costs are associated with small social weights.

The present paper shows how intuitions derived from finite types models extend to the standard optimal taxation framework with a continuum of types, which is more appropriate to practical situations. In this setup, we revisit the analysis of Saez (2002), making explicit the determinants of the social weights and linking them to the agents' heterogeneous characteristics. In particular, we build on the result that negative marginal tax rates are only optimal at the bottom of the skill distribution if these individuals have below average social weights, a rather drastic condition. We show that this condition may hold rather naturally if (i) there is heterogeneity in utilities, holding income constant, and (ii) the heterogeneity reduces the social weights put on the low income types, relatively to those of the larger income types.

We consider a standard model à la Mirrlees where the workers' choices can be represented by a utility function which is separable in consumption and leisure.<sup>1</sup> Although agents differ in productivity and work opportunity cost, their behavior is assumed to only depend on a unidimensional combination of the two underlying parameters, which avoids the technicalities that typically go with multidimensional heterogeneity. A similar shortcut has been used by Brett and Weymark (2003) in a continuum environment, while Beaudry, Blackorby, and Szalay (2006) tackles the full multidimensional issue in a model where there are a finite number of types.

The paper is in three parts. First, without any a priori assumptions on the relative weights given to the various agents in the economy, we describe the set of second best allocations. Each allocation is associated with a tax schedule. Under regularity assumptions, we show that the first order condition relating the

---

<sup>1</sup>Our framework is more general than some of the studies in the previous literature: we allow for income effects, we do not assume linearity in leisure.

marginal tax rate to the social weights of the agents, familiar from optimal tax theory, also applies here with arbitrary exogenous social weights.

The second part of the paper relates the social weights to the underlying distribution of characteristics in the economy. It describes situations where the Mirrlees result, everywhere positive marginal tax rates, holds. This is the case, for instance, when the distribution of opportunity costs is independent of that of productivities, whatever the impact of these costs on the agents utilities.

The third part builds on Saez (2002) to show when negative tax rates are optimal. We provide a theoretical example where agents with low productivities exhibit a large spectrum of opportunity costs, and are better off, the larger their costs. We give a general formula for the social weight of the low skilled, as a function of the distribution of the heterogenous characteristics and of the derivative of marginal utility with respect to the heterogeneity parameter. Finally we illustrate the potential importance of heterogeneity on a simulation exercise grounded on UK data. We show that a small change in heterogeneity, from a standard error of the work opportunity cost of £ 100 at the bottom of the earnings distribution to £ 200 at the first quartile, is enough to warrant negative marginal tax rates. This indicates that more work and attention should be devoted to heterogeneity when designing optimal tax schemes.

## 2 Second best optimality in the Mirrlees model

We consider an economy with a continuum of agents. Agents differ by their opportunity costs of work and their abilities, respectively measured by the non-negative numbers  $\theta^c$  and  $\theta^p$ . The overall utility  $U$  of an agent that earns a before tax income  $y$  leading to a final consumption (or after tax income)  $c$  is

$$u(c) + \tilde{v}(y, \theta),$$

where  $\theta = (\theta^c, \theta^p)$  designates the agent's type. The utility of consumption  $u(c)$  is increasing and concave in  $c$ . The second term measures the disutility from working. It is decreasing and concave in  $y$ , increasing in  $\theta^p$ , decreasing in  $\theta^c$ : the larger productivity, the less work is needed to achieve a given production level; given  $y$ , the larger the cost of work, the smaller the utility.

The government does not observe the agents' types  $\theta$  and has to base redistri-

bution solely on observed earnings, using a nonlinear tax scheme. A citizen who earns a before tax income  $y$  is left with an after-tax income  $c = R(y) = y - T(y)$ , where  $T(y)$  denotes taxes net of benefits. A measurable function  $\tilde{y}(\theta)$  is called an *allocation*. Given a consumption schedule  $R$ , an allocation such that  $\tilde{y}(\theta)$  maximizes

$$u(R(y)) + \tilde{v}(y, \theta)$$

is *incentive compatible*. Such an allocation will be denoted  $\tilde{y}_R(\theta)$  to make explicit its relationship with  $R$ .

An allocation is *feasible* when aggregate production, the sum of the  $\tilde{y}_R(\theta)$ , is equal to aggregate consumption, the sum of the  $R(\tilde{y}_R(\theta))$ . The government problem is to choose the incentive compatible feasible allocation which is optimal given his redistributive motives.

We restrict our analysis to the case where the consumers' behaviors depend effectively on a unidimensional parameter  $\alpha = A(\theta)$ . In general when there are several dimensions of heterogeneity and the government has only one dimension of observation (income), a major difficulty is to identify the shape of the set of types that are associated with a given level of income, since this shape typically depends on the tax schedule. Here, the shape is given by the level curves of the function  $A$ , independently of  $R$ . Formally, the above utility function reduces to  $u(c) + v(y, A(\theta)) = u(c) + v(y, \alpha)$ , where  $v(y, A(\theta)) \equiv \tilde{v}(y, \theta)$ .<sup>2</sup> The function  $A$  is supposed to be increasing in productivity  $\theta^p$  and decreasing in the onerousness of work  $\theta^c$ . Then  $v$  is increasing in its second argument. We assume that the cross derivative  $v_{y\alpha}$  is everywhere positive, which implies that the indifference curves of the agents in the  $(c, y)$  space are single-crossing.<sup>3</sup> In this circumstance, it is known that incentive compatibility is equivalent to a simple monotonicity condition (for completeness we provide the proof in the Appendix). The value of utility at the maximum, or indirect utility, is denoted  $U_R(\alpha)$ .

**Lemma 1.** *A necessary and sufficient condition for an allocation  $y$  to be incentive compatible is that it be nondecreasing in  $\alpha$ . Furthermore, the indirect utility  $U_R(\alpha)$*

---

<sup>2</sup>To distinguish the functions of  $\alpha$  from those of  $\theta$ , we put a tilde on the latter.

<sup>3</sup>It would be of interest to know whether our analysis extends to situations where utility is not separable between consumption and leisure, as in Mirrlees (1971) or Hellwig (2007). This would probably require a general version of the single-crossing assumption in the spirit of Edlin and Shannon (1998).

is differentiable at all but at most countably many points  $\alpha$ , with

$$U'_R(\alpha) = v_\alpha(y(\alpha), \alpha).$$

Both  $R(y_R(\alpha))$  and  $U_R(\alpha)$  are nondecreasing functions of  $\alpha$ .

Our formulation encompasses the standard model in the literature. The link can be made precise as follows. Let the single heterogeneity parameter be  $\alpha = \theta^p$ , and take the function  $v$  as  $-\hat{v}(y/\theta^p)$ , where  $y/\theta^p$  is the quantity of labor that the agent must provide to get income  $y$ , given his ability  $\theta^p$ . The function  $\hat{v}$  measures the cost of providing labor in utility terms. It is increasing and convex. In general, when the agents not only have different productivities but also differ in their costs of providing a given labor supply  $y/\theta^p$ , the Mirrlees function  $\hat{v}(y/\theta^p)$  becomes  $\bar{v}(y/\theta^p, \theta^c)$ . The unidimensional restriction here is the assumption that in fact  $-\bar{v}$  reduces to some  $v$  as above.<sup>4</sup> With a well designed incentive scheme, the government may infer the characteristic  $\alpha$  of an agent from his income  $y$ , but it is unable, say for a large income and a large  $\alpha$ , to know whether it comes from a high productivity type or a low cost of work type.

Given the information available to the government, all the agents of type  $\theta = (\theta^c, \theta^p)$  with the same  $\alpha = A(\theta)$  must be treated equally. The *second best optimal allocations* are incentive compatible allocations which maximize a weighted sum of the agents utility functions, with positive weights  $\pi(\alpha)$  summing up to 1, subject to the government budget constraint.<sup>5</sup> Let  $G$  be the cumulative distribution function of the parameter  $\alpha$ . We assume that  $G$  admits a density  $g$  which is continuous and positive on the interior<sup>6</sup> of the support  $[\underline{\alpha}, \bar{\alpha}]$ ,  $\underline{\alpha} < \bar{\alpha} < \infty$ . The Lagrangian of the second best program is

$$\mathcal{L} = \int_{\underline{\alpha}}^{\bar{\alpha}} \{\pi(\alpha)U_R(\alpha) + \lambda[y(\alpha) - R(y(\alpha))]\} dG(\alpha), \quad (1)$$

---

<sup>4</sup>In an earlier version of the paper, we studied the situation where the disutility of work is isoelastic,  $\hat{v}(y) = y^{1+1/e}$  and the two parameters  $\theta^c$  and  $\theta^p$  combine in a one-dimensional parameter  $\alpha$  as follows:  $v(y, \alpha) = -\hat{v}(y)/\alpha$  with  $\alpha = (\theta^p)^{1+1/e}/\theta^c$ .

<sup>5</sup>To be precise, one gets all second best allocations when the weights describe the set of probability measures on the space of types. For notational simplicity, we write the programs below for probability measures that have a density with respect to the measure on types. The results are valid in the general case.

<sup>6</sup>As pointed out by a referee, the multidimensional construction typically implies  $g(\underline{\alpha}) = g(\bar{\alpha}) = 0$ . This is discussed in Brett and Weymark (2003), p.2565.



where  $\lambda$  is the multiplier of the budget constraint, i.e. the marginal cost of public funds.

A marginal admissible transformation, that does not change the tax payers' labor supplies, consists in uniformly increasing utility by some amount  $dU$  through a change in consumption  $dR(y) = dU/u'(R(y))$ , for all  $y$ . Since the Lagrangian should be left unchanged after this transformation, it follows that

$$\int_{\underline{\alpha}}^{\bar{\alpha}} \frac{\lambda}{u'(R(y_R(\alpha)))} dG(\alpha) = 1.$$

Therefore  $g^*(\alpha) \equiv \lambda g(\alpha)/u'(R(y_R(\alpha)))$  is the density of a probability measure  $G^*$ . To give a simple statement of the first order condition for second best optimality, we associate with an allocation  $y_R$  the modified weights:

$$\pi^*(\alpha) = \pi(\alpha)u'(R(y_R(\alpha))). \quad (2)$$

Computed with the probability measure  $G^*$ , the average modified social weight equals the cost of public funds:

$$\int_{\underline{\alpha}}^{\bar{\alpha}} \pi^*(\alpha) dG^*(\alpha) = \lambda \int_{\underline{\alpha}}^{\bar{\alpha}} \pi(\alpha) dG(\alpha) = \lambda.$$

Let  $p^*(\alpha)$  be the average value of the social weights of all the agents with idiosyncratic characteristics larger than  $\alpha$ :

$$p^*(\alpha) = \frac{1}{1 - G^*(\alpha)} \int_{\alpha}^{\bar{\alpha}} \pi^*(x) dG^*(x) = E_{G^*}(\pi^*(x)|x \geq \alpha). \quad (3)$$

By construction, any second best allocation satisfies, for all  $\alpha$

$$[1 - G^*(\alpha)]E_{G^*}(\pi^*(x)|x \geq \alpha) + G^*(\alpha)E_{G^*}(\pi^*(x)|x \leq \alpha) = \lambda.$$

If the average weights of the agents above  $\alpha$  is smaller than  $\lambda$ , the weights of the agents below is larger than  $\lambda$ , and conversely. Also, at the lower boundary,  $p^*(\underline{\alpha}) = \lambda$ . The first order condition for the optimal tax rate at  $\alpha$ , similar to formula (21) of Mirrlees (1971), can then be expressed as follows.<sup>7</sup>

---

<sup>7</sup>In a previous version of this paper, Choné and Laroque (2007), in a setup where the utility function is linear in consumption, we provide a full characterization of the set of second best

**Lemma 2.** *Let  $y$  be a second best allocation. Consider  $\alpha$  in  $(\underline{\alpha}, \bar{\alpha})$  where  $y$  is continuous and there is no bunching. Then*

$$\lambda \frac{u'(R(y_R(\alpha)))}{v_{y\alpha}(y_R(\alpha), \alpha)} T'(y_R(\alpha)) = \frac{1 - G^*(\alpha)}{g^*(\alpha)} [\lambda - p^*(\alpha)]. \quad (4)$$

The proof in appendix derives the first order condition at all points, including discontinuity points of  $y$ , and shows, under the additional assumption  $v_{\alpha yy} \geq 0$ , that any second best allocation is continuous in the region where marginal tax rates are nonnegative.<sup>8</sup>

As a consequence of (4), the marginal tax rate supported by type  $\alpha$  has the same sign as  $\lambda - p^*(\alpha)$ . The formula, as discussed in Diamond (1998) for an economy with quasi-linear preferences, is associated with an increase  $dT'$  in the marginal tax rate on the interval  $[\alpha - d\alpha, \alpha]$ , which reduces consumption by  $-dT' d\alpha$  for  $\alpha$  in  $[\alpha, \bar{\alpha}]$ . The left hand side of (4), once multiplied by  $g^*(\alpha) dT' d\alpha$ , corresponds to the distortion in labor supply that takes place in the interval  $[\alpha - d\alpha, \alpha]$ . The right hand side, multiplied by  $g^*(\alpha) dT' d\alpha$ , is the change in the Lagrangian associated with the decrease in consumption above  $\alpha$ : the government collects  $(1 - G^*(\alpha)) dT' d\alpha$ , which is valued at the marginal cost of public funds  $\lambda$ , but the utilities of the concerned agents fall, which taking into account their social weights reduces welfare by  $(1 - G^*(\alpha)) p^*(\alpha) dT' d\alpha$ . This marginal transformation must leave the Lagrangian unchanged, which yields (4).

The sign of the marginal tax rate indicates how the incentive constraints bind. The intuition, largely taken from Boadway, Marchand, Pestieau, and Racionero (2002), is as follows. There is no distortion at  $\alpha$ , and the incentive constraints are not binding when  $R' = 1$ , i.e. from (4), when the average social weights of the agents with a higher income is equal to that of the agents with a smaller income,  $p^*(\alpha) = E_{G^*}(\pi^*(x)|x \leq \alpha) = \lambda$ . When the weight of the agents with lower  $\alpha$ 's is larger than the average weight (i.e.,  $p^*(\alpha) < \lambda < E_{G^*}(\pi^*(x)|x \leq \alpha)$ ), the optimal income tax redistributes in their favor, increasing after tax income at the bottom, lowering it at the top: the slope of the after tax income curve becomes smaller than 1, in effect creating a positive marginal tax rate. The

---

allocations, without the no-bunching condition of Lemma 2. Werning (2007) also studies the set of second best allocations.

<sup>8</sup>We are grateful to an anonymous referee who urged us to investigate the continuity properties of the optimal allocations.

redistribution is limited by the fact that the high skilled would fake being low skilled: the incentive constraints binds downwards. Conversely when the social weights of the agents above  $\alpha$  is larger than  $\lambda$ , redistribution lowers after tax incomes of types below  $\alpha$  (compared to laissez faire) and increases that of types above  $\alpha$ : this gives the after tax income curve a slope larger than 1, the marginal tax rate is negative and incentive constraints bind upwards.

### 3 Utilitarianism and positive marginal tax rates

Lemma 2 links the optimal tax rate with the social weights at a regular second best allocation. From (4), the marginal tax rate supported by an agent of type  $\alpha_0$  depends on the average social weight of the agents of type  $\alpha$  larger than  $\alpha_0$ .

Therefore, to study the sign of the marginal tax rate, we relate the values of the social weights to the objective maximized by the government. Utilitarianism requires a specification of cardinal utilities. A priori the cardinal utility can be any function  $K[U, \theta]$ .<sup>9</sup> The only requirement for consistency with private choices is that it increases in its first argument. Any difference between types can be taken into account through the second argument of  $K$ . Since by Lemma 1, incentive compatibility implies that the indirect utility function is nondecreasing in  $\alpha$ ,  $K$  is nondecreasing in  $\alpha$  through its first argument. To keep in line with the literature, we assume that productivity  $\theta^p$  only enters cardinal utility by its impact on  $U$ , through  $\alpha$ : it is enough for our purpose to restrict ourselves to cardinal utilities of the form  $K[U, \theta^c]$ .<sup>10</sup>

A utilitarian planner maximizes the sum of the cardinal utility functions

$$\iint K[U_R(\alpha), \theta^c] dF(\theta^c|\alpha) dG(\alpha),$$

---

<sup>9</sup>We adopt a utilitarian viewpoint: the cardinal utility is an objective measure of the agents' well being. All the results of course apply to a situation where the function  $K$  reflects society's value judgments.

<sup>10</sup>As pointed out by a referee, formally one does not need the extra degree of freedom given by the parameter  $\theta^c$  to reverse traditional conclusions. It is enough to have a utilitarian function  $K[U, \alpha]$  with a first derivative  $K_U[U_R(\alpha), \alpha]$  which is sufficiently increasing in its second argument to counteract the effect of the first argument, so that  $K_U[U_R(\alpha), \alpha]$  is increasing in  $\alpha$ . From an economic viewpoint, however, one would need some empirical justification for this unconvincing assumption. The shape of the (non-degenerate) distribution of  $\theta^c$  conditional on  $\alpha$  may provide a rationale for the type of phenomena under study, as illustrated in the last section of the paper.

under its budget constraint, where  $F(\theta^c|\alpha)$  denotes the cumulative distribution function of  $\theta^c$ , conditional on the equality  $\alpha = A(\theta)$ . Identifying the Lagrangian with its second best counterpart (1), the associated weights  $\pi(\alpha, \theta^c)$  are equal, up to a multiplicative normalization factor, to  $K_U[U_R(\alpha), \theta^c]$ . Under the assumption that labor supply is driven by the one-dimensional parameter  $\alpha$ , the optimal schedule only depends on the aggregate weights

$$\pi(\alpha) = \int \tilde{\pi}(\alpha, \theta^c) dF(\theta^c|\alpha). \quad (5)$$

To account for income effects, we introduce the modified weights as explained in Section 2:

$$\tilde{\pi}^*(\alpha, \theta^c) = \tilde{\pi}(\alpha, \theta^c)u'(R(y_R(\alpha))), \quad \pi^*(\alpha) = \pi(\alpha)u'(R(y_R(\alpha))). \quad (6)$$

The above formulas make clear how the cardinal representation plays a role, through its impact on the marginal social value of a change in utility  $K_U[U, \theta^c]$ . When  $K_U$  is increasing in its second argument, large work opportunity costs, holding  $\alpha$  constant, go with large social weights: this can be interpreted as if a large opportunity cost comes from a handicap that deserves some social compensation. When  $K_U$  is decreasing in its second argument, a large opportunity cost of work reduces the social weight, perhaps because non-market time allows enjoyable leisure. All things considered, the quantity of interest is the integral of  $K_U$  with respect to the conditional distribution  $F(\theta^c|\alpha)$ .

### 3.1 The unidimensional case

Consider the standard Mirrlees case where  $\theta^c$  is constant across the population, and  $\theta^p$  has a continuous distribution on  $[\underline{\theta}^p, \bar{\theta}^p]$ . Productivity, as well as cardinal utility, increases with  $\theta^p$ . Then we recover the optimal taxation result of non-negative marginal tax rates everywhere, provided the function  $K$  is concave.

**Proposition 1.** *Assume that  $K[U]$  is concave. Then the optimal utilitarian optimum exhibits positive marginal tax rates, at any regular point below the top productivity where there is no bunching.*

**Proof:** From the definition of the weights (2), since after tax income and the indirect utility both increase with  $\alpha$  along any incentive compatible allocation,

social weights decrease with  $\theta^p$ , and from (3),  $p^*(\theta^p)$  decreases as well. Since  $p^*(\underline{\theta}^p) = \lambda$ ,  $p^*(\theta^p) < \lambda$  for all  $\theta^p < \bar{\theta}^p$ , and (4) gives the standard result: the marginal tax rate is always positive. ■

When  $K$  is concave in  $U$ , the cardinal utility specification makes society more redistributive than the initial separable  $U$ : the social weight  $K'(U)$  is decreasing in  $U$ . This is the situation considered in Mirrlees of a redistributive government.

Note that a regressive government would *not* implement positive tax rates. Consider the case where  $K[U]$  is strictly convex and where the original utility function is linear in consumption ( $u(R(y)) = R(y)$ ). Then the weight  $\pi_a^*(\alpha) = K'(U_R(\alpha))$  is increasing in  $\alpha$ , so that  $p^*(\alpha)$  is increasing as well. The optimal marginal tax rate is negative everywhere, except at the boundaries! Heterogeneity is not necessary for negative marginal tax rates to be optimal.

### 3.2 Non-negative marginal tax rates and heterogeneity

There are a variety of cases where tax rates are non-negative, even in the presence of heterogeneity. As should be clear from the previous section, a sufficient condition is that the average social weights of the agent of characteristics  $\alpha$  decreases with  $\alpha$ . If the only parameter entering the social utility function is  $\alpha$ , this is warranted by the concavity of the welfare index.<sup>11</sup> Otherwise one must put restrictions that bear simultaneously on how the social weight  $K_U$  varies with  $\theta^c$  and on how the conditional distribution of  $\theta^c$  given  $\alpha$  changes with  $\alpha$ .

**Proposition 2.** *Assume that the cardinal utility  $K[U, \theta^c]$  is concave in  $U$ . Assume furthermore that  $K_U[U, \theta^c]$  is nondecreasing in  $\theta^c$  and that the distribution of  $\theta^c$ , conditional on  $\alpha$ , is first order stochastically decreasing in  $\alpha$ .*

*Then under utilitarianism, the weights  $\pi^*(\alpha)$  are decreasing and marginal tax rates are everywhere non-negative.*

**Proof:** Let

$$\Psi(a, b) = \int \tilde{\pi}(a, \theta^c) dF(\theta^c|b).$$

---

<sup>11</sup>From the initial assumption of Section 1 and Lemma 1 labor supply only depends on  $\alpha$  and the indirect utility level is nondecreasing in  $\alpha$ . Given  $\alpha$ , the only effect of heterogeneity is to change social valuations and social weights.

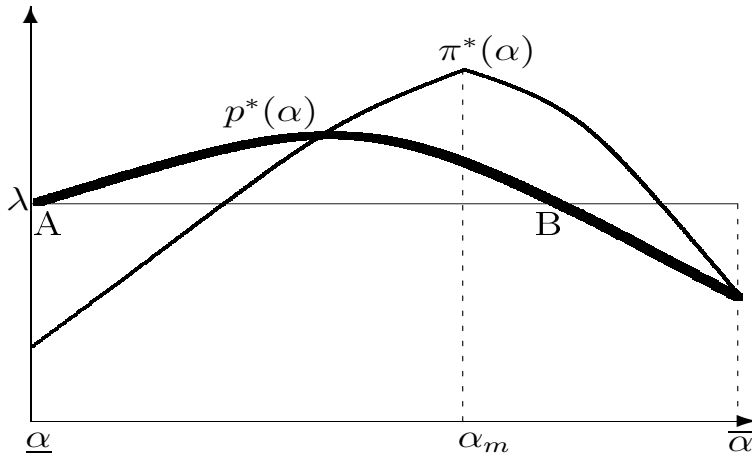


Figure 1: Social weights and negative marginal tax rates

Then the function  $\Psi$  is decreasing in  $a$ , since  $\tilde{\pi}$ , proportional to  $K_U$ , is. It is decreasing in  $b$  by first order stochastic dominance. It follows that  $\pi(\alpha) = \Psi(\alpha, \alpha)$  is also decreasing in its argument. Multiplying by  $u'(R(y_R(\alpha)))$ , which is decreasing in  $\alpha$ , yields a decreasing  $\pi^*(\alpha)$ . ■

The result also holds in the opposite situation:  $K_U[U, \theta^c]$  decreasing in  $\theta^c$ , and  $\theta^c$  first order stochastically nondecreasing in  $\alpha$ . It is simpler to state when the opportunity cost,  $\theta^c$ , is independent from the income level.

**Corollary 1.** *Assume that the cardinal utility  $K[U, \theta^c]$  is concave in  $U$ . When  $\theta^c$  is independent from  $\alpha$ , marginal tax rates are non-negative.*

**Proof:** The function  $\Psi$  defined above does not depend on its second argument. As a result,  $\pi(\alpha) = \Psi(\alpha, \alpha)$  is decreasing, irrespective of the way the opportunity cost enters the cardinal utility. ■

## 4 Negative marginal tax rates

### 4.1 A simple theoretical example

The following example illustrates the analysis. At the lowest productivity  $\underline{\theta}^p$ , the agents exhibit a variety of work opportunity costs  $\theta^c$ , a continuous distribution on  $[\underline{\theta}^c, \bar{\theta}^c]$ . For all productivities above the minimum, a continuous distribution on  $(\underline{\theta}^p, \bar{\theta}^p]$ , there is a unique value of  $\theta^c$ , equal to  $\underline{\theta}^c$ . The labor supply of the

agent of type  $(\theta^p, \theta^c)$  depends on a parameter  $\alpha = A(\theta^p, \theta^c)$ , where  $A$  is a function increasing in  $\theta^p$  and decreasing in  $\theta^c$ . Therefore:

$$\underline{\alpha} = A(\underline{\theta}^p, \bar{\theta}^c) < \alpha_m = A(\underline{\theta}^p, \underline{\theta}^c) < \bar{\alpha} = A(\bar{\theta}^p, \underline{\theta}^c).$$

All the agents of the segment  $[\alpha_m, \bar{\alpha}]$  differ only by their productivities. All the agents in  $[\underline{\alpha}, \alpha_m]$  have the same low productivity  $\underline{\theta}^p$ , but have different, decreasing, opportunity costs.<sup>12</sup>

Suppose now that the cardinal utility is  $K[U_R(\alpha) + k\theta^c]$ , where  $K$  is concave increasing and  $k$  is a well chosen scalar. With a positive  $k$ , holding  $\alpha$  constant, a larger  $\theta^c$  yields a higher utility level, obtained for instance through activities at home or on the black market, and therefore low marginal utilities of income. With this parameterization,  $\pi^*(\alpha) = K'[U_R(\alpha) + k\theta^c]$  is decreasing on  $[\alpha_m, \bar{\alpha}]$ . Suppose, possibly for a large enough  $k$ , that it increases on  $[\underline{\alpha}, \alpha_m]$ . One then gets Figure 1. The agent with the largest social weight is the  $\alpha_m$  person with lowest productivity and opportunity cost to work. The associated function  $p^*(\alpha)$ , which measures the average height of  $\pi^*(x)$  for  $x$  larger than  $\alpha$ , is represented on the Figure:  $p^*(\alpha)$  increases whenever it lies above the graph of  $\pi^*$ , decreases when it is below the graph, and has an horizontal tangent when it crosses the  $\pi^*$  curve. Also, we know that  $p^*(\underline{\alpha}) = \lambda$ . From (4), in the situation depicted on Figure 1, all the agents in the segment AB then face negative tax rates.

More generally, there are a number of cases where specific welfare functions would lead to (locally) negative marginal tax rates. Suppose for simplicity that the utility function  $u$  is linear in consumption. Consider a situation where productivity  $\theta^p$  is an increasing function of  $\alpha$  while the opportunity cost of work  $\theta^c$  is a decreasing function of  $\alpha$ : a larger income indicates a larger productivity and a smaller opportunity cost of work. Assume that the cardinal utility puts a high value on the willingness to work, meaning that the derivative  $K_{U\theta^c} d\theta^c / d\alpha$  is negative and large in absolute value, larger than the absolute value of  $K_{UU} dU_R / d\alpha$ . A high enough value would make the optimal income tax regressive.

---

<sup>12</sup>Technically, given  $\alpha$ , the distribution of  $\theta^c$  is degenerate: for  $\alpha \geq \alpha_m$ ,  $\theta^c$  is equal to  $\underline{\theta}^c$  while for  $\alpha \leq \alpha_m$ ,  $\theta^c$  is equal to the unique root of the equation  $\alpha = A(\underline{\theta}^p, \theta^c)$ .

## 4.2 May negative marginal tax rates be optimal for low skilled workers?

While the preceding example is illustrative, the distribution of the work opportunity cost conditional on  $\alpha$  is degenerate. It is interesting to more generally identify the circumstances where negative marginal tax rates on low incomes are likely to be optimal.

First a theoretical remark is useful. As noted by Saez (2002), p.1054, negative marginal tax rates at the bottom of the income distribution can only occur if the social weights of the concerned agents are smaller than the average social weight. Indeed, to see negative marginal tax rates bearing on the low incomes, i.e. in a neighborhood of  $\underline{\alpha}$ , from (4), in the absence of pooling<sup>13</sup>  $p^*(\alpha)$  has to be larger than  $\lambda$  in this neighborhood. Since by construction  $p^*(\underline{\alpha}) = \lambda$ , assuming differentiability, this amounts to

$$\frac{dp^*}{d\alpha} > 0 \text{ in a neighborhood of } \underline{\alpha}.$$

Since by (3)  $dp^*/d\alpha = [p^*(\alpha) - \pi^*(\alpha)]g^*(\alpha)/[1 - G^*(\alpha)]$ , this occurs when  $\pi^*$ , the social weight of the agents, is smaller than  $\lambda$ , the average social weight in the economy, in this neighborhood. In the absence of pooling, a necessary and sufficient condition for the low income agents, with  $\alpha$  close to  $\underline{\alpha}$ , to face negative marginal tax rates at the optimum is that the social weight  $\pi^*(\underline{\alpha})$  be smaller than the average social weight in the economy. An expression for this condition can be given, where both the dependence of the social weights on the heterogeneity characteristics and the distribution of  $\theta^c$  conditionally on  $\alpha$  interact.

**Proposition 3.** *When the functions  $\tilde{\pi}^*(\alpha, \theta^c)$ ,  $F(\theta^c|\alpha)$  and  $G^*(\alpha)$  are twice continuously differentiable with respect to their arguments, one has*

$$\pi^*(\underline{\alpha}) - \lambda = \iint \{\nabla F \times \nabla \tilde{\pi}^*\} [1 - G^*(\alpha)] d\theta^c d\alpha, \quad (7)$$

---

<sup>13</sup>Under the assumption that the marginal cost of work is zero at the origin,  $v_y(0, \alpha) = 0$ , it is easy to show that everybody works at an optimum. Indeed adapting the proof of Lemma 2, a small increase  $dy$  in income at the origin does not change the utility but increases the Lagrangian by  $\lambda dy$ . This assumption is satisfied in the standard example described in footnote 4.



where

$$\nabla F \times \nabla \tilde{\pi}^* = \frac{\partial \tilde{\pi}^*}{\partial \theta^c}(\alpha, \theta^c) \frac{\partial F}{\partial \alpha}(\theta^c | \alpha) - \frac{\partial \tilde{\pi}^*}{\partial \alpha}(\alpha, \theta^c) \frac{\partial F}{\partial \theta^c}(\theta^c | \alpha)$$

is the cross-product<sup>14</sup> of the two gradient vectors  $\nabla F$  and  $\nabla \tilde{\pi}^*$ .

Proposition 3 gives an expression for the weight  $\pi^*(\underline{\alpha})$  in the situation where the distributions are smooth. There are two terms in the formula:

- The second term is positive, from the standard motive of aversion to income inequality. Indeed, from (6),  $\tilde{\pi}^*(\alpha, \theta^c) = \tilde{\pi}(\alpha, \theta^c)u'(R(y_R(\alpha)))$  and the derivative  $\partial \tilde{\pi}^*/\partial \alpha$  is typically negative. It is only equal to zero when there are both no desire for redistribution across the  $\alpha$  characteristic (the cardinal utility is linear in  $U$  and  $\pi(\alpha)$  is constant), and no income effect ( $u$  is linear in  $c$ ).
- The first term,  $\partial \tilde{\pi}^*/\partial \theta^c \partial F/\partial \alpha$ , cannot be signed in general. It is equal to zero in a number of cases, for instance if the two parameters are independently distributed ( $\partial F/\partial \alpha = 0$ ), or if the social weight does not depend on  $\theta^c$ . Then the marginal tax rate is positive at the bottom of the income distribution. When  $\theta^c$  is first order stochastically decreasing in  $\alpha$  ( $\partial F/\partial \alpha \geq 0$ ) and  $\tilde{\pi}^*$  is nondecreasing in  $\theta^c$ , then the term is positive, which yields the analog to Proposition 2 at the point  $\underline{\alpha}$ .

In practice, for negative tax rates to be optimal, the first term must be negative and larger in absolute value than the second one. A special case where this is easier to achieve is when society has no aversion to income inequality ( $u$  linear in consumption and  $K$  linear in  $U$ ). The only redistribution motive then is linked to the  $\theta^c$  parameter, i.e.  $K[U, \theta^c] = K(\theta^c)U$ . In this circumstance, for all  $\alpha$   $\tilde{\pi}^*(\alpha, \theta^c) = K(\theta^c)$  and

$$\pi(\alpha) - \lambda = \int_{\underline{\theta}^c}^{\bar{\theta}^c} K(\theta^c) f(\theta^c | \alpha) d\theta^c - \int_{\underline{\theta}^c}^{\bar{\theta}^c} K(\theta^c) \phi(\theta^c) d\theta^c, \quad (8)$$

where  $\phi$  denotes the density of the marginal distribution of the parameter  $\theta^c$  in the economy. Suppose that  $K$  is increasing in the work opportunity cost, so that

<sup>14</sup>The cross-product of two vectors of  $\mathbb{R}^2$  is the product of their modulus, multiplied by the sinus of their oriented angle. A simple diagram, noting that  $\partial F/\partial \theta^c > 0$  and  $\partial \tilde{\pi}^*/\partial \alpha < 0$ , shows that the angle is between 0 and  $\pi$  under the conditions of Proposition 2.

the government wants to transfer income to the agents with the larger cost. Then the marginal tax rate is negative at  $\underline{\alpha}$  when the conditional distribution of  $\theta^c$  at  $\underline{\alpha}$  is first order stochastically smaller than the unconditional distribution of  $\theta^c$ .

### 4.3 An illustration using UK data

The following computations are based on UK data. We use annual earnings and marginal tax rate information<sup>15</sup> in the UK for the year 2003. Our starting point is the specification of Brewer, Saez, and Shephard (2008), which served to compute optimal taxes in the Mirrlees review of the UK tax system. The basic utility function is of the form

$$U_R(\alpha) = \max_y R(y) - \frac{y^{1+1/e}}{(1 + 1/e)\alpha^{1/e}},$$

with  $e = .25$ , and the first order condition  $R'(y) = (y/\alpha)^{1/e}$  allows us to recover the parameter  $\alpha$  from earnings and the marginal tax rate. Note that with this specification  $\alpha$  is equal to the income that would be observed under laissez-faire. The density of  $\alpha$  is obtained from standard kernel estimation, with a rather large bandwidth (£ 3000) to get smooth drawings. Let  $\theta^c$  stand for a measure of the work opportunity cost. We take as cardinal utility:

$$K[U, \theta^c] = -\exp[-(U - \theta^c)],$$

so that the government objective is to maximize

$$-\iint \exp[-U_R(\alpha) + \theta^c] dF(\theta^c|\alpha) dG(\alpha) = -\int \left[ \int \exp(\theta^c) dF(\theta^c|\alpha) \right] \exp[-U_R(\alpha)] dG(\alpha).$$

With the exponential specification the effect of heterogeneity factors out, and the optimization is easy. Also the social weight of agent  $(\alpha, \theta^c)$  is proportional to  $\exp[-U_R(\alpha) + \theta^c]$ : it is increasing in the work opportunity cost.

We have to make choices on the shape of the distribution of  $\theta^c$  conditional on  $\alpha$ , for which we have no guidance from empirical observations. The results that we present are the ones that we found most suggestive, out of a number of experiments. We assume the distribution of  $\theta^c$ , conditional on  $\alpha$ , to be normal

---

<sup>15</sup>We thank Andrew Shephard for giving us access to the data set.

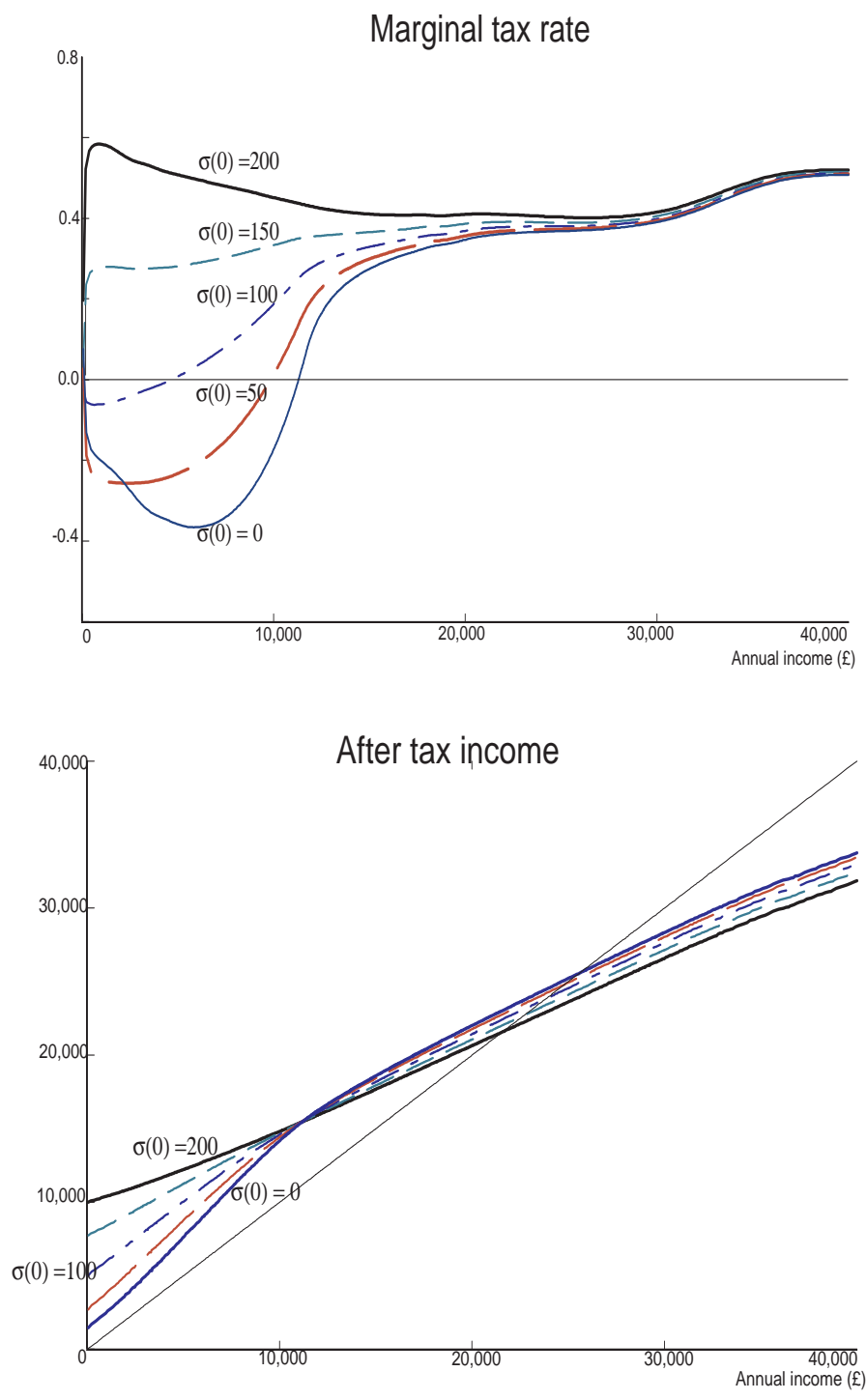


Figure 2: Optimal tax schedules and heterogeneity in work opportunity costs

with zero mean. In the baseline case, we take its standard error to be equal to £ 200, independently of  $\alpha$ . This is a small number (the median earnings in 2003 is equal to £ 16,500). In variants, we keep the conditional distribution of  $\theta^c$  unchanged for the top three quartiles, but we suppose that low  $\alpha$  types, i.e. wage earners in the first quartile of the distribution ( $\alpha \leq \text{£ } 11,850$ ), have less dispersed opportunity costs:

$$\sigma(\alpha) = 200k \left(1 - \frac{\alpha}{11850}\right) + 200 \left(\frac{\alpha}{11850}\right),$$

where  $k$  takes the values 0, .25, .50, .75 and 1, corresponding to a standard error  $\sigma(0)$  respectively equal to £ 0, 50, 100, 150 and 200.

It is not implausible that the dispersion of work opportunity costs increases with earnings, but we have no evidence to support this assumption. It certainly pushes towards negative marginal tax rates at the bottom of the income distribution. Since

$$\int \exp(\theta^c) dF(\theta^c|\alpha) = \exp\frac{\sigma(\alpha)^2}{2}$$

increases with  $\alpha$  for all  $k$ ,  $k < 1$ , the weights  $\pi(\alpha)$  will take the shape of Figure 1, for well chosen parameters. Indeed, as shown in the upper panel of Figure 2, the marginal tax rate increases with  $\sigma(0)$ , and negative tax rates are optimal for  $\sigma(0)$  equal to or smaller than £ 100. For small values of  $k$  and  $\sigma(0)$ , lower weights imply lower transfers to the low income population ( $R(0)$  is increasing with  $\sigma(0)$  as shown in the lower panel).

This example makes clear how lowering the social weights  $\pi(\alpha)$  at the bottom of the income distribution modifies the optimal tax schedule: it leads to redistribute away from the bottom towards the middle and the top of the distribution. This involves reducing the income *level* of the low  $\alpha$  types while increasing the slope of the after tax schedule (and therefore the labor supply of these agents) to maintain or increase the incomes of the middle and upper class.

Notice that, in this example, the distribution of  $\theta^c$  is symmetric ( $F$  increasing with  $\sigma(\alpha)$  for negative  $\theta^c$ , decreasing for positive  $\theta^c$ ) and the derivative  $\partial\tilde{\pi}^*/\partial\theta^c$  is increasing in  $\theta^c$ . Under such circumstances, the first term of equation (7), once integrated over  $\theta^c$ , is negative for all  $\alpha$ :

$$\int_{-\infty}^{\infty} \frac{\partial\tilde{\pi}^*}{\partial\theta^c}(\alpha, \theta^c) \frac{\partial F}{\partial\alpha}(\theta^c|\alpha) d\theta^c = \int_0^{\infty} \left[ \frac{\partial\tilde{\pi}^*}{\partial\theta^c}(\alpha, \theta^c) - \frac{\partial\tilde{\pi}^*}{\partial\theta^c}(\alpha, -\theta^c) \right] \frac{\partial F}{\partial\alpha}(\theta^c|\alpha) d\theta^c < 0.$$

Finally, while the numbers chosen for the dispersion of the work opportunity costs look small (£ 200), they are magnified through the expectation of the exponential of the normal variable:  $\sigma^2/2 = 20000$ . Indeed, one would get identical results in a model where  $\theta^c$  would be a deterministic function of  $\alpha$ , equal to  $20000 - \sigma^2(\alpha)/2$ . The certainty equivalent gain associated with the change in the distribution of work opportunity costs in the first quartile is  $\mathcal{L} (20000 - \sigma^2(\alpha)/2)$ , of the order of magnitude of average income. Further work is needed to better understand how to calibrate the effect of heterogeneity in the social choice function.

## 5 Conclusion

Even keeping with a setup where labor supply is driven by an exogenously predetermined unidimensional combination of the agents' characteristics, heterogeneity in utilities may play an important role in the determination of optimal redistributive schemes. The optimal allocation depends on the average social weights of the agents in the economy, computed conditionally on observable income. Heterogeneity enters through two channels: the individual social weight which depends on the microeconomic characteristics, and the distribution of characteristics conditional on income. We have spelled out this relationship, and applied it to find conditions under which negative marginal tax rates may be optimal at the bottom of the income distribution. A numerical example indicates that small variations with income of (unobserved) heterogeneity may induce large changes in optimal tax schedules.

## References

- BEAUDRY, P., C. BLACKORBY, AND D. SZALAY (2006): "Taxes and Employment Subsidies in Optimal Redistribution Programs (revised version)," Discussion paper, Warwick Economic Research Papers.
- BOADWAY, R., M. MARCHAND, P. PESTIEAU, AND M. D. M. RACIONERO (2002): "Optimal Redistribution with Heterogeneous Preferences for Leisure," *Journal of Public Economic Theory*, 4(4), 475–498.

- BRETT, C., AND J. A. WEYMARK (2003): “Financing Education using Optimal Redistributive Taxation,” *Journal of Public Economics*, 87, 2549–2569.
- BREWER, M., E. SAEZ, AND A. SHEPHARD (2008): “Means testing and tax rates on earnings,” Discussion paper, The Institute for Fiscal Studies.
- CHONÉ, P., AND G. LAROQUE (2007): “Negative Marginal Tax Rates and Heterogeneity,” Discussion paper, INSEE-CREST.
- CHONÉ, P., AND G. LAROQUE (2008): “Optimal taxation in the extensive model,” Discussion paper, Institute for Fiscal Studies.
- CUFF, K. (2000): “Optimality of Workfare with Heterogeneous Preferences,” *The Canadian Journal of Economics*, 33(1), 149–174.
- DIAMOND, P. (1980): “Income Taxation with Fixed Hours of Work,” *Journal of Public Economics*, 13, 101–110.
- (1998): “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates,” *American Economic Review*, 88, 83–95.
- EDLIN, A., AND C. SHANNON (1998): “Strict Single Crossing and the Strict Spence-Mirrlees Condition: A Comment on Monotone Comparative Statics,” *Econometrica*, 66(6), 1417–1425.
- HELLWIG, M. (2007): “A Contribution to the Theory of Optimal Utilitarian Income Taxation,” *Journal of Public Economics*, 91, 1449–1471.
- MIRRELES, J. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 38, 175–208.
- (1976): “Optimal Tax Theory: A Synthesis,” *Journal of Public Economics*, 6, 327–358.
- ROCHET, J.-C. (1987): “A necessary and sufficient condition for rationalizability in a quasi-linear context,” *Journal of Mathematical Economics*, 16, 191–200.
- SAEZ, E. (2002): “Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses,” *Quarterly Journal of Economics*, 117, 1039–1073.

SEADE, J. (1977): “On the Shape of Optimal Tax Schedules,” *Journal of Public Economics*, 7, 203–236.

——— (1982): “On the Sign of the Optimum Marginal Income Tax,” *The Review of Economic Studies*, 49, 637–643.

STIGLITZ, J. (1982): “Self-selection and Pareto Efficient Taxation,” *Journal of Public Economics*, 17, 213–240.

WERNING, I. (2000): “An Elementary Proof of Positive Optimal Marginal Taxes,” Discussion paper, MIT.

——— (2007): “Pareto Efficient Income Taxation,” Discussion paper, MIT.

## Appendix

### Proof of Lemma 1

We adapt the simple argument of Rochet (1987). (The property holds more generally under the single crossing condition.) First, suppose  $y$  is incentive compatible. Then, using the indirect utility  $U_R$ , incentive compatibility implies for  $\alpha_0 < \alpha_1$ :

$$v(y(\alpha_0), \alpha_1) - v(y(\alpha_0), \alpha_0) \leq U_R(\alpha_1) - U_R(\alpha_0) \leq v(y(\alpha_1), \alpha_1) - v(y(\alpha_1), \alpha_0). \quad (9)$$

It follows that:

$$v(y(\alpha_1), \alpha_1) - v(y(\alpha_1), \alpha_0) - v(y(\alpha_0), \alpha_1) + v(y(\alpha_0), \alpha_0) = \int_{\alpha_0}^{\alpha_1} \int_{y(\alpha_0)}^{y(\alpha_1)} v_{y\alpha} \geq 0,$$

which yields  $y(\alpha_1) \geq y(\alpha_0)$ .

Conversely, suppose  $y$  is nondecreasing. Then define  $U_R(\cdot)$  by its derivative  $U'_R(\alpha) = v_\alpha(y(\alpha), \alpha)$  and  $u(R(y(\alpha))) = U_R(\alpha) - v(y(\alpha), \alpha)$ , the general level of  $U_R$  and  $R$  being given by the government budget constraint. Then, for  $\alpha_0 < \alpha_1$ ,

using  $v_{y\alpha} > 0$ ,

$$\begin{aligned} U_R(\alpha_1) - U_R(\alpha_0) &= \int_{\alpha_0}^{\alpha_1} v_\alpha(y(x), x) dx \\ &\geq \int_{\alpha_0}^{\alpha_1} v_\alpha(y(\alpha_0), x) dx \\ &= v(y(\alpha_0), \alpha_1) - v(y(\alpha_0), \alpha_0), \end{aligned}$$

which yields incentive compatibility.

The allocation  $y(\alpha)$  being nondecreasing admits a right and left limit at every point  $\alpha$ . These limits are noted  $y(\alpha^+)$  and  $y(\alpha^-)$  respectively. Dividing the right inequality of (9) by  $\alpha_1 - \alpha_0 > 0$  shows that the difference quotient  $[U_R(\alpha_1) - U_R(\alpha_0)]/[\alpha_1 - \alpha_0]$  is bounded below  $v_\alpha(y(\alpha_0^+), \alpha_0)$  as  $\alpha_1$  tends to  $\alpha_0$  from above. Now permuting the indexes 0 and 1 in the left inequality of (9) and dividing by  $\alpha_0 - \alpha_1 > 0$  shows that the difference quotient is bounded above  $v_\alpha(y(\alpha_0^-), \alpha_0)$  as  $\alpha_1$  tends to  $\alpha_0$  from below. It follows that the indirect utility function  $U_R$  is everywhere continuous and differentiable whenever  $y$  is continuous. But  $y$  being a nondecreasing function is continuous everywhere except at most at countably many  $\alpha$ 's. This yields the desired result.

It is worth examining discontinuity points  $\alpha_0$  where  $y(\alpha_0^-) < y(\alpha_0^+)$ . Since  $U_R$  is continuous at  $\alpha_0$ , the agent  $\alpha_0$  is indifferent between  $y(\alpha_0^-)$  and  $y(\alpha_0^+)$ . So the left inequality of (9) holds with  $y(\alpha_0^+)$  instead of  $y(\alpha_0)$ . It follows that the difference quotient tends to  $v_\alpha(y(\alpha_0^+), \alpha_0)$  as  $\alpha_1$  tends to  $\alpha_0$  from above. Similarly the difference quotient tends  $v_\alpha(y(\alpha_0^-), \alpha_0)$  as  $\alpha_1$  tends to  $\alpha_0$  from below. Thus, the indirect utility everywhere admits a right and a left derivative, which coincide at all but at most countably many points.

Finally, incentive compatibility and  $v_y < 0$  yield for  $\alpha_1 > \alpha_0$

$$u(R(y(\alpha_1))) - u(R(y(\alpha_0))) \geq v(y(\alpha_0), \alpha_1) - v(y(\alpha_1), \alpha_1) \geq 0,$$

implying that  $R$  is nondecreasing in  $\alpha$  (and in  $y$ ). ■

## Proof of Lemma 2

We present here a heuristic derivation of the first order condition (4). To simplify notation, we denote the indirect utility by  $U$  rather than  $U_R$ . According



to Lemma 1,  $U'(\alpha) = v_\alpha(y(\alpha), \alpha)$  at all but at most countably many  $\alpha$ 's. As explained in the above proof, discontinuity points of  $y$  correspond to kinks of  $U$ . Since  $v_\alpha(y, \alpha)$  is increasing in  $y$ , the inverse function  $Y(\cdot, \alpha) = [v_\alpha(\cdot, \alpha)]^{-1}$  is well defined for all  $\alpha$ . Except at discontinuity points of  $y$ ,  $y(\alpha) = Y(U'(\alpha), \alpha)$  and

$$R(y_R(\alpha)) = \bar{R}(U, U', \alpha),$$

where  $\bar{R}(U, U', \alpha) = u^{-1}[U - v(Y(U', \alpha), \alpha)]$ . Note that  $Y_{U'} = 1/v_{y\alpha}$  and  $\bar{R}'_U = 1/u'(\bar{R}(U, U', \alpha))$ . Using  $U$  as the unknown function, the Lagrangian can be rewritten as

$$\mathcal{L} = \int_{\underline{\alpha}}^{\bar{\alpha}} \{ \pi(\alpha)U(\alpha) + \lambda[Y(U'(\alpha), \alpha) - \bar{R}(U(\alpha), U'(\alpha), \alpha)] \} dG(\alpha).$$

Consider small changes  $\Delta U(x)$  and  $\Delta U'(x)$  to the optimal solution  $U(x)$  and  $U'$ . Pick a point  $\alpha$  and keep  $U'$  unchanged outside a small interval  $[\alpha, \alpha + a]$  where  $y$  is locally strictly increasing (no bunching). Let  $\Delta U'(x)$  be a small continuous positive function on  $[\alpha, \alpha + a]$ , with  $\Delta U'(\alpha) = \Delta U'(\alpha + a) = 0$ . Define then

$$\Delta U(x) = \int_{\underline{\alpha}}^x \Delta U'(x) dG(x).$$

For sufficiently small  $a$  and  $\Delta U'$ , the variation  $\Delta U$  is admissible, i.e. the allocation  $Y(U' + \Delta U', \alpha)$  is increasing in  $\alpha$ . At the first order, the change in the Lagrangian is proportional to  $\Delta U(\bar{\alpha})$ . Supposing  $y$  continuous at  $\alpha$ ,  $\Delta \mathcal{L}/\Delta U(\bar{\alpha})$  tends to

$$\lambda \frac{g(\alpha)}{v_{\alpha y}} \left[ 1 + \frac{v_y(y_R(\alpha), \alpha)}{u'(R(y_R(\alpha)))} \right] + \int_{\alpha}^{\bar{\alpha}} \left( \pi(x) - \frac{\lambda}{u'(R(y_R(x)))} \right) dG(x) \quad (10)$$

as  $a$  goes to 0. From the agent's first order condition  $u'R' + v'_y = 0$ , the term in square brackets is equal to the marginal tax rate  $T'(y_R(\alpha))$ . By definition,  $g^*(\alpha) = \lambda g(\alpha)/u'(R(y_R(\alpha)))$ , so that using (2) the last term of 10 is equal to

$$\int_{\alpha}^{\bar{\alpha}} \left( \frac{\pi(x)u'(R(y_R(x)))}{\lambda} - 1 \right) g^*(x) dx = \int_{\alpha}^{\bar{\alpha}} \left( \frac{\pi^*(x)}{\lambda} - 1 \right) dG^*(x).$$

Using (3) and rearranging terms, it is easy to check that  $\lim \Delta \mathcal{L}/\Delta U(\bar{\alpha}) = 0$  is equivalent to (4).

At a discontinuity point  $\alpha$  of the allocation, the limits of  $\Delta\mathcal{L}/\Delta U(\bar{\alpha})$  as  $a$  tends to 0 from above and from below differ: they are given by (10) where  $y_R(\alpha^+)$  and  $y_R(\alpha^-)$  respectively replace  $y_R(\alpha)$ . Since  $U_R$  is continuous, as shown in the proof of Lemma 1, the agent  $\alpha$  is indifferent between the two values of  $y$ . The indifference curves in the plane  $(c, y)$  have a positive slope, equal to  $-u'/v_y$ , which decreases with  $y$ . Therefore the term in square brackets evaluated at  $y_R(\alpha^+)$  is lower than or equal to the same term evaluated at  $y_R(\alpha^-)$ . If this term is nonnegative (that is, if the marginal tax rates at the two points are nonnegative) and if  $v_{\alpha yy} \geq 0$ , the same property holds for the first term of (10) as a whole. Since the second term is continuous and the sum is zero, it must be the case that the allocation  $y_R$  is continuous:  $y_R(\alpha^-) = y_R(\alpha^+)$ .<sup>16</sup>

■

### Proof of Proposition 3

By integration by parts

$$\pi^*(\underline{\alpha}) - \lambda = \pi^*(\underline{\alpha}) - \int \pi^*(\alpha) dG^*(\alpha) = - \int \frac{\partial \pi^*}{\partial \alpha}(\alpha) [1 - G^*(\alpha)] d\alpha.$$

Differentiating  $\pi^*(\alpha) = \int_{\theta^c} \tilde{\pi}^*(\alpha, \theta^c) dF(\theta^c|\alpha)$  and integrating again by parts, we get

$$\begin{aligned} \frac{\partial \pi^*}{\partial \alpha}(\alpha) &= \int \left\{ \frac{\partial \tilde{\pi}^*}{\partial \alpha}(\alpha, \theta^c) f(\theta^c|\alpha) + \tilde{\pi}^*(\alpha, \theta^c) \frac{\partial f}{\partial \alpha}(\theta^c|\alpha) \right\} d\theta^c \\ &= \int \left\{ \frac{\partial \tilde{\pi}^*}{\partial \alpha}(\alpha, \theta^c) f(\theta^c|\alpha) - \frac{\partial \tilde{\pi}^*}{\partial \theta^c}(\alpha, \theta^c) \frac{\partial F}{\partial \alpha}(\theta^c|\alpha) \right\} d\theta^c + \\ &\quad \left[ \tilde{\pi}^*(\alpha, \bar{\theta}^c) \frac{\partial F}{\partial \alpha}(\bar{\theta}^c|\alpha) - \tilde{\pi}^*(\alpha, \underline{\theta}^c) \frac{\partial F}{\partial \alpha}(\underline{\theta}^c|\alpha) \right]. \end{aligned}$$

Noting that the last line is equal to zero since  $F(\bar{\theta}^c|\alpha) = 1$  and  $F(\underline{\theta}^c|\alpha) = 0$  and rearranging terms yield (7).

■

---

<sup>16</sup>It is easy to check that the assumption  $v_{\alpha yy} \geq 0$  yields the concavity of the Lagrangian with respect to  $U'$  in the region where marginal tax rates are nonnegative. We thank an anonymous referee for suggesting this property to us.