CESifo WORKING PAPERS

8443
2020

July 2020

Bunching and Rank-Dependent Optimal Income Taxation

Laurent Simula, Alain Trannoy

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8443*

Bunching and Rank-Dependent Optimal Income Taxation

Abstract

We consider optimal non-linear income tax problems when the social welfare function only depends on ranks as in Yaari (1987) and weights agree with the Lorenz quasi-ordering. Gini, S-Gini, and a class putting more emphasis on inequality in the upper part of the distribution belong to this set. Adopting a first-order approach, we establish marginal tax formula assuming a continuous population framework, and derive conditions on the primitives of the model for which the socially optimal allocation is either fully separating or involves some bunching. For all log-concave survival functions, bunching is precluded for the maximin, Gini, and ”illfare-ranked single-series Ginis”. We then turn to a discrete population setting, and provide ”ABC” formulas for optimal marginal tax rates, which are related to those for a continuum of types but remain essentially distinct.

JEL-Codes: D630, D820, H210.

Keywords: rank dependence, Gini, optimal income taxation, bunching, log-concavity.

Laurent Simula
ENS de Lyon & UMR GATE
University of Lyon / France
laurent.simula@ens-lyon.fr

Alain Trannoy
Aix-Marseille University (AMSE)
CNRS and EHESS
Marseille / France
alain.trannoy@univ-amu.fr

June 26, 2020
We thank Etienne Lehmann for helpful comments and discussion, and especially for his insights on the derivation of optimal tax formulas. We acknowledge financial support from IDEX-University of Lyon within the Programme Investissements d’Avenir (ANR-16-IDEX-0005).

I. Introduction

Among many other topics, John Weymark has contributed to the literature on optimal income taxation (see, e.g., Weymark (1986a,b, 1987)) as well as on the better understanding of inequality indexes of the Gini family (see, e.g., Weymark (1981)). The present article articulates the former and the latter, and more specifically explore some properties of optimal non-linear income taxation for rank-dependent welfare functions à la Yaari (1987). In that case, individual welfare weights depend on the *position* of individuals in the distribution of indirect utility. We consider individual weights such that social welfare can be written as the egalitarian benchmark, deflated by inequality measured by the Gini, S-Gini or ″illfare-ranked single-series Ginis″ introduced by Donaldson and Weymark (1980) as well as Bossert (1990). The illfare-ranked single series Ginis are referred to as the ″A-family″ below, given that they have extensively been studied by Aaberge (2000). Beyond Lorenz agreement, the S-Gini family favors a reduction of inequality in the lower part of the distribution, while the A-family places more weight on a reduction of inequality in the upper part.

Following Mirrlees (1971), the productivity of each agent is private information, and the benevolent policymaker only knows how productivity is distributed within the population. As a result, the optimal allocation maximizes rank-dependent welfare subject to the tax-revenue *and* incentive-compatibility constraints. The latter are necessary and sufficient to provide each agent with the right incentive for them to truthfully report their private information, i.e., to behave in the way chosen by the policy-designer. Incentive-compatible allocations verify two conditions: on the one hand, indirect utility must increase at a sufficient rate to induce truth-telling; on the other hand, gross income (or equivalently net income) must be non-decreasing with productivity. A situation in which the ″monotonicity″ condition on gross income is binding is referred to in the literature as ″bunching″. In other words, ″bunching″ is said to occur in a social optimum when agents with different productivity levels choose the same gross-income/net-income combination. Absent bunching, ranks in terms of indirect utility, gross income, net income and productivity are all the same. In addition, providing agents behave in a rational way, these ranks corresponds to those in the *actual* allocation. Ranks then become ″invariant″ and provide a standard with respect to which tax rates may be expressed, as suggested by Trannoy (2019).

To cast light on the main mechanisms and intuitions, we consider that individual utility is quasi-linear, linear with respect to net income. Such preferences are widely used in the literature following Piketty (1997) and Diamond (1998), even though assuming that there is no income effect on gross income is obviously restrictive. For a continuous population and relying on the two families of parameterized rank-dependent welfare functions presented above (with Gini social welfare function as a common element), we are able to obtain clear and intuitive formulas for optimal income tax rates absent bunching. In addition, we derive tax formulas for a discrete population, and find expressions as close as possible to the ″ABC″ formula in the continuous-population case.

We then pay special attention to situations in which the optimal allocation is not ″fully separating″ in a continuous setting. Thanks to rank-dependent welfare weights, we are able to obtain several conditions on the *parameters* of the model, under which bunching in the social optimum is either present or precluded. These conditions are derived exploiting the implicit expression for the optimal gross income function, coming out from the optimization process.

This expression is particularly simple due to the combined assumptions of quasi-linear individual preferences and rank-dependent social welfare weights. To study bunching, we restrict the class of preferences to those for which the marginal dis-utility of gross income is convex, but less convex the higher the agent's productivity. Under this assumption, the discussion about bunching only depends on the shape of the productivity distribution and the class of rank-dependent social welfare functions (S-Gini or A-families).

We in particular resort to log-concavity of the type's survival function or log-convexity of the type's probability density function (pdf). Under log-concavity of survival function of the productivity distribution and for the whole A-family (including Gini), the optimal allocation does not involve any bunching whatsoever; if the focus is on reducing inequality in the top of the income distribution, bunching is precluded. The same assumption of log-concavity does not guarantee the absence of a pooling equilibrium when we pay closer attention to reducing inequalities in the lower part of the distribution, as for the S-Gini family. We then are only able to prove that the optimal gross income function is increasing above a productivity threshold.

We also examine specific distributions as did Boadway et al. (2000) for quasi-linear-in-leisure preferences and weighted utilitarianism and Boadway and Jacquet (2008) for separable preferences and the "Rawlsian" maximin. The findings in Boadway et al. (2000) regarding the absence of bunching for any exponential distribution of productivity are extended to our setting. In addition, the result in Boadway and Jacquet (2008) regarding the absence of bunching under the maximin, quasi-linear-in-consumption preferences, and a Weibull probability density for types (with scaling parameter above 1), appears as a subcase of a result we obtain for all productivity distributions with a log-concave survival function. Indeed, the maximin appears as a limit case to the S-Gini family, and in this situation, bunching does not occur in the social optimum under the considered assumptions.

Weymark (1987) and Simula (2010) already considered optimal income taxation for a social welfare function with exogenous individual weights, diminishing with productivity. The parametric families we consider, consistent with Lorenz second-order stochastic dominance, are specific examples of such decreasing weights. One of their main interest is to base such welfare weights by relying upon the literature on inequality measurement. In addition, the various patterns of decreasing welfare weights they generate allow us to go one step further and obtain both simpler expressions for optimal marginal tax rates and clearer conditions regarding bunching. This findings complements Simula and Trannoy (2020) through a more technical exploration in the theory of rank-depend optimum income taxation.

Many articles on optimal taxation, among which several of us Simula and Trannoy (2010) or Lehmann et al. (2014), adopt the "first-order" approach, ignoring the monotonicity constraints for gross income, and checking it ex post in numerical simulations. Brito and Oakland (1977) and Lollivier and Rochet (1983) were the first to explain how to formally account for this difficulty which already had been noted by Mirrlees (1971). For continuous populations, Lollivier and Rochet (1983) considered an example in which bunching occured in the social optimum, showing the limitations of the first-order approach. In addition, Ebert (1992) provided an other example of bunching and examined whether the general features of the "optimal" schedule obtained when ignoring the possibility of bunching were preserved in the "full" optimum. For a discrete population, Weymark (1986b) –focusing on quasilinear-in-leisure preferences– provided an example in which the solution to the "relaxed problem" (ignoring possible bunching) cannot be the solution to the full problem, before showing how to solve the full problem through ”ironing” techniques. Also for a discrete population, Simula (2010) provided conditions for bunching to occur in the social optimum when preferences are quasi-linear, linear with respect to consumption.

The article is organized as follows. Section II introduces rank-dependent welfare functions. Section III examines optimal marginal tax rates in the absence of bunching. Section IV characterizes the bunching pattern, deriving conditions on the primitives of the model. Section V focuses on a discrete population of types to show similarities and differences with regards to the continuous setting. Section VI provides concluding comments.

II. FROM WEIGHTS DEPENDING ON RANKS TO OPTIMAL TAXATION

This section introduces rank-dependent social welfare function and shows how they connect to the literature on inequality measurement. We present the concepts assuming a continuous population of individuals, differing with a single dimension of heterogeneity. This setting will be maintain in Sections III and IV. Section V focuses on similarities and differences between the continuous and discrete population settings.

II.1. Rank-Dependent Social Welfare

We consider a population of individuals, heterogeneous with respect to a variable x. For simplicity, we assume that the latter is uni-dimensional and smoothly distributed according to the cumulative distribution function (CDF) denoted $F(x)$, with support $X \subseteq \mathbb{R}_+$. We call $f(x)$ the corresponding probability density function (pdf). The average value of x within the population is $\mu = \int_X x f(x) dx$.

We define the quantile function as $F^{-1}(p) = x$, where $p \in [0,1]$ stands for the rank or ”position”, and introduce weights to capture the social planner’s aversion to inequality. The marginal weights are denoted $\lambda(p)$ and the cumulated weights $\Lambda(p) = \int_0^p \lambda(\rho) d\rho$. In the whole article, we focus on weights consistent with second-order stochastic dominance, belonging to the set:

$$\mathcal{L} = \{\forall p \in (0,1), \lambda(p) > 0 \text{ and } \lambda'(p) < 0; \Lambda(0) = \lambda(1) = 0; \Lambda(1) = 1\}. \tag{1}$$

The assumption that $\lambda(p)$ is positive and decreasing means that every individual counts, but to a lower extent the higher the rank. The other assumptions are normalizations. On this basis, *rank-dependent social welfare* (Yaari, 1987, 1988) is defined as:

$$\mathcal{W} = \int_0^1 \lambda(p) F^{-1}(p) dp. \tag{2}$$

II.2. From Rank-Dependent Social Welfare to Inequality Indexes

By definition, F is an *egalitarian* distribution if and only if $\mathcal{W} = \mu$. Otherwise, there is a positive gap $\Delta \equiv \mu - \mathcal{W}$ between social welfare $\mathcal{W}$ and the equality benchmark. Dividing this gap by μ, we obtain the mean-invariant inequality index, $I \equiv \Delta/\mu$. Using these definitions, we can rewrite social welfare (2) in *abbreviated form*:

$$\mathcal{W} = \mu(1 - I). \tag{3}$$

This expression illustrates the close connection between rank-dependent social welfare functions and inequality indexes: $\mathcal{W}$ is equal to the egalitarian benchmark deflated by inequality as measured by I.

Specifying the weights $\Lambda(p)$, we focus on two important families of rank-dependent social welfare functions:

- The S-Gini family (Donaldson and Weymark, 1980) for $\Lambda(p) = 1 - (1-p)^{\delta}$ and $\delta \geq 2$.
- The "A" family for $\Lambda(p) = (\delta p - p^{\delta})/(\delta - 1)$ and $\delta \geq 2$.[1] When $\delta \to 1$, $\Lambda(p) = p(1 - \log(p))$ which corrresponds to the rank-dependent welfare function $\mathcal{W}$ in which I is the Bonferroni index of inequality.[2]

For both families, the weights coincide when $\delta = 2$, with $\Lambda(p) = p(2-p)$. In that case, the inequality measure I is the Gini coefficient and $\mathcal{W}$ the Gini social welfare function introduced by Sen (1974). The marginal weights $\lambda(p) = 2(1-p)$ are then linear with respect to rank. By contrast, for any $\delta > 2$, the weights $\lambda(p)$ are convex when considering the S-Gini family, and concave for the A-family. To gain further insights, let us consider a fixed transfer taking place between two agents with equal difference in ranks. For the S-Gini family, convexity implies that the equalizing effect of the transfer becomes larger the lower the ranks considered. The focus is thus on poverty. On the contrary, for the A family, the higher the ranks the stronger the equalizing effect. The emphasis is layed on inequalities at the top of the distribution. When δ goes up, the S-Gini and A-families tend to two important benchmarks, the Rawlsian maximin on the one hand and pure utilitarianism on the other hand. In addition, for every interior p, the cumulated weights are larger in the Bonferroni case than in the Gini one, implying that the former stresses poverty more than the latter.

III. OPTIMAL MARGINAL TAX RATES

This Section is devoted to the characterization of optimal marginal income tax rates for a population of agents differing with respect to productivity θ belonging to a compact subset $[\underline{\theta}, \overline{\theta}]$ of the positive real line. The parameter of heterogeneity θ is smoothly distributed, with cumulative density function H and probability density function $h = H'$ (satisfying $h > 0$ over its support). Its distribution is common knowledge; but the exact productivity of a given agent is only known to herself. The connection with the distribution of x introduced in Section II will be made clear below.

[1] Aaberge (2000) refers to this class as the Lorenz family for $\delta = 2, 3, ...$ It corresponds to the integer subfamily of the "illfare-ranked single-series Ginis" discussed by Donaldson and Weymark (1980) and Bossert (1990). We refer to the latter as the "A" family. Aaberge (2009) and Aaberge et al. (2020) have indeed shown the usefulness of this family for analyzing inequality when Lorenz curves intersect.

[2] Let $\mu(x) = \left(\int_0^x x dF(x)\right) / F(x)$ and $r(x) = (\mu - \mu(x))/\mu$. The Bonferroni inequality index is $\int_X r(x) dF(x)$.

III.1. Formulation of the Optimal Income Tax Problem

The starting point is the utility maximization problem of an agent with given productivity θ. The latter chooses consumption c and taxable income z so as to maximize:

$$u(c,z;\theta) = c - v(z;\theta) = z - T(z) - v(z;\theta), \tag{4}$$

where $T(z)$ stands for the nonlinear income tax schedule. It should be noted that the dis-utility of gross income $v(z;\theta)$ is written in a more general way than the more commonly-used multiplicative form where $v(z;\theta) = \nu(z/\theta)$. In that case, gross income is obtained as the product of the wage rate and working hours. The specification that we use is not just a theoretical refinement. Indeed, empirical findings do not bolster the case for a multiplicative form, except for shift work. In addition, we let the dis-utility of gross income $v(z;\theta)$ be increasing and convex in gross income z. Hence, $v'_z(z;\theta) > 0$ and $v''_{zz}(z;\theta) > 0$. In addition, we assume that the Spence-Mirrlees condition is satified.

ASSUMPTION 1 (Spence-Mirrlees Condition). $v''_{z\theta}(z;\theta) < 0$.

Assumption 1 has the following interpretation. When productivity goes up, it becomes easier for an agent to increase gross income by a small given amount dz. We do not mention it explicitly in Propositions and Lemmas below, but it holds in the rest of the article.

The first-order condition of Agent θ's utility maximization program is:

$$1 - T'(z) = v'_z(z;\theta). \tag{5}$$

We call $c(\theta)$ and $z(\theta)$ the optimal consumption and taxable income, and $V(\theta) = c(\theta) - v(z(\theta);\theta)$ the corresponding indirect utility. Because of the taxation principle, a tax function is equivalent to the specification of a $(c(\theta), z(\theta))$-allocation subject to incentive-compatibility constraints. The latter ensure that a θ-individual has an incentive to disclose private information, i.e., her productivity level θ. This is the case providing:

$$V(\theta) = \max_{\theta'} c(\theta') - v(z(\theta');\theta) \text{for any } (\theta,\theta') \in \left[\underline{\theta}, \overline{\theta}\right]^2. \tag{6}$$

As is well-known (see, e.g., Salanié (2005)), this is equivalent to:

$$V'(\theta) = -v'_\theta(z(\theta);\theta), \tag{7}$$

$$z'(\theta) \geq 0. \tag{8}$$

Equation (7) is the first-order condition for incentive compatibility. It ensures that indirect utility $V(\theta)$ increases at a sufficient rate to prevent agents from downward mimicking. Inequality (8) is the second-order condition for incentive compatibility. It requires gross income $z(\theta)$ to be non-decreasing when productivity θ goes up. It is equivalent to $c'(\theta) \geq 0$.

Because $T(z(\theta)) = z(\theta) - c(\theta)$, the government's budget constraint can be written as:

$$\int_{\underline{\theta}}^{\overline{\theta}} (z(\theta) - c(\theta))dH(\theta) \geq E, \tag{9}$$

where E is an exogenous amount of public expenditures. We focus on the case in which the tax policy is purely redistributive. Hence, $E = 0$ from now on. Moreover, in any social optimum, (9) will be binding. Using (6), we obtain $c(\theta) = V(\theta) + v(z(\theta);\theta)$, which can be substituted into the binding government's budget constraint (9) to get:

$$\int_{\underline{\theta}}^{\overline{\theta}} (z(\theta) - V(\theta) - v(z(\theta);\theta))dH(\theta) = 0. \tag{10}$$

In reference to Section II, we let $x = V$. Consequently, $p = F(V(\theta)) = H(\theta)$, which also implies: $F^{-1}(p) = V(\theta)$. The social objective is to maximize a weighted sum $\mathcal{W}$ of indirect utilities $V(\theta)$, with weights given by $\lambda(p) = \lambda(H(\theta))$, i.e.,

$$\mathcal{W} = \int_{\underline{\theta}}^{\overline{\theta}} \lambda(H(\theta))V(\theta)dH(\theta). \tag{11}$$

The optimal income tax problem can therefore be stated as follows:

PROBLEM 1 (Full Problem). Choose $z(\theta)$ and $c(\theta)$ to maximize $\mathcal{W}$ subject to the incentive-compatibility constraints (7) and (8) as well as the tax-revenue constraint (10).

We also formulate a variation of Problem 1 from which the second-order condition for incentive compatibility (8) has been removed.

PROBLEM 2 (Relaxed Problem). Choose $z^*(\theta)$ and $c^*(\theta)$ to maximize $\mathcal{W}$ subject to the first-order condition for incentive compatibility (7) and the tax-revenue constraint (10).

The so-called "first-order approach" consists in solving Problem 1 without explicitly accounting for the second-order condition for incentive compatibility (8). The idea is to solve the Relaxed Problem and verify in a second step whether its solution generates a non-decreasing pattern of taxable income $\theta \to z^*(\theta)$. If this is the case, it is also the solution to the Full Problem, and $z(\theta) = z^*(\theta)$ at any productivity θ in $[\underline{\theta}, \overline{\theta}]$.

III.2. Solution to the Relaxed Problem

The method of derivation follows the same steps as Brett and Weymark (2017) in a political-economy setting.[3] The main advantage of this procedure is to solve the optimal income tax problem in a clear and transparent manner, without resorting to optimal control theory. In addition, this derivation accommodates the possibility of mass points in the distribution of social weights. On this basis, the Rawlsian maximin in particular can be obtained as a limit case of the $\mathcal{L}$-class. More generally, cumulated social weights Λ are allowed to be defined as a step function, which is continuous almost everywhere, with a finite number of discontinuity points.

This Subsection is devoted to the solution to the Relaxed Problem. We first used the first-order condition for incentive compatibility (7). Integrating between $\underline{\theta}$ and any given θ, we

[3]Brett and Weymark (2017) characterize the income tax schedule maximizing the utility of a given agent subject to incentive-compatibility constraints and budget balancedness. Brett and Weymark (2017) themselves rely on Lollivier and Rochet (1983) who present a simple non-linear taxation model in which the agents, with quasi-linear in labor preferences, are indexed by a one-dimensional parameter.

obtain:

$$V(\theta) = V(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} v'_{\theta}(z(t);t)dt. \tag{12}$$

Replacing the latter into $\mathcal{W}$ given in (11) yields:

$$\mathcal{W} = V(\underline{\theta}) \int_{\underline{\theta}}^{\overline{\theta}} \lambda(H(\theta))dH(\theta) - \int_{\underline{\theta}}^{\overline{\theta}} \lambda(H(\theta)) \left[\int_{\underline{\theta}}^{\theta} v'_{\theta}(z(t);t)dt \right] dH(\theta). \tag{13}$$

Because $\Lambda(1) = 1$ for any $\theta \to \lambda(\theta)$ in $\mathcal{L}$, the first term on the right-hand side of (13) is equal to $V(\underline{\theta})$. Therefore, (13) may be rearranged as:

$$\mathcal{W} = V(\underline{\theta}) - \iint_{\underline{\theta} \le t \le \theta \le \overline{\theta}} \lambda(H(\theta)) v'_{\theta}(z(t),t)dH(\theta)dt. \tag{14}$$

Making use of Fubini's Theorem, this simplifies into:

$$\mathcal{W} = V(\underline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} v'_{\theta}(z(\theta),\theta) \left[1 - \Lambda(H(\theta))\right] d\theta. \tag{15}$$

It is worth noting that the above derivation does not require the weight function λ to be continuous. We only need them to be Lebesgue integrable. Therefore, they remain valid for Rawlsian weights, with a mass point at $\underline{\theta}$ and thereafter equal to zero.

The next step is to plug in $V(\theta)$ as defined in (12) into the government's budget constraint (10). We obtain:

$$\int_{\underline{\theta}}^{\overline{\theta}} (z(\theta) - v(z(\theta);\theta) - V(\underline{\theta}))dH(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} v'_{\theta}(z(t);t)\, dt\, dH(\theta) = 0, \tag{16}$$

which is equivalent to:

$$V(\underline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}} (z(\theta) - v(z(\theta);\theta))dH(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} v'_{\theta}(z(t);t)\, dt\, dH(\theta) \tag{17}$$

because $\int_{\underline{\theta}}^{\overline{\theta}} \lambda(\theta)dH(\theta) = 1$. We then employ Fubini's Theorem to evaluate the double integral and get:

$$\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} v'_{\theta}(z(t);t)\, dt\, dH(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} v'_{\theta}(z(\theta);\theta)(1 - H(\theta))d\theta, \tag{18}$$

which remains valid even if H is a step function. Consequently, (17) may be rearranged as:

$$V(\underline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}} (z(\theta) - v(z(\theta);\theta))dH(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} v'_{\theta}(z(\theta);\theta)(1 - H(\theta))d\theta \tag{19}$$

The above expression may now be substituted into (15), which yields:

$$\mathcal{W} = \int_{\underline{\theta}}^{\overline{\theta}} \left[\left(z(\theta) - v(z(\theta);\theta) \right) \right] dH(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} v'_{\theta}(z(\theta);\theta) \left[\Lambda(H(\theta)) - H(\theta) \right] d\theta. \tag{20}$$

The Relaxed Problem amounts to choosing $\theta \to z(\theta)$ such that $\mathcal{W}$ defined in (20) is maximum. The first-order condition is:

$$1 - v'_z(z(\theta);\theta))h(\theta) = -v''_{\theta z}(z(\theta);\theta) \left[\Lambda(H(\theta)) - H(\theta) \right] \text{ for any } \theta \in \left[\underline{\theta}, \overline{\theta} \right], \tag{21}$$

which may also be rearranged as follows:

$$1 - v'_z(z(\theta);\theta)) = -v''_{\theta z}(z(\theta);\theta) \frac{\Lambda(H(\theta)) - H(\theta)}{h(\theta)} \text{ for any } \theta \in \left[\underline{\theta}, \overline{\theta} \right], \tag{22}$$

providing $h(\theta) > 0$, as assumed in this Section. We obtain the following Lemma.

LEMMA 1. Optimal gross income levels are determined by:

$$-\frac{1 - v'_z(z(\theta);\theta))}{v''_{\theta z}(z(\theta);\theta)} = \frac{\Lambda(H(\theta)) - H(\theta)}{h(\theta)} \text{ for any } \theta \in \left[\underline{\theta}, \overline{\theta} \right]. \tag{23}$$

The latter equation is of particular interest for the study of bunching because it provides the optimal gross income function implicitly. Its LHS is related to individual preferences, while the RHS is determined by the distributions of both productivity and social weights.

Thanks to the first-order condition of the individual utility maximization program (5), we see that the right-hand side of (21) is equal to the marginal tax rate solution to the Relaxed Problem, that we call T'^{*}. The z-function solution to (21) is also that obtained for the Relaxed Problem, that we denote by z^{*}. The following Proposition summarizes the results.

PROPOSITION 1. The tax function T^{*} and gross income z^{*} solution to the Relaxed Problem are such that:

$$T'^{*}(z^{*}(\theta)) = -v''_{\theta z}(z^{*}(\theta);\theta) \frac{\Lambda(H(\theta)) - H(\theta)}{h(\theta)} \text{ for any } \theta \in \left[\underline{\theta}, \overline{\theta} \right]. \tag{24}$$

In Proposition 1, $\frac{\Lambda(H(\theta)) - H(\theta)}{h(\theta)}$ only depends on the distribution of productivity within the population, and therefore only on rank $p = H(\theta)$ and its derivative,

$$p' = \pi. \tag{25}$$

If we define $G(p,\pi)$ as:

$$G(p(\theta), \pi(\theta)) = \frac{\Lambda(p(\theta)) - p(\theta)}{\pi(\theta)}, \tag{26}$$

Formula (24) may be rewritten as:

$$T'^{*}(z^{*}(\theta)) = -v''_{\theta z}(z^{*}(\theta);\theta)\, G(p(\theta), \pi(\theta)). \tag{27}$$

III.3. *Interpretation*

First, we see that Formulas (24) and, equivalently, (27) correspond to a joint determination of the optimal marginal tax rates and the optimal gross income levels. They do not, in particular, incorporate any term related to consumption. This feature allows a sequential determination of the optimal allocation: in a first step, gross incomes and marginal tax rates are jointly obtained. Optimal consumption levels can then be characterized in a second step. This two-step procedure is a byproduct of quasi-linear preferences as well as weights which only depend on ranks. It was emphasized in a discrete population setting by Simula (2010), who develops a sequential procedure to construct the optimal allocation. In other words, Formulas (24) and (27) capture an interaction between two endogenous variables: the marginal tax rate, on the one hand, and gross income, on the other hand. Applying the implicit function theorem to $T'^{*}(z) + v''_{\theta z}(z^{*};\theta)G(p,\pi) = 0$, the effect of marginal tax rates on gross income are given by:

$$\frac{\partial z^{*}}{\partial T'^{*}} = -\frac{1}{v'''_{zz\theta}(z^{*};\theta)}G(p,\pi). \tag{28}$$

Therefore, the following Corollary is obtained:

COROLLARY 1. Gross income z^{*} decreases with the marginal tax rate if and only iff $v'''_{zz\theta} > 0$.

Second, following the insights of Saez (2001), it has become usual to pay a lot of attention to the part played by behavioral elasticities in optimal income tax formulas. In Proposition 1, the behavioral term is equal to $-v''_{\theta z}(z^{*};\theta)$, which –as expected– is directly related to the Spence-Mirlees single-crossing condition. This term tells us how the slope of the indifference curve in the gross-income/net-income space should vary with productivity. It is is non-negative under the Spencer-Mirrlees condition, implying that the dis-utility of working an extra hour decreases with productivity. The more it does so, the steeper the profile of marginal tax rates will be. Indeed, the potential downward mimicking behavior then becomes less stringent. Everything else being equal, the optimal tax schedule can thus be more marginally progressive.

Third, the above Formulas may also be written as an "ABC formula". This allows a more direct comparison with standard optimal income tax formulas, and in particular Diamond's (1998) one. Making use of the first-order condition of agent θ's utility maximization program (5), we obtain:

$$\frac{T'^{*}(z^{*}(\theta))}{1 - T'^{*}(z^{*}(\theta))} = -\frac{v''_{\theta z}(z^{*}(\theta);\theta)}{v'_{z}(z^{*}(\theta);\theta))}G(p(\theta),\pi(\theta)) \tag{29}$$

It remains to understand the intuition behind the fraction on the right-hand side. To this aim, it should be noted that the elasticity of taxable income with respect to the marginal retention rate is equal to:

$$\epsilon_{1-T'}(z,\theta) = \frac{\partial z}{\partial(1-T')}\frac{1-T'}{z} = \frac{1-T'}{zv''_{zz}(z;\theta)} = \frac{v'_{z}(z;\theta)}{zv''_{zz}(z;\theta)}, \tag{30}$$

where we made use of the implicit function theorem applied to (5). Following Jacquet et al. (2013) (see also Lehmann et al. (2014)), we also introduce the elasticity of taxable income with respect to productivity (i.e., gross wage in the laissez-faire) for a linearized income tax schedule with the same slope as T at the gross income/consumption combination chosen by taxpayer θ. It is given by:

$$\epsilon_{\theta}(z,\theta) = \frac{\partial z}{\partial \theta}\frac{\theta}{z} = -\frac{\theta v''_{z\theta}(z;\theta)}{zv''_{zz}(z;\theta)}, \tag{31}$$

where the implicit function theorem was employed again to the taxpayer's first-order condition (5), neutralizing any change in the marginal tax rate due to the assumed linearization. The ratio of both elasticities, denoted $e(z,\theta)$, is equal to:

$$e(z,\theta) \equiv \frac{\epsilon_{1-T'}(z,\theta)}{\epsilon_{\theta}(z,\theta)} = -\frac{v'_z(z;\theta)}{\theta v''_{z\theta}(z;\theta)}. \tag{32}$$

Using (32), Formula (29) may be rearranged as:

$$\frac{T'^*(z^*(\theta))}{1-T'^*(z^*(\theta))} = \frac{1}{e(z^*(\theta);\theta)}\frac{1}{\theta}G(p(\theta),\pi(\theta)), \tag{33}$$

or, equivalently, as:

$$\frac{T'^*(z^*)}{1-T'^*(z^*)} = \frac{1}{e(z^*(\theta);\theta)}\frac{1-p(\theta)}{\theta\pi(\theta)}\frac{\Lambda(p(\theta))-p(\theta)}{1-p(\theta)}. \tag{34}$$

We see that the elasticity of taxable income with respect to the marginal retention rate is "normalized" by the elasticity of taxable income with respect to productivity. The latter provides a benchmark to which the former is compared. However, even though "ABC" formulas offer a nice and insightful economic interpretation, they do not offer a direct route to find out the optimum allocation, except when elasticities are independent of gross income z.

IV. BUNCHING

The previous Section provided a formulation of the Full Problem and then focused on the Relaxed Problem to offer a characterization and interpretation of optimal marginal tax rates. This characterization is valid providing $\theta \to z^*(\theta)$ is non-decreasing in productivity, so that the monotonicity constraint (8) is satisfied. Otherwise, the solution to the Relaxed Problem is not solution to the Full Problem. In this Section, we examine cases in which $\theta \to z^*(\theta)$ is non-decreasing, so that the allocation solution to the Full Problem does not involve any bunching of types; it is then "fully separating". We also examine the opposite situation, in which some types are bunched together.

IV.1. General Properties

We let θ_p be the solution in θ to: $p = F(\theta)$. The latter increases with p because F is increasing. Given this definition, note that $z^*(\theta_p)$ being strictly increasing with rank p is a sufficient condition for bunching not to arise, so that $z(\theta_p) = z^*(\theta_p)$ for any rank p. Moreover, a necessary and sufficient condition for bunching *to arise* is that $z^*(\theta_p)$ be non-increasing. In that case, $z(\theta) \neq z^*(\theta)$ on at least of subset of $\left[\underline{\theta},\overline{\theta}\right]$ with positive measure.

The first strategy we follow is to look at conditions on the primitives of the model for which $z^*(\theta)$ is increasing over $\left[\underline{\theta},\overline{\theta}\right]$ and therefore equal to $z(\theta)$. Its starting point is to consider the Relaxed Problem's first-order condition (23). We rewrite it as:

$$1 - v'_z(z) + v''_{\theta z}(z;\theta)\Gamma(\theta) = 0, \tag{35}$$

where

$$\Gamma(\theta) = G(p(\theta), \pi(\theta)) = \frac{\Lambda(p(\theta)) - p(\theta)}{\pi(\theta)}, \tag{36}$$

only depends on how types are distributed within the population. Applying the implicit function theorem, we obtain:

$$z^{*\prime}(\theta) = -\frac{-v''_{\theta z} + v'''_{z\theta\theta}\Gamma(\theta) + v''_{\theta z}\Gamma'(\theta)}{-v''_{zz} + v'''_{zz\theta}\Gamma(\theta)} = -\frac{v'''_{z\theta\theta}\Gamma(\theta) + v''_{\theta z}(\Gamma'(\theta) - 1)}{-v''_{zz} + v'''_{zz\theta}\Gamma(\theta)} \tag{37}$$

where some arguments have been omitted. We know that $\Gamma(\theta) > 0$, $v''_{z}(z;\theta) > 0, v''_{\theta z}(z;\theta) < 0$. If we add: $v'''_{zz\theta}(z;\theta) \leq 0$, $v'''_{z\theta\theta}(z;\theta) > 0$ and $\Gamma'(\theta) \leq 1$, we obtain a sufficient condition for $z^{*\prime}(\theta) > 0$. The following Proposition summarizes the result.

PROPOSITION 2. Assume $v'''_{zz\theta}(z;\theta) \leq 0$, $v'''_{z\theta\theta}(z;\theta) > 0$ and $\Gamma'(\theta) \leq 1$ for any θ. Then, the solution to the Relaxed Problem is the solution to the Full Problem. In particular:

(i) $z(\theta) = z^*(\theta)$ for any $\theta \in [\underline{\theta}, \overline{\theta}]$;

(ii) The solution is fully separating, i.e., does not involve any bunching.

The assumptions on the third derivatives of $v(z;\theta)$ made in Proposition 2 are additional requirements to the Spence-Mirrlees condition, $v''_{z\theta}(z;\theta) < 0$ (see Assumption 1 above). The first one, $v'''_{zz\theta}(z;\theta) \leq 0$, already appeared in Corollary 1. The marginal dis-utility of highly productive individuals is then less convex than that of less productive agents. The second one, $v'''_{z\theta\theta}(z) > 0$, requires the marginal dis-utility of gross income to decrease with productivity at a decreasing rate. Both conditions are verified when an iso-elastic specification is used for the dis-utility of gross income $v(z;\theta)$. They are also satisfied when the dis-utility of gross income takes the often used multiplicative form, $v(z;\theta) = \phi(z/\theta)$, as soon as we add $\phi''' > 0$ to the classical conditions $\phi' > 0$ and $\phi'' > 0$. The marginal dis-utility of gross income is then increasing and convex. In the rest of this Section, we assume:

ASSUMPTION 2. The marginal dis-utility of gross income is convex, but less convex the higher the productivity: $v'''_{zz\theta}(z;\theta) < 0$ and $v'''_{z\theta\theta}(z) > 0$.

On this basis, our strategy is to check whether $\Gamma'(\theta) \leq 1$ for the different scenarios we consider. Starting from the definition of $\Gamma(\theta)$ in (36) and differentiating, we obtain:

$$\Gamma'(\theta) - 1 = (\lambda(p(\theta)) - 2) - \frac{(\Lambda(p(\theta)) - p(\theta))\pi'(\theta)}{\pi^2(\theta)}. \tag{38}$$

First, note that a sufficient condition for the second term on the right-hand side (RHS) to be non-positive is $\pi'(\theta) \geq 0$. Second, the first term on the RHS is always negative for the Aaberge Family including Gini because $\lambda(0) = 0 < 2$ and $\lambda'(p) < 0$. For the two other families, S-Gini and Bonferroni respectively, the first term is negative providing $\lambda(p(\theta)) < 2$. Because $\lambda(\cdot)$ is monotone decreasing while p is monotone increasing, then $(\lambda \circ p)\,(.)$ is monotone decreasing. In addition, $\lambda(0) > 2$ and $\lambda(\overline{\theta}) = 0$, there exists θ^* in $(\underline{\theta}, \overline{\theta})$ such that (i) $\lambda(p(\theta^*)) = 2$ and (ii) $\lambda(p(\theta)) < 2$ if and only if $\theta \geq \theta^*$. Consequently, the first term in (38) is negative for all $\theta > \theta^*$. The following Proposition follows from these observations.

PROPOSITION 3. Let Assumption 2 and $\pi'(\theta) \geq 0$ be verified. Then:

(i) For any Aaberge social welfare function (including Gini, i.e., $\delta = 2$), the solution to the Relaxed Problem is the solution to the Full Problem, i.e., the optimal solution.

(ii) For the S-Gini and Bonferroni families, the unconstrained solution $\theta \to z(\theta)$ is increasing beyond θ^* such that $\lambda(p(\theta^*)) = 2$.

If we instead assume that the pdf of productivity $\pi(\theta)$ is single-peaked, reaching its mode at θ^M, we obtain:

PROPOSITION 4. Let Assumption 2 hold and π' be single-peaked, with mode θ^M. Then:

(i) For any Aaberge social welfare function (including Gini, i.e., $\delta = 2$), the solution to the Relaxed Problem $\theta \to z(\theta)$ is increasing for all $\theta < \theta^M$.

(ii) For the S-Gini and Bonferroni families, the unconstrained solution $\theta \to z(\theta)$ is increasing for all θ in $\left[\theta^*, \theta^M\right]$, providing θ^* defined by $\lambda(p(\theta^*)) = 2$ is below θ^M.

IV.2. Log-Concave and Log-Convex Distributions

In this Subsection, we maintain Assumption 2 and establish properties of the optimal solution using log-concavity or log-convexity. Any given function f is said to be log-concave on an interval if and only if $log(f)$ is concave on this interval. It turns out that if the probability density function of productivity $h(\theta)$ is log-concave on its support, then the corresponding survival function, $1 - H(\theta)$, is also log-concave. To see that, let us consider $\pi' > 0$. Then $p\pi' < \pi^2 \Rightarrow -\pi' + p\pi' < \pi^2 \Rightarrow -(1-p)\pi' < \pi^2$, implying a log-concave survival function. If $\pi' < 0$ instead, then $-(1-p)\pi' < \pi^2 \Rightarrow -\pi' + p\pi' < \pi^2 \Rightarrow p\pi' < \pi^2$, implying again a log-concave survival function. In addition, many commonly used distributions, such as the normal, exponential, Weibull (with scaling parameter above 1) or power distributions, generate a log-concave $h(\theta)$ and, thus, a log-concave $1 - H(\theta)$. See Table 1 in Bagnoli and Bergstrom (2005). Because log-concavity of the survival function is less demanding than log-concavity of the pdf, we resort to the former rather than the latter whenever possible.

PROPOSITION 5. Let Assumption 2 be verified and consider any Aaberge social welfare function (including Gini, i.e., $\delta = 2$). If the survival function $1 - H(\theta)$ is log-concave, then:

(i) $\Gamma'(\theta) - 1 < 0$, so that $z^*(\theta) = z(\theta)$ for any θ in $\left[\underline{\theta}, \overline{\theta}\right]$;

(ii) There is no bunching whatsoever.

Proof. For Aaberge social weights, note that:

$$\Gamma'(\theta) - 1 = \frac{2 - \delta - \delta p^{\delta-1}}{\delta - 1} + \frac{p}{\delta - 1} \frac{-(1 - p^{\delta-1})\pi'}{\pi^2} \tag{39}$$

where arguments on the RHS have been removed for notational convenience and $\delta \geq 2$. There are two cases to consider:

- For θ such that $\pi' \geq 0$: $\Gamma'(\theta) \leq 1$ given that $\delta \geq 2$ and p belongs to $[0, 1]$.
- For θ such that $\pi' < 0$: Log-concavity of $1 - H(\theta) = 1 - p$ implies that $\frac{-(1-p)\pi'}{\pi^2} < 1$. Therefore,

$$\Gamma'(\theta) - 1 < \frac{2 - \delta - \delta p^{\delta-1}}{\delta - 1} + \frac{p}{\delta - 1} \equiv \Upsilon(\delta). \tag{40}$$

It turns out that $\Upsilon'(\delta) = -1 - (p^{\delta-1} + \delta(\delta-1)p^{\delta-2}) < 0$ for $\delta \geq 2$. Because $\Upsilon(2) = -p$, we conclude that, for $\delta \geq 2$, $\Gamma'(\theta) - 1 < 0$. □

We now turn to the S-Gini family of social welfare functions. We are not able to show that there is no bunching whatsoever for this whole family. Assuming a log-concave survival function $1 - H(\theta)$, we are able to show that the solution $\theta \to z^*(\theta)$ is increasing above some threshold denoted $\hat{\theta}$.

PROPOSITION 6. Let Assumption 2 be verified and consider any S-Gini social welfare function (including Gini, i.e., $\delta = 2$). If the survival function $1 - H(\theta)$ is log-concave, then:

- There is a productivity threshold $\hat{\theta}$ defined by $\hat{p} = p(\hat{\theta}) = 1 - [1/(\delta-1)]^{1/(\delta-1)}$;
- For any $\theta > \hat{\theta}$, $z^*(\theta)$ is strictly increasing.

Proof. Given S-Gini weights, one obtains:

$$\Gamma'(\theta) - 1 = \delta(1-p)^{\delta-1} - 2 - \frac{(1-p)\left[1-(1-p)^{\delta-1}\right]\pi'(\theta)}{\pi^2(\theta)}, \tag{41}$$

$$= \left[\delta(1-p)^{\delta-1} - 2\right] + \left[1-(1-p)^{\delta-1}\right]\left[-(1-p)\frac{\pi'(\theta)}{\pi^2(\theta)}\right]. \tag{42}$$

Because $\delta \geq 2$, the first square bracket is non-positive. There are two cases to consider:

- At any θ for which $\pi'(\theta) \geq 0$: the product of the two square brackets on the RHS of (42) is negative except for $p = 1$ where it is zero. Consequently, $\Gamma'(\theta) < 0$ for any θ in $(\underline{\theta}, \overline{\theta})$. Then $\hat{\theta} = \underline{\theta}$ and $z^*(\theta)$ is increasing, implying $z(\theta) = z(\theta)$ for all θ in $[\underline{\theta}, \overline{\theta}]$.
- At any θ for which $\pi'(\theta) < 0$: We can rely on log-concavity of the survival function, implying $\Gamma'(\theta) - 1 < (\delta(1-p)^{\delta-1} - 2) + (1-(1-p)^{\delta-1}) = (\delta-1)(1-p)^{\delta-1} - 1 \equiv \Upsilon(\delta)$. In addition:

$$\Upsilon(\delta) < 0 \Leftrightarrow (1-p)^{\delta-1} < \frac{1}{\delta-1} \Leftrightarrow \hat{p} = p(\hat{\theta}) \equiv 1 - \left(\frac{1}{\delta-1}\right)^{\frac{1}{\delta-1}} < p. \tag{43}$$

  Consequently, $\Gamma'(\theta) < 1$ for any $p \in (\hat{p}, 1)$. □

For example, when $\delta = 3$, we obtain $\widehat{p} = 1 - \sqrt{1/2} \approx 0.29$.

When δ tends to infinity, so that the S-Gini social welfare function tends to the Rawlsian maximin, $[1/(\delta-1)]^{1/(\delta-1)}$ tends to 1, so that $\hat{p}$ tends to 1 and $\hat{\theta}$ to $\underline{\theta}$. Consequently, $z^*(\theta)$ is increasing for all θ and $z(\theta) = z^*(\theta)$. The following Proposition summarizes this result.

PROPOSITION 7. Let Assumption 2 be verified and consider the Rawlsian social welfare function (maximin). If the survival function $1 - H(\theta)$ is log-concave, then the optimal allocation does not involve any bunching of types.

Four points are worth noting. First, it follows from Proposition 7 that, under Assumption 2, adopting the "first-order approach" consisting in solving the Relaxed Problem instead of the Full Problem implies no loss of generality whatsoever to solve for the maximin allocation as soon as the survival function of productivity $1 - H(\theta)$ is log-concave. Second, the findings in Boadway et al. (2000) regarding the absence of bunching for any exponential distribution of productivity and quasi-linear in leisure preferences are extended to quasi-linear in consumption preferences. Third, the result in Boadway and Jacquet (2008) regarding the absence of bunching under a Weibull probability density for types (with scaling parameter above 1) and quasi-linear in consumption preferences are a subcase of Proposition 7. Last, it is insightful to compare $\widehat{p}$ to the value of p^* introduced in Proposition 3 and for which $\lambda(p^*) = 2$. It turns out that: $\hat{p} > p^* \Leftrightarrow (1/(\delta-1))^{1/(\delta-1)} < (2/\delta)^{1/(\delta-1)} \Leftrightarrow 2 \leq \delta$, which is always satisfied. This is natural because of the positivity of the second term of (38) when the conditions of Proposition 7 are verified.

We were able to provide results for the S-Gini and A families. The following remark examines what would happen for the Bonferroni social welfare function.

REMARK 1 (Bonferroni Social Welfare Function). With Bonferroni social weights, $\Lambda(p) - p = -p \log p$, $\lambda(p) - 1 = -1 - \log p$ and thus:

$$\Gamma'(\theta) - 1 = -2 + \log p \left[\frac{p\pi'(\theta)}{\pi^2(\theta)} - 1 \right]. \tag{44}$$

If $\log p$ is concave, then $\frac{p\pi'(\theta)}{\pi^2(\theta)} < 1$ and, thus, $\frac{p\pi'(\theta)}{\pi^2(\theta)} - 1 < 0$. Because $\log p < 0$, we then have: $\log p \left[\frac{p\pi'(\theta)}{\pi^2(\theta)} - 1 \right] > 0$. We therefore cannot say anything about the sign of $\Gamma'(\theta) - 1$.

A second remark focuses on the variation of $\Gamma'(\theta)$.

REMARK 2. Computing $\Gamma''(\theta)$, we obtain:

$$\Gamma'' \quad = \quad \lambda'\pi - \frac{(\lambda-1)\pi'}{\pi} - (\Lambda - p)(\frac{\pi''\pi - 2\pi'^2}{\pi^3}), \tag{45}$$

in which the (θ) arguments were removed for notational convenience. We see that the sign of $\Gamma''(\theta)$ depends on the sign of π' and π''.

- Because $\lambda'(\theta) < 0$ in $\mathcal{L}$, the first term on the RHS of (45) is strictly negative.
- There exists θ^{**} such that $\lambda(p(\theta^{**})) = 1$, $\lambda(p(\theta)) > 1$ for $\theta < \theta^{**}$ and $\lambda(p(\theta)) < 1$ for $\theta > \theta^{**}$. Then, the second term on the RHS of (45) is always strictly negative when the sign of π' switches from positive to negative exactly at θ^{**}, which is the case for any single-peaked distribution the mode of which coincides with θ^*.
- We always have $\Lambda(p) - p > 0$ for the S-Gini, A- and Bonferroni families. Moreover, a log-concave probability density function satisfies $\pi''(\theta)\pi(\theta) < [\pi'(\theta)]^2$ for any θ. This implies that the third term on the RHS of (45) is always positive for a log-concave probability density function, and therefore for any concave probability density function (concavity implying log-concavity).
- Therefore, we cannot conclude about the sign of $\Gamma''(\theta)$ for log-concave probability density functions. There is indeed a conflict between the sign of the above terms. On the contrary, if the probability density function $h(\theta)$ is log-convex. Then, $\pi''(\theta)\pi(\theta) > [\pi'(\theta)]^2$, so that the third term on the RHS of (45) is negative, as the first two terms.

The above remark allows us to formulate the following Proposition.

PROPOSITION 8. Let Assumption 2 be satisfied and social weights belong to the A-family (including Gini). Let the probability density function be (i) single-peaked, (ii) log-convex and such that (iii) its mode θ^M coincides with the value of productivity for which $\lambda'(\theta) = 1$, i.e., $\theta^M = \theta^{**}$. Then, $\theta \to z^*(\theta)$ is increasing on it support, so that $z(\theta) = z^*(\theta)$. The optimal allocation is therefore full separating.

Proof. Given the above conditions, the function $\Gamma'(\theta)$ is always decreasing. In addition, $\Gamma'(0) < 1$. Consequently, $\Gamma'(\theta) - 1 < 0$ for all θ in $[\underline{\theta}, \overline{\theta}]$. □

Proposition 8 applies to *log-convex* probability *density* functions. According to Table 3 in Bagnoli and Bergstrom (2005), the Weibull (for scaling parameter below 1), the Gamma (for scaling parameter between 0 and 1) and the Pareto distributions have this property. However, Proposition 8 also requires the mode to be in the interior of the support, a condition that is not satisfied for a Pareto distribution (often used to describe the distribution of productivities at the top). For the log-normal distribution, often used to describe the distribution of productivities (except at the top), we know that the cumulative probability function is log-concave, but the log-concavity/log-convexity of the probability density function is undetermined in general. In that case, we cannot sign $\Gamma'(0) - 1$ and can only rely on Proposition 4 making use of unimodality.

IV.3. Local Properties

In this Subsection, we establish a few local conditions regarding the solution in gross income to the Relaxed Problem. First, note that $\Gamma''(0) = \lambda'(0)\pi(0) - \frac{(\lambda(0)-1)\pi'(0)}{\pi(0)}$. For all parametric families introduced above, we have $\lambda(0) - 1 > 0$. Therefore, if we add the following Assumption, we are able to sign $z'(\theta)$ in the lower part of the distribution.

ASSUMPTION 3. $\Gamma''(0) < 0$ when $\pi'(0) > 0$.

We already know that $\Gamma'(0) - 1 < 0$. Therefore, if we make Assumption 3 and let $\pi'(0) > 0$, the following Lemma is obtained:

LEMMA 2. Let Assumptions 2 and 3 be satisfied. If $\pi'(0) > 0$, then there is a $\tilde{\theta} > \underline{\theta}$ (with $\tilde{\theta}$ possibly arbitrarily large) such that the solution in gross income to the Relaxed Problem is strictly increasing on $[\underline{\theta}, \tilde{\theta}]$.

For the upper bound of the support, $\overline{\theta}$, we have: $\Gamma''(\overline{\theta}) = \lambda'(\overline{\theta})\pi(\overline{\theta}) + \frac{\pi'(\overline{\theta})}{\pi(\overline{\theta})}$. We know that $\lambda'(\overline{\theta}_-) < 0$ for any weights in $\mathcal{L}$ (where the notation θ_- stands for the limit to the left of θ). In addition, it is likely that $\pi'(\overline{\theta}_-) \leq 0$. We can therefore formulate the following Lemma.

LEMMA 3. Let Assumption 2 be verified. If $\pi'(\overline{\theta}_-) \leq 0$, the solution to the Relaxed Problem $\theta \to z(\theta)$ is increasing on an interval with upper bound $\overline{\theta}$.

IV.4. Results for Specific Distributions

We now consider specific distributions of productivity to see whether we may say more regarding the bunching or separating pattern of the optimal allocation.

PROPOSITION 9. Let Assumption 2 be verified, and productivity be described by a Pareto distribution with parameters $A > 0$ and $k > 1$. For the Gini welfare function, the optimal allocation is fully separating.

Proof. We have: $H(\theta) = (A/\theta)^k = E > 0$, implying $p = 1-(A/\theta)^k = 1-E$, $1-p = E$, $\pi(\theta) = k\,(A)^k/\theta^{k+1} = kE/\theta$, $\pi^2(\theta) = k^2E^2/\theta^2$, $\pi'(\theta) = -k(k+1)A^k/\theta^{k+2} = -k(k+1)E/\theta^2$. Consequently,

$$\Gamma'(\theta) = \lambda(p(\theta)) - 1 + \frac{(\Lambda(p(\theta)) - p(\theta))k(k+1)\frac{E}{\theta^2}}{\pi^2(\theta)}. \tag{46}$$

Replacing with Gini weights, and using $1-p = E$,

$$\Gamma'(\theta) = \frac{1}{k} + E(1-\frac{1}{k}) = \frac{1}{k} + (1-p)(\frac{k-1}{k}) < 1 \tag{47}$$

for $k > 1$ and $p \in [0,1]$. □

PROPOSITION 10. Let Assumption 2 be verified, and productivity be described by a Pareto distribution with parameters $A > 0$ and $k > 1$. For social weights belonging to the A-family, the optimal allocation is fully separating.

Proof. Starting from (46), and plugging weights for the A-family,

$$\Gamma'(\theta) = \frac{1}{(\delta-1)}(1-\delta p^{\delta-1}) + \frac{\frac{1}{(\delta-1)}(p(1-p^{\delta-1})\frac{(k+1)}{k}}{E}. \tag{48}$$

Using a majoration argument, we establish that:

$$\Gamma'(\theta) < \frac{1}{\delta-1}(1-\delta p(k)^{\delta-1}) + \frac{\frac{1}{\delta-1}(p(k)(1-p(k)^{\delta-1})^2}{1-pk} \tag{49}$$

for any $k > 1$. Moreover, we check that: $\Gamma'(\theta, k=1) < 1$ for any p and deduce that:

$$\frac{1}{\delta-1}(1-\delta p(k)^{\delta-1}) + \frac{\frac{1}{\delta-1}(p(k)(1-p(k)^{\delta-1})\frac{k+1}{k}}{1-p(k)} < 1, \tag{50}$$

implying $\Gamma'(\theta) < 1$. □

PROPOSITION 11. Let Assumption 2 be verified, and productivity be described by a Pareto distribution with parameters $A > 0$ and $k > 1$. For S-Gini (including Gini) welfare weights, then the sign of $\Gamma'(\theta) - 1$ follows that of $\widetilde{p} - p(\theta)$ with $\widetilde{p} = 1 - (\frac{k-1}{1+k+k\delta})^{\frac{1}{\delta-1}}$.

Proof. With S-Gini weights:

$$\Lambda(p) - p = 1-(1-p)^\delta - p = (1-p)(1-(1-p)^{\delta-1}) \tag{51}$$

and $\lambda(p) - 1 = \delta(1-p)^{\delta-1} - 1$. Using $1-p = E$,

$$\Gamma'(\theta) = \frac{1}{k} + E^{\delta-1}(\delta + \frac{(k+1)}{k}) = \frac{1}{k} + E^{\delta-1}(\frac{1+k+k\delta}{k}). \tag{52}$$

Consequently,

$$\Gamma'(\theta) \lesseqgtr 1 \;\; \Leftrightarrow E^{\delta-1}(\frac{1+k+k\delta}{k}) \lesseqgtr \frac{k-1}{k}. \tag{53}$$

Rearranging, the latter is equivalent to: $p \lesseqgtr 1 - (\frac{k-1}{1+k+k\delta})^{\frac{1}{\delta-1}}$. □

V. THE DISCRETE POPULATION CASE

The objective of this Section is to highlight the similarities between the previously examined situations and the case in which the population of agents is discrete. Optimal income taxation with a finite number of types has been examined by Stiglitz (1982, 1987), providing key insights in a two-type model. This setting was extended to any given number of agents by Röell (1985) under a very weak redistributive assumption, by Weymark (1987, 1986a,b) when agents have quasilinear-in-consumption preferences, and then by Simula (2010) when preferences are quasilinear, linear in gross income, as in Equation (4). See also Simula and Trannoy (2011), in which a discrete-type optimal income taxation model is used to decentralize the so-called "ELIE" transfers proposed by Kolm (2004). To a large extent, these studies pay attention to the characterization of the socially optimal gross-income/net-income allocation, which contrasts with the literature considering a continuum of types, the main focus of which is the derivation of optimal marginal tax rates (for a survey see, e.g., Piketty and Saez (2013)).

V.1. Setting

The size of the population is normalized to 1. We define the type distribution by a discrete probability measure $\{\theta_k, \pi_k; k = 1, ..., K\}$, where the support is a finite set of K points in a compact subset $\mathcal{A}$ of the positive real line. The probability of any subset $\mathcal{S}$ of $\mathcal{A}$ is given by

$$p(\mathcal{S}) = \sum_{\theta_k \in \mathcal{S}} \pi_k \delta_k \tag{54}$$

with δ_k the Dirac measure at θ_k. We let $F(x)$ be the CDF, i.e., the step function corresponding to this discrete probability measure. Namely, $F : \mathcal{A} \to [0, 1]$ with $p = F(\theta)$ gives the cumulative proportion of people for whom $\theta_k \leq \theta$ for any θ in $\mathcal{S}$. By construction, we have:

$$p_k = \sum_{j=1}^{k} \pi_j, \tag{55}$$

which corresponds to the rank of agents with productivity θ_k. For later use, note that $p_k - p_{k-1} = \pi_k$. In addition, it should be highlighted that $p = F(\theta)$ is constant between two consecutive steps (θ_k, θ_{k+1}) of the CDF. We define the quantile function as the right inverse of F, i.e., $F^{-1}(p) = \sup_{F(\theta) \leq p} \theta_k$. The marginal weights $\lambda(p)$ and cumulative weights $\Lambda(p)$ are step functions, with an upward jump at each θ_k, and a plateau for any θ in $[\theta_k, \theta_{k+1}[$. The rank dependent social welfare function is therefore given by:

$$W = \int_0^1 \lambda(p) F^{-1}(p) dp. \tag{56}$$

Evaluating the integral for the discrete distribution, we obtain:

$$\mathcal{W} = \theta_1 \int_0^{p_1} \lambda(p) dp + \theta_2 \int_{p_1}^{p_2} \lambda(p) dp + ...\theta_K \int_{p_{K-1}}^{p_K} \lambda(p) dp = \sum_k \theta_k \left[\Lambda(p_k) - \Lambda(p_{k-1})\right]. \tag{57}$$

If we denote by $g(p_k)$ the per-capita social weight of group k, i.e.,

$$g(p_k) = \frac{\Lambda(p_k) - \Lambda(p_{k-1})}{p_k - p_{k-1}} \tag{58}$$

the social objective may be rewritten as:

$$\mathcal{W} = \sum_{k=1}^{K} \theta_k \cdot g(p_k) \cdot \pi_k. \tag{59}$$

Applying the mean value theorem, there exists for every k, a cumulative probability p in $[0,1]$, with $p_{k-1} < p < p_k$, such that $g(p_k) = \lambda(p)$. Given that $\lambda(p)$ is strictly decreasing in $\mathcal{L}$, the per-capita social weight $g(p_k)$ is also strictly decreasing with k.

V.2. Marginal Tax Rate Formula

The way in which the optimal marginal tax rates are derived is formally similar to Simula (2010). The novelty is to show the similarities with the derivation of formulas for optimal marginal tax rates in the continuous case. In this way, we complement the important article by Hellwig (2007), which shows that optimal income taxation in the tradition of Mirrlees (1971) is built on a single rock, irrespective of whether the population is continuous or discrete. However, Hellwig (2007) does not provide a comparison of optimal income tax formulas for either a continuous or a discrete population.

The proof exploits the fact that, when Assumption 1 is satisfied, only the local and downward incentive-compatibility conditions are binding. This implies that an incentive-compatible allocation must be such that $v(z_{k+1};\theta_{k+1}) = v(z_k;\theta_{k+1})$ for any $k = 1, ..., K-1$. This corresponds to the first-order condition for incentive compatibility (7) highlighted above. The equivalent to the second-order condition for incentive compatibility (8) is now: $z = (z_1, ..., z_K)$ non-decreasing in the sense that $z_{k+1} \geq z_k$ for any $k = 1, ..., K-1$. See Guesnerie and Seade (1982) for a detailed presentation. On this basis, and following the same steps as in Simula (2010), it is straightforward to establish that the Relaxed Problem formulated above is now equivalent to:

PROBLEM 3 (Relaxed Discrete-Population Problem). Find gross incomes $z_1, ..., z_K$ maximizing:

$$\mathcal{W} = \sum_{k=1}^{K} \theta_k \cdot g(p_k) \cdot \pi_k - \sum_{k=1}^{K} (\Lambda_k - p_k) \left[v(z_k;\theta_k) - v(z_k;\theta_{k+1}) \right]. \tag{60}$$

It is worth noting that (60) is analogue, in the discrete setting, to (20) in the continuous case. The Full Discrete-Population Problem is similar, with the extra requirement that the solution vector z is non-decreasing. From now on, we focus on the "first-order approach" and assume that the monotonicity condition on z is verified. When $\mathcal{W}$ as defined in (60) is maximized:

$$\left(1 - v'_z(z_k;\theta_k)\right) \pi_k - \left(\Lambda(p_k) - p_k\right) \left(v'_z(z_k;\theta_k) - v'_z(z_k;\theta_{k+1})\right) = 0, \forall k = 1, ..., K-1. \tag{61}$$

In the continuous-population case, the above condition corresponds to (21). We let:

$$T'(z_j;\theta_k) = 1 - v'_z(z_j;\theta_k). \tag{62}$$

This corresponds to the "implicit marginal tax rate" faced by the θ_k-individuals at gross income z_j. In particular, $T'(z_k;\theta_{k+1})$ corresponds to the marginal tax rate an agent with productivity θ_{k+1} would face when choosing the gross-income level z_k designed for the θ_k-individual. Note that $T'(z_j;\theta_k)$ only depends on gross income because, in the gross-income/net-income space, any given θ_k-individual's indifference curves are obtained from one another through vertical displacements, due to quasilinear-in-consumption preferences. Using the above definition, Equation (61) may be rewritten as:

$$\begin{aligned} T'(z_k;\theta_k) &= \frac{\Lambda(p_k)-p_k}{\pi_k}\left[v'(z_k;\theta_k)-v'(z_k;\theta_{k+1})\right], \\ &= \frac{1-p_k}{\pi_k}\frac{\Lambda(p_k)-p_k}{1-p_k}\left[v'(z_k;\theta_k)-v'(z_k;\theta_{k+1})\right]. \end{aligned} \tag{63}$$

Following Simula (2010), we define the *Spence-Mirrlees wedge* as:

$$\begin{aligned} SM(z_k;\theta_k;\theta_{k+1}) = T'(z_k;\theta_{k+1}) - T'(z_k;\theta_k) &= MRS(z_k;\theta_k) - MRS(z_k;\theta_{k+1}), \\ &= v'_z(z_k;\theta_k) - v'_z(z_k;\theta_{k+1}), \end{aligned} \tag{64}$$

where $MRS(z_j;\theta_k)$ stands for the marginal rate of substitution of a θ_k-agent at gross income z_j. Under Assumption 1, the Spence-Mirrlees wage $SM(z_k;\theta_k;\theta_{k+1})$ is positive. Using this definition, (64) becomes:

$$T'(z_k;\theta_k) = \frac{1-p_k}{\pi_k}\frac{\Lambda(p_k)-p_k}{1-p_k}SM(z_k;\theta_k;\theta_{k+1}). \tag{65}$$

In this form, the optimal marginal tax rate appears as the product of three factors: the "demographic" factor $(1-p_k)/\pi_k$ depends on the distribution of the population across ranks; the "ethical" factor $(\Lambda(p_k)-p_k)/(1-p_k)$ depends on the society's preference for redistributing incomes across ranks, as captured by $p \to \Lambda(p)$; and the last factor $SM(z_k;\theta_k;\theta_{k+1})$ ensures that marginal tax rates are adjusted in such a way that indirect utility increases at a sufficient rate to prevent downward mimicking of θ_k-indivuals by θ_{k+1}-individuals. Formula (65) shows in a transparent way the role played by the Spence-Mirrlees condition (here Assumption 1), which was key for the solution to the optimal income tax problem as emphasized in Mirrlees's Nobel Lecture (Mirrlees, 1997).

Dividing (65) by $1-T'(z_k;\theta_k) = v'(z_k;\theta_k)$, we obtain:

$$\frac{T'(z_k;\theta_k)}{1-T'(z_k;\theta_k)} = \frac{1-p_k}{\pi_k}\frac{\Lambda(p_k)-p_k}{1-p_k}\frac{SM(z_k;\theta_k;\theta_{k+1})}{v'_z(z_k;\theta_k)}. \tag{66}$$

For a discrete population, gross income is not a function, but a vector taking K values, one for each productivity level. Each of these k values implicitly depends on the marginal tax rate faced by the θ_k- individuals, and thus on the retention tax rate, as well as on productivity θ_k. Using the differentiability of indifference curves, the elasticities of gross income with respect to the retention rate (assuming a constant marginal tax rate) and productivity, defined in (30) and (31) respectively, directly translate to the discrete population setting. To make use of these definitions, we may rearrange the last factor in the right-hand side of Formula (66) as follows:

$$\begin{aligned}\frac{SM(z_k;\theta_k;\theta_{k+1})}{v'_z(z_k;\theta_k)} &= \frac{1}{z_k v''_{zz}(z_k;\theta_k)}\frac{MRS(z_k;\theta_k)-MRS(z_k;\theta_{k+1})}{v'_z(z_k;\theta_k)/(z_k v''_{zz}(z_k;\theta_k))} \\ &= \frac{\theta_k-\theta_{k+1}}{z_k v''_{zz}(z_k;\theta_k)}\frac{1}{\epsilon_{1-T'}(z_k;\theta_k)}\frac{MRS(z_k;\theta_k)-MRS(z_k;\theta_{k+1})}{\theta_k-\theta_{k+1}} \\ &\approx -\frac{\theta_{k+1}-\theta_k}{z_k v''_{zz}(z_k;\theta_k)}\frac{1}{\epsilon_{1-T'}(z_k;\theta_k)}MRS'_\theta(z_k;\theta_k), \end{aligned} \tag{67}$$

where the approximation is valid for θ_{k+1} sufficiently close to θ_k. Because $MRS'_\theta(z_k;\theta_k) = v''_{z\theta}(z_k;\theta_k)$, we obtain:

$$\begin{aligned}\frac{T'(z_k;\theta_k)}{1-T'(z_k;\theta_k)} &\approx \frac{1-p_k}{\pi_k}\frac{\Lambda(p_k)-p_k}{1-p_k}\frac{1}{\epsilon_{1-T'}(z_k;\theta_k)}\frac{-\theta_k v''_{z\theta}(z_k;\theta_k)}{z_k v''_{zz}(z_k;\theta_k)}\frac{\theta_{k+1}-\theta_k}{\theta_k}, \\ &\approx \frac{1-p_k}{\pi_k}\frac{\Lambda(p_k)-p_k}{1-p_k}\frac{\epsilon_\theta(z_k;\theta_k)}{\epsilon_{1-T'}(z_k;\theta_k)}\frac{\theta_{k+1}-\theta_k}{\theta_k}, \\ &\approx \frac{1}{e(z_k;\theta_k)}\frac{1-p_k}{\theta_k\pi_k}\frac{\Lambda(p_k)-p_k}{1-p_k}\left(\theta_{k+1}-\theta_k\right), \end{aligned} \tag{68}$$

where the behavioral term $e(z_k;\theta_k)$ is equal to the ratio $\epsilon_{1-T'}(z_k;\theta_k)/\epsilon_\theta(z_k;\theta_k)$. In this form, the first-order approximation of the optimal marginal tax rates obtained for a discrete population is given by the same ABC terms as in the continuous-population case, multiplied by $\theta_{k+1}-\theta_k$. When θ_{k+1} tends to θ_k, the RHS of (68) tends to zero. This implies that this approximation does not converge to the continuous-population formula when θ_{k+1} and θ_k become arbitrarily close. Indeed, by (65), two consecutive marginal tax rates do not correspond to the same discrete change, with:

$$T'(z_k;\theta_k) = \frac{1-p_k}{\pi_k}\frac{\Lambda(p_k)-p_k}{1-p_k}SM(z_k;\theta_k;\theta_{k+1}), \tag{69}$$

$$T'(z_{k+1};\theta_{k+1}) = \frac{1-p_{k+1}}{\pi_{k+1}}\frac{\Lambda(p_{k+1})-p_{k+1}}{1-p_{k+1}}SM(z_{k+1};\theta_{k+1};\theta_{k+2}). \tag{70}$$

Taking the limit of (68) does not provide the adequate limit.

VI. Conclusion

This article relies upon rank-dependent welfare functions to connect the theory of optimal nonlinear income taxation and the literature on the measurement of inequality. Paying special attention to two families, generating welfare weights depending on a single-dimension parameter, we were able to obtain remarkably simple tax formulas, as well as conditions on the primitives of the model precluding or implying bunching in the social optimum. In particular, we obtain general results for bunching not to occur for either log-concave type's survival functions or log-convex type's pdf, providing the dis-utility of gross income verifies an assumption on the third derivatives. We also obtain conditions for specific distributions of productivity. Another development would be to characterize optimal solutions involving bunching, and obtain the corresponding tax formulas.

This article also exhibits new "ABC" formulas for optimal marginal tax rates in a discrete population setting. If the formulas look like the ones obtained for a continuous population, we show that the former do not converge to the latter when the distance between two consecutive productivity levels tend to zero.

Because each population setting has some irreductible specificities, it would be of interest to explore how comparative statics techniques and properties for discrete-population optimal income-tax models introduced in Weymark's (1987) seminal article should be adjusted when a continuum of agents is considered. This would necessitate a conceptual adjustment of what is meant by a change in productivity such as those considered for discrete populations by Brett and Weymark (2008) and Simula (2010).

REFERENCES

Rolf Aaberge. Characterization of Lorenz curves and income distributions. *Social Choice and Welfare*, 17:639–653, 2000.

Rolf Aaberge. Ranking intersecting Lorenz curves. *Social Choice and Welfare*, 33(2):235–259, 2009.

Rolf Aaberge, Tarjei Havnes, and Magne Mogstad. A theory for ranking distribution functions. *Journal of Applied Econometrics*, 2020.

Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. In *Economic Theory*, volume 26, pages 445–469. Springer, aug 2005. doi: 10.1007/s00199-004-0514-4.

Robin Boadway and Laurence Jacquet. Optimal marginal and average income taxation under maximin. *Journal of Economic Theory*, 143:425–441, 2008.

Robin Boadway, Kate Cuff, and Maurice Marchand. Optimal income taxation with quasi-linear preferences revisited. *Journal of Public Economic Theory*, 2(4):435–460, 2000.

Walter Bossert. An axiomatization of the single-series Ginis. *Journal of Economic Theory*, 50(1): 82–92, February 1990.

Craig Brett and John A. Weymark. The impact of changing skill levels on optimal nonlinear income taxes. *Journal of Public Economics*, 92:1765–1771, 2008.

Craig Brett and John A. Weymark. Voting over selfishly optimal nonlinear income tax schedules. *Games and Economic Behavior*, 101:172–188, jan 2017. ISSN 10902473. doi: 10.1016/j.geb.2016.02.008.

Dagobert L. Brito and William H. Oakland. Some properties of the optimal income-tax. *International Economic Review*, 18:407–423, 1977.

Peter A. Diamond. Optimal income taxation: An example with U-shaped pattern of optimal marginal tax rates. *American Economic Review*, 88(1):83–95, 1998.

David Donaldson and John A. Weymark. A single-parameter generalization of the Gini indices of inequality. *Journal of Economic Theory*, 22(1):67–86, 1980.

Udo Ebert. A reexamination of the optimal nonlinear income tax. *Journal of Public Economics*, 49 (1):47–73, 1992.

Roger Guesnerie and Jesus Seade. Nonlinear pricing in a finite economy. *Journal of Public Economics*, 57:157–179, 1982.

Martin Hellwig. A contribution to the theory of optimal utilitarian income taxation. *Journal of Public Economics*, 91:1449–1477, 2007.

Martin F. Hellwig. Incentive Problems With Unidimensional Hidden Characteristics: A Unified Approach. *Econometrica*, 78(4):1201–1237, 07 2010.

Laurence Jacquet, Etienne Lehmann, and Bruno Van der Linden. Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory*, 148(5):1770 – 1805, 2013.

Serge-Christophe Kolm. *Macrojustice: The Political Economy of Fairness*. Cambridge University Press, Cambridge, 2004.

Etienne Lehmann, Laurent Simula, and Alain Trannoy. Tax me if you can! Optimal nonlinear income tax between competing governments. *Quarterly Journal of Economics*, 129(4):1995–2030, 2014.

Stéphane Lollivier and Jean-Charles Rochet. Bunching and second-order conditions: A note on optimal tax theory. *Journal of Economic Theory*, 31:392–400, 1983.

James A. Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38:175–208, 1971.

James A. Mirrlees. Information and incentives: The economics of carrots and sticks. *The Economic Journal*, 107:1311–1329, 1997.

Thomas Piketty. La redistribution fiscale face au chômage. *Revue Française d'Economie*, 12:157–201, 1997.

Thomas Piketty and Emmanuel Saez. Optimal labor income taxation. In Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, editors, *Handbook of Public Economics*, volume 5, pages 391–474, 2013.

Ailsa Röell. A note on the marginal tax rate in a finite economy. *Journal of Public Economics*, 28: 267–272, 1985.

Emmanuel Saez. Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68:205–229, 2001.

Bernard Salanié. *The Economics of Contracts*. MIT Press, Cambridge, MA, 2005.

Amartya Sen. Informational bases of alternative welfare approaches. *Journal of Public Economics*, 3(4):387–403, 1974.

Laurent Simula. Optimal nonlinear income tax and nonlinear pricing: Optimality conditions and comparative static properties. *Social Choice and Welfare*, 35:199–220, 2010.

Laurent Simula and Alain Trannoy. Optimal income tax under the threat of migration by top-income earners. *Journal of Public Economics*, 94:163–173, July 2010.

Laurent Simula and Alain Trannoy. Incentive-Compatible ELIE Transfers. In Claude Gamel and Michel Lubrano, editors, *On Kolm's Theory of Macrojustice*, chapter 7, pages 207–231. Springer-Verlag, 2011.

Laurent Simula and Alain Trannoy. Gini and optimal income taxation by rank. Technical Report 8141, CESifo WP, March 2020.

Joseph E. Stiglitz. Self-selection and pareto efficient taxation. *Journal of Public Economics*, 17(2): 213–240, March 1982.

Joseph E. Stiglitz. Pareto efficient and optimal taxation and the new welfare economics. In A. Auerbach and M. Feldstein, editors, *Handbook of Public Economics, vol.II*, volume II, chapter 15, pages 991–1042. Elsevier, 1987.

Alain Trannoy. Talent, equality of opportunity and optimal non-linear income tax. *Journal of Economic Inequality*, 17(1):5–28, 2019.

John A. Weymark. Generalized Gini inequality indices. *Mathematical Social Sciences*, 1(4):409–430, 1981.

John A. Weymark. A reduced-form optimal nonlinear income tax problem. *Journal of Public Economics*, 30:199–217, 1986a.

John A. Weymark. Bunching properties of optimal nonlinear income taxes. *Social Choice and Welfare*, 3:213–232, 1986b.

John A. Weymark. Comparative static properties of optimal nonlinear income taxes. *Econometrica*, 55:1165–1185, 1987.

Menahem E. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.

Menahem E. Yaari. A controversial proposal concerning inequality measurement. *Journal of Economic Theory*, 44(2):381–397, 1988.