Optimal Non-Welfarist Income Taxation for Inequality and Polarization Reduction

Vincenzo Prete, Alessandro Sommacal, Claudio Zoli.

University of Verona*

25/01/2019

Abstract

We characterize the property of an optimal 3 bracket piecewise linear tax system adopting a non-welfarist objective, namely we consider inequality and income polarization reduction objectives. When the elasticity of labour supply is positive, the optimal tax schedule always entails an inverse U-shape relationship between marginal tax rates and income brackets. However, quantitatively, there are striking differences in the optimal tax schedule depending on the specific distributive objective. When the objective is inequality, the last bracket includes few tax payers (about the 98th percentile) and implies marginal tax rates well above zero. When the objective is polarization the third bracket includes a larger amount of tax payers (about the top quartile) and entails a marginal tax rate equal to zero. In special case in which the elasticity of labor supply is equal to zero, the optimal tax scheme when the objective of the government is inequality reduction has only two brackets, with a marginal tax rate equal to zero for the first bracket.

Keywords: Non welfarism, Rank-dependent social evaluation function, Optimal Taxation, Inequality, Polarization.

JEL classification: D31, D63, H21.

*Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona Italy. E-mail: vincenzo.prete@univr.it, alessandro.sommacal@univr.it, claudio.zoli@univr.it.

1 Introduction

In this paper we adopt a non-welfarist approach and analyze how a piecewise linear tax system should be designed in order to reduce income inequality or income polarization.

In line with Kanbur *et al.* (2006) a government is non-welfarist if its social welfare function is defined on individuals' incomes instead of their utilities. Individual preferences do not play a direct role into the social welfare, but still play a role in the design of the optimal tax system in that they shape individuals' reactions in terms of consumption and labour supply to different tax schemes. We assume that the non-welfarist government maximizes, given a revenue requirement constraint, a *rank-dependent Social Evaluation Function* (*SEF*) defined on individuals' incomes, which are linearly aggregated and weighted according to their position in the income ranking. By suitable modifications of the positional weighting function, it is possible to move within the same social evaluation model from evaluations based on inequality to those relying on income polarization. Our focus on non-welfarist objectives is not motivated by the fact that we regard them as superior with respect to the standard social welfare function: we do not take any stand in the debate between welfarist and non-welfarist approaches to social justice. As Kanbur *et al.* (1994 and 2018), we simply think that the study of non-welfarist optimal taxation is interesting because, in many instances, the policy debate is *de facto* centred more around income redistribution than around utilities and social welfare.

In our analysis we focus on piecewise linear tax system, which is the most commonly adopted tax schedule. Moreover, we restrict our attention to the case where only three brackets are present, because, as we will show later, a tax scheme with three brackets is the minimal set-up needed to highlight the different implications of the two social objectives we consider, i.e. inequality reduction and polarization reduction.

We show that redistributive objective matters as the properties of the optimal tax schedule changes depending on whether the policy maker is inequality or polarization sensitive. In particular we focus on marginal tax rate progressivity (regressivity), i.e. the fact that marginal tax rates are increasing (decreasing) with income. We label as convex a tax system which is never marginal tax rate regressive. We call non-convex a tax system that for at least one level of income is marginal tax rate regressive. With fixed labour supply, we show that the optimal tax system is convex when the objective of the government is inequality reduction and non-convex with reduced marginal tax rate for the upper income bracket when the aim of the government is polarization reduction. When individuals' reactions to taxation are taken into account, the optimal tax system is always non-convex both for inequality and polarization reduction. However, there is still a striking difference between the two distributive objectives and it relates with the size of the highest income bracket and the reduced marginal tax rate which is applied to the part of income falling within that bracket. More specifically, when the objective is inequality, the third bracket includes few tax payers (about the 98th percentile) and implies marginal tax rates well above zero. When the objective is polarization the third bracket includes a larger amount of tax payers (about the top quartile) and entails a marginal tax rate equal to zero. In addition, when labour supply elasticity is large the optimal tax system reducing polarization requires a sort of lump-sum taxation: all tax payers fall in the third bracket, to which a zero marginal tax rate is associated.

Our paper is related to the literature on optimal income taxation. Many analysis has been conducted in the welfarist tradition. Here, we are particularly interested in those analysis that develop models of piecewise linear optimal taxation. Sheshinski (1989) shows that the optimal piecewise linear tax system is convex in the sense that higher tax rates are associated with higher income brackets. Slemrod *et al.* (1994) argue that in his analysis Sheshinski ignored the discontinuity in the tax revenue function and they use numerical simulation to show that the optimal tax structure could be non-convex. Apps *et al.* (2014) show that, the results of Slemrod *et al.* (1994) are not robust to changes in the distribution of wages used for the numerical analysis: they find that under assumptions that better describe the current wage distribution, the tax system is essentially convex unless when labour elasticities are high. Using a microeconometric model of labour supply, Aaberge *et al.* (2013) also find that the optimal piecewise tax system is convex. Recently, Andrienko *et al.* (2016) analyse the effect of wage/income inequality on the structure of a piecewise linear tax system, showing that the higher the inequality, the more progressive the tax system should be, with the highest marginal tax rate associated with the top 1%.

To the best of our knowledge, there are only few papers in the non-welfarist tradition which deal with the issue of optimal taxation. In particular, Kanbur *et al.* (1994 and 2018) study optimal income taxation when the objective of the government is the reduction of poverty: while the first paper focuses on a fully non-linear income tax, the second one considers the other extreme case, i.e. a linear tax.[1]

Our approach extends the existing literature on non-welfarist taxation, whose focus has been poverty alleviation, by looking at inequality and polarization reduction objectives. Moreover, we consider the case of piecewise linear tax function which is intermediate between the two polar cases of linear taxation and fully non-linear taxation analysed in the non-welfarist literature. With respect to the welfarist literature on optimal piecewise linear taxation, we show if and how the shape of the tax function is affected by government's objectives that differs from the maximization of a standard social welfare function defined on individual utilities.

The remainder of the paper proceeds as follows. Section 2 introduces the notion of linear rank-dependent *SEF* and describes the two different weighting schemes adopted in the paper to capture inequality and polarization reduction objectives. Section 3 formalizes the optimal tax problem faced by the non-welfarist government. Section 4

[1] Recently Saez and Stantcheva (2016) propose to evaluate tax reforms according to an alternative approach based on a social evaluation model where individuals' net income are weighted differently according to the specific distributive view of the policy-maker.

presents some theoretical results under the assumption of exogenous labour supply. The case of endogenous labour supply is analyzed in Section 5 through the use of numerical simulations. Section 6 concludes.

2 Setting

2.1 Rank-dependent social evaluation functions

To assess alternative taxation policies, we consider the family of linear rank-dependent evaluation functions that aggregate individuals' net incomes weighted according to their position in the income ranking.

Let $F(y)$ denote the cumulative distribution function of income y of a population with bounded support $(0, y^{\max})$ and finite mean $\mu(F) = \int_0^{y^{\max}} y \; dF(y)$. The left inverse continuous distribution function or quantile function, showing the income level of an individual that covers position $p \in (0,1)$ in the distribution of incomes ranked in ascending order, is defined as $F^{-1}(p) := \inf\{y : F(y) \geq p\}$. For expositional purposes, in the remainder of the paper we will also equivalently denote with $y(p)$ the quantile function. The average income could then be calculated as $\mu(F) = \int_0^1 F^{-1}(p)\, dp$.

Consider a set of positional weights $v(p) \geq 0$ for $p \in [0,1]$ such that $V(p) = \int_0^p v(t)\, dt$, with $V(1) = 1$. A *rank-dependent SEF* where incomes are weighted according to the individuals' position in the income ranking is formalized as

$$W_v(F) = \int_0^1 v(p)\, F^{-1}(p)\, dp \tag{1}$$

where $v(p) \geq 0$ is the weight attached to the income of individual ranked p. The normative basis for this evaluation function have been introduced in Yaari (1987) for risk analysis and in Weymark (1982) and Yaari (1988) for income distribution analysis and recently have been discussed as measures of the desirability of redistribution in society by Bennett and Zitikis (2015).[2] This representation model is dual to the utilitarian additively decomposable model. According to W_v the evaluation of income distributions is based on the weighted average of incomes ranked in ascending order and weighted through transformations of the cumulated frequencies (namely the individuals' position). The social evaluation expressed by (1) can be summarized by the mean income of the distribution $\mu(F)$ and a linear index of "dispersion" $I_v(F)$ dependent on the choice of the weighting function v. This "*abbreviated form*" of social evaluation[3] is defined as $W_v(F) = \mu(F)\left[1 - I_v(F)\right]$.

[2] See also Aaberge (2000), Aaberge *et al.* (2013) and Maccheroni *et al.* (2005).

[3] For general details see Lambert (2001).

2.2 Weighting functions

The specific non-welfarist government's objective is formalized by the particular form of the weighting function $v(p)$. We consider two different non-welfarist objectives that combine the average income evaluation with different distributional objectives, namely the reduction of inequality and the reduction of polarization.

2.2.1 Inequality sensitive SEFs

A non-welfarist government aimed at reducing inequality attaches to each quantile $F^{-1}(p)$ of the income distribution a weight according to the following function $v(p) = 2(1-p)$, which is consistent with the Gini index. Then, we can rewrite (1) as

$$W_{\delta}(F) = \int_0^1 2(1-p) F^{-1}(p)\, dp,$$

which can be further rewritten as $W_2(F) = \mu(F)[1 - G(F)]$,where $\mu(F) G(F)$ denotes the absolute version of the Gini index that is invariant with respect to addition of the same amount to all individual incomes.

Moreover, the weighting function consistent with the Gini index can be written as

$$v_G(p) = \begin{cases} 1 - \left[-2\left(\frac{1}{2} - p\right)\right] \text{ if } p \leq \frac{1}{2} \\ 1 - 2\left(p - \frac{1}{2}\right) \text{ if } p \geq \frac{1}{2} \end{cases}. \quad (2)$$

That is, to the weight 1 associated with the average income is subtracted the weight associated to the absolute Gini index that captures the inequality concerns. The weights in (2) are linearly decreasing in the individuals' position moving from poorer to richer individuals (see panel (a) of Figure 1).

With a "non-traditional" interpretation of the absolute Gini index, inequality could be measured by considering the difference between incomes covering equal positional distance from the median weighted with linear weights that increase moving from the median position ($p = 1/2$) to the extreme positions 0 and 1. For instance, take the incomes that are either t positions above the median and t positions below the median, the index considers the difference between these incomes $F^{-1}\left(\frac{1}{2} + t\right) - F^{-1}\left(\frac{1}{2} - t\right)$ and weights it with the weight $2t$. That is

$$\mu(F) G(F) = \int_{1/2}^{1} 2\left|\frac{1}{2} - p\right| F^{-1}(p)\, dp - \int_0^{1/2} 2\left|\frac{1}{2} - p\right| F^{-1}(p)\, dp.$$

The weights attached to the income differences increase as the position of the individuals moves away from the median position. Therefore, any rank-preserving income transfer from individuals above the median to poorer individuals below the median reduces inequality in that it reduces the income distances between individuals covering symmetric positions with respect to the median. Rank-preserving transfers from richer to poorer individuals positioned on the same side with respect to the median, also reduce inequality because it increases the income difference between the incomes that are closer to the median and decreases of the same amount the income difference of the incomes that are in the tails of the distribution. That is, the inequality index gives lower weight to the income differences between individuals closer to the median, therefore the effect for individuals that are more distant from the median is dominant and inequality is reduced.

2.2.2 Polarization sensitive SEFs

When the non-welfarist objective is the reduction of polarization, the distributive concern is for reducing inequality between richer individuals and poorer ones but not necessarily reducing the inequality within the rich and within the poor individuals. In line with the seminal works of Esteban and Ray (1994) and Duclos *et al.* (2004) the polarization measurement combines two components: the *isolation* between economic/social groups and the *identification* between individuals belonging to a group. The first component decreases if the distance between richer and poorer individuals is reduced. In the case of the measurement of income bipolarization, the two social groups are delimited by the median income. The higher is the degree of identification within each group, the higher is the effect of their isolation on polarization. In this case the identification decreases as more disperse is the distribution within one group. Thus, reducing inequality between individuals that are on the same side of the median increases their identification and then increases the overall polarization.

Here, we adopt the bipolarization measurement model introduced in Aaberge and Atkinson (2013).[4] The associated *SEF* is rank-dependent with a weighting function that can be formalized as:

$$v_P(p) = \begin{cases} 2p + 1 \text{ if } p \leq \frac{1}{2} \\ 2p - 1 \text{ if } p \geq \frac{1}{2} \end{cases}, \tag{3}$$

where weights $v_P(p)$ are linear and increasing both below and above the median and exhibit a jump at the median, with higher (lower) weights below (above) the median (see panel (*b*) of Figure 1). In line with the formalization presented for inequality measurement, these weights can be written as

$$v_P(p) = \begin{cases} 1 - \{-[1 - 2(\frac{1}{2} - p)]\} \text{ if } p \leq \frac{1}{2} \\ 1 - [1 - 2(p - \frac{1}{2})] \text{ if } p \geq \frac{1}{2} \end{cases}, \tag{4}$$

where the polarization component is subtracted from the weight 1 associated with the average income.

We focus primarily on this weighting function as it constitutes the counterpart

[4] An alternative approach to the construction of polarization sensitive *SEFs* is presented in Rodriguez (2015).

of the Gini weighting function for the (bi-)polarization measures. It is also possible to derive an associated abbreviated *SEF* where polarization reduces welfare for a given average income level, i.e. $W_P(F) = \mu(F)[1 - P(F)]$, with $P(F)$ denoting a polarization index. An absolute polarization index can be formalized similarly to the inequality index, by considering the difference between the incomes with equal positional distance from the median weighted with linear weights that *decrease* moving from the median position ($p = 1/2$) to the extreme positions 0 and 1. For instance, for the incomes that are either t positions above the median and t positions below the median, the index considers the difference between these incomes $F^{-1}\left(\frac{1}{2}+t\right) - F^{-1}\left(\frac{1}{2}-t\right)$ and weights it with the weight $1-2t$. That is

$$\mu(F)\,P(F) = \int_{1/2}^{1}\left(1 - 2\left|\frac{1}{2} - p\right|\right)F^{-1}(p)\,dp - \int_{0}^{1/2}\left(1 - 2\left|\frac{1}{2} - p\right|\right)F^{-1}(p)\,dp.$$

This representation guarantees that income transfers from richer to poorer individuals on the same side of the median income increase polarization.[5]

An elementary normative implication of the weighting function (4)is that, in order to maximize the welfare, redistribution should be from the individuals above the median to those below. However, when tax schedules are set over few brackets defined in terms of incomes and not positions, then the implications arising from moving from an inequality reducing objective to a polarization reducing one are more subtle.

From Figure 1 it appears evident that the two weighting functions weight more individuals below the median than those one above. However, for inequality (polarization) concerns the weight decreases (increases) for the individuals on the same side of the median as their income increases. In the remainder of the paper we will show how the optimal tax formula changes according to the choice of the weighting function.

[5] The construction of this family of polarization indices is also consistent with the rank-dependent generalization of the Foster–Wolfson polarization measure (see Wolfson, 1994) presented in Wang and Tsui (2000). The main difference between the two approaches is that the Wang and Tsui paper normalizes the index by dividing it by the median instead of the mean income.

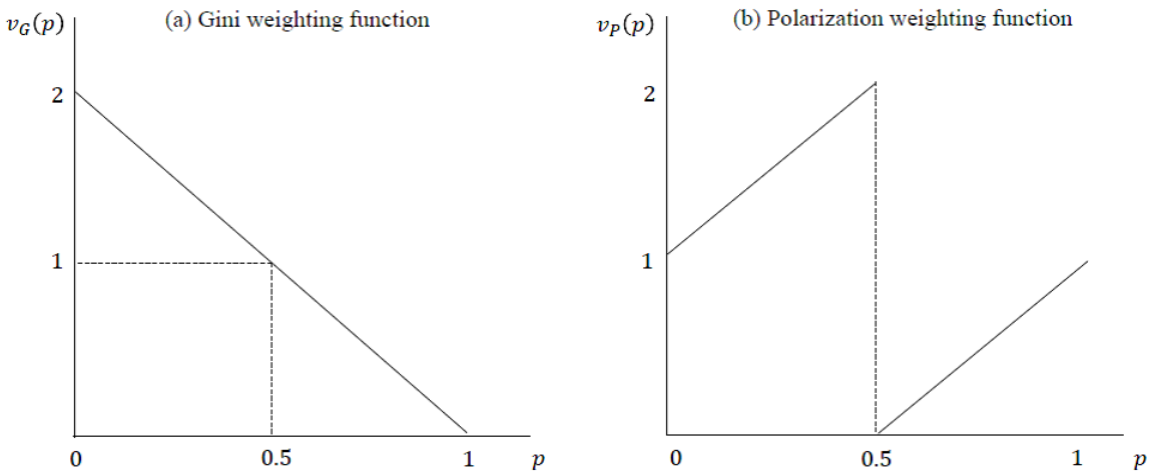


Fig. 1: Weighting function for Gini and Polarization based SEF

3 Non-welfarist optimal piecewise linear taxation

Let (t_i, y_1, y_2) the parameters of a three brackets piecewise linear tax system, where t_i denotes the marginal tax rate in the i^{th} income bracket, with $i = 1, 2, 3$, while y_1 and y_2 are the two income thresholds, with $y_1 < y_2$. These two thresholds are such that $p_1 := \sup\{p : y(p) = y_1\}$ and $p_2 := \sup\{p : y(p) = y_2\}$, with $y(p_1) = y_1$ and $y(p_2) = y_2$, where $F(y_1) = p_1$ and $F(y_2) = p_2$. The non-welfarist government maximizes a rank-dependent *SEF* defined over individuals' net incomes

$$W_v = \int_0^1 v(p)\left[y(p) - T(y(p))\right] dp, \tag{5}$$

subject to the revenue requirement constraint

$$\int_0^1 T(y(p))\, dp = \overline{R} \tag{6}$$

where $\overline{R}$ represents the per capita revenue requirement, while $T(y)$ denotes the three brackets linear tax function, which is defined as follows

$$T(y) := t_1 y + (t_2 - t_1) \cdot \max\{y - y_1, 0\} + (t_3 - t_2) \cdot \max\{y - y_2, 0\}. \tag{7}$$

In our analysis the focus is only on the socially desirable mechanism that guarantees to collect a given level of per capita revenue. That is, government transfers are not allowed and collected revenues are used to finance public expenditures that do not affect neither the individual budget constraint nor their utility function. We consider situations where gross incomes are unequally distributed across individuals and derive results that hold under the assumption of bounded maximal marginal tax rate whose admissible upper level is $\bar{\tau} \in (0, 1]$.

4 The solution with fixed labour supply

The socially optimal taxation design is first illustrated under the assumption of exogenous fixed labour supply. This approach is in line with the literature on the redistributive effect of taxation pioneered by the works of Fellman (1976) Jakobsson (1976) and Kakwani (1977).[6] We derive the results for the three brackets piecewise linear taxation in order to compare the effects on taxation of an inequality reducing sensitive *SEF* with the ones of a polarization reducing sensitive *SEF*.

The derivation of the solutions is illustrated in details in Appendix A both for inequality sensitive and for polarization sensitive *SEFs.* Here, we summarize and comment the main findings and the qualitative features of the optimal taxation design that hold for *any distribution* of pre-tax gross income and for a large class of inequality sensitive and polarization sensitive *SEFs.*

4.1 Inequality concerns

Let $\mathcal{W}_I$ denote the set of all linear rank-dependent *SEFs* with *decreasing non-negative weights* $v(p)$. These *SEFs* are sensitive to inequality reducing transformations of the distributions through rank-preserving progressive transfers from richer to poorer individuals. For instance, the Gini based social weighting function in (2) satisfies this condition.

The set of all three brackets piecewise linear taxation schemes is denoted by $\mathcal{T}_{\bar{\tau}}$, with maximal marginal tax rate $\bar{\tau} \in (0, 1]$ s.t. $\overline{R} \leq \bar{\tau} \cdot \mu(F)$, then we can derive the statement highlighted in the next proposition.

Proposition 1 *A solution of the optimal taxation problem with fixed labour supply for tax schedules in $\mathcal{T}_{\bar{\tau}}$ maximizing SEFs in $\mathcal{W}_I$ is: $t_1 = 0$ and $t_3 = t_2 = \bar{\tau}$, with y_1 s.t. the revenue constraint is satisfied.*[7]

A more detailed specification of the above proposition is proved in Appendix A as Proposition 5.

[6] See also the review in Lambert (2001).

[7] Many equivalent taxation schemes could solve the optimization problem. The presented scheme, indeed, is not affected by the choice of $y_2 > y_1$, moreover an equivalent scheme could be derived where $t_3 = \bar{\tau}$, $t_1 = t_2 = 0$ and the relevant income threshold is y_2.

That is, the optimal taxation design requires only two income brackets with the maximal admissible proportional tax burden in the higher brackets and no taxation for bottom incomes. When $\bar{\tau} = 100\%$ then the solution involves reducing to y_1 all incomes that are above this value.

This result holds not only for the *SEFs* in $\mathcal{W}_I$ but could be shown to hold for any strictly inequality averse *SEF* not necessarily belonging to the family of those that are linearly rank-dependent. It is well known, indeed, that all such *SEFs* for comparisons of distributions with the same average income are consistent with the partial order induced by the Lorenz curve or equivalently by the criterion of second order stochastic dominance [see Atkinson (1970), and Lambert (2001)]. Then, the result in Proposition 1 could be generalized to all *SEFs* that are consistent with the Principle of Transfers, that is are such that any income transfer from a richer individual to a poorer one does not decrease the social evaluation of the distribution. In mathematical terms these functions are Schur-concave [see Dasgupta *et al.* (1973) and Marshall *et al.* (2011)]. Here, we provide the generalization of the result in Proposition 1. Its proof is illustrated in the Appendix of the paper and it is obtained following a different strategy than the one adopted for the proof of Proposition 1. More specifically, we consider a larger set of tax functions that include $\mathcal{T}_{\bar{\tau}}$. Let $T_{\bar{\tau}}$ denote the set of all non-negative and non-decreasing taxation schemes with maximal marginal tax rate $\bar{\tau} \in (0,1]$, such that $T(y) \geq 0$ and $\bar{\tau} \geq \frac{T(y)-T(y')}{y-y'} \geq 0$ for all y, y' such that $y > y'$, then we derive the following statement

Proposition 2 *The solution of the optimal taxation problem with fixed labour supply involving tax schedules in $T_{\bar{\tau}}$ maximizing all the Schur-Concave evaluation functions of the post-tax income distribution obtained under a given revenue constraint involves a two brackets linear taxation scheme where $t_1 = 0$, and $t_2 = \bar{\tau}$, with y_1 s.t. the revenue constraint is satisfied.*

The results in Proposition 2 could also be interpreted in term of progressivity comparisons of the alternative tax schemes considered. It clarifies that the tax scheme in the proposition is the more progressive among all tax schemes that guarantee the same revenue [see, Keen *et al.* (2000) and references therein, and Lambert (2001) Ch. 8]. Thus, the Lorenz curve of tax burden under the taxation scheme considered is more unequal (and then more disproportional) in terms of Lorenz dominance than the one of any alternative tax scheme in $T_{\bar{\tau}}$ giving the same revenue, as originally suggested in Suits (1977) as a criterion to assess the progressivity of a tax schedule.

4.2 Polarization concerns

Let $\mathcal{W}_P$ the set of all polarization sensitive linear rank-dependent *SEFs*, where $v(p)$ is increasing below the median and above the median and weights are larger in the first interval than in the second with $v(0) = v(1) = 1$ and $\lim_{p\to 1/2-} v(p) = 2 \neq \lim_{p\to 1/2+} v(p) = 0$, (see right panel of Figure 1).

To specify the solution we need to consider two hypothetical two brackets tax schemes with marginal tax rates t_1 and t_2 and whose threshold between the two brackets is set at the median income level $y(1/2) = y_M$. Under the first tax scheme the first bracket is not taxed $(t_1 = 0)$, while the second bracket is taxed at the maximal tax rate $t_2 = \bar{\tau}$. We denote with R^+ the revenue arising from such taxation. Under the second tax scheme the first bracket is taxed at the maximal tax rate $t_1 = \bar{\tau}$, while the second bracket exhibits zero marginal tax rate $(t_2 = 0)$ and so all income recipients above the median are taxed with a lump-sum tax equal to $\bar{\tau} y_M$. We denote with R^- the revenue arising from this latter taxation scheme. We can now formalize the results in next the proposition.

Proposition 3 *The solution of the optimal taxation problem with fixed labour supply for tax schedules in $\mathcal{T}_{\bar{\tau}}$ maximizing linear SEFs in $\mathcal{W}_P$ is:*

(i) If $\overline{R} \leq \min\{R^+, R^-\}$, $p_1 < 1/2 < p_2$ where $\frac{1-V_P(p_1)}{1-p_1} = \frac{1-V_P(p_2)}{1-p_2}$ and such that the revenue constraint is satisfied with $t_1 = t_3 = 0$ and $t_2 = \bar{\tau}$.

(iia) If $\overline{R} > R^+$, solution (i) should be compared with $p_1 < 1/2$, $t_1 = 0$ and $t_2 = t_3 = \bar{\tau}$, where p_1 [and so also y_1] is such that the revenue constraint is satisfied.

(iib) If $\overline{R} > R^-$, solution (i) should be compared with $p_1 > 1/2$, $t_1 = \bar{\tau}$ and $t_2 = t_3 = 0$, where p_1 [and so also y_1] is such that the revenue constraint is satisfied.

(iii) If $\overline{R} > \max\{R^+, R^-\}$, all three solutions (i), (iia) and (iib) should be compared.

A more detailed specification of the above proposition is proved in Appendix A as Proposition 7.

The result in Proposition 3 highlights the fact that under standard revenue requirements, i.e. $\overline{R} \leq \min\{R^+, R^-\}$, the marginal tax rate is maximal within the central bracket that includes the median income, while for very large revenue requirements maximal marginal tax rates are applied in the tail brackets. However, note that solution (iib) involves also a lump-sum taxation for those individuals in the higher bracket. While solution (iia) coincides with the optimal solution for inequality sensitive *SEFs.* In all cases the median income is subject to the maximal marginal tax rate. Moreover, it should be pointed out that solution (i) is associated with a local maximum of the optimization problem under any condition on the level of revenue. While solution (i) always exists, solutions (iia) and (iib) may lead to local maxima and the conditions $\overline{R} > R^+$ and $\overline{R} > R^-$ are only necessary for this result and in any case they need to be compared with solution (i).

The comparison between the results in Proposition 1 and Proposition 3 highlights the striking role of the distributive objective in determining the qualitative shape of the optimal taxation scheme. While for inequality sensitive *SEFs* the optimal scheme considers increasing marginal tax rates, for polarization sensitive *SEFs* it requires to tax heavily the "middle class". These two results act as benchmarks for the analysis of optimal taxation with elastic labour supply presented in the next section.

5 The solution with elastic labour supply

In Section 4 we provide an analytical solution to the optimal taxation problem, under the assumption of exogenous labour supply. We now rely on numerical analysis to study the optimal taxation problem under the assumption of endogenous labour supply. First, we specify how agents take their labour/leisure choices. Second, we state the optimal taxation problem of a government who take agents' behavioral responses into account when designing the optimal tax system. Then, we set the parameters of the model and lastly we run numerical simulations to derive the properties of the optimal tax schedule.

As to labour/leisure choices, we consider the following static model of labour supply. There is a continuum of agents and each of them maximizes a quasi-linear utility function[8]:

$$U(x,l) = x - \frac{1}{\alpha} \cdot l^{\alpha}, \tag{8}$$

subject to the budget constraint:

$$x = y - T(y) \tag{9}$$

in which x denotes the net disposable income/consumption, $l \in \left[0, \bar{L}\right]$ (where $\bar{L}$ is the time endowment) is the labour supply, $y = wl$ (where w is the wage rate) is the gross labor income, $T(y)$ is the tax function given by equation (7), the parameter α determines the wage elasticity of labour supply, which is constant and equal to:[9]

$$\varepsilon = \frac{1}{(\alpha - 1)}. \tag{10}$$

The solution of the agent's optimization problem gives the chosen value $y^{\star}$ of gross income as a function of the wage rate, the policy variables $(t_1, t_2, t_3, y_1, y_2)$ and the parameter α:

$$y^{\star} = g(w, t_1, t_2, t_3, y_1, y_2, \alpha). \tag{11}$$

Each agent receives a different wage rate and the distribution of wage rates is denoted by ξ_w. This distribution, the policy variables and the parameter α, determine, through equation (11), the distribution of gross income which is denoted by $\xi_{g(w,t_1,t_2,t_3,y_1,y_2,\alpha)}$.

[8]The quasi-linearity of the utility function is often done in the optmal taxation literature: it rules out the income effect on labour supply and allows to focus on the substitution effect only.

[9]More precisely, the wage elasticity of labour supply is the one specified in equation (10) only when there is no income tax. When we introduce the tax function (7), the wage elasticity of labour supply is given by equation (10) only if gross income falls in the interior of the income brackets of the tax schedule. When gross income is equal to the thresholds of the income brackets, the wage elasticity of labour supply depends on the fact that an increase or a decrease of the wage rate is considered, and it is no longer always equal to the expression in (10). Details are available upon request.

Once agents' behavior has been specified, the government optimal taxation problem can be stated. Given the distribution of wages ξ_w and a specific value for the parameter α, the government chooses the policy variables $(t_1, t_2, t_3, y_1, y_2)$ in order to maximize the *SEF* (5) subject to the government budget constraint (6), taking into account that the distribution of gross incomes is given by $\xi_{g(w,t_1,t_2,t_3,y_1,y_2,\alpha,)}$. To compute the optimal tax schedule we resort to numerical simulations. For the sake of computational time, we constraint the government to choose $t_1 \leq t_2$ while the choice of t_3 is unrestricted: when $t_3 \geq t_2$ the tax schedule is convex; when $t_3 < t_2$ the tax schedule is non-convex.[10]

To compute the solution, we need to specify the distribution ξ_w of individual wages, the value of the parameter α of the utility function, the exogenous government's revenue requirement $\overline{R}$ and the maximal admissible tax rate $\bar{\tau}$.

As to the distribution ξ_w of individual wages, we follow Apps *et al.* (2014) and we consider a truncated Pareto distribution ranging from 20 to 327, with mean (μ) and median (m) equal to 48.05 and 32.36 respectively.[11]

As to α, we consider different values which, in their turn, imply different values of the wage elasticity of labour supply. Our purpose is to use α to perform a robustness analysis in which we explore the impact, for the optimal tax schedule, of changing the sensitiveness to taxation of a given reference distribution. As a reference distribution, we choose the gross income distribution when there is no income tax and labour supply is completely inelastic, i.e. $\alpha \to \infty$ (i.e. $\varepsilon = 0$). We denote this reference distribution by $\xi_{g(w,0,0,0,0,0,\infty)}$.[12] Then, we consider lower (higher) values of α (ε), namely α equal to $3, 6$ and 11 (and ε equal to 0.10, 0.20 and 0.50).[13] However, we need to take into account that changing α not only affects how the distribution of gross income reacts to taxation (see equation (10)), it also affect the distribution itself (see equation (11)). In order to neutralize this latter implication of changing α, and

[10] We use a grid search method. More specifically, we define the grids for t_1, t_2, y_1 and y_2, with $t_1 \leq t_2$ and $y_1 \leq y_2$. For each quadruple (t_1, t_2, y_1, y_2) we compute first the value of t_3 which keeps the government budget constraint balanced and then the associated value of the *SEF*. Last, we identify the combination of policy parameters delivering the highest value of the *SEF*. The results that we report in the two tables at the end of this section are obtained by iterating such procedure three times, by considering at each stage a different specification of the grids of the thresholds. In particular, in the first round grids range from 0 to 260 with step-size equal to 1. In the second round, we define grids on a neighborhood (+/-10) of the optimal thresholds obtained in the previous round with step-size equal to 0.5. Lastly, in the third round grids range in a neighborhood (+/-5) of the second round optimal solution with step-size equal to 0.1. Only for the case of fixed labour supply (i.e. $\varepsilon = 0$) we run a further simulation where grids are defined on a neighborhood (+/-1) of the third round optimal solution, with step-size equal to 0.01. The grids for tax rates always range from 0 to 0.75 with step-size equal to 0.01.

[11] More specifically, this distribution corresponds to the distribution (1.a) considered by Apps *et al.* (2014).

[12] Note that, when $\alpha \to \infty$, the gross income distribution when there is no income tax is equal to the wage distribution, i.e. $\xi_{g(w,0,0,0,0,0,\infty)} = \xi_w$

[13] These values of the labor supply elasticity are broadly consistent with the empirical estimates provided by the literature (see Meghir and Philips (2008), Saez *et al.* (2009)).

keep the distribution of gross income in the absence of taxation equal to the reference distribution $\xi_{g(w,0,0,0,0,0,\infty,)}$ even when $\alpha < \infty$, we rescale the wage distribution: in particular wages need to be raised to the power of $\frac{\alpha}{\alpha-1}$.[14]

As to the exogenous revenue requirement for the government $\overline{R}$, we consider different values, expressed as fraction of the average gross income μ in absence of taxation: $0.10, 0.15$ and 0.20. Lastly, the maximal admissible marginal tax rate $\bar{\tau}$ that the government can set is equal to 0.75.

Table 1 and Table 2 present the results when the government has inequality concerns or polarization concerns respectively. In each Table, for different values of ε and of $\bar{R}$, we compute the optimal convex and non-convex tax schedule: the comparison of social welfare in these two tax regimes give us the optimal tax system.

More specifically, from Table 1 we may observe that with fixed labour supply ($\varepsilon = 0$) the socially optimal tax schedule is as described in Proposition 1: two income brackets with maximal admissible proportional taxation in the higher bracket and zero for bottom incomes. The income threshold between the two brackets is such that the revenue constraint is satisfied. That is, the higher is the revenue requirement, the lower is the income threshold and the no-taxation area. By introducing labour supply elasticity, the socially optimal tax schedule entails three brackets and becomes non-convex, with a reduced marginal tax rate for the highest income bracket, which includes the extremely right tail of the income distribution. The reason for choosing to reduce the tax rate on top incomes, whose social weight is very low[15], is related to a Laffer-curve type effect and is reminiscent of the classical result obtained for welfarist optimal non-linear income taxation, i.e. zero marginal tax rate for the top income. That is, by setting $t_3 < t_2$, it is possible to collect more revenues from top incomes and thus to widen the first income bracket, reducing the fiscal burden for people in the lower tail of the income distribution. In summary, the welfare gains due to the fact that more people belong to the first income bracket (and to the fact that top incomes face a lower marginal tax rate), offset the welfare loss determined by the higher marginal tax rate on the incomes belonging to the central bracket.

Numerical results for polarization based *SEF* are presented in Table 2. With fixed labour supply the socially desirable tax schedule is the one described by Proposition 3, i.e. a central bracket with the maximal admissible marginal tax rate and zero marginal tax rates in the tail brackets. The median income falls within the central bracket, while the two income thresholds are such that the revenue constraint is satisfied. The higher is the revenue requirement, the larger is the central bracket. With elastic labour supply the optimal tax scheme is always non-convex. However, differently than the case of inequality sensitive *SEF*, to reduce polarization a positive marginal tax rate is applied in the first income bracket, while for top incomes the marginal tax rate is zero. Moreover, the highest bracket is larger than in the case of inequality based *SEF*. When the revenue requirement increases, differently from the case of fixed

[14] Details are available upon request.

[15] See the Gini weighting function in the left panel of Figure 1.

labour supply, the central bracket is almost unchanged and the revenue constraint is satisfied with higher marginal tax rate in the first bracket. However, when elasticity is large enough (i.e. $\varepsilon = 0.20$) an increase of the revenue requirement from 15% to 20% of the mean, requires a "lump-sum" tax taxation. In other words, when elasticity is large the optimal tax system is such that the two thresholds are lower than the lowest income level (recall that we are considering a truncated Pareto distribution with lowest income equal to 20) and taxation is the maximal admissible within the central bracket. Simulations for the convex regime show that when elasticity is high the optimal convex configuration reduces to a proportional tax system, however this possibility is never socially preferred compared to the non-convex regime.

Table 1: Optimal tax system with inequality (Gini) based SEF

Gini		Convex tax system ($t_1 \leq t_2 \leq t_3$)						Non – Convex tax system ($t_1 \leq t_2$ and $t_3 \leq t_2$)					
ε	Revenue	t_1	t_2	t_3	y_1	y_2	SW	t_1	t_2	t_3	y_1	y_2	SW
0	0.10 x μ	**0**	**0**	**75%**	**90.26 0.897 (-)**	**90.26 0.897 (-)**	**29.965**	**0**	75%	75%	**90.26 0.897 (-)**	**90.26 0.897 (-)**	**29.965**
	0.15 x μ	**0**	**0**	**75%**	**66.07 0.830 (-)**	**66.07 0.830 (-)**	**29.640**	**0**	75%	75%	**66.07 0.830 (-)**	**66.07 0.830 (-)**	**29.640**
	0.20 x μ	**0**	**0**	**75%**	**50.72 0.744 (-)**	**50.72 0.744 (-)**	**29.132**	**0**	75%	75%	**50.72 0.744 (-)**	**50.72 0.744 (-)**	**29.132**
0.1	0.10 x μ	0	58%	73%	69.4 0.863 (2.1%)	78.9 0.896 (1.1%)	29.679	**0**	**75%**	**46%**	**71.6 0.880 (3.0%)**	**232 0.991 (-)**	**29.685**
	0.15 x μ	0	59%	75%	49.9 0.771 (3.4%)	57 0.827 (1.4%)	29.011	**0**	**74%**	**60%**	**50.8 0.792 (4.8%)**	**203.4 0.986 (-)**	**29.015**
	0.20 x μ	0	61%	75%	37.7 0.652 (5.2%)	43.1 0.734 (1.9%)	27.958	**0**	**75%**	**44%**	**38.5 0.684 (7.2%)**	**232 0.991 (-)**	**27.970**
0.2	0.10 x μ	0	47%	61%	54.1 0.809 (4.1%)	61.9 0.860 (1.5%)	29.276	**0**	**62%**	**33%**	**56 0.837 (5.7%)**	**229.5 0.992 (-)**	**29.294**
	0.15 x μ	0	49%	64%	37.5 0.670 (7.3%)	44.2 0.769 (2.7%)	28.032	**0**	**66%**	**41%**	**39 0.725 (10.6%)**	**186.9 0.984 (-)**	**28.059**
	0.20 x μ	0	53%	67%	27.7 0.497 (12.3%)	34.3 0.668 (3.6%)	26.029	**0**	**68%**	**49%**	**28.8 0.575 (16.7%)**	**172.1 0.981 (-)**	**26.072**
0.5	0.10 x μ	0	31%	43%	32.2 0.616 (12.0%)	38.1 0.742 (4.1%)	27.396	**0**	**45%**	**28%**	**33.8 0.698 (16.8%)**	**145.8 0.974 (-)**	**27.446**
	0.15 x μ	9%	32%	43%	33.6 0.643 (8.5%)	39.4 0.755 (3.6%)	24.395	**8%**	**46%**	**23%**	**33.3 0.696 (14.8%)**	**183.2 0.986 (-)**	**24.463**
	0.20 x μ	18%	34%	43%	34.8 0.669 (5.8%)	40.5 0.765 (2.8%)	21.317	**17%**	**46%**	**17%**	**34 0.705 (11.1%)**	**191.6 0.987 (-)**	**21.397**

Note:

Reported results are obtained through a three rounds simulation procedure, where the grids for the thresholds are set differently at each round. More specifically, in the first round grids range from 0 to 260 with step-size equal to 1. In the second round, we define grids on a neighborhood (+/-10) of the optimal thresholds obtained in the previous round with step-size equal to 0.5. Lastly, in the third round grids range in a neighborhood (+/-5) of the second round optimal solution with step-size equal to 0.1. Only for the case of fixed labour supply (i.e. ε=0) we run a further simulation where grids are defined on a neighborhood (+/-1) of the third round optimal solution, with step-size equal to 0.01. The grids for tax rates always range from 0 to 0.75 with step-size equal to 0.01.

Tax rates have been rounded to the unit, while for thresholds we report both the income level (first row) and the associated CDF (second row). The value within brackets denote the bunching occurred at that level.

With no taxation the mean and the median income are respectively 48.05 and 32.36, while Inequality (Gini) is 0.37.

Table 2: Optimal tax system with Polarization based SEF

Polarization		Convex tax system ($t_1 \leq t_2 \leq t_3$)						Non – Convex tax system ($t_1 \leq t_2$ and $t_3 \leq t_2$)					
ε	**Revenue**	t_1	t_2	t_3	y_1	y_2	SW	t_1	t_2	t_3	y_1	y_2	SW
0	0.10 x μ	0	0	22%	27.61 0.370 (-)	27.61 0.370 (-)	39.076	**0**	**75%**	**0**	**29.41 0.426 (-)**	**44.61 0.688 (-)**	**39.961**
	0.15 x μ	0	0	33%	27.61 0.370 (-)	27.61 0.370 (-)	37.197	**0**	**75%**	**0**	**27.98 0.383 (-)**	**53.25 0.762 (-)**	**38.217**
	0.20 x μ	0	0	44%	27.61 0.370 (-)	27.61 0.370 (-)	35.317	**0**	**75%**	**0**	**26.69 0.339 (-)**	**64.91 0.825 (-)**	**36.308**
0.10	0.10 x μ	0	23%	24%	29 0.436 (2.2%)	29.1 0.439 (0.1%)	38.306	**0**	**65%**	**0**	**29.5 0.509 (8.1%)**	**52.8 0.774 (-)**	**39.117**
	0.15 x μ	0	12%	36%	27.9 0.391 (1.1%)	28.2 0.428 (2.7)%	35.914	**3%**	**66%**	**0**	**28.4 0.483 (8.5%)**	**62.9 0.828 (-)**	**36.705**
	0.20 x μ	0	8%	38%	0 0.0 (-)	28.0 0.424 (3.3%)	33.437	**12%**	**64%**	**0**	**28.5 0.482 (7.2%)**	**65 0.836 (-)**	**34.224**
0.20	0.10 x μ	0	9%	12%	0 0.0 (-)	31.5 0.499 (0.5%)	37.639	**5%**	**50%**	**0**	**29.7 0.537 (9.5%)**	**52.5 0.778 (-)**	**38.395**
	0.15 x μ	16%	16%	16%	0 0.0 (-)	0 0.0 (-)	34.990	**13%**	**50%**	**0**	**29.5 0.532 (8.1%)**	**54.8 0.792 (-)**	**35.770**
	0.20 x μ	21%	21%	21%	0 0.0 (-)	0 0.0 (-)	32.300	**9%**	**75%**	**0**	**0.1 0.0 (-)**	**12.9 0.0 (-)**	**33.226**
0.50	0.10 x μ	11%	11%	11%	0 0.0 (-)	0 0.0 (-)	36.224	**4%**	**75%**	**0**	**0.1 0.0 (-)**	**6.5 0.0 (-)**	**38.030**
	0.15 x μ	16%	16%	16%	0 0.0 (-)	0 0.0 (-)	32.739	**7%**	**75%**	**0**	**0.1 0.0 (-)**	**9.7 0.0 (-)**	**35.628**
	0.20 x μ	23%	23%	23%	0 0.0 (-)	0 0.0 (-)	29.080	**9%**	**75%**	**0**	**0.1 0.0 (-)**	**12.9 0.0 (-)**	**33.226**

Note:

Reported results are obtained through a three rounds simulation procedure, where the grids for the thresholds are set differently at each round. More specifically, in the first round grids range from 0 to 260 with step-size equal to 1. In the second round, we define grids on a neighborhood (+/-10) of the optimal thresholds obtained in the previous round with step-size equal to 0.5. Lastly, in the third round grids range in a neighborhood (+/-5) of the second round optimal solution with step-size equal to 0.1. Only for the case of fixed labour supply (i.e. ε=0) we run a further simulation where grids are defined on a neighborhood (+/-1) of the third round optimal solution, with step-size equal to 0.01. The grids for tax rates always range from 0 to 0.75 with step-size equal to 0.01.

Tax rates have been rounded to the unit, while for thresholds we report both the income level (first row) and the associated CDF (second row). The value within brackets denote the bunching occurred at that level.

6 Concluding remarks

In this paper we adopt a non-welfarist approach to analyze how the optimal labor income tax schedule changes according to the government's redistributive objective, expressed using a linear rank-dependent social evaluation function (SEF) which can alternatively incorporate concerns for the reduction of inequality or polarization. More precisely, we consider a three brackets linear piecewise tax schedule.

Our results reveal that redistributive objectives matter. The optimal tax schedule, indeed, substantially changes depending on whether the government is inequality or polarization sensitive.

With fixed labour supply, the optimal tax schedule maximizing an inequality sensitive SEF requires only two income brackets with the maximal admissible proportional tax burden in the higher bracket and no taxation for bottom incomes. The socially desiderable tax schedule reducing polarization is such that taxation is the maximal admissible within the central bracket, which includes the median income. While the marginal tax rates applied in the two tail brackets are set equal to zero.

With positive labour supply elasticity, the optimal tax schedule is non-convex both for inequality and polarization reduction, with reduced marginal tax rate for the highest bracket. However, while for polarization sensitive SEF this bracket includes about the top quartile of the distribution, for inequality sensitive SEF the reduced marginal tax rate is applied only on the extreme tail of the distribution (about the 98^{th} percentile).

Appendix

Solutions for the constrained optimization problems for inequality and polarization sensitive *SEFs*

Recall the constrained optimization problem faced by the non-welfarist government

$$\max_{t_1,t_2,t_3,y_1,y_2} \mathcal{L} = W_v + \lambda \left[\overline{R} - \int_0^1 T\left(y\left(p\right)\right) dp \right], \tag{12}$$

with $t_i \in [0,1]$ and $y_1 < y_2$. The associated partial derivatives with respect the three tax rates are t_i for $i = 1,2,3$ are respectively

$$\frac{\partial \mathcal{L}}{\partial t_1} = -\int_0^{p_1} v\left(p\right) y\left(p\right) dp - \int_{p_1}^1 v\left(p\right) y_1 dp - \lambda \left[\int_0^{p_1} y\left(p\right) dp + \int_{p_1}^1 y_1 dp \right], \tag{13}$$

$$\frac{\partial \mathcal{L}}{\partial t_2} = -\int_{p_1}^1 v\left(p\right) \min\left\{y\left(p\right), y_2\right\} dp + \int_{p_1}^1 v\left(p\right) y_1 dp - \lambda \left[\int_{p_1}^1 \min\left\{y\left(p\right), y_2\right\} dp - \int_{p_1}^1 y_1 dp \right], \tag{14}$$

and

$$\frac{\partial \mathcal{L}}{\partial t_3} = -\int_{p_2}^{1} v(p)\left[y(p) - y_2\right] dp - \lambda \int_{p_2}^{1} \left[y(p) - y_2\right] dp. \tag{15}$$

The two first order conditions (*FOCs*) with respect the income thresholds y_1 and y_2 are:

$$\frac{\partial \mathcal{L}}{\partial y_1} = -\int_{p_1}^{1} v(p)\left[t_1 - t_2\right] dp - \lambda \int_{p_1}^{1} (t_1 - t_2)\, dp = 0 \tag{16}$$

and

$$\frac{\partial \mathcal{L}}{\partial y_2} = -\int_{p_2}^{1} v(p)\left[t_2 - t_3\right] dp - \lambda \left[\int_{p_2}^{1} (t_2 - t_3)\, dp\right] = 0. \tag{17}$$

The *FOC* with respect to the Lagrangian multiplier is

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \overline{R} - \int_{0}^{1} T(y(p))\, dp = 0. \tag{18}$$

Derivation and simplification of *FOCs*

The associated Kuhn-Tucker *FOCs* for the marginal tax rates are either $\left.\frac{\partial \mathcal{L}}{\partial t_i}\right|_{t_i=0} \leq 0$, or $\left.\frac{\partial \mathcal{L}}{\partial t_i}\right|_{t_i\in(0,1)} = 0$, or $\left.\frac{\partial \mathcal{L}}{\partial t_i}\right|_{t_i=1} \geq 0$ for $i = 1, 2, 3$. While the *FOCs* for the income bracket thresholds are $\frac{\partial \mathcal{L}}{\partial y_1} = 0$ and $\frac{\partial \mathcal{L}}{\partial y_2} = 0$,with $y_2 > y_1 > 0$, and for the multiplier λ the *FOC* requires that $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$. The derivatives with respect to the three marginal tax rates can be rewritten as:

$$\frac{\partial \mathcal{L}}{\partial t_i} = -\int_{0}^{1} v(p)\, h_i(p)\, dp - \lambda \left[\int_{0}^{1} h_i(p)\, dp\right] \tag{19}$$

for $i = 1, 2, 3$, where

$$\begin{aligned}
h_1(p) &:= \begin{cases} y(p) & if\ p < p_1 \\ y_1 & if\ p \geq p_1 \end{cases}; \\
h_2(p) &:= \begin{cases} 0 & if\ p < p_1 \\ y(p) - y_1 & if\ p \in [p_1, p_2) \\ y_2 - y_1 & if\ p \geq p_2 \end{cases}; \\
h_3(p) &:= \begin{cases} 0 & if\ p < p_2 \\ y(p) - y_2 & if\ p \geq p_2 \end{cases}.
\end{aligned}$$

The associated *cdfs* of these three inverse functions are denoted with H_i. The partial derivatives with respect to the thresholds of the income brackets are

$$\frac{\partial \mathcal{L}}{\partial y_1} = [t_2 - t_1]\left[1 - V(p_1) + (1-p_1)\lambda\right] \quad (20)$$

$$\frac{\partial \mathcal{L}}{\partial y_2} = [t_3 - t_2]\left[1 - V(p_2) + (1-p_2)\lambda\right] \quad (21)$$

and the derivative with respect to the Lagrangian multiplier is

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \overline{R} - \sum_{i=1}^{3} t_i \int_0^1 h_i(p)\, dp. \quad (22)$$

Recall that each *SEF* can be decomposed into an abbreviated social evaluation where the average of a distribution is multiplied by 1 minus a linear measure of dispersion $I_v(.)$, that is $W_v(F) = \mu(F)[1 - I_v(F)]$. In our case $I_v(F)$ could be for instance the Gini index or a polarization index as those illustrated in Section 2. Moreover, let $\phi_i(p)$ denote the quantile function at position p of distribution Φ_i where incomes are equal to 0 for all individuals whose position is lower than p_i and are constant with value $z > 0$ for all individuals in positions $p \geq p_i$, with $\mu(\Phi_i) = z \cdot (1 - p_1)$. The next remark summarizes the partial derivatives of the social optimization problem

Remark 4 *The partial derivatives of the Lagrangian optimization problem in (12) are:*

$$\frac{\partial \mathcal{L}}{\partial t_i} = -\mu(H_i) \cdot [1 - I_v(H_i) + \lambda] \text{ for } i \in \{1,2,3\},$$

$$\frac{\partial \mathcal{L}}{\partial y_1} = [t_2 - t_1] \cdot \mu(\Phi_1) \cdot [1 - I_v(\Phi_1) + \lambda],$$

$$\frac{\partial \mathcal{L}}{\partial y_2} = [t_3 - t_2] \cdot \mu(\Phi_2) \cdot [1 - I_v(\Phi_2) + \lambda],$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \overline{R} - \sum_{i=1}^{3} t_i \cdot \mu(H_i).$$

Note that if we let $\frac{\partial \mathcal{L}}{\partial y_i} = 0$, then either $t_{i+1} = t_i$ holds or $\lambda = -[1 - I_v(\Phi_i)]$.

Inequality concerns

We derive here the qualitative features of the socially optimal tax schedule that hold for any distribution of pre-tax gross incomes, for the class $\mathcal{W}_I$ of linear rank-dependent *SEFs* with decreasing non-negative weights $v(p)$, and for the set $\mathcal{T}_{\bar{\tau}}$ of three brackets piecewise linear tax functions whose marginal tax rates could not exceed $\bar{\tau} \in (0, 1]$.

Derivation of optimal tax scheme for SEFs in $\mathcal{W}_I$. Consider the results in Remark 4. If we consider SEFs where $v(p)$ is decreasing as is the case for the Gini based *SEF* and in general for all *SEFs* that are sensitive to inequality reductions through rank preserving progressive transfers from richer to poorer individuals, then $I_v(\Phi_1) < I_v(\Phi_2)$ [with $I_v(\Phi_1) = I_v(\Phi_2)$ only if $p_1 = p_2$]. This is the case because once the distributions Φ_1 and Φ_2 are normalized by their respective means, then it is possible to move from the latter to the former through a series of progressive transfers from the richer individuals with those poorest with normalized income 0.

It then follows that either (i) $[t_3 = t_2 = t_1 = t]$ or (ii) $\lambda = -[1 - I_v(\Phi_1)]$ and $[t_3 = t_2 = \tau]$.

The case (i) is not consistent with the solution because according to the revenue constraint we should obtain $t = \sum_{i=1}^{3} \mu(H_i) / \overline{R} \in (0,1)$. In this case it should be

$$\frac{\partial \mathcal{L}}{\partial t_i} = -\mu(H_i) \cdot [1 - I_v(H_i) + \lambda] = 0$$

for all $i = 1, 2, 3$. Given that $I_v(H_i)$ could be different for all i, then $\lambda = 1 - I_v(H_i)$ could not hold for all i.

The solution associated to case (ii) then should hold. It then follows that, given that $\lambda = I_v(\Phi_1) - 1$, we obtain

$$\frac{\partial \mathcal{L}}{\partial t_i} = -\mu(H_i) \cdot [1 - I_v(H_i) + \lambda] = -\mu(H_i) \cdot [I_v(\Phi_1) - I_v(H_i)].$$

It can be proved that $I_v(H_3) > I_v(H_2) > I_v(\Phi_1) > I_v(H_1)$ for any *SEF* where $v(p)$ is decreasing and there is positive density both below y_1, in between y_1 and y_2, and above y_2 [that is if $0 < p_1 < p_2 < 1$]. In order to make these comparisons one has to normalize all incomes by the total income of the respective distribution and therefore make the comparisons by looking at the distribution of the shares of total income. Once the income shares are compared the distribution with the smaller dispersion evaluated by any rank-dependent *SEF* with decreasing positional weights is the one where the cumulated income shares are larger for any p. In fact in H_1 income shares are larger than those in Φ_1 at the bottom of the distribution for all $p \leq p_1$ and are constant and smaller than those in Φ_1 for $p > p_1$. As a result the cumulated income shares are larger in H_1 than in Φ_1 for any $p \in (0,1)$. Following an analogous logic it could be proved also that $I_v(H_3) > I_v(H_2) > I_v(\Phi_1)$.

From the condition $I_v(H_3) > I_v(H_2) > I_v(\Phi_1) > I_v(H_1)$ then follows that: $\frac{\partial \mathcal{L}}{\partial t_1} < 0$, $\frac{\partial \mathcal{L}}{\partial t_2} > 0$, and $\frac{\partial \mathcal{L}}{\partial t_3} > 0$. As a result we obtain then that $t_1 = 0$, $t_3 = t_2 = \tau = 1$, where y_1 and y_2 are set such that $\overline{R} = \sum_{i=2}^{3} \mu(H_i)$.

Given the above result, the only threshold that matters for the solution is y_1. Moreover, given the sign of the partial derivatives $\frac{\partial \mathcal{L}}{\partial t_1} < 0$, $\frac{\partial \mathcal{L}}{\partial t_2} > 0$, and $\frac{\partial \mathcal{L}}{\partial t_3} > 0$ then for any given value of y_1 we have that the choice of $t_1 = 0$, $t_3 = t_2 = 1$ identifies a maximum point of the objective function. However, for $t_1 = 0$, $t_3 = t_2 = 1$ the value of the threshold y_1 is identified by the revenue constraint, in this case we have that y_1 should be such that $\overline{R} = \mu(H_2) + \mu(H_3)$. As a result the solution is a global maximum for the constrained optimization problem.

The above result could be generalized in order to take into account tax functions whose upper marginal tax rate is not necessarily 100%. To summarize, if we assume that the maximal marginal tax rate is $\bar{\tau} \in (0, 1]$ s.t. $\overline{R} \leq \bar{\tau} \cdot \mu(F)$ we can derive the statement highlighted in the next proposition.

Proposition 5 (1A) *A solution of the optimal taxation problem with fixed labour supply for tax schedules in $\mathcal{T}_{\bar{\tau}}$ maximizing linear SEFs in $\mathcal{W}_I$ is: $t_1 = 0$ and $t_3 = t_2 = \bar{\tau}$, with y_1 s.t. $\overline{R} = \bar{\tau}[\mu(H_2) + \mu(H_3)]$.*

Proof proposition 2

Proof. Dominance of the tax scheme presented in the proposition over all alternative schemes in $T_{\bar{\tau}}$ that satisfy the revenue constraint for all social evaluation functions that are Schur-Concave requires to check that the obtained post-tax net income distribution dominates in terms of Lorenz any of the alterative post-tax distributions [see Marshall et al. 2011]. That is, let T^0 denote the optimal tax function then the Lorenz curve of the post tax income distribution is obtained as $L_{T^0}(p) = \frac{1}{\mu_{T^0}} \int_0^p [y(q) - T^0(y(q))]\, dq$ where $\mu_{T^0} = \int_0^1 [y(q) - T^0(y(q))]\, dq$ denotes the average post-tax net income under taxation T^0.

It then follows that Lorenz dominance of this tax scheme over all alternative schemes T in $T_{\bar{\tau}}$ requires that $L_{T^0}(p) = \frac{1}{\mu_{T^0}} \int_0^p [y(q) - T^0(y(q))]\, dq \geq L_T(p) = \frac{1}{\mu_T} \int_0^p [y(q) - T(y(q))]\, dq$ for all $T \in T_{\bar{\tau}}$ and all $p \in [0, 1]$. Recalling that all the alternative tax schemes should guarantee the same revenue, the condition could be simplified as $\int_0^p [y(q) - T^0(y(q))]\, dq \geq \int_0^p [y(q) - T(y(q))]\, dq$, that is after simplifying for $y(q)$ we obtain

$$\int_0^p T^0(y(q))\, dq \leq \int_0^p T(y(q))\, dq \tag{23}$$

for all $T \in T_{\bar{\tau}}$ and all $p \in [0, 1]$, where by construction the revenue constraint requires that $\int_0^1 T^0(y(q))\, dq = \int_0^1 T(y(q))\, dq = \overline{R}$.

Recall that by construction (i) $T^0(y(p)) = 0$ for all $p \leq p_1$, and that (ii) $\bar{\tau} = \frac{T^0(y) - T^0(y')}{y - y'} \geq \frac{T(y) - T(y')}{y - y'}$ for all $y > y'$ and all $T \in T_{\bar{\tau}}$. By combining the conditions (i) and (ii) and the revenue constraint condition it follows that $T^0(y(p)) \leq T(y(p))$ for all $p \leq p_1$ (with strict inequality for some p), $T^0(y(1)) > T(y(1))$ and the tax schedule $T^0(y)$ crosses once each schedule $T(y)$ from below.

As a result the condition in (23) holds for all $T \in T_{\bar{\tau}}$ and all $p \in [0, 1]$. ■

Polarization concerns.

In order to derive the optimal three brackets linear tax scheme for polarization sensitive evaluation measures we will take as starting point the results in Remark 4.

We consider polarization sensitive linear rank-dependent SEFs where $v(p)$ is increasing below the median and above the median and weights are larger in the first interval than in the second with $v(0) = v(1) = 1$ and $\lim_{p \to 1/2-} v(p) = 2 \neq \lim_{p \to 1/2+} v(p) = 0$ as for the polarization P index illustrated in the previous section. We denote with $\mathcal{W}_P$ the set of all these SEFs.

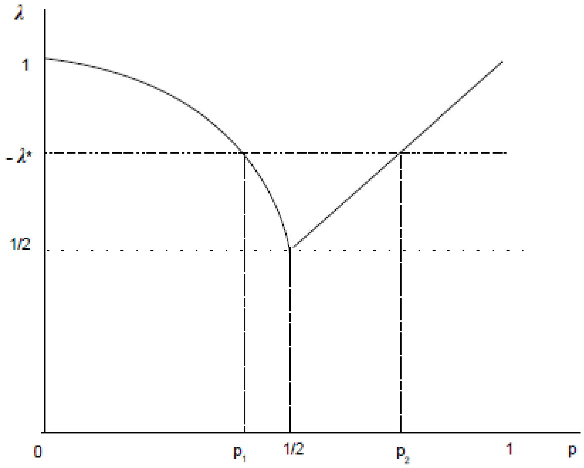
For these SEFs it is possible to derive p_1 and p_2 such that $I_v(\Phi_1) = I_v(\Phi_2)$. This is the case for instance for the SEF whose weights are represented in (4). For these measures it is possible to derive the associated $V(p)$ and compute $\frac{1-V(p)}{1-p}$. They are respectively:

$$V_P(p) = \begin{cases} p^2 + p & if \ p \leq 1/2 \\ p^2 + 1 - p & if \ p > 1/2 \end{cases},$$

with

$$\frac{1 - V_P(p)}{1 - p} = \begin{cases} 1 - \frac{p^2}{1-p} & if \ p \leq 1/2 \\ p & if \ p > 1/2 \end{cases}.$$

Which can be represented as in the following figure



$\frac{1-V_P(p)}{1-p}$

Note that for this specific SEF we have that $\frac{\partial \mathcal{L}}{\partial y_1} = \frac{\partial \mathcal{L}}{\partial y_2} = 0$ if $-\lambda = \frac{1-V_P(p_1)}{1-p_1} = \frac{1-V_P(p_2)}{1-p_2}$. The above function $\frac{1-V_P(p)}{1-p}$ is continuous and is decreasing for $p \leq 1/2$, and increasing for $p > 1/2$, with the minimum in $p = 1/2$ where it takes the value of $1/2$, and the maxima in $p = 0$ and $p = 1$ where it takes the value of 1. It then follows that there exist $p_1 < 1/2$ and $p_2 > 1/2$ such that $-\lambda = \frac{1-V_P(p_1)}{1-p_1} = \frac{1-V_P(p_2)}{1-p_2}$ for $-\lambda > 1/2$.

In this case

$$\begin{aligned}-\lambda &= 1 - I_v(\Phi_1) = 1 - \frac{p_1^2}{1-p_1} \\ &= 1 - I_v(\Phi_2) = p_2\end{aligned}$$

thus $\frac{p_1^2}{1-p_1} = I_v(\Phi_1) = I_v(\Phi_2) = 1 - p_2$.

More generally for all SEFs in $\mathcal{W}_P$ the associated function $1 - V(p)$ is continuous and strictly decreasing [from 1 to 0] for all p, and is concave for $p \leq 1/2$ and for $p \in (1/2, 1]$, with slope -1 for $p = 0$ and $p = 1$. By computing the derivative of $\frac{1-V(p)}{1-p}$, its sign depends on the sign of $-v(p)(1-p) + 1 - V(p)$, by construction of the weighting function it turns out that in line with what shown for the bi-polarization weighting $V_P(p)$, we have that for all SEFs in $\mathcal{W}_P$ the value of $\frac{1-V(p)}{1-p}$ is decreasing for $p \leq 1/2$, and increasing for $p > 1/2$, with the minimum in $p = 1/2$.

Following the same logic presented for the inequality sensitive *SEFs* the optimal solution for *SEFs* in $\mathcal{W}_P$ excludes the case where $[t_3 = t_2 = t_1 = t]$.

We can then consider three cases: (i) $t_3 \neq t_2$; $t_1 \neq t_2$, (ii) $t_3 = t_2$; $t_1 \neq t_2$, and (iii) $t_3 \neq t_2$; $t_1 = t_2$. Where cases (ii) and (iii) can be analyzed symmetrically.

Consider first case (i) where

$$\frac{\partial \mathcal{L}}{\partial y_1} = \frac{\partial \mathcal{L}}{\partial y_2} = 0 \rightarrow \lambda = -1 + I_v(\Phi_1) = -1 + I_v(\Phi_2). \tag{24}$$

By substituting λ into the formula for $\frac{\partial \mathcal{L}}{\partial t_i}$ one obtains

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t_i} &= -\mu(H_i) \cdot [I_v(\Phi_1) - I_v(H_i)] \\ &= -\mu(H_i) \cdot [I_v(\Phi_2) - I_v(H_i)]\end{aligned}$$

for all $i = 1, 2, 3$, with $p_1 < 1/2 < p_2$.

Note that for any polarization measure $I_v(\Phi_2) > I_v(H_3)$, that is $\frac{\partial \mathcal{L}}{\partial t_3} < 0$, implying that $t_3 = 0$. This result is obtained because the difference between Φ_2 and H_3 is that the latter distribution is more disperse for realizations that take place in positions above $p_2 > 1/2$, while in Φ_2 all incomes covering these positions are equal. As we have argued, moving from H_3 to Φ_2 increases polarization because this transformation increases the identification effect reducing the inequality between the individuals on the same side of the median.

It is possible also to show that for dispersion measures that are sensitive to polarization we have that $I_v(\Phi_1) > I_v(H_1)$ that is $\frac{\partial \mathcal{L}}{\partial t_1} < 0$, implying that $t_1 = 0$.

This result could be obtained by properly defining distributions Φ_1 and H_1 so that $\mu(\Phi_1) = \mu(H_1)$. By construction it follows that these distributions cross once for $p = p_1$ and for all $p > p_1$ with $p_1 < 1/2$, incomes are larger in Φ_1 with a constant difference compared to those in H_1, while for $p < p_1$ incomes are larger in H_1. It then follows that H_1 can be obtained from Φ_1 by transferring all the income differences for $p > p_1$ in order to compensate the differences of opposite sign for $p < p_1$. Note that the average weight in the *SEF* for income in position $p > p_1$ is lower than the minimal weight [that corresponds to 1] for all the incomes in position $p < p_1$. As a result the *SEF* value increases when moving from Φ_1 to H_1 and given that $\mu(\Phi_1) = \mu(H_1)$ then $I_v(\Phi_1) > I_v(H_1)$.

In order to verify the condition related to the sign of $\frac{\partial \mathcal{L}}{\partial t_2}$, it is possible to combine distributions Φ_1 and Φ_2 whose linear measures of polarization are the same in order to obtain a new distribution Φ_{12} with the same value for the measure of polarization but such that its quantile function intersects from above the one of H_2 for $p = 1/2$.

In this case it can be shown that for polarization sensitive dispersion measures we have that $I_v(\Phi_1) = I_v(\Phi_2) < I_v(H_2)$, thus we obtain $\frac{\partial \mathcal{L}}{\partial t_2} > 0$ and therefore $t_2 = 1$.

This is the case because by construction Φ_{12} can be obtained from H_2 by transferring incomes from above the median to below the median and transferring incomes from positions that are above the median and close to it to individuals in the upper tail. Both operations reduce the polarization and thus $I_v(H_2) > I_v(\Phi_{12})$.

We then obtain $t_2 = 1$ and $t_1 = t_3 = 0$, with $p_1 < 1/2 < p_2$ where $I_v(\Phi_1) = I_v(\Phi_2)$ and such that $\overline{R} = \mu(H_2)$.

In order to verify that such conditions are associated to a *constrained maximum*, note first that given the sign of the partial derivatives $\frac{\partial \mathcal{L}}{\partial t_3} < 0$, $\frac{\partial \mathcal{L}}{\partial t_1} < 0$, and $\frac{\partial \mathcal{L}}{\partial t_2} > 0$, then for given values of p_1 and p_2 (and so also for given values of y_1 and y_2) satisfying the revenue constraint $\overline{R} = \mu(H_2)$ we have that the combination $t_2 = 1$ and $t_1 = t_3 = 0$ is associated to a maximum. Consider now the population shares $p_1^* < 1/2 < p_2^*$ associated to the solution that satisfy the condition (24) and the revenue constraint that is such that $\lambda = -1 + I_v(\Phi_1) = -1 + I_v(\Phi_2)$ and $\overline{R} = \mu(H_2)$. Our aim is to show that under the condition $t_2 = 1$ and $t_1 = t_3 = 0$ these population shares (and the associated values of y_1 and y_2) correspond to a maximum of the constrained optimization problem.

Associated to these shares we have the value λ^* and the dispersion indices $I_v(\Phi_1^*) = I_v(\Phi_2^*)$ such that $1 - I_v(\Phi_1^*) + \lambda^* = 0$ and $1 - I_v(\Phi_2^*) + \lambda^* = 0$.

Consider a generic pair of shares $p_1 < 1/2 < p_2$ (with associated values of y_1 and y_2) in the neighborhood of p_1^* and p_2^* that satisfies the revenue constraint. By construction, given that the revenue constraint has to satisfied it should be either that (I) $p_1 < p_1^* < 1/2 < p_2 < p_2^*$ or that (II) $p_1^* < p_1 < 1/2 < p_2^* < p_2$. That is, a reduction (increase) in y_1 should be paired with a reduction (increase) in y_2 in order to continue to satisfy the revenue constraint. Substituting the condition $t_2 = 1$ and $t_1 = t_3 = 0$ in the *SEF* and making use of the calculations leading to (16) and (17) we have that $\frac{\partial W_v}{\partial y_1} = \int_{p_1}^1 v(p)\,dp = 1 - V(p_1)$ and $\frac{\partial W_v}{\partial y_2} = -\int_{p_2}^1 v(p)\,dp = 1 - V(p_2)$. Moreover, denoting with R the revenue $\int_0^1 T(y(p))\,dp$ we obtain also that $\frac{\partial R}{\partial y_1} = -\int_{p_1}^1 dp = -(1-p_1)$ and $\frac{\partial R}{\partial y_2} = \int_{p_2}^1 dp = (1-p_2)$. It follows that by taking the differential of the revenue we have $dR = -(1-p_1)dy_1 + (1-p_2)dy_2$, so under the assumption that the revenue constraint is satisfied $R = \overline{R}$, we have that $dR = 0$ and so

$$(1 - p_1)dy_1 = (1 - p_2)dy_2. \tag{25}$$

Analogously the differential of the *SEF* is

$$dW_v = [1 - V(p_1)]\, dy_1 - [1 - V(p_2)]\, dy_2. \tag{26}$$

Substituting for dy_2 from (25) we obtain

$$dW_v = (1 - p_1) \cdot \left[\frac{1 - V(p_1)}{1 - p_1} - \frac{1 - V(p_2)}{1 - p_2} \right] dy_1. \tag{27}$$

Recall that the value of $\frac{1-V(p)}{1-p}$ is decreasing for $p \leq 1/2$, and increasing for $p > 1/2$, with the minimum in $p = 1/2$. As a result under case (I) we have that $dy_1 < 0$ and that p_1 and p_2 decrease w.r.t. p_1^* and p_2^*. As a result $\frac{1-V(p_1)}{1-p_1} > \frac{1-V(p_2)}{1-p_2}$ and so $dW_v < 0$. Similarly we have that if $dy_1 > 0$ then p_1 and p_2 increase w.r.t. p_1^* and p_2^*, and so $\frac{1-V(p_1)}{1-p_1} < \frac{1-V(p_2)}{1-p_2}$ leading to $dW_v < 0$ according to (27). As a result the combination of p_1^* and p_2^* where $\frac{\partial \mathcal{L}}{\partial y_1} = \frac{\partial \mathcal{L}}{\partial y_2} = 0$ identifies a maximum for the constrained optimization.

Consider now case (ii) where $t_3 = t_2$; $t_1 \neq t_2$ implying that in order to obtain $\frac{\partial \mathcal{L}}{\partial y_1} = 0$ necessarily it is required that $\lambda = -1 + I_v(\Phi_1)$.

Note that $t_3 = t_2$ guarantees that $\frac{\partial \mathcal{L}}{\partial y_2} = 0$ irrespective of the value of p_2, that in any case has to satisfy $p_2 > p_1$.

Substituting for λ into $\frac{\partial \mathcal{L}}{\partial t_i}$ we obtain

$$\frac{\partial \mathcal{L}}{\partial t_i} = -\mu\,(H_i) \cdot [I_v(\Phi_1) - I_v(H_i)]\,.$$

Recall that $t_3 = t_2$ implies that the sign of $I_v(\Phi_1) - I_v(H_2)$ according to the polarization sensitive dispersion measures $I_v(\cdot)$ should be the same as the sign of $I_v(\Phi_1) - I_v(H_3)$, and this result should hold for any $p_2 > p_1$.

We leave aside for the moment the case where $I_v(\Phi_1) - I_v(H_2) = I_v(\Phi_1) - I_v(H_3) = 0$.

We can then have two cases, either $t_3 = t_2 = 1$ and $t_1 = 0$, or $t_3 = t_2 = 0$ and $t_1 = 1$.

Note that in the first case the revenue constraints require that $\overline{R} = \mu\,(H_1) + \mu\,(H_2)\,,$ while in the second case it is required that $\overline{R} = \mu\,(H_1)\,.$

As $\overline{G}$ increases $-\lambda$ should increase, therefore in consideration that $-\lambda = 1 - I_v(\Phi_1)$ we have that:

(iia) either $p_1 < 1/2$, $t_3 = t_2 = 1$ and $t_1 = 0$,

(iib) or $p_1 > 1/2$, $t_3 = t_2 = 0$ and $t_1 = 1$.

In fact for (iia) we have that as $\overline{R}$ increases then p_1 should be reduced to increase the tax base in order to collect the required tax revenue, at the same time as Φ_1 changes we have that also $-\lambda$ increases. Given the definition of Φ_1 this will not be the case if $p_1 > 1/2$.

For (iib) we have the symmetric argument where the value of $p_1 > 1/2$ should increase in order to guarantee to collect the required revenue and this will lead to an increase of $-\lambda$ because $p_1 > 1/2$.

As for the previous case (i), given the shape of Φ_1, we can either have $p_1 < 1/2$, or $p_1 > 1/2$, and therefore both (iia) and (iib) are admissible cases.

Suppose we take $p_1 < 1/2$.

Substituting for $\lambda = -1+I_v(\Phi_1)$ into $\frac{\partial \mathcal{L}}{\partial t_i}$ we obtain $\frac{\partial \mathcal{L}}{\partial t_i} = -\mu(H_i)\cdot[I_v(\Phi_1) - I_v(H_i)]$. As for the analysis in case (i) we can show that $I_v(\Phi_1) > I_v(H_1)$ giving $t_1 = 0$. Note that we obtain $t_3 = t_2 = 1$ if the signs of $I_v(\Phi_1) - I_v(H_2)$ and of $I_v(\Phi_1) - I_v(H_3)$ are negative, it should also be that $I_v(\Phi_1) < I_v(H_2)$ when p_2 is set equal to 1. However, it is not possible here to derive a clear-cut conclusion on the sign of $I_v(\Phi_1) - I_v(H_2)$, and in general for a given weighting function and a given distribution the possibility of obtaining $I_v(\Phi_1) > I_v(H_2)$ when $p_2 = 1$ cannot be ruled out.

Consider now case (iib) where $p_1 > 1/2$. Again, referring to the analysis developed for case (i) we can show that $I_v(\Phi_1) > I_v(H_2)$ and $I_v(\Phi_1) > I_v(H_3)$ giving $t_3 = t_2 = 0$. Similarly to what argued for the previous case (iia) it is not possible now to derive a clear-cut conclusion on the sign of $I_v(\Phi_1) - I_v(H_1)$, and in general for a given weighting function and a given distribution the possibility of having $I_v(\Phi_1) > I_v(H_1)$ and therefore that it should not hold $t_1 = 1$ cannot be ruled out.

Going back now to the case where $I_v(\Phi_1) - I_v(H_2) = I_v(\Phi_1) - I_v(H_3) = 0$. If this is the case, then $t_3 = t_2$ may not reach the maximal value. However, as the revenue requirement increases then $-\lambda$ should also increase, then p_1 changes and accordingly also Φ_1 changes, it follows that $I_v(\Phi_1)$ is modified and given that H_2 and H_3 are not affected then the signs of $I_v(\Phi_1) - I_v(H_2)$ and $I_v(\Phi_1) - I_v(H_3)$ change leading either to $t_3 = t_2 = 1$ or $t_3 = t_2 = 0$. Thus, the solutions where tax rates take the extreme values as in (iia) or (iib) are admissible only for cases related to specific revenue values, and in general are not guaranteed as the solution at point (i). If these latter solutions are identified they are associated to local maxima of the constrained optimization problem (see the arguments discussed for the solution related to the inequality sensitive *SEF* case) and should be compared to the solution at point(i).

If we consider case (iii) we can note that it is analogous to case (ii) because both cases will require to consider essentially two brackets with maximal marginal tax rate within one bracket and minimal marginal tax rate in the other.

A remark for cases (iia) and (iib). Before summarizing the results we make the following remark that is motivated by the fact that cases (iia) and (iib) hold only if the revenue requirement is "sufficiently high". In fact for case (iia) we have $p_1 < 1/2$, and the maximal tax rates are $t_3 = t_2 = 1$ with $t_1 = 0$, and for case (iib) we have

$p_1 > 1/2$, with $t_3 = t_2 = 0$ and maximal tax rate set at $t_1 = 1$. Analogous results hold also if we assume that the maximal marginal tax rate is $\overline{\tau} \in (0, 1]$. Let $y(1/2) = y_M$ denote the median income. Then, let H^- denote the distribution whose quantile function is

$$h^-(p) = \begin{cases} y(p) & if\ p < 1/2 \\ y_M & if\ p \geq 1/2 \end{cases};$$

and let H^+ denote the distribution whose quantile function is

$$h^+(p) = \begin{cases} 0 & if\ p < 1/2 \\ y(p) - y_M & if\ p \geq 1/2 \end{cases}$$

The associated averages of these two distributions are respectively $\mu(H^-)$ and $\mu(H^+)$ such that by construction their sum coincides with the overall per-capita gross income, that is $\mu(H^-) + \mu(H^+) = \mu(F)$. The next remark holds

Remark 6 *Case (iia) may hold only if* $\overline{R} > \overline{\tau}[\mu(H^+)]$. *Case (iib) may hold only if* $\overline{R} > \overline{\tau}[\mu(H^-)]$.

Recall that the condition in the remark are only necessary for (iia) or (iib) to hold, while if they do not hold this is sufficient to guarantee that case (i) holds.

We can now summarize the results in the next proposition.

Proposition 7 (3A) *The solution of the optimal taxation problem with fixed labour supply for tax schedules in* $\mathcal{T}_{\overline{\tau}}$ *maximizing linear SEFs in* $\mathcal{W}_P$ *is:*

(i) $p_1 < 1/2 < p_2$ *where* $I(\Phi_1) = I(\Phi_2)$ *and such that* $\overline{R} = \overline{\tau}\mu(H_2)$ *with* $t_1 = t_3 = 0$ *and* $t_2 = \overline{\tau}$, *if* $\overline{R} \leq \min\{\overline{\tau}\mu(H^+), \overline{\tau}\mu(H^-)\}$.

(iia) If $\overline{R} > \overline{\tau}\mu(H^+)$ *solution (i) should be compared with* $p_1 < 1/2$, $t_1 = 0$ *and* $t_2 = t_3 = \overline{\tau}$ *where* $\overline{R} = \overline{\tau}[\mu(H_2) + \mu(H_3)]$.

(iib) If $\overline{R} > \overline{\tau}\mu(H^-)$ *solution (i) should be compared with* $p_1 > 1/2$, $t_1 = \overline{\tau}$ *and* $t_2 = t_3 = 0$,*where* $\overline{R} = \overline{\tau}\mu(H_1)$.

(iii) If $\overline{R} > \max\{\overline{\tau}\mu(H^+), \overline{\tau}\mu(H^-)\}$ *all three solutions should be compared.*

References

[1] Aaberge, R. (2000). Characterizations of Lorenz curves and income distributions. Social Choice and Welfare, 17 (4): 639-653.

[2] Aaberge, R. and Atkinson, A.B. (2013). The median as watershed. Discussion Papers No. 749, (2013). Statistics Norway, Research Department.

[3] Aaberge, R. and Colombino, U. (2013). Using a microeconometric model of household labour supply to design optimal income taxes. Scandinavian Journal of Economics, 115 (2): 449-475.

[4] Andrienko, Y., Apps, P. and Rees, R. (2016). Optimal Taxation and Top Incomes. International Tax Public and Finance, 23 (6): 981-1003.

[5] Apps, P., Long, N. and Rees, R. (2014). Optimal Piecewise Linear Income Taxation. Journal of Public Economic Theory, 16 (4): 523-545.

[6] Atkinson, A.B. (1970). The measurement of inequality. Journal of Economic Theory 2: 244-263.

[7] Bennett, C. J. and Zitikis, R. (2015). Ignorance, lotteries, and measures of economic inequality. Journal of Economic Inequality, 13, 309-316.

[8] Dasgupta, P., Sen. A. K. and D. Starrett (1973). Notes on the Measurement of Inequality. Journal of Economic Theory, 6, 180-187.

[9] Donaldson, D. and Weymark, J.A. (1983). Ethically flexible Gini indices for income distributions in the continuum. Journal of Economic Theory, 29: 353-358.

[10] Duclos, J. Y., Esteban, J. and Ray, D. (2004). Polarization: Concepts, Measurement, Estimation. Econometrica, 72: 1737-1772.

[11] Esteban, J. and Ray, D. (1994). On the Measurement of Polarization. Econometrica, 62: 819-852.

[12] Fellman, J. (1976). The effect of transformation on Lorenz curves. Econometrica, 44: 823-824.

[13] Jakobsson, U. (1976). On the measurement of the degree of progression. Journal of Public Economics, 5: 161-168.

[14] Kakwani, N. C. (1977). Applications of Lorenz curves in economic analysis. Econometrica, 87: 71-80.

[15] Kanbur, R, Keen, M. and Tuomala, M. (1994). Optimal non-linear income taxation for the alleviation of income-poverty. European Economic Review, 38: 1613-1632.

[16] Kanbur, R, Paukkeri, T., Pirtilla, J. and Tuomala, M. (2018). Optimal taxation and public provision for poverty reduction. International Tax and Public Finance, 25: 64-98.

[17] Kanbur, R, Pirtilla, J. and Tuomala, M. (2006). Non-Welfarist Optimal Taxation and Behavioural Public Economics. Journal of Economic Surveys, 20 (5): 849-868.

[18] Keen, M., Papapanagos H. and A. Shorrocks (2000): Tax Reform and Progressivity. The Economic Journal, Vol. 110, No. 460, pp. 50-68

[19] Lambert, P.J. (2001). The Distribution and Redistribution of Income. Manchester University Press.

[20] Maccheroni, F., Muliere P., Zoli C. (2005). Inverse stochastic orders and generalized Gini functionals. Metron, LXIII, (3): 529-559.

[21] Marshall, A. W., Olkin, I. and Arnold, B. C. (2011). Inequalities: Theory of Majorization and Its Applications, Springer.

[22] Meghir, C. and Phillips, D. (2008). Labour Supply and Taxes. Iza Discussion Paper no. 3405.

[23] Mirrlees, J.A. (1971). An Exploration in the Theory of Optimum Income Taxation. Review of Economic Studies, 38: 175-208.

[24] Rodriguez, J. G. (2015). A class of Social Welfare Functions that Depends on Mean Income and Income Polarization. The Review of Income and Wealth, 61-3, 422-439.

[25] Saez, E., Slemrod, J.B. and Giertz, S.H. (2009). The elasticity of taxable income with respect to marginal tax rates: a critical review. National Bureau of Economic Research Working Paper, no. 15012.

[26] Saez, E. and Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. American Economic Review. 106 (1): 24-45.

[27] Sheshinski, E. (1989). Note on the shape of the optimum income tax schedule. Journal of Public Economics, 40: 201-215.

[28] Slemrod, J., Yitzhaki, S., Mayshar, J. and Lundholm, M. (1994). The optimal two-bracket linear income tax. Journal of Public Economics, 53: 269-290.

[29] Suits, D. (1977): Measurement of Tax Progressivity. American Economic Review, 67: 747-752.

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

[31] Wang, Y. and Tsui, K. (2000). Polarization orderings and new classes of polarization indices. Journal of Public Economic Theory, 2, 349-363.

[32] Weymark, J.A. (1981). Generalized Gini inequality indices. Mathematical Social Sciences, 1: 409-430.

[33] Wolfson, M. C. (1994). When inequalities diverge. American Economic Review, 84(2), 353–358.

[34] Yaari, M.E. (1987). The dual theory of choice under risk. Econometrica 55: 99-115.

[35] Yaari, M.E. (1988). A controversial proposal concerning inequality measurement. Journal of Economic Theory 44: 381-397.

[36] Yitzhaki, S. (1983). On an extension of the Gini inequality index. International Economic Review 24: 617-628.