

Optimal Non-Welfarist Income Taxation for Inequality and Polarization Reduction

Vincenzo Prete, Alessandro Sommacal, Claudio Zoli.

University of Verona*

February 2018

Abstract

We adopt a non-welfarist approach to characterize the property of an optimal piecewise linear tax system. We consider inequality and income polarization reduction objectives and we formalize them through the use of a rank-dependent social evaluation function defined on net incomes. In the case of inequality concerns the optimal tax system is mainly convex exhibiting increasing marginal tax rates. In case of polarization concerns the optimal tax scheme is non-convex with a reduced marginal tax rate for the upper income 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 analyze how the optimal tax system should be designed in order to reduce income inequality or income polarization. For this purpose we adopt a non-welfarist approach and consider a piecewise linear tax system.

In line with Kanbur et al. (2006) a government is said non-welfarist if its social welfare function is defined over individuals' incomes instead of their utilities. Individual preferences still play a role in the design of an optimal tax system in that they shape the reaction in terms of consumption and labour supply of the individuals to different tax schemes, but do not play a direct role into the social welfare. In particular, we assume that the non-welfarist government maximizes, given a revenue requirement constraint, a rank-dependent social evaluation function defined over individuals' incomes. According to this social evaluation function, incomes are aggregated linearly and are 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 the polarization of the incomes. Our focus on non-welfarist objectives of the government 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 Kanbur et al. (2017), 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 three brackets piecewise linear tax systems. We regard the study of an optimal piecewise linear tax as particularly relevant, since this tax schedule is the most commonly adopted in practice. 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, when the goal of the government is to reduce inequality, the optimal tax system is mainly convex exhibiting increasing marginal tax rates. When the objective is the reduction of polarization, the optimal tax scheme is non-convex with reduced marginal tax rate for the upper income bracket.

Our paper is obviously related to the literature on optimal income taxation. Many analysis has been conducted in the welfarist tradition. We are here 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. Recently, 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 microeconomic model of labor supply, Aaberge et al. (2013) also find that the optimal piecewise tax system is convex.

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 Kanbur et al (2017) 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.

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 a piecewise linear tax function which is intermediate between the two extremes of a fully non-linear and a linear tax schedule. With respect to the welfarist literature on optimal piecewise linear taxation, we show if and how the shape (convexity or non-convexity) of the tax function is affected by government's objectives that differs from the maximization of a standard social welfare function defined over individual utilities.

The remainder of the paper proceeds as follows. Section 2 introduces the notion of linear rank-dependent social evaluation function 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 presents some theoretical results under the assumption of exogenous labor supply. The case of endogenous labor supply is analysed in Section 5 through the use of numerical simulation. Section 6 concludes.

2 Rank-dependent social evaluation functions

In order to assess alternative taxation policies we consider the family of linear rank-dependent evaluation functions that aggregate the net incomes of the individuals weighting them according to the 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 increasing 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 Social Evaluation Function* [SEF] where incomes are weighted according to individuals' position in the income ranking is formalized as

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

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).¹ 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 according to their positions. Incomes are therefore linearly aggregated across individuals and weighted through transformations of the cumulated frequencies (the individuals' position).

The specific non-welfarist objective of the government can be 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.

When taking into account inequality considerations the social evaluation can be summarized by the mean income of the distribution $\mu(F)$ and a linear index of inequality $I_v(F)$ dependent on the choice of the weighting function v . This "abbreviated form" of social evaluation² is defined as

$$W_v(F) = \mu(F) [1 - I_v(F)].$$

For instance, by defining $v(p) = \delta(1-p)^{\delta-1}$ we can rewrite (1) as

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

which is the class of Generalized Gini SEF parameterized by $\delta \geq 1$ introduced by Donaldson and Weymark (1983) and Yitzhaki (1983). The parameter δ is a measure of the degree of inequality aversion, for $\delta = 1$ we obtain the mean income $\mu(F)$ and therefore inequality neutrality, while for $\delta = 2$ the SEF is associated with the Gini index $G(F)$ and becomes as³

$$W_2(F) = \mu(F) [1 - G(F)].$$

¹See also Aaberge (2000), Aaberge *et al.* (2013) and Maccheroni *et al.* (2005).

²For general details see Lambert (2001).

³The single parameter family of relative Gini index of inequality parameterized by δ is expressed as $G(F) = \frac{1}{\mu(F)} \int_0^1 [1 - \delta(1-p)^{\delta-1}] F^{-1}(p) dp$, which becomes the standard Gini coefficient for $\delta = 2$.

The SEF could also be interpreted as $W_2(F) = \mu(F) - \mu(F) G(F)$ 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.

2.1 Weighting functions

2.1.1 Inequality sensitive SEFs

A non-welfarist government aimed at reducing inequality, once individual incomes are ranked in ascending order, when expresses evaluations consistent with the Gini index attaches to each quantile $F^{-1}(p)$ of the income distribution a weight according to the following function $v_G(p) = 2(1-p)$. These weights are linearly decreasing in the position of the individuals moving from poorer to richer individuals. Alternatively we can write these weights as

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

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, this weight is

$$w_G(p) = \begin{cases} -2(\frac{1}{2} - p) & \text{if } p \leq \frac{1}{2} \\ 2(p - \frac{1}{2}) & \text{if } p \geq \frac{1}{2} \end{cases}. \quad (3)$$

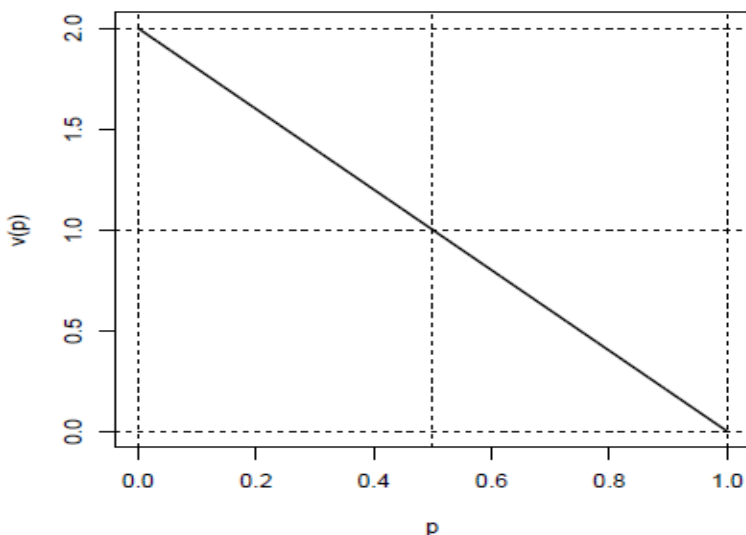
With a "non-traditional" interpretation of the absolute Gini index, inequality could be measured by considering the difference between incomes covering equal positional distance from the median weighted with linear weights that increase moving from the median position = 1/2 to the extreme positions 0 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 $F^{-1}(\frac{1}{2} + t) - F^{-1}(\frac{1}{2} - t)$ 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. In this case any rank-preserving transfer of income 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. However, the inequality index

gives lower weight to the income differences between individuals closer to the median and therefore the effect for the individuals that are more distant from the median is dominant and inequality is reduced.

The next figure shows the weighting function v_G , and as we can see the weights attached to the lowest and to the highest income are respectively equal to two and zero, while the median income receives a weight equal to one. This equivalent representation of the SEF makes clear the positive social effect of a progressive transfer from richer to poorer individuals given that the incomes are transferred from individuals with lower social weight to individuals with higher weight.



The weighting function for the Gini based SEF

2.1.2 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 an isolation component that decreases if the distance between richer and poorer individuals is reduced. The second relevant component in the measurement of polarization is the identification between the individuals belonging to an economic/social class. 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.

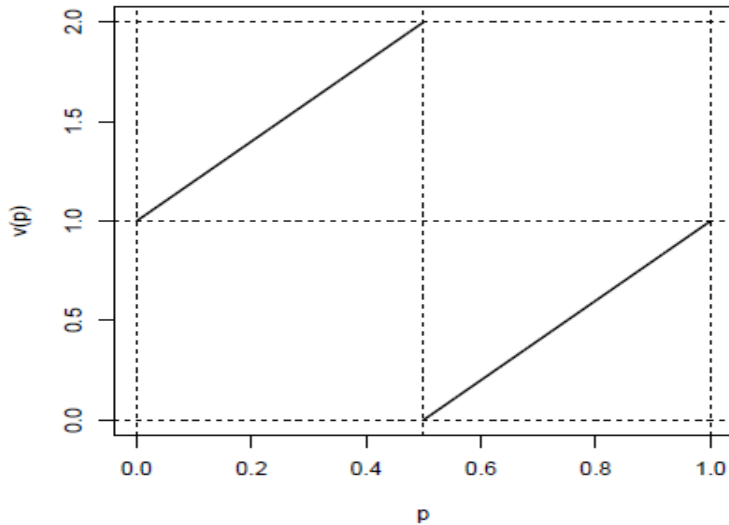
We adopt here the bipolarization measurement model introduced in Aaberge and Atkinson (2013).⁴ The associated SEF is rank-dependent with a weighting function that can be formalized as:

$$v_{P(\beta,\delta)}(p) = \begin{cases} 1 + \beta(2p)^{\delta-1} & \text{if } p \leq \frac{1}{2} \\ 1 - \beta(2 - 2p)^{\delta-1} & \text{if } p \geq \frac{1}{2} \end{cases} \quad (4)$$

Where $\beta \geq 0$ quantifies the relative relevance of polarization with respect to the average income in the overall social evaluation. Moreover $\delta \geq 1$ is a measure of the relative sensitivity of polarization to changes in incomes that occurs at different positions p around the median. For $\beta = 1$ and $\delta = 2$ the weights $v_P(p)$ are linear and increasing,

$$v_P(p) = \begin{cases} 2p + 1 & \text{if } p \leq \frac{1}{2} \\ 2p - 1 & \text{if } p \geq \frac{1}{2} \end{cases} \quad (5)$$

We focus primarily on this weighting function as it constitutes the counterpart of the Gini weighting function for the (bi-)polarization measures. The shape of the weighting function in (5) is illustrated in the following figure.



The weighting function for the Polarization based SEF.

The weights are linearly increasing both below and above the median and exhibit a jump at the median, with higher weights below the median and lower above the median.

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

It is also possible to derive an associated abbreviated SEF where polarization reduces welfare for a given average income level

$$W_P(F) = \mu(F) [1 - P(F)]$$

with $P(F)$ denoting a polarization index. In the case of the linear polarization measure we have that the polarization index can be derived from the condition

$$\mu(F) P(F) = - \int_0^{1/2} 2pF^{-1}(p) dp + \int_{1/2}^1 2(1-p)F^{-1}(p) dp. \quad (6)$$

In line with the formalization presented for inequality measurement, the SEF weighting function can be formalized as

$$v_P(p) = \begin{cases} 1 - \{-[1 - 2(\frac{1}{2} - p)]\} & \text{if } p \leq \frac{1}{2} \\ 1 - [1 - 2(p - \frac{1}{2})] & \text{if } p \geq \frac{1}{2} \end{cases}. \quad (7)$$

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

$$w_P(p) = \begin{cases} -[1 - 2(\frac{1}{2} - p)] & \text{if } p \leq \frac{1}{2} \\ [1 - 2(p - \frac{1}{2})] & \text{if } p \geq \frac{1}{2} \end{cases}. \quad (8)$$

The polarization index can then 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 where $p = 1/2$ 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 $F^{-1}(\frac{1}{2} + t) - F^{-1}(\frac{1}{2} - t)$ and weights it with the weight $1 - 2t$. That is

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

The weights attached to the income differences decrease linearly as the position of the individuals moves away from the median position. 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 polarization based welfare weighting function is that in order to maximize the welfare, redistribution should be from the

⁵The 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.

individuals above the median to those below. However, when tax schedules are set over few brackets that are 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 the two figures above it appears evident that the two weighting functions give more weight to individuals below the median with respect to those above the median. However, for inequality concerns the weight decreases for the individuals on the same side of the median as their income increases, while it increases for polarization concerns.

The associated non-welfarist objectives will lead to different profiles of the income taxation. Our aim is to see how the optimal tax formula changes according to the choice of the weighting function.

3 Non-welfarist optimal piecewise linear taxation

In this section we formalize the optimal tax problem faced by a non-welfarist government. The social evaluation function considered is a general rank-dependent function W with generic non-negative positional weights $v(p)$ with

$$W_v = \int_0^1 v(p) [y(p) - T(y(p))] dp, \quad (9)$$

where $y(p)$ denotes the quantile function or the inverse of the income distribution. Let $p_1 := \sup\{p : y(p) = y_1\}$ and $p_2 := \sup\{p : y(p) = y_2\}$ with $y(p_1) = y_1$ and $y(p_2) = y_2$ denoting the two income thresholds of the considered tax system, where $F(y_1) = p_1$ and $F(y_2) = p_2$. The tax function is denoted by $T(y)$, where taxation is non-negative. The per capita government budget constraint is

$$\int_0^1 T(y(p)) dp = \bar{G}$$

where \bar{G} represents the per capita revenue requirement. We consider a three brackets linear tax function, with $T(y)$ defined as follows

$$T(y) := \begin{cases} t_1 y & \text{if } y \leq y_1 \\ t_1 y_1 + t_2 (y - y_1) & \text{if } y_1 < y \leq y_2 \\ t_1 y_1 + t_2 (y_2 - y_1) + t_3 (y - y_2) & \text{if } y > y_2 \end{cases} \quad (10)$$

or in equivalent terms

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

In our analysis we consider situations where the gross incomes are unequally distributed across individuals. Moreover, we will derive results that hold under the

assumption of bounded maximal marginal tax rate whose admissible upper level is $\bar{\tau} \in (0, 1]$.

The social optimization problem requires to maximize W_v with respect to the three tax rates t_i with $i = 1, 2, 3$, and the two income thresholds y_1 and y_2 where $y_1 < y_2$. As a result the final net incomes distribution could lead to configurations where groups of individuals exhibit the same net income. These distributions could substantially differ depending on whether the social objective is concerned about reducing inequality or with reducing polarization.

4 The solution with fixed labour supply

The taxation design that is socially optimal is first illustrated under the assumption of exogenous fixed labour supply. This first 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 one of a polarization reducing sensitive SEF. Our aim will be to maximize the social evaluation under the revenue constraint that collects the per-capita value \bar{G} .

The constrained optimization Lagrangian function for this problem is as follows

$$\max_{t_1, t_2, t_3, y_1, y_2} \mathcal{L} = W_v + \lambda \left[\bar{G} - \int_0^1 T(y(p)) dp \right], \quad (11)$$

with $t_i \in [0, 1]$ for $i = 1, 2, 3$, and $y_1 < y_2$.

The SEF W_v is presented in (9). As argued in the previous section the shape of the positional social weights $v(p)$ could make the SEF consistent with different distributional objectives, and in particular it could be made sensitive to either inequality or polarization reduction concerns.

The derivation of the solutions for the constrained optimization problem are illustrated in details in Appendix A both for inequality sensitive and for polarization sensitive SEFs. Here we summarize and comment the main findings.

4.1 Inequality concerns

We present here the qualitative features of the optimal taxation problem that hold for *any distribution* of pre-tax gross income and for a large class of inequality sensitive SEFs. Our results hold for piecewise linear three brackets tax functions whose upper marginal tax rate is 100% and are generalized in order to consider maximal marginal tax rates that could not exceed $\bar{\tau} \in (0, 1]$.

⁶See also the review in Lambert (2001).

The family of SEFs considered is denoted by \mathcal{W}_I that represents 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.

Let $\mathcal{T}_{\bar{\tau}}$ denote the set of all piecewise linear taxation schemes with three brackets with maximal marginal tax rate $\bar{\tau} \in (0, 1]$.

We assume that the maximal marginal tax rate $\bar{\tau}$ is s.t. $\bar{G} \leq \bar{\tau} \cdot \mu(F)$ 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 linear SEFs in \mathcal{W}_I is*

$$\begin{aligned} t_1 &= 0, \\ t_3 &= t_2 = \bar{\tau}, \end{aligned}$$

with y_1 s.t. the revenue constraint is satisfied.

A more detailed specification of the above proposition is illustrated and proved in Appendix A as Proposition 4.

All the SEF in \mathcal{W}_I are maximized under the revenue constraint by the taxation schemes presented in Proposition 1. Thus only two income brackets are required to derive the result. Many equivalent taxation schemes could solve the optimization problem. In fact the scheme presented 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 .

To summarize, the optimal taxation problem involves the maximal admissible proportional tax burden in the higher bracket 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 social evaluation function not necessarily belonging to the family of those that are linearly rank-dependent.

In fact it is well known that all such social evaluation functions 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). The result in Proposition 1 could then be generalized to all social evaluation functions 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, Sen and Starrett 1973 and Marshall et al. 2011]. We provide here the generalization of the result in Proposition 1. Its proof is obtained following a different strategy than the

one adopted for the proof of Proposition 1. We present both proofs because the one of Proposition 1 allows more direct comparisons with the results that will be presented for SEFs that are polarization sensitive. To derive the desired result we also consider a larger set of tax functions that include $\mathcal{T}_{\bar{\tau}}$. We denote by $T_{\bar{\tau}}$ the set of all non-negative and non-decreasing taxation schemes with maximal marginal tax rate $\bar{\tau} \in (0, 1]$, that is all tax functions 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 $y > y'$.

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*

$$t_1 = 0, \text{ and } t_2 = \bar{\tau},$$

with y_1 s.t. the revenue constraint is satisfied.

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 alternative 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

$$\int_0^p T^0(y(q)) dq \leq \int_0^p T(y(q)) dq \tag{12}$$

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

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 (12) holds for all $T \in T_{\bar{\tau}}$ and all $p \in [0, 1]$. ■

The above results 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). The result shows that 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

We now move to 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 $\lim_{p \rightarrow 1/2^-} v(p) = 2 \neq \lim_{p \rightarrow 1/2^+} v(p) = 0$ as for the polarization P index illustrated in the previous section. We denote with \mathcal{W}_P the set of all these SEFs.

In order 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, that is $t_1 = 0$, and the second bracket is taxed at the maximal tax rate $t_2 = \bar{\tau}$. We denote with G^+ 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 the income recipients above the median are taxed with a lump-sum tax equal to $\bar{\tau}y_M$. We denote with G^- the revenue arising from this latter taxation scheme. We can now formalize the results in next proposition.

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

(i) $p_1 < 1/2 < p_2$ 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

$$\begin{aligned} t_1 &= t_3 = 0, \\ t_2 &= \bar{\tau}, \end{aligned}$$

if $\bar{G} \leq \min\{G^+, G^-\}$.

(iia) If $\bar{G} > G^+$ solution (i) should be compared with $p_1 < 1/2$, and

$$\begin{aligned} t_1 &= 0, \\ t_2 &= t_3 = \bar{\tau} \end{aligned}$$

where p_1 [and so also y_1] is such that the revenue constraint is satisfied

(iib) If $\bar{G} > G^-$ solution (i) should be compared with $p_1 > 1/2$, and

$$\begin{aligned} t_1 &= \bar{\tau}, \\ t_2 &= t_3 = 0, \end{aligned}$$

where p_1 [and so also y_1] is such that the revenue constraint is satisfied.

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

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

The proposition highlights the fact that under standard revenue requirements $\bar{G} \leq \min\{G^+, G^-\}$ 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 the 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. It should be pointed out that solution (i) is under associated to a local maximum of the optimization problem under any condition on the level of revenue. While solution (i) always exists, as also highlighted in the proof of the proposition, solutions (iia) and (iib) may lead to local maxima and the conditions $\bar{G} > G^+$ and $\bar{G} > G^-$ 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 the polarization sensitive SEFs it requires to tax heavily the "middle class". These two results act as benchmarks for the analysis of optimal taxation with variable labour supply developed in the next section.

5 The solution with variable labour supply

In this section we first describe the agents optimization problem, then we provide numerical results about the optimal tax schedule reducing income inequality and income polarization, with fixed and elastic labor supply. Here we assume that redistribution is not allowed and the focus is on the socially desirable mechanism that ensures to collect a given level of per-capita revenue.

5.1 The agents optimization problem

Agents make labour supply decisions based on the constrained optimization of the quasi-linear utility function

$$U(x, l) = x - \phi(l)$$

where $x \in \mathbb{R}$ denotes net disposable income/consumption and $l \in [0, L]$ denotes labour supply. The function $\phi : [0, L] \rightarrow \mathbb{R}$ is continuous, convex and increasing in l with $\phi'(0) = 0$ where ϕ' denotes the marginal disutility of labour. The utility function could also be expressed in terms of disposable income and leisure ℓ , where $\ell = L - l$. In this case given the above assumptions the function is strictly quasi-concave in x and ℓ .

We will consider an utility specification where ϕ is isoelastic, taking the form

$$\phi(l) = k \cdot l^\alpha \quad (13)$$

with $\alpha > 1$, $k > 0$.

Each agent is endowed with a productivity level formalized by the exogenous wage $w > 0$. The agents in the economy earn a gross income $y \geq 0$ obtained only through labour supply, that is $y = wl$. Agents are subject to taxation $T(y) \geq 0$ formalized by (10), that leads to the net disposable income, considered in their utility function, obtained as $x = y - T(y)$.

Quasi linearity of the utility function rules out income effects in agents decisions and allows to focus only on substitution effects on labour supply.

We can equivalently re-express the problem in the space (x, y) for each agent. In this case the utility function becomes

$$u(x, y) = U(x, y/w) = x - \phi(y/w)$$

and the relation between x and y is

$$x := y - T(y) = \begin{cases} (1 - t_1)y & \text{if } y \in Y_1 \equiv [0, y_1) \\ (t_2 - t_1)y_1 + (1 - t_2)y & \text{if } y \in Y_2 \equiv [y_1, y_2) \\ (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)y & \text{if } y \in Y_3 \equiv [y_2, \infty) \end{cases} \quad (14)$$

Where Y_i denotes the income set associated to the i^{th} income bracket. The set $Y \setminus y_{i-1}$ will instead denote the set Y_i net of its lower element y_{i-1} , where $y_0 = 0$.

The marginal rate of substitution between y and x is $MRS_{yx} = \phi'(y/w)/w$. For levels of gross income that do not coincide with the thresholds $y_1 < y_2$ it should hold that $MRS_{yx} = (1 - t_i)$ when $y \in Y_i$. That is

$$y^* = w \cdot \phi'^{-1}[(1 - t_i)w]$$

when $y^* \in Y_i \setminus y_{i-1}$, where the function $\phi'^{-1}(\cdot)$ by construction is positive and strictly increasing. Given the definition of $y = wl$, one obtains also the associated optimal labour supply

$$l^* = [(1 - t_i)w]$$

when $wl^* \in Y_i \setminus y_{i-1}$.

Given the assumptions, y^* and l^* are continuous and strictly increasing w.r.t. w within the sets $Y_i \setminus y_{i-1}$.

We consider now in details the issues when $\phi(l) = k \cdot l^\alpha$ with $\alpha > 1$. Thereby leading to

$$\begin{aligned} y^* &= w \cdot \left[\frac{(1-t_i)w}{k\alpha} \right]^{\frac{1}{\alpha-1}} = w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_i)}{k\alpha} \right]^{\frac{1}{\alpha-1}} \\ l^* &= \left[\frac{(1-t_i)w}{k\alpha} \right]^{\frac{1}{\alpha-1}} \end{aligned} \quad (15)$$

when $y^* \in Y_i \setminus y_{i-1}$. Note that within the sets $Y_i \setminus y_{i-1}$ the elasticity ε of labour supply w.r.t. the wage is constant and equals $\frac{1}{(\alpha-1)}$.⁷ In this paper we will consider as a reference distribution the gross income distribution in absence of taxation. Then, by setting $t_i = 0$ from (15) we obtain $y^* = w^{\frac{\alpha}{\alpha-1}} \left[\frac{1}{k\alpha} \right]^{\frac{1}{\alpha-1}}$ and $l^* = \left[\frac{w}{k\alpha} \right]^{\frac{1}{\alpha-1}}$. Let $w(p)$ denote the gross wage of the individual in position $p \in [0, 1]$ in the distribution of gross wages ranked in non-decreasing order. Then, the following monotonically increasing transformation of the wage

$$y(p) := w(p)^{\frac{\alpha}{\alpha-1}} \left[\frac{1}{k\alpha} \right]^{\frac{1}{\alpha-1}} = w(p)^{1+\varepsilon} \left[\frac{\varepsilon}{k(\varepsilon+1)} \right]^\varepsilon \quad (16)$$

represents the gross income of the individual covering position p under the assumption of no-taxation, with the associated labor supply $l(p) = \left[\frac{w(p)}{k\alpha} \right]^{\frac{1}{\alpha-1}} = \left[\frac{w(p)\varepsilon}{k(\varepsilon+1)} \right]^\varepsilon$. The gross income distribution in absence of taxation, formalized by the quantile (or inverse distribution) function $y(p)$, is the reference distribution in our analysis.

In order to simplify the exposition, and in line with the results obtained with fixed labour supply we focus only on tax schedules where $t_1 \leq t_2$, and assume two possible regimes, i.e. convex (case A) and non-convex (case B) of tax rates depending on the ranking of t_2 and t_3 . Case A, is such that $t_1 \leq t_2 \leq t_3$, while case B considers the configuration where $t_1 \leq t_3 < t_2$.

Depending on what case is considered we could either have as in case A that some agents experience the same gross income coinciding with one of the thresholds y_1 and y_2 , or as under case B that this could happen for y_1 while around y_2 the map of y^* w.r.t. w is discontinuous, but still increasing.

To simplify the exposition in the next two subsections we express the gross income distribution in terms of intervals of quantiles $y(p)$, while in the Appendix B we show the gross income distribution also in terms of wages intervals.

⁷In the case with fixed labor supply elasticity is set equal to zero, hence labor supply reduces to one and gross incomes and wages coincide.

5.1.1 Case A: $t_1 \leq t_2 \leq t_3$

Under case A the above optimality conditions hold if $y_{i-1} < y^* < y_i$ that is, if

$$\frac{y_{i-1}}{(1-t_{i-1})^\varepsilon} < y(p) < \frac{y_i}{(1-t_i)^\varepsilon}$$

for $i \in \{1, 2, 3\}$ where $y_3 = +\infty$. The three sets of values can then be expressed in terms of intervals of gross incomes such that

$$\begin{aligned} 0 &< y(p) < \frac{y_1}{(1-t_1)^\varepsilon}; \\ \frac{y_1}{(1-t_2)^\varepsilon} &< y(p) < \frac{y_2}{(1-t_2)^\varepsilon}; \\ \frac{y_2}{(1-t_3)^\varepsilon} &< y(p). \end{aligned}$$

Note that by construction it follows that $\left[\frac{y_i}{(1-t_i)^\varepsilon} \right] < \left[\frac{y_{i+1}}{(1-t_{i+1})^\varepsilon} \right]$, and therefore we obtain:

$$y_t(p) = \begin{cases} y(p)(1-t_1)^\varepsilon & \text{if } y(p) < \frac{y_1}{(1-t_1)^\varepsilon} \\ y_1 & \text{if } \frac{y_1}{(1-t_1)^\varepsilon} \leq y(p) \leq \frac{y_1}{(1-t_2)^\varepsilon} \\ y(p)(1-t_2)^\varepsilon & \text{if } \frac{y_1}{(1-t_2)^\varepsilon} < y(p) \leq \frac{y_2}{(1-t_2)^\varepsilon} \\ y_2 & \text{if } \frac{y_2}{(1-t_2)^\varepsilon} \leq y(p) \leq \frac{y_2}{(1-t_3)^\varepsilon} \\ y(p)(1-t_3)^\varepsilon & \text{if } y(p) > \frac{y_2}{(1-t_3)^\varepsilon} \end{cases} \quad (17)$$

where $y_t(p)$ denotes the post tax gross income of an individual that covers position p in the distribution of $y(p)$.

5.1.2 Case B: $t_1 \leq t_3 \leq t_2$

Under case B (non-convex regime) we assume that the optimal labour supply and gross income are the same for all incomes that are in the first bracket and at the first threshold, the result changes for the income levels in the second and third brackets. In particular, if $t_2 > t_3$ then there exists a threshold level \hat{y} in the gross income distribution such that all incomes above \hat{y} are such that $y \in Y_3 \setminus y_2$, while all incomes below are such that $y \in Y_2 \setminus y_1$. Appendix B illustrates the derivation of \hat{y} that is

$$\hat{y} := (1+\varepsilon) \frac{(t_2 - t_3) y_2}{(1-t_3)^{(1+\varepsilon)} - (1-t_2)^{(1+\varepsilon)}}$$

with $t_2 > t_3$, while if $t_2 \rightarrow t_3$ then $\hat{y} = \frac{y_2}{(1-t_2)^\varepsilon}$. It follows that

$$y_t(p) = \begin{cases} y(p)(1-t_1)^\varepsilon & \text{if } y(p) < \frac{y_1}{(1-t_1)^\varepsilon} \\ y_1 & \text{if } \frac{y_1}{(1-t_1)^\varepsilon} \leq y(p) < \frac{y_1}{(1-t_2)^\varepsilon} \\ y(p)(1-t_2)^\varepsilon & \text{if } \frac{y_1}{(1-t_2)^\varepsilon} \leq y(p) \leq \hat{y} \\ y(p)(1-t_3)^\varepsilon & \text{if } y(p) > \hat{y} \end{cases} \quad (18)$$

where the after tax gross income $y_t(p)$ is discontinuous at $y(p) = \hat{y}$.

The presentation of the further case where the optimal labour supply choice is such that after tax no gross incomes belong to the second income bracket is discussed in Appendix B.

5.2 Numerical results

The optimal taxation problem described in the previous section is solved numerically.⁸ To this end, we need to assign a value to the parameters α and k of the utility function (13) and to specify the distribution ξ_w of individual wages and the exogenous revenue requirement \bar{G} .

The parameter α determines the wage elasticity ε of labor supply, which is constant throughout the entire wage distribution and equal to $\frac{1}{(\alpha-1)}$. The parameter k is a scale parameter which is set equal to $1/\alpha$. We simulate the model for three different values of ε , i.e. 0.1, 0.2, 0.5, and accordingly we set α respectively equal to 11, 6 and 3.⁹ For a given distribution of wages, different values of ε have two effects: first they impact on the distribution of gross income in the absence of taxation; second they determine how this distribution reacts to the tax system. We want to get rid of the first effect in order to focus on how the optimal tax structure is affected by the strength of the agents' reaction to the tax system. Accordingly, when ε changes, we keep the distribution of gross income in the absence of taxation constant, by an appropriate rescaling of the wage distribution. This constant distribution of gross income in the absence of taxation is chosen to be equal to the distribution implied by a wage elasticity ε that tends to zero.¹⁰ In turn, given that ε tends to zero, it is possible to show that such a distribution is equal to the distribution of wages. With regard to this wage distribution we assume that it is a Pareto distribution, as in Apps et al. (2014), Andrienko et al. (2016), and Slack (2015). More specifically, we follow Apps et al. (2014) and consider a truncated Pareto distribution, with mean μ and median m respectively equal to 48.07 and 32.3, and wages ranging from 20 to 327.¹¹

⁸We 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 combination of these policy parameters we compute the value of t_3 which keeps the government budget constraint balanced and then we compute the value of the social evaluation function. Last, we identify the combination of policy parameters that delivers the highest value of the social evaluation function.

⁹The values of the labor supply elasticity we consider are broadly consistent with the empirical estimates provided by the literature (see Giertz 2004, Meghir and Philips 2008, Saez, Slemrod and Giertz 2009 and Creedy 2009).

¹⁰Given a reference distribution $\hat{\xi}_w$ of wages, an ε that tends to zero and the implied gross income distribution $\hat{\xi}_y$, it is possible to show that the distribution of income is equal to $\hat{\xi}_y$ even when the ε is positive if all wages are raised to the power of $(1 + \varepsilon)$ (see equation (15) and set $t = 0$ and $k = 1/\alpha$).

¹¹The cdf of a Pareto distributed variable x is $F(x) = 1 - (\frac{L}{x})^\alpha$, where L is a scale parameter, denoting the lowest value of the distribution, while the coefficient $\alpha > 1$ represents the Pareto index, which is a measure of the degree of inequality within the distribution. As in Apps et al. (2014)

Finally we consider different values of the exogenous revenue requirement \bar{G} , namely we alternatively set \bar{G} equal to 10%, 15%, 20%, 25% of the average gross income computed in the absence of taxation.

As to the tax system, we assume two different regimes (convex and non-convex) depending on the ranking between t_2 and t_3 . The convex tax regime is such that $t_1 \leq t_2 \leq t_3$, while the non-convex tax regime considers the configuration where $t_1 \leq t_3 \leq t_2$. We always assume that there is an upper limit $\bar{\tau}$ to the value of the marginal tax rates and we set $\bar{\tau} = 50\%$.

Before we present the results of the simulations for the values of $\varepsilon > 0$ mentioned above, we report in Table 1 the optimal values of the policy parameters in the case in which ε tends to zero and accordingly labor supply is fixed. The Table provides a quantitative illustration of the theoretical analysis that has been performed in Section

(case (1.a)), we assume $L=20$, $\alpha = 1.4$ and we truncate the distribution at the 98th percentile which corresponds to a value of 327.

4.

Table 1. Optimal tax systems with fixed labor supply

Panel A: Gini Social Evaluation Function.

Initial social welfare: 30.21, Inequality before taxes: 0.37.

\bar{G}	t_1	t_2	t_3	y_1	y_2	W_G
$0.1 \times \mu$	0	0	50%	66.07 0.83	66.07 0.83	29.83
$0.15 \times \mu$	0	0	50%	45.02 0.69	45.02 0.69	29.27
$0.2 \times \mu$	0	0	50%	32.66 0.51	32.66 0.51	28.31
$0.25 \times \mu$	0	0	50%	24.68 0.26	24.68 0.26	26.85

Panel B: Polarization Social Evaluation Function.

Initial social welfare: 42.83, Polarization before taxes: 0.11.

\bar{G}	t_1	t_2	t_3	y_1	y_2	W_P
$0.1 \times \mu$	0	50%	0	27.5 0.37	52.00 0.75	39.75
$0.15 \times \mu$	0	50%	0	26.0 0.31	71.50 0.85	37.81
$0.2 \times \mu$	0	50%	0	24.25 0.25	106.50 0.92	35.62
$0.25 \times \mu$	0	50%	0	22.5 0.15	171.50 0.97	33.40

Note. The two values reported in the columns y_1 and y_2 express the thresholds in terms of the level of income and the associated percentile in the income distribution.

More specifically panel A illustrates the first two propositions, and shows that the optimal tax system reducing income inequality, is such that there is a no-taxation area ($t_1 = t_2 = 0$) until a given threshold ($y_1 = y_2$). Above this cutoff, the tax rate is set to its upper bound, i.e. ($t_3 = 50%$). The higher the amount of collected taxes (\bar{G}) is, the lower is the income threshold and the no-taxation area. For example, when the required revenue doubles from 10% to 20% of the average income, the fraction of incomes falling in the taxation area increases from 17% to 50%, (compare rows 1 and 3 of panel A).

Simulation results panel B of Table 1 illustrate Proposition 3. The optimal tax system aimed at reducing polarization, envisages a central bracket with the maximum admissible tax rate and no taxation in the two external brackets. The median income falls within this central bracket which widens as the amount of required tax revenue

increases. For example, comparing the first and the fourth row of panel B, we see that, when the amount of collected taxes increases, the fraction of people belonging to the central bracket changes from about 40% to about 80%.

We now present the results of the simulations for positive values of the labor supply elasticity. Tables 2A and 2B show the optimal policy aimed at reducing inequality both under the convex and the non-convex tax regime.

Table 2A. Optimal convex tax-system: Gini based SEF.

\bar{G}	ε	t_1	t_2	t_3	y_1	y_2	W
$0.1 \times \mu$	0.1	0	18%	49.9%	55.00 0.78	60.00 0.82	29.57
$0.15 \times \mu$	0.1	0	9%	49.85%	35.00 0.56	40.00 0.67	28.65
$0.2 \times \mu$	0.1	0	18%	49.79%	25.00 0.29	30.00 0.49	27.06
$0.25 \times \mu$	0.1	6%	7%	49.97%	20.00 0.01	25.00 0.34	24.79
$0.1 \times \mu$	0.2	0	39%	49.8%	50.00 0.77	55.00 0.82	29.23
$0.15 \times \mu$	0.2	0	10%	49.89%	30.00 0.46	35.00 0.64	27.80
$0.2 \times \mu$	0.2	2%	42%	49.91%	25.00 0.38	30.00 0.54	25.47
$0.25 \times \mu$	0.2	12%	42%	49.87%	25.00 0.38	30.00 0.54	22.91
$0.1 \times \mu$	0.5	1%	36%	43.30%	35.00 0.68	40.00 0.76	27.39
$0.15 \times \mu$	0.5	9%	11%	42.43%	30.00 0.49	35.00 0.70	24.38
$0.2 \times \mu$	0.5	18%	34%	43.10%	35.00 0.67	40.00 0.76	21.32
$0.25 \times \mu$	0.5	27%	41%	44.29%	35.00 0.70	40.00 0.76	18.14

Table 2B. Optimal non-convex tax-system: Gini based SEF.

\bar{G}	ε	t_1	t_2	t_3	y_1	y_2	W
$0.1 \times \mu$	0.1	0	47%	46.34%	55.00 0.79	60.00 0.82	29.54
$0.15 \times \mu$	0.1	1%	49%	48.59%	40.00 0.67	45.00 0.72	28.47
$0.2 \times \mu$	0.1	2%	50%	49.83%	30.00 0.49	35.00 0.60	26.97
$0.25 \times \mu$	0.1	7%	49%	48.80%	25.00 0.34	30.00 0.49	24.66
$0.1 \times \mu$	0.2	0	49%	48.36%	50.00 0.79	55.00 0.82	29.21
$0.15 \times \mu$	0.2	1%	50%	49.12%	35.00 0.64	40.00 0.70	27.70
$0.2 \times \mu$	0.2	6%	50%	49.75%	30.00 0.54	35.00 0.64	25.43
$0.25 \times \mu$	0.2	15%	50%	49.50%	30.00 0.54	35.00 0.64	22.86
$0.1 \times \mu$	0.5	1%	41%	20.53%	35.00 0.70	165.00 0.98	27.41
$0.15 \times \mu$	0.5	9%	42%	21.86%	35.00 0.70	200.00 0.99	24.45
$0.2 \times \mu$	0.5	17%	50%	25.46%	35.00 0.73	165.00 0.99	21.36
$0.25 \times \mu$	0.5	27%	48%	29.61%	40.00 0.78	155.00 0.98	18.21

The comparison between Table 1A and Table 2B shows that, when the wage elasticity of labor supply is positive, the optimal convex tax system aimed at reducing income inequality always requires a central bracket with positive marginal tax rate (t_2). The tax rate on the third income bracket (t_3) is approximately equal to its upper bound and it declines as the elasticity increases. As to the tax rate (t_1) on the first income bracket, it is zero when the wage elasticity of labor supply and the exogenous revenue requirement are low. However, when the amount of collected taxes or the wage elasticity of labor supply rise, this no-taxation area may disappear.

Table 2B shows the optimal tax system for inequality reducing social objectives under the non-convex regime, that is when $t_3 \leq t_2$. It always happens that the optimal value of t_3 is strictly below t_2 . In particular the difference between the two tax rates is sizeable when the wage elasticity of labor supply is equal to 0.5.

Table 3 compares the values of the social evaluation function associated to the

two different tax regimes for each combination of \overline{G} and ε . The comparison shows that, to reduce income inequality, the convex system is socially preferred to the non-convex one for low level of the wage elasticity of labor supply. When the elasticity is equal to 0.5 the optimal tax system becomes the non-convex one and top incomes face lower marginal tax rates than incomes in the central part of the distribution. The reason for choosing to reduce the tax rate on top incomes, whose weight in the social evaluation function is low¹², is related to a Laffer curve type effect and is reminiscent of the classical result for optimal non linear income taxation by Mirrlees (1971) of zero marginal tax rate for the top income. Setting t_3 below t_2 , it is possible to collect more revenues from top incomes and thus to reduce the fiscal burden for people in the lower tail of the income distribution. The argument can be understood looking at the last row of Tables 2A and 2B. Under the convex tax regime (Table 2A), the first income threshold is around the 70th percentile and the marginal tax rate in this income bracket is equal to 27%. Then, there is a narrow central bracket with a marginal tax rate equal to 41% and including about the 7% of population. The marginal tax rate on the remaining 23% of the population is equal to 44%. The non-convex tax regime (table 2B) entails a remarkable reduction of the marginal tax rate in the last bracket which however includes only the 2% of the population. The marginal tax rate in the central bracket increases to 48%. Finally the marginal tax rate within the first bracket is the same as in the convex case but the first bracket is however larger (it includes the 78% of the population) than the corresponding bracket in the convex tax system. In summary, when the wage elasticity is equal to 0.5, 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.

¹²See the Gini weighting function in figure 1.

Table 3. Convex vs. Non-convex regime: Gini SEF.

\bar{G}	ε	W_C	W_{NC}	Socially preferred tax regime
$0.1 \times \mu$	0.1	29.57	29.54	<i>Convex</i>
$0.15 \times \mu$	0.1	28.65	28.47	<i>Convex</i>
$0.2 \times \mu$	0.1	27.06	26.97	<i>Convex</i>
$0.25 \times \mu$	0.1	24.79	24.66	<i>Convex</i>
$0.1 \times \mu$	0.2	29.23	29.21	<i>Convex</i>
$0.15 \times \mu$	0.2	27.80	27.70	<i>Convex</i>
$0.2 \times \mu$	0.2	25.47	25.43	<i>Convex</i>
$0.25 \times \mu$	0.2	22.91	22.86	<i>Convex</i>
$0.1 \times \mu$	0.5	27.39	27.41	<i>Non Convex</i>
$0.15 \times \mu$	0.5	24.38	24.45	<i>Non Convex</i>
$0.2 \times \mu$	0.5	21.32	21.36	<i>Non Convex</i>
$0.25 \times \mu$	0.5	18.14	18.21	<i>Non Convex</i>

Table 4A. Optimal convex tax-system for polarization based SEF.

\bar{G}	ε	t_1	t_2	t_3	y_1	y_2	W
$0.1 \times \mu$	0.1	0	0	24.53%	30.00 0.46	30.00 0.46	38.30
$0.15 \times \mu$	0.1	1%	1%	36.53%	20.00 0	30.00 0.48	35.89
$0.2 \times \mu$	0.1	9%	9%	37.81%	20.00 0.01	30.00 0.48	33.42
$0.25 \times \mu$	0.1	18%	18%	37.59%	30.00 0.48	30.00 0.48	30.94
$0.1 \times \mu$	0.2	10.52%	10.52%	10.52%			37.63
$0.15 \times \mu$	0.2	15.51%	15.51%	15.51%			34.99
$0.2 \times \mu$	0.2	20.96%	20.96%	20.96%			32.30
$0.25 \times \mu$	0.2	26.59%	26.59%	26.59%			29.56
$0.1 \times \mu$	0.5	10.57%	10.57%	10.57%			36.22
$0.15 \times \mu$	0.5	16.40%	16.40%	16.40%			32.74
$0.2 \times \mu$	0.5	22.75%	22.75%	22.75%			29.08
$0.25 \times \mu$	0.5	29.85%	29.85%	29.85%			25.17

Table 4B. Optimal non-convex tax system: polarization based SEF.

\bar{G}	ε	t_1	t_2	t_3	y_1	y_2	W
$0.1 \times \mu$	0.1	0	49%	0	30.00 0.49	65.00 0.84	39.04
$0.15 \times \mu$	0.1	3%	50%	3%	30.00 0.49	85.00 0.90	36.52
$0.2 \times \mu$	0.1	16%	48%	16.32%	25.00 0.34	75.00 0.87	33.99
$0.25 \times \mu$	0.1	23%	50%	23.41%	25.00 0.34	70.00 0.86	31.41
$0.1 \times \mu$	0.2	2%	50%	2%	30.00 0.54	60.00 0.84	38.33
$0.15 \times \mu$	0.2	9%	46%	9.40%	30.00 0.53	60.00 0.84	35.55
$0.2 \times \mu$	0.2	16%	48%	16.32%	30.00 0.54	55.00 0.81	32.76
$0.25 \times \mu$	0.2	23%	50%	23.41%	30.00 0.54	50.00 0.79	29.90
$0.1 \times \mu$	0.5	8%	28%	8.13%	30.00 0.56	50.00 0.80	36.58
$0.15 \times \mu$	0.5	14%	34%	14.18%	30.00 0.59	50.00 0.81	32.97
$0.2 \times \mu$	0.5	22%	32%	22.15%	30.00 0.58	40.00 0.71	29.18
$0.25 \times \mu$	0.5	29%	39%	29.11%	25.00 0.49	35.00 0.69	25.21

Tables 4A and 4B show the optimal tax schedule when the government objective is the reduction of polarization. Table 4a illustrates the case of a convex tax regime. When the wage elasticity of labor supply is equal to 0.1, the optimal tax system envisages two income brackets, identified by an income threshold which is close to the median income. The marginal tax rate in the income bracket above that threshold is higher than that in the first income bracket. When the exogenous revenue requirement increases, the income threshold remains constant while the marginal tax rates, both above and below the threshold, rise. When the wage elasticity of labor supply is higher than 0.1, the optimal tax system aimed at reducing polarization requires a proportional taxation, which is increasing in the amount of revenues required.

Simulations under the non-convex tax regime are reported in table 4B. In this case, the optimal tax system requires a central bracket with a high marginal tax rate (t_2). For low values of the wage elasticity of labor supply, i.e. ε equal to 0.1 or to 0.2, t_2 is almost equal to its upper bound, while it sharply reduces when ε raises to 0.5. As to the marginal tax rates on the two external brackets, they increase with

the exogenous revenue requirement.

Finally Table 5 compares, for different combinations of \bar{G} and ε , the values of the social evaluation function associated to the the convex and the non-convex tax regimes when the aim of the government is to reduce polarization. The comparison shows that the non-convex tax system is always socially preferred to the convex one and therefore the optimal tax schedule is such that $t_2 > t_3 > t_1$. Thus Proposition 3, which has been proved under the assumption of fixed labor supply, also holds qualitatively when labor supply is endogenous, with the important qualification that marginal tax rates in the first and the third bracket are no longer always equal to zero and the marginal tax rate in the second bracket is no longer always equal to its upper bound.

Table 5. Convex vs. Non Convex regime:Polarization SEF.

\bar{G}	ε	W_C	W_{NC}	Socially preferred tax regime
$0.1 \times \mu$	0.1	38.30	39.04	<i>Non Convex</i>
$0.15 \times \mu$	0.1	35.89	36.52	<i>Non Convex</i>
$0.2 \times \mu$	0.1	33.42	33.99	<i>Non Convex</i>
$0.25 \times \mu$	0.1	30.94	31.41	<i>Non Convex</i>
$0.1 \times \mu$	0.2	37.63	38.33	<i>Non Convex</i>
$0.15 \times \mu$	0.2	34.99	35.55	<i>Non Convex</i>
$0.2 \times \mu$	0.2	32.30	32.76	<i>Non Convex</i>
$0.25 \times \mu$	0.2	29.56	29.90	<i>Non Convex</i>
$0.1 \times \mu$	0.5	36.22	36.58	<i>Non Convex</i>
$0.15 \times \mu$	0.5	32.74	32.97	<i>Non Convex</i>
$0.2 \times \mu$	0.5	29.08	29.18	<i>Non Convex</i>
$0.25 \times \mu$	0.5	25.17	25.21	<i>Non Convex</i>

6 Concluding remarks

In this paper we adopt a non-welfarist approach to analyze how the optimal income tax schedule changes according to the government's redistributive objective expressed in terms of either inequality or polarization reduction. More specifically, the focus is on the socially desirable mechanism collecting a given level of per-capita revenue, when redistribution is not allowed. We consider a piecewise linear income tax schedule with three income brackets. As in the optimal taxation literature, the tax problem is formalized as a constrained optimization exercise. The interesting aspect of our work is the formalization of the government's redistributive objective, which is expressed by a rank-dependent social evaluation function. In particular, in line with the literature on income inequality measurement we have considered two families of rank-dependent evaluation functions that incorporate either concerns for inequality

reduction or concerns for polarization reduction.

Our results reveal that redistributive objectives matter. The optimal tax schedule substantially changes depending on whether the government is inequality or polarization sensitive. In particular, with fixed labor supply, the optimal tax schedule maximizing an inequality sensitive SEF requires an income threshold above which the tax burden is the maximal admissible, and below which there is no taxation. In other words, to reduce income inequality the optimal tax system suggests to reduce the income distance between incomes within the second bracket and between these incomes and those in the first bracket that are not taxed. As to polarization reduction, the optimal tax schedule envisages a central interval where the tax rate is the maximum admissible and it is set equal to zero outside this interval. That is, the way to face polarization is to reduce the distance between the incomes in the central bracket so to create a sort of less disperse middle class. At the same time the income in the higher bracket are taxed according to a lump-sum taxation that is keeping their absolute dispersion unaffected.

In order to make explicit the optimal tax system and to highlight differences in the redistributive objective numerical simulations are performed. In addition, simulations are implemented for different levels of wage labor supply elasticity and by considering two different tax regimes, depending on the ranking of t_2 and t_3 , i.e. convex scheme where $t_2 \leq t_3$ and non-convex where $t_3 \leq t_2$.

Simulations shows that in order to reduce inequality the convex regime is socially preferred to the non-convex one for low levels of wage labor supply elasticity. In addition, the optimal tax schedule always requires a central bracket exhibiting a positive tax rate. When elasticity is high the optimal tax schedule is non-convex and the reason is related to a Laffer curve type argument. With regard to the polarization reduction, the socially desirable tax configuration is always non-convex. In this case the result derived with fixed labour supply that requires for a lower marginal taxation for the upper income bracket is also combined with the Laffer type effect that is exhibited also when considering inequality sensitive SEFs.

Appendix A

Solutions of the constrained optimization problems for inequality and polarization sensitive SEFs.

Recall the SEF constrained optimization problem where

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

with $t_i \in [0, 1]$, $y_1 < y_2$. The associated partial derivatives are $\frac{\partial \mathcal{L}}{\partial t_i}$ for $i = 1, 2, 3$, $\frac{\partial \mathcal{L}}{\partial y_i}$ for $i = 1, 2$, and $\frac{\partial \mathcal{L}}{\partial \lambda}$.

More specifically

$$\frac{\partial \mathcal{L}}{\partial t_i} = - \int_0^1 v(p) \frac{\partial T(y(p))}{\partial t_i} dp - \lambda \int_0^1 \frac{\partial T(y(p))}{\partial t_i} dp \text{ for } i = 1, 2, 3.$$

Given the tax function $T(y)$, the term $\frac{\partial T(y)}{\partial t_i}$ is

$$\begin{aligned} \frac{\partial T(x)}{\partial t_1} &= \min \{y, y_1\}, \\ \frac{\partial T(y)}{\partial t_2} &= \begin{cases} 0 & \text{if } y \leq y_1 \\ y - y_1 & \text{if } y_1 < y \leq y_2 \\ y_2 - y_1 & \text{if } y > y_2 \end{cases} \end{aligned}$$

and

$$\frac{\partial T(y)}{\partial t_3} = \max \{y - y_2, 0\}.$$

Hence the partial derivatives with respect the three tax rates t_i are respectively

$$\frac{\partial \mathcal{L}}{\partial t_1} = - \int_0^{p_1} v(p) y(p) dp - \int_{p_1}^1 v(p) y_1 dp - \lambda \left[\int_0^{p_1} y(p) dp + \int_{p_1}^1 y_1 dp \right], \quad (19)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t_2} &= - \int_{p_1}^{p_2} v(p) [y(p) - y_1] dp - \int_{p_2}^1 v(p) [y_2 - y_1] dp \\ &\quad - \lambda \left[\int_{p_1}^{p_2} [y(p) - y_1] dp + \int_{p_2}^1 (y_2 - y_1) dp \right] \end{aligned} \quad (20)$$

or equivalently, after rearranging, $\frac{\partial \mathcal{L}}{\partial t_2}$ could be written as

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

and

$$\frac{\partial \mathcal{L}}{\partial t_3} = - \int_{p_2}^1 v(p) [y(p) - y_2] dp - \lambda \int_{p_2}^1 [y(p) - y_2] dp. \quad (21)$$

The two *F.O.Cs* with respect the income thresholds y_1 and y_2 are:

$$\frac{\partial \mathcal{L}}{\partial y_1} = - \int_0^1 v(p) \frac{\partial T(y)}{\partial y_1} dp - \lambda \left[\int_0^1 \frac{\partial T(y)}{\partial y_1} dp \right] = 0$$

and

$$\frac{\partial \mathcal{L}}{\partial y_2} = - \int_0^1 v(p) \frac{\partial T(y)}{\partial y_2} dp - \lambda \left[\int_0^1 \frac{\partial T(y)}{\partial y_2} dp \right] = 0$$

where the derivatives of the tax function with respect to the income threshold are respectively

$$\frac{\partial T(y)}{\partial y_1} = \begin{cases} 0 & \text{if } y \leq y_1 \\ t_1 - t_2 & \text{if } y > y_1 \end{cases}$$

and

$$\frac{\partial T(y)}{\partial y_2} = \begin{cases} 0 & \text{if } y \leq y_2 \\ t_2 - t_3 & \text{if } y > y_2 \end{cases}.$$

These two associated *F.O.Cs* can then be rewritten as

$$\frac{\partial \mathcal{L}}{\partial y_1} = - \int_{p_1}^1 v(p) [t_1 - t_2] dp - \lambda \int_{p_1}^1 (t_1 - t_2) dp = 0 \quad (22)$$

and

$$\frac{\partial \mathcal{L}}{\partial y_2} = - \int_{p_2}^1 v(p) [t_2 - t_3] dp - \lambda \left[\int_{p_2}^1 (t_2 - t_3) dp \right] = 0. \quad (23)$$

The *F.O.C.* with respect to the Lagrangian multiplier is

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \bar{G} - \int_0^1 T(y(p)) dp = 0. \quad (24)$$

Derivation and simplification of F.O.Cs.

The associated Kuhn-Tucker first order conditions (F.O.Cs) are for the marginal tax rates, either

$$\left. \frac{\partial \mathcal{L}}{\partial t_i} \right|_{t_i=0} \leq 0, \text{ or } \left. \frac{\partial \mathcal{L}}{\partial t_i} \right|_{t_i \in (0,1)} = 0, \text{ or } \left. \frac{\partial \mathcal{L}}{\partial t_i} \right|_{t_i=1} \geq 0$$

for $i = 1, 2, 3$. While the F.O.Cs for the income bracket thresholds are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial y_1} &= 0, \\ \frac{\partial \mathcal{L}}{\partial y_2} &= 0 \end{aligned}$$

with $y_2 > y_1 > 0$, and for the multiplier λ the F.O.C. requires that

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0.$$

We provide here first a proof of the optimization result for inequality sensitive SEFs, then we will prove the result for the polarization sensitive SEFs.

The first simplifications of the F.O.Cs. are expanded here below.

As shown above, the derivatives of the Lagrangian function in (11) are:

$$\frac{\partial \mathcal{L}}{\partial t_i} = - \int_0^1 v(p) h_i(p) dp - \lambda \left[\int_0^1 h_i(p) dp \right] \quad (25)$$

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

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

The associated cdfs of these three inverse functions are denoted with H_i .

The partial derivatives w.r.t. the thresholds of the income brackets are also

$$\frac{\partial \mathcal{L}}{\partial y_1} = [t_2 - t_1] [1 - V(p_1) + (1 - p_1)\lambda] \quad (26)$$

$$\frac{\partial \mathcal{L}}{\partial y_2} = [t_3 - t_2] [1 - V(p_2) + (1 - p_2)\lambda] \quad (27)$$

and the derivative w.r.t. Lagrange multiplier is

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \bar{G} - \int_0^1 T(y(p)) dp = \bar{G} - \sum_{i=1}^3 t_i \int_0^1 h_i(p) dp. \quad (28)$$

Recall that each SEF can be decomposed into an abbreviated social evaluation where the average of a distribution is multiplied by 1 minus a measure $D_v(\cdot)$ of the degree of dispersion quantified by a linear index. That is $W_v(F) = \mu(F) [1 - D_v(F)]$, in our case $D_v(F)$ could be for instance the Gini index or a polarization index as those illustrated in Section 2. It follows that

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

Moreover, denote with $\phi_i(p)$ the quantile function at position p of distribution of 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 $p \geq p_i$, that is

$$\phi_i(p) := \begin{cases} 0 & \text{if } p < p_1 \\ z & \text{if } p \geq p_1 \end{cases} .$$

Note that $\mu(\Phi_i) = z \cdot (1 - p_1)$. It follows that:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial y_1} & = [t_2 - t_1] [1 - V(p_1) + (1 - p_1)\lambda] \\ & = [t_2 - t_1] [\mu(\Phi_1) \cdot [1 - D_v(\Phi_1)] + \mu(\Phi_1)\lambda], \end{aligned} \quad (29)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial y_2} & = [t_3 - t_2] [1 - V(p_2) + (1 - p_2)\lambda] \\ & = [t_3 - t_2] [\mu(\Phi_2) \cdot [1 - D_v(\Phi_2)] + \mu(\Phi_2)\lambda], \end{aligned} \quad (30)$$

and

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \bar{G} - \sum_{i=1}^3 t_i \cdot \mu(H_i). \quad (31)$$

The partial derivatives for the social optimization problem are summarized in the next remark.

Remark 1 *The partial derivatives of the Lagrangian optimization problem in (11) are:*

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t_i} &= -\mu(H_i) \cdot [1 - D_v(H_i) + \lambda] \text{ for } i \in \{1, 2, 3\}, \\ \frac{\partial \mathcal{L}}{\partial y_1} &= [t_2 - t_1] \cdot \mu(\Phi_1) \cdot [1 - D_v(\Phi_1) + \lambda], \\ \frac{\partial \mathcal{L}}{\partial y_2} &= [t_3 - t_2] \cdot \mu(\Phi_2) \cdot [1 - D_v(\Phi_2) + \lambda], \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= \bar{G} - \sum_{i=1}^3 t_i \cdot \mu(H_i). \end{aligned}$$

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

Inequality concerns.

We derive here the qualitative features of the optimal taxation problem that hold for any distribution of pre-tax gross incomes, for the class of inequality sensitive SEFs \mathcal{W}_I given by the set of all linear rank-dependent SEFs with decreasing weights $v(p)$, and for the set $\mathcal{T}_{\bar{\tau}}$ of piecewise linear three brackets 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 1. 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 $D_v(\Phi_1) < D_v(\Phi_2)$ [with $D_v(\Phi_1) = D_v(\Phi_2)$ 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) $\lambda = -[1 - D_v(\Phi_1)]$ and $[t_3 = t_2 = \tau]$.

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

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

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

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

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

It can be proved that $D_v(H_3) > D_v(H_2) > D_v(\Phi_1) > D_v(H_1)$ 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 $D_v(H_3) > D_v(H_2) > D_v(\Phi_1)$.

From the condition $D_v(H_3) > D_v(H_2) > D_v(\Phi_1) > D_v(H_1)$ 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 $\bar{G} = \sum_{i=2}^3 \mu(H_i)$.

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 $\bar{G} = \mu(H_2) + \mu(H_3)$. 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. $\bar{G} \leq \bar{\tau} \cdot \mu(F)$ we can derive the statement highlighted in the next proposition.

Proposition 4 (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*

$$\begin{aligned} t_1 &= 0, \\ t_3 &= t_2 = \bar{\tau}, \end{aligned}$$

with y_1 s.t. $\bar{G} = \bar{\tau} [\mu(H_2) + \mu(H_3)]$.

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 1.

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 $\lim_{p \rightarrow 1/2^-} v(p) = 2 \neq \lim_{p \rightarrow 1/2^+} v(p) = 0$ as for the polarization P index illustrated in the previous section. We denote with \mathcal{W}_P the set of all these SEFs.

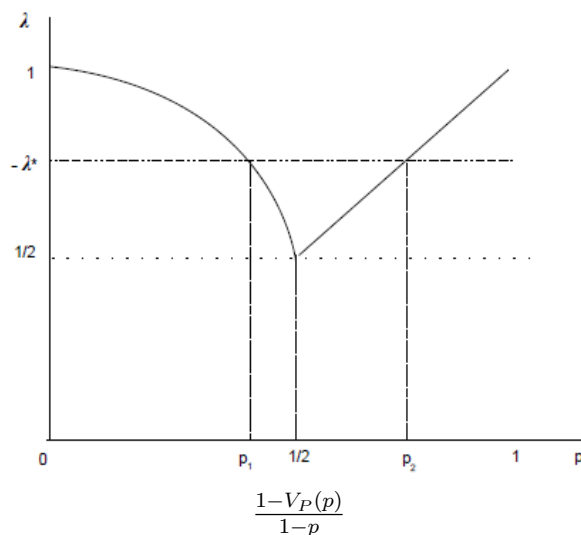
For these SEFs it is possible to derive p_1 and p_2 such that $D_v(\Phi_1) = D_v(\Phi_2)$. This is the case for instance for the SEF whose weights are represented in (5). For these measures it is possible to derive the associated $V(p)$ and compute $\frac{1-V(p)}{1-p}$. They are respectively:

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

with

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

Which can be represented as in the following figure



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

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

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

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

$$\frac{\partial \mathcal{L}}{\partial y_1} = \frac{\partial \mathcal{L}}{\partial y_2} = 0 \rightarrow \lambda = -1 + D_v(\Phi_1) = -1 + D_v(\Phi_2). \quad (32)$$

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

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

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

Note that for any polarization measure $D_v(\Phi_2) > D_v(H_3)$, 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 $D_v(\Phi_1) > D_v(H_1)$ 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 $D_v(\Phi_1) > D_v(H_1)$.

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 $D_v(\Phi_1) = D_v(\Phi_2) < D_v(H_2)$, 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 $D_v(H_2) > D_v(\Phi_{12})$.

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

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 $\bar{G} = \mu(H_2)$ 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 (32) and the revenue constraint that is such that $\lambda = -1 + D_v(\Phi_1) = -1 + D_v(\Phi_2)$ and $\bar{G} = \mu(H_2)$. 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 $D_v(\Phi_1^*) = D_v(\Phi_2^*)$ such that $1 - D_v(\Phi_1^*) + \lambda^* = 0$ and $1 - D_v(\Phi_2^*) + \lambda^* = 0$.

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 be 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 (22) and (23) 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 G the revenue $\int_0^1 T(y(p)) dp$ we obtain also that $\frac{\partial G}{\partial y_1} = - \int_{p_1}^1 dp = -(1 - p_1)$ and $\frac{\partial G}{\partial y_2} = \int_{p_2}^1 dp = (1 - p_2)$. It follows that by taking the differential of the revenue we have $dG = -(1 - p_1)dy_1 + (1 - p_2)dy_2$, so under the

assumption that the revenue constraint is satisfied $G = \bar{G}$, we have that $dG = 0$ and so

$$(1 - p_1)dy_1 = (1 - p_2)dy_2. \quad (33)$$

Analogously the differential of the SEF is

$$dW_v = [1 - V(p_1)] dy_1 - [1 - V(p_2)] dy_2. \quad (34)$$

Substituting for dy_2 from (33) we obtain

$$dW_v = (1 - p_1) \cdot \left[\frac{1 - V(p_1)}{1 - p_1} - \frac{1 - V(p_2)}{1 - p_2} \right] dy_1. \quad (35)$$

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 (35). 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 $\lambda = -1 + D_v(\Phi_1)$.

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

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

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

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

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 $\bar{G} = \mu(H_1) + \mu(H_2)$, while in the second case it is required that $\bar{G} = \mu(H_1)$.

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

Going back now to the case where $D_v(\Phi_1) - D_v(H_2) = D_v(\Phi_1) - D_v(H_3) = 0$. 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 $D_v(\Phi_1)$ is modified and given that H_2 and H_3 are not affected then the signs of $D_v(\Phi_1) - D_v(H_2)$ and $D_v(\Phi_1) - D_v(H_3)$ 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 $\bar{\tau} \in (0, 1]$. Let $y(1/2) = y_M$ denote the median income. Then, let H^- denote the distribution whose quantile function is

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

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

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

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 2 *Case (iia) may hold only if $\bar{G} > \bar{\tau}[\mu(H^+)]$. Case (iib) may hold only if $\bar{G} > \bar{\tau}[\mu(H^-)]$.*

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 5 (3A) *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) $p_1 < 1/2 < p_2$ where $I(\Phi_1) = I(\Phi_2)$ and such that $\bar{G} = \bar{\tau}\mu(H_2)$ with

$$\begin{aligned} t_1 &= t_3 = 0, \\ t_2 &= \bar{\tau}, \end{aligned}$$

if $\bar{G} \leq \min\{\bar{\tau}\mu(H^+), \bar{\tau}\mu(H^-)\}$.

(iia) If $\bar{G} > \bar{\tau}\mu(H^+)$ solution (i) should be compared with $p_1 < 1/2$, and

$$\begin{aligned} t_1 &= 0, \\ t_2 &= t_3 = \bar{\tau} \end{aligned}$$

where $\bar{G} = \bar{\tau}[\mu(H_2) + \mu(H_3)]$

(iib) If $\bar{G} > \bar{\tau}\mu(H^-)$ solution (i) should be compared with $p_1 > 1/2$,

$$\begin{aligned} t_1 &= \bar{\tau}, \\ t_2 &= t_3 = 0, \end{aligned}$$

where $\bar{G} = \bar{\tau}\mu(H_1)$.

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

Appendix B

The derivation of the optimal gross income distribution for the non-convex tax schedule.

In this appendix we present all computations underlying the derivation of the gross income distribution for the non-convex tax schedule case. We first derive the gross income distribution in the space of wages w , then we express such distribution in terms of quantiles $y(p)$. More specifically, we start the analysis by first assuming that under the non-convex regime the optimal labour supply and gross income are the same for all incomes that are in the first bracket and at the first threshold, the result changes for the income levels in the second and third brackets. In particular, if $t_2 > t_3$ then there exists a threshold level \widehat{w} in the wage distribution such that all wages above \widehat{w} are such that the associated $y \in Y_3 \setminus y_2$, while for all wages in $\left(\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} ; \widehat{w} \right)$ the associated gross income is such that $y \in Y_2 \setminus y_1$.

For all $w > y_1^{\alpha-1} \left(\frac{k\alpha}{(1-t_2)} \right)^{\frac{1}{\alpha}}$ the optimal gross income is $y^* > y_1$. If $t_2 > t_3$, the conditions in (15) could identify two potential levels of incomes one in $Y_2 \setminus y_1$ and one in $Y_3 \setminus y_2$ where the MRS_{xy} and the slope of the net income function $y - T(y)$ coincide. The optimal choice should then correspond to the one that exhibits larger utility.

Let $y_i^* = w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_i)}{k\alpha} \right]^{\frac{1}{\alpha-1}}$ with $l_i^* = \left[\frac{(1-t_i)w}{k\alpha} \right]^{\frac{1}{\alpha-1}}$ for $i = 2, 3$. Recall from (14) that the associated net incomes x_i^* are $x_2^* = (t_2 - t_1)y_1 + (1 - t_2)y_2^*$ and $x_3^* = (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)y_3^*$, then the utility levels associated to the pairs (x_i^*, l_i^*) for $i = 2, 3$ are respectively

$$\begin{aligned} U_2 &= U(x_2^*, l_2^*) = x_2^* - k \cdot l_2^{*\alpha} = (t_2 - t_1)y_1 + (1 - t_2)wl_2^* - kl_2^{*\alpha}, \\ U_3 &= U(x_3^*, l_3^*) = x_3^* - k \cdot l_3^{*\alpha} = (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)wl_3^* - kl_3^{*\alpha}. \end{aligned}$$

It then follows that $l^* = l_2^*$ when $w > y_1^{\alpha-1} \left(\frac{k\alpha}{(1-t_2)} \right)^{\frac{1}{\alpha}}$ if and only if $U_2 \geq U_3$, otherwise we have $l^* = l_3^*$.

That is, $l^* = l_2^*$ holds whenever

$$(t_2 - t_1)y_1 + (1 - t_2)wl_2^* - kl_2^{*\alpha} \geq (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)wl_3^* - kl_3^{*\alpha},$$

which can be simplified as

$$(1 - t_2)wl_2^* - kl_2^{*\alpha} \geq (t_3 - t_2)y_2 + (1 - t_3)wl_3^* - kl_3^{*\alpha}.$$

After substituting for l_i^* one obtains

$$\begin{aligned} & (1-t_2)w \left[\frac{(1-t_2)w}{k\alpha} \right]^{\frac{1}{\alpha-1}} - k \left[\frac{(1-t_2)w}{k\alpha} \right]^{\frac{\alpha}{\alpha-1}} \\ & - (1-t_3)w \left[\frac{(1-t_3)w}{k\alpha} \right]^{\frac{1}{\alpha-1}} + k \left[\frac{(1-t_3)w}{k\alpha} \right]^{\frac{\alpha}{\alpha-1}} \\ & \geq (t_3-t_2)y_2, \end{aligned}$$

that is

$$\left[\frac{(1-t_2)w}{k\alpha} \right]^{\frac{\alpha}{\alpha-1}} k(\alpha-1) - \left[\frac{(1-t_3)w}{k\alpha} \right]^{\frac{\alpha}{\alpha-1}} k(\alpha-1) \geq (t_3-t_2)y_2,$$

leading to

$$\begin{aligned} w^{\frac{\alpha}{\alpha-1}} k(\alpha-1) \left(\left[\frac{(1-t_3)w}{k\alpha} \right]^{\frac{\alpha}{\alpha-1}} - \left[\frac{(1-t_2)w}{k\alpha} \right]^{\frac{\alpha}{\alpha-1}} \right) & \leq (t_2-t_3)y_2, \\ w^{\frac{\alpha}{\alpha-1}} \frac{(\alpha-1)}{k^{\frac{1}{\alpha-1}} \alpha^{\frac{\alpha}{\alpha-1}}} \left((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}} \right) & \leq (t_2-t_3)y_2. \end{aligned}$$

It follows that

$$w^{\frac{\alpha}{\alpha-1}} \leq \frac{k^{\frac{1}{\alpha-1}} \alpha^{\frac{\alpha}{\alpha-1}}}{(\alpha-1)} \frac{(t_2-t_3)y_2}{\left((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}} \right)},$$

or expressing the condition in terms of w one obtains that the wage should be lower than a threshold \hat{w} , that is

$$w \leq \hat{w} := k^{\frac{1}{\alpha}} (\alpha-1)^{\frac{1}{\alpha}} \frac{\alpha}{(\alpha-1)} \left[\frac{(t_2-t_3)y_2}{\left((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}} \right)} \right]^{\frac{\alpha-1}{\alpha}}.$$

Recall that in order to obtain that y^* is in $Y_2 \setminus y_1$ it should hold that

$$w \in \left(\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} ; \left[y_2^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} \right),$$

we can then show that $\hat{w} < \left[y_2^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}}$.

To prove this condition consider the equivalent constraint $\hat{w}^{\frac{\alpha}{\alpha-1}} < y_2 \left[\frac{(1-t_2)}{k\alpha} \right]^{-\frac{1}{\alpha-1}}$, that is

$$\frac{k^{\frac{1}{\alpha-1}} \alpha^{\frac{\alpha}{\alpha-1}}}{(\alpha-1)} \frac{(t_2-t_3)y_2}{\left((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}} \right)} < y_2 \left[\frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha-1}}.$$

After a series of simplifications and rearrangements one obtains

$$\begin{aligned} \frac{\alpha}{(\alpha-1)} \frac{(t_2-t_3)}{\left((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}}\right)} &< \frac{1}{(1-t_2)^{\frac{1}{\alpha-1}}}, \\ \frac{\alpha}{(\alpha-1)}(t_2-t_3) &< \left(\frac{1-t_3}{1-t_2}\right)^{\frac{1}{\alpha-1}} (1-t_3) - (1-t_2), \\ \frac{\alpha}{(\alpha-1)} \frac{(t_2-t_3)}{(1-t_2)} &< \left(\frac{1-t_3}{1-t_2}\right)^{\frac{1}{\alpha-1}} \frac{(1-t_3)}{(1-t_2)} - 1, \\ 1 + \frac{\alpha}{(\alpha-1)} \frac{(t_2-t_3)}{(1-t_2)} &< \left(\frac{1-t_3}{1-t_2}\right)^{\frac{\alpha}{\alpha-1}}. \end{aligned}$$

Let $\delta = \frac{t_2-t_3}{1-t_2} > 0$, $\frac{1-t_3}{1-t_2} = 1+\delta$ and $\frac{\alpha}{(\alpha-1)} = \beta > 1$, the condition can then be rewritten as

$$1 + \beta\delta < (1 + \delta)^\beta.$$

This condition holds for all $\delta > 0$ and $\beta > 1$. Making use of the Hopital rule one can also prove that as $(t_2 - t_3)$ tends to 0 for positive values, the level of \widehat{w} converges to $\left[y_2^{\alpha-1} \frac{k\alpha}{(1-t_2)}\right]^{\frac{1}{\alpha}}$ from below.

We summarize these finding with the following remark, where the condition (ii) could be derived by taking the derivative of \widehat{w} w.r.t. t_3 .

Remark 3 *If $t_2 > t_3$, (i) $\widehat{w} < \left[y_2^{\alpha-1} \frac{k\alpha}{(1-t_2)}\right]^{\frac{1}{\alpha}}$, (ii) \widehat{w} is increasing in t_3 , and (iii) $\lim_{t_3 \rightarrow t_2^-} \widehat{w} = \left[y_2^{\alpha-1} \frac{k\alpha}{(1-t_2)}\right]^{\frac{1}{\alpha}}$.*

It could however be possible that $\widehat{w} < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)}\right]^{\frac{1}{\alpha}}$, that is the threshold \widehat{w} is below the infimum of the interval of wages leading to optimal choices of post tax gross incomes in $Y_2 \setminus y_1$. If this is the case no post tax gross income is in the interval $Y_2 \setminus y_1$. All gross incomes are therefore in the non adjacent intervals $Y_1 \setminus y_0$ and $Y_3 \setminus y_2$.

Given that $t_1 \leq t_3$ then in accordance with case A for all $w < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)}\right]^{\frac{1}{\alpha}}$ we have $l^* = l_1^* = \left[\frac{(1-t_1)w}{k\alpha}\right]^{\frac{1}{\alpha-1}}$ and $y^* = y_1^* = w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_1)}{k\alpha}\right]^{\frac{1}{\alpha-1}}$ with $y_1^* \in Y_1 \setminus y_0$.

If $\widehat{w} < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)}\right]^{\frac{1}{\alpha}}$ then for all wages where $w \geq \widehat{w}$ we have that $l^* = l_3^* = \left[\frac{(1-t_3)w}{k\alpha}\right]^{\frac{1}{\alpha-1}}$ and $y^* = y_3^* = w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_3)}{k\alpha}\right]^{\frac{1}{\alpha-1}}$. This is the case because the indifference curve that for these wages is tangent to the net income function in $Y_3 \setminus y_2$, lies above the one that is passing through the kink of the function associated to $y = y_1$.

However, there could be also other wage levels lower than \widehat{w} that lead to l_3^* and y_3^* as optimal solutions.

In order to identify them we need to investigate the case where $\widehat{w} < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}}$ and $w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} ; \widehat{w} \right)$.

In this case agents should choose between setting either $y^* = y_1$ or $y^* = y_3^* = w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_3)}{k\alpha} \right]^{\frac{1}{\alpha-1}}$.

The utility comparison then becomes

$$\begin{aligned} U_1 &= U((1-t_1)y_1, y_1/w) = (1-t_1)y_1 - k \cdot (y_1/w)^\alpha, \\ U_3 &= U(x_3^*, l_3^*) = (t_2-t_1)y_1 + (t_3-t_2)y_2 + (1-t_3)wl_3^* - kl_3^{*\alpha}. \end{aligned}$$

with $y^* = y_1$ if and only if $U_1 \geq U_3$, that is

$$(1-t_1)y_1 - k \cdot (y_1/w)^\alpha \geq (t_2-t_1)y_1 + (t_3-t_2)y_2 + (1-t_3)w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_3)}{k\alpha} \right]^{\frac{1}{\alpha-1}} - k \left[\frac{(1-t_3)w}{k\alpha} \right]^{\frac{\alpha}{\alpha-1}}.$$

The condition can be simplified into

$$y_1 - k \cdot (y_1/w)^\alpha \geq t_2y_1 + (t_3-t_2)y_2 + (1-t_3)w^{\frac{\alpha}{\alpha-1}} \left[\frac{1}{k\alpha} \right]^{\frac{1}{\alpha-1}} \left(\frac{\alpha-1}{\alpha} \right),$$

that is

$$(1-t_2)y_1 + (t_2-t_3)y_2 \geq k \cdot y_1^\alpha \cdot w^{-\alpha} + (1-t_3)w^{\frac{\alpha}{\alpha-1}} \left[\frac{1}{k\alpha} \right]^{\frac{1}{\alpha-1}} \left(\frac{\alpha-1}{\alpha} \right).$$

A wage level \tilde{w} could then be derived such that the above condition is solved with equality, that is such that

$$(1-t_2)y_1 + (t_2-t_3)y_2 = k \cdot y_1^\alpha \cdot \tilde{w}^{-\alpha} + (1-t_3)\tilde{w}^{\frac{\alpha}{\alpha-1}} \left[\frac{1}{k\alpha} \right]^{\frac{1}{\alpha-1}} \left(\frac{\alpha-1}{\alpha} \right).$$

Case B.1. Let $\widehat{w} := k^{\frac{1}{\alpha}} (\alpha-1)^{\frac{1}{\alpha}} \frac{\alpha}{(\alpha-1)} \left[\frac{(t_2-t_3)y_2}{((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}})} \right]^{\frac{\alpha-1}{\alpha}}$, and assume that

$\widehat{w} \geq \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}}$. It follows that

$$y^* = \begin{cases} w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_1)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} \\ y_1 & \text{if } w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} ; \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} \right) \\ w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_2)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} ; \widehat{w} \right) \\ w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_3)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w > \widehat{w} \end{cases} \quad (36)$$

where the post tax gross income y^* is discontinuous at $w = \hat{w}$. With the associated optimal labour supply levels

$$l^* = \begin{cases} \left[w \frac{(1-t_1)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} \\ y_1/w & \text{if } w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} ; \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} \right) \\ \left[w \frac{(1-t_2)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} ; \hat{w} \right] \\ \left[w \frac{(1-t_3)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w > \hat{w} \end{cases}.$$

By applying the following monotonically increasing transformation of the wage threshold \hat{w} we obtain the gross income threshold \hat{y} derived in the paper. In fact taking the definition of \hat{w} one obtains that

$$\begin{aligned} \frac{\hat{w}^{\frac{\alpha}{\alpha-1}}}{k^{\frac{1}{\alpha-1}}} &= (\alpha-1)^{\frac{1}{\alpha-1}} \left(\frac{\alpha}{(\alpha-1)} \right)^{\frac{\alpha}{\alpha-1}} \left[\frac{(t_2-t_3)y_2}{\left((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}} \right)} \right] \\ &= (\alpha-1)^{\left(\frac{1}{\alpha-1} - \frac{\alpha}{\alpha-1} \right)} \alpha^{\left(\frac{1}{\alpha-1} + 1 \right)} \left[\frac{(t_2-t_3)y_2}{\left((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}} \right)} \right]. \end{aligned}$$

By using the definition of the gross income in (16) where $y(p) := w(p)^{\frac{\alpha}{\alpha-1}} \left[\frac{1}{k\alpha} \right]^{\frac{1}{\alpha-1}}$ we obtain that the gross income threshold satisfy $\hat{y}\alpha^{\frac{1}{\alpha-1}} = \frac{\hat{w}^{\frac{\alpha}{\alpha-1}}}{k^{\frac{1}{\alpha-1}}}$ that is after substituting

$$\begin{aligned} \hat{y} &= \frac{\alpha}{\alpha-1} \left[\frac{(t_2-t_3)y_2}{\left((1-t_3)^{\frac{\alpha}{\alpha-1}} - (1-t_2)^{\frac{\alpha}{\alpha-1}} \right)} \right] \\ &= (1+\varepsilon) \left[\frac{(t_2-t_3)y_2}{\left((1-t_3)^{(1+\varepsilon)} - (1-t_2)^{(1+\varepsilon)} \right)} \right]. \end{aligned}$$

It then follows that the post tax gross income distribution is

$$y_t(p) = \begin{cases} y(p)(1-t_1)^\varepsilon & \text{if } y(p) < \frac{y_1}{(1-t_1)^\varepsilon} \\ y_1 & \text{if } \frac{y_1}{(1-t_1)^\varepsilon} \leq y(p) < \frac{y_1}{(1-t_2)^\varepsilon} \\ y(p)(1-t_2)^\varepsilon & \text{if } \frac{y_1}{(1-t_2)^\varepsilon} \leq y(p) \leq \hat{y} \\ y(p)(1-t_3)^\varepsilon & \text{if } y(p) > \hat{y} \end{cases}$$

Note that as explained before with this configuration of the tax system ($t_2 \geq t_3$) there is no bunching of incomes at the second income threshold.

Case B.2. Suppose that $\widehat{w} < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}}$. Let \tilde{w} denote the solution of

$$(1-t_2)y_1 + (t_2-t_3)y_2 = k \cdot y_1^\alpha \cdot \tilde{w}^{-\alpha} + (1-t_3)^{\frac{\alpha}{\alpha-1}} \cdot \tilde{w}^{\frac{\alpha}{\alpha-1}} \left(\frac{1}{k\alpha} \right)^{\frac{1}{\alpha-1}} \left(\frac{\alpha-1}{\alpha} \right) \quad (37)$$

such that $\tilde{w} \in \left(\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}}, \widehat{w} \right)$. The optimal levels are:

$$y^* = \begin{cases} w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_1)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} \\ y_1 & \text{if } w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}}; \tilde{w} \right] \\ w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_3)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w > \tilde{w} \end{cases},$$

where the gross income is discontinuous at $w = \tilde{w}$ with no gross income in the second income bracket Y_2 , and

$$l^* = \begin{cases} \left[w \frac{(1-t_1)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} \\ y_1/w & \text{if } w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}}; \tilde{w} \right] \\ \left[w \frac{(1-t_3)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w > \tilde{w} \end{cases}.$$

At the same time, by using (16) and substituting in the implicit definition of \tilde{w} we have that \tilde{y} is the solution of:

$$\begin{aligned} (1-t_2)y_1 + (t_2-t_3)y_2 &= y_1^\alpha \cdot \frac{\tilde{y}^{1-\alpha}}{\alpha} + (1-t_3)^{\frac{\alpha}{\alpha-1}} \cdot \tilde{y} \left(\frac{\alpha-1}{\alpha} \right) \\ (1-t_2)y_1 + (t_2-t_3)y_2 &= y_1^{\left(\frac{\varepsilon+1}{\varepsilon}\right)} \left(\frac{\varepsilon}{\varepsilon+1} \right) \tilde{y}^{\left(-\frac{1}{\varepsilon}\right)} + (1-t_3)^{\varepsilon+1} \cdot \tilde{y} \left(\frac{1}{\varepsilon+1} \right). \end{aligned}$$

Then the post tax gross income distribution is

$$y_t(p) = \begin{cases} y(p) (1-t_1)^\varepsilon & \text{if } y(p) < \frac{y_1}{(1-t_1)^\varepsilon} \\ y_1 & \text{if } \frac{y_1}{(1-t_1)^\varepsilon} \leq y(p) \leq \tilde{y} \\ y(p) (1-t_3)^\varepsilon & \text{if } y(p) > \tilde{y} \end{cases}.$$

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 microeconomic model of household labour supply to design optimal income taxes. *Scandinavian Journal of Economics*, 115 (2): 449-475.
- [4] Aaberge, R. and Colombino, U. and Wennemo, T. (2012). Heterogeneity in the elasticity of labour supply in Italy and some policy implications.
- [5] Aaberge, R., Dagsvik, J.K and Strom, S. (1995). Labour supply responses and welfare effects of tax reforms.
- [6] Aaberge, R., Havnes, T. and Mogstad, M. (2013). A theory for ranking distribution functions, Discussion Papers 763, Statistics Norway, Research Department.
- [7] Andrienko, Y., Apps, P. and Rees, R. (2016). Optimal Taxation and Top Incomes. *International Tax Public and Finance*, 23 (6): 981-1003.
- [8] Apps, P., Long, N. and Rees, R. (2014). Optimal Piecewise Linear Income Taxation. *Journal of Public Economic Theory*, 16 (4): 523-545.
- [9] Apps, P. and Rees, R. (2009). *Public Economics and Household*. Cambridge University Press.
- [10] Atkinson, A.B. (1970). The measurement of inequality. *Journal of Economic Theory* 2: 244-263.
- [11] Atkinson, A.B. and Stiglitz, J. (1980). *Lectures on Public Economics*. McGraw-Hill, New York-London.
- [12] Bargain, O. and Peichl, A. (2016). Own-wage labor supply elasticities: variation across time and estimation methods. *IZA Journal of Labor Economics*, (2016) 5:10.
- [13] Bargain, O. Orsini, K. and Peichl, A. (2014). Comparing labor supply elasticities in Europe and the US: new results. *Journal Human Resources* 49 (3): 723-838.
- [14] Bennett, C. J. and Zitikis, R. (2015). Ignorance, lotteries, and measures of economic inequality. *Journal of Economic Inequality*, 13, 309-316.
- [15] Brzezinski, M. (2013). *Income Polarization and Economic Growth*. National Bank of Poland Working Paper, no. 147.
- [16] Creedy, J. (2009). *The Elasticity of Taxable Income: A Non-Technical Summary*. Paper prepared for the New Zealand Treasury.

- [17] Dasgupta, P., Sen. A. K. and D. Starrett (1973). Notes on the Measurement of Inequality. *Journal of Economic Theory*, 6, 180-187.
- [18] Deaton, A. (1983). An explicit solution to an optimal tax problem. *Journal of Public Economics*, 20: 333-346.
- [19] Donaldson, D. and Weymark, J.A. (1983). Ethically flexible Gini indices for income distributions in the continuum. *Journal of Economic Theory*, 29: 353-358.
- [20] Duclos, J. Y., Esteban, J. and Ray, D. (2004). Polarization: Concepts, Measurement, Estimation. *Econometrica*, 72: 1737-1772.
- [21] Ebert, U. and Moyes, P. (2003). The difficulty of income redistribution with labour supply. *Economics Bulletin*, 8 (2): 1-9.
- [22] Ebert, U. and Moyes, P. (2007). Income taxation with labor responses. *Journal of Public Economic Theory*, 9 (4): 653-682.
- [23] Esteban, J. and Ray, D. (1994). On the Measurement of Polarization. *Econometrica*, 62: 819-852.
- [24] Fellman, J. (1976). The effect of transformation on Lorenz curves. *Econometrica*, 44: 823-824.
- [25] Giertz, S.H. (2004). Recent literature on taxable income elasticities. Tax Analysis Division Congressional Budget Office, Washington D.C., paper 2004-16.
- [26] Jakobsson, U. (1976). On the measurement of the degree of progression. *Journal of Public Economics*, 5: 161-168.
- [27] Kakwani, N. C. (1977). Applications of Lorenz curves in economic analysis. *Econometrica*, 87: 71-80.
- [28] 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.
- [29] Kanbur, R, Pirtilla, J. and Tuomala, M. (2006). Non-Welfarist Optimal Taxation and Behavioural Public Economics. *Journal of Economic Surveys*, 20 (5): 849-868.
- [30] Kanbur R, and Tuomala, M. (1994). Inherent Inequality and the Optimal Graduation of Marginal Tax Rates. *Scandinavian Journal of Economics* 96 (2): 275-282.
- [31] Keen, M., Papapanagos H. and A. Shorrocks (2000): Tax Reform and Progressivity. *The Economic Journal*, Vol. 110, No. 460, pp. 50-68

- [32] Lambert, P.J. (2001). *The Distribution and Redistribution of Income*. Manchester University Press.
- [33] Maccheroni, F., Muliere P., Zoli C. (2005). Inverse stochastic orders and generalized Gini functionals. *Metron*, LXIII, (3): 529-559.
- [34] Mankiw, N. G., Weinzierl, M. and Yagan, D. (2009). Optimal Taxation in Theory and Practice. *The Journal of Economic Perspectives*, 23 (4): 147-174.
- [35] Marshall, A. W., Olkin, I. and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*, Springer.
- [36] Meghir, C. and Phillips, D. (2008). *Labour Supply and Taxes*. Iza Discussion Paper no. 3405.
- [37] Mirrlees, J.A. (1971). An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, 38: 175-208.
- [38] Mirrlees, J.A., Adam, S., Besley, T. et al. (2011). *Tax by design*. Oxford University Press.
- [39] OECD (2011). *Revenues Statistics, 1965-2010*. OECD, Paris, 2011 Edition.
- [40] OECD (2015). *Revenues Statistics, 1965-2014*. OECD, Paris, 2015 Edition.
- [41] O'Donoghue, T. and Rabin, M. (2003). Studying optimal paternalism, illustrated by a model of sin taxes. *American Economic Review*, 93: 186-191.
- [42] Onrubia, J., Salas, R. and Sanz, J.F. (2005). Redistribution and labour supply. *Journal of Economic Inequality*, (2005) 3: 109-124.
- [43] Peter, K.S., Buttrick, S. and Duncan D. (2009). *Global Reform of Personal Income Taxation, 1981-2005: Evidence from 189 Countries*. Iza Discussion Paper no. 4228.
- [44] Piketty, T. and Saez, E. (2012). *Optimal Labor Income Taxation*. NBER Wp series 18521.
- [45] Piketty, T., Saez, E., and Stantcheva, S. (2014). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy*, 6 (1): 230-271.
- [46] 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.
- [47] Sadka, E. (1976). On Income Distribution, Incentive Effects and Optimal Income Taxation. *Review of Economic Studies*, 43: 261-268.

- [48] Saez, E. (2002). Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses. *Quarterly Journal of Economics* 117, 1039-1073.
- [49] 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.
- [50] Saez, E. and Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*. 106 (1): 24-45.
- [51] Schroyen, F. (2005). An alternative way to model merit good arguments. *Journal of Public Economics*, 89: 957-966.
- [52] Seade, J. (1977). On the Shape of Optimal Tax Schedules. *Journal of Public Economics*, 7: 203-235.
- [53] Sen, A. (1985). *Commodities and Capabilities*. Amsterdam: North-Holland.
- [54] Sheshinski, E. (1972). The optimal linear income tax. *Review of Economic Studies*, 39: 297-302.
- [55] Sheshinski, E. (1989). Note on the shape of the optimum income tax schedule. *Journal of Public Economics*, 40: 201-215.
- [56] Slack, S.E. (2015). Revisiting the optimal linear income tax with categorical transfers. *Economics Letters*, 134: 73-77.
- [57] Slemrod, J., Yitzhaki, S., Mayshar, J. and Lundholm, M. (1994). The optimal two-bracket linear income tax. *Journal of Public Economics*, 53: 269-290.
- [58] Stern, N. H. (1976). On the Specification of Models of Optimum Income Taxation. *Journal of Public Economics*, 6: 123-162.
- [59] Suits, D. (1977): Measurement of Tax Progressivity. *American Economic Review*, 67: 747-752.
- [60] Tuomala, M. (1984). On Optimal Income Taxation. Some further numerical results. *Journal of Public Economics*, 23: 351-366.
- [61] Tuomala, M. (1985). Simplified Formulae for Optimal Linear Income Taxation. *Scandinavian Journal of Economics* 87 (4): 668-672.
- [62] Tuomala, M. (1990). *Optimal Income Tax and Redistribution*. Clarendon Press, Oxford.
- [63] Wang, Y. and Tsui, K. (2000). Polarization orderings and new classes of polarization indices. *Journal of Public Economic Theory*, 2, 349-363.

- [64] Weymark, J.A. (1981). Generalized Gini inequality indices. *Mathematical Social Sciences*, 1: 409-430.
- [65] Wolfson, M. C. (1994). When inequalities diverge. *American Economic Review*, 84(2), 353-358.
- [66] Yaari, M.E. (1987). The dual theory of choice under risk. *Econometrica* 55: 99-115.
- [67] Yaari, M.E. (1988). A controversial proposal concerning inequality measurement. *Journal of Economic Theory* 44: 381-397.
- [68] Yitzhaki, S. (1983). On an extension of the Gini inequality index. *International Economic Review* 24: 617-628.