Optimal Non-Welfarist Income Taxation for Inequality and Polarization Reduction

Vincenzo Prete
University of Verona
PRELIMINARY

September 1, 2016

Abstract

We adopt a non-welfarist approach to compute the optimal tax income schedule reducing income inequality and polarization.

We consider the case of three brackets piecewise linear taxation and formalize the government’s redistributive objective as the maximization of rank-dependent social evaluation function, where incomes are weighted according to individuals’ position in the income ranking. The social evaluation function takes into account inequality and polarization concerns.

We derive and compare the optimal tax schedule for both redistributive objectives under fixed labor supply in order to highlight the crucial differences between them. Results reveal that the two tax systems differ qualitatively. When we consider the reduction of income inequality the optimal taxation problem involves the maximal admissible proportional tax burden in the higher bracket and no taxation for bottom income. If the government is aimed at reducing income polarization the optimal tax schedule requires to apply the maximal admissible tax rate to the class between the two thresholds and no taxation for the ends. When the required revenue increases the solution is to move the two thresholds toward the ends of the distribution.

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


1 Introduction

Income taxation represents the most used fiscal tool to achieve redistributive purposes. However, the total tax burden is not equally distributed among the various sources of income. In particular three quarters of the total amount of income taxes
fall on labor income (OECD (2011)). Given that labor income taxation affects individuals labor supply decisions, it follows that governments can not ignore these distortive effects of taxation when they design the tax system. Then, a good tax schedule should satisfy the government’s redistributive aim by minimizing distortions on individuals’ work effort.

In this paper we are interested in analyzing the effects on the optimal income tax schedule of two specific redistributive objectives, such as inequality and polarization reduction. Therefore, we refer to the optimal income tax literature pioneered by Mirrlees (1971), whose work represents the first attempt to explicitly introduce efficiency concerns in the design of the optimal income taxation. In particular, Mirrlees formalizes the optimal tax problem as the maximization of a social welfare function defined over individuals’ utility, subject to the revenue requirement constraint and taking into account individuals’ reaction to taxation.

In this paper we consider a piecewise linear income tax, which represents the most common tax schedule, as it is relatively simple to implement. Individual gross incomes are divided into a given number of brackets and marginal tax rates change between brackets but are constant within. The aim of this paper is to contribute to the existing literature, by providing a model of piecewise linear taxation for the reduction of income inequality and polarization and to show the crucial differences in terms of optimal tax schedule between these two redistributive objectives. More specifically, we adopt a three brackets piecewise linear tax model with fixed labor supply. The use of three income brackets represents the most appropriate and easy way to show how the optimal tax structure changes when government moves from inequality to polarization concerns. We argue, indeed, that the analysis of a more complex tax system with more than three brackets makes computations more demanding without adding substantial improvements in the results.

Usually, optimal tax models following Mirrlees or Sheshinki’s approach assume that the government’s objective is the maximization of a social welfare function defined over individuals’ utility. Kanbur et al. (1994) introduce an alternative approach, based on the assumption that governments’ evaluation of social welfare relies on criteria other than individuals’ preferences. Therefore, as argued by Kanbur et al. (2006) a government is non-welfarist if it evaluates social welfare by using a different criterion than individuals’ utility. The focus on incomes rather than on utilities is also justified by the widespread use in the policy discussions of social indicators based on incomes. To this regard, for instance, Kanbur et al. (1994) consider a particular form of non-welfarism, namely the reduction of poverty.

This paper is related to this strand of literature. Specifically, we are interested in analyzing the effects of two specific non-welfarist objectives, such as inequality or polarization reduction, on the optimal income tax schedule. The main novelty of this work lies in our approach to formalize the government’s non-welfarist objective as the

---

1 See Piketty and Saez (2013) for a description of the actual tax systems.
maximization of a rank-dependent social evaluation function defined over individuals’ income. Incomes are weighted according to individuals’ position in the income ranking and the specific non-welfarist objective is described by the choice of the weighting function. Therefore, the purpose of this paper is to highlight qualitatively and quantitatively by using numerical simulation, the crucial differences between the two different non-welfarist objective.

This work proceeds as follows: the second section reviews the literature on optimal income taxation, distinguishing between welfarist and non-welfarist approach. In the third section we introduce the notation and present the rank-dependent social welfare function and the different weighting functions. Section 4 consider the case of piecewise linear income taxation when individuals labor supply is fixed. The results of numerical simulations are presented in section 5. Section 6 concludes and discusses policy implications.

2 Literature Review

2.1 Optimal Welfarist Income Taxation

The theoretical foundations of the standard welfarist approach are provided in Mirrlees (1971) and Sheshinski (1972) respectively for non linear and linear taxation. In the first article, the optimal tax problem is formulated as a problem of mechanism design. The government collects a given amount of taxes from a population of individuals heterogeneous in their productivity level expressed by an exogenous wage. The government’s redistributive objective is represented by a standard utilitarian social welfare function defined over individuals’ utility. The optimal tax problem is expressed as the maximization of this social welfare function under two constraints: the first is the classical government’s budget constraint, while the second, known as incentive compatibility constraint, deals with the fact that each individual prefers his allocated marginal tax rate and subsidy.

Mirrlees’ model outlines two main results: first, marginal tax rates are non negative and lower than 100 percent. Second, the marginal tax rate on the top of the income distribution is zero. Moreover, Seade (1977) shows that, if the income distribution is bounded from below, the optimal marginal tax rate on the lowest income is also zero.

However, the complexity of the non-linear tax model led the theoretical literature to formulate the optimal tax problem in a simpler way, by considering the case of linear taxation. Sheshinski (1972) was the pioneer of this approach. He considers a population of individuals with different levels of working ability and then different gross incomes. The government pursues a redistributive objective by maximizing a Bergson-Samuelson social welfare function whose arguments are individuals’ utility. The optimal tax problem requires to choose a tax rate and the level of lump-sum subsidy by maximizing the social welfare function under the budget constraint.
Following Sheshinski’s approach, Tuomala (1985) provides a simplified formula for the optimal linear income tax, where the equity-efficiency trade-off appears very clearly. Specifically, the higher the elasticity of labor supply, the higher the efficiency costs of taxation and the lower the optimal tax rate. Likewise, the optimal tax rate decreases with higher levels of inequality in the distribution of individuals’ wage.²

Therefore, both non-linear and linear optimal taxation models highlight the impact of the equity-efficiency trade-off on the design of the optimal tax schedule. However, actual tax systems diverge from those considered by the Mirrlees and Sheshinski’s models. In particular, most countries adopt a piecewise linear tax system, with few income brackets and marginal tax rates vary between brackets but are constant within.

Given this consideration, theoretical literature developed models of piecewise linear optimal taxation. To this regard, the works by Sheshinski (1989) and Slemrod et al. (1994) represent the starting point of this literature. More specifically, Sheshinski’s paper shows that the optimal tax system is convex in the sense that high tax rates are associated with high income brackets. Slemrod et al. (1994) challenge Sheshinski’s result arguing that the optimal tax structure could be non-convex. This is because Sheshinski ignored the discontinuity in the tax revenue function. Recently, Apps et al. (2014) provide a simple model of piecewise linear taxation with just two income brackets. They consider two systems, namely concave and convex, and by using numerical simulations they analyze the conditions under which each system is optimal.

2.2 Optimal Non-Welfarist Income Taxation

All the models described in the previous section adopt the standard welfarist approach, where the government’s redistributive objective is represented by the maximization of a Bergson-Samuelson social welfare function. Then, an utilitarian government weights all individuals’ utility equally and maximizes their sum irrespective of distributional aspects. If some conditions are satisfied, the optimal tax system envisages a confiscatory tax rate and lump-sum redistribution of the tax revenues. The extreme nature of this result leads the theoretical literature to criticize the welfarism

²Another easily understandable alternative expression for the optimal linear income tax rate has been provided by Dixit and Sadno (1977), Atkinson and Stiglitz (1980) and it is known as the covariance rule. In this formulation the equity-efficiency trade-off is readily interpretable, because the optimal tax rate is expressed as a ratio where the numerator captures the distributional aspect of taxation, while the denominator represents a measure of the distortions of taxation. Specifically, the numerator is the covariance between individuals income and the net social marginal valuation of this income, while the denominator is the weighted sum of individuals’ compensated elasticities of labor supply with respect to the tax rate, where weights are individuals incomes. Therefore, the higher the covariance (numerator), the higher the inequality in the income distribution, the higher the optimal tax rate. At the same time, the lower the compensated elasticity of labor supply, the lower the distortive effect of taxation, the higher the tax rate.
and to develop an alternative approach.

Kanbur et al. (1994) are the proponents of this alternative approach. Relying on Sen (1985)’s claim about the limits of welfarism implications in terms of policy evaluation, Kanbur et al. (1994) stress the importance of aspects not directly related to utility and to this regard they argue that policy discussions are based on individuals’ income and do not consider the disutility from working experienced by individuals. As a consequence, the optimal tax schedule differs with the two approaches. For example, they derive the optimal income tax schedule for a non-welfarist government whose object is the reduction of poverty. In this case the optimal tax problem is expressed as the minimization of an income based poverty index. They show as the non-welfarist tax formula differs from the welfarist one. In particular, the former is given by the sum of two components. The first is exactly the same as in the welfarist approach and represents the second best motive for taxation. While the second component is novel and measures the difference between private and social preferences. In other words, this component is the non-welfarist motive for taxation and it is also called a first best motive for taxation, to distinguish it from the previous one. From a numerical point of view, Kanbur et al. (1994) use numerical simulations in order to show that the optimal non-welfarist tax rules envisage lower marginal tax rates than the welfarist approach.

Another criticism of the use of individuals’ utility as in the welfarist approach is that in some situations individuals’ preferences can be manipulated and individuals’ behaviors are not optimal from the society’s point of view. Therefore, in these situations there is the need of a government’s corrective intervention, who uses taxes and transfer to correct individuals’ behavior. To this regard some papers develop models of paternalistic taxation, for example O’Donoghue and Rabin (2003) consider the case of taxation for the reduction of the consumption of harmful goods, while Schroyen (2005) provides a non-welfarist characterization of the merit goods provision.

However, despite the effort of the non-welfarist literature there is no a clear delimiting line between the two approaches. To this regard, the main indication arising from theoretical literature is that a government is "non-welfarist" when evaluates social welfare by using a different criterion than that one used by individuals to evaluate their own welfare. To some extent, as pointed by Kanbur et al. (2006), one could argue that redistribution itself represents an example of non welfarism, since government evaluates individuals welfare by using a different criterion than the individuals’ preferences. Conventionally, following Kanbur et al. (1994) a non-welfarist government is one that goes beyond utility and maximizes a social welfare function whose argument differ from individuals’ utility.

To summarize, the non-welfarist literature provides models to analyze how the impact of non-welfarist objectives on the optimal tax formula. These changes depend on the specific non welfarist objective, which is formalized in a particular form of the social evaluation function. However, in order to quantify the differences in terms of marginal tax rates between welfarist and non-welfarist approach numerical simulation
are needed.

3 Rank-dependent social evaluation function

Let $F (y)$ denote the cumulative distribution function of a population with bounded support $(0, y_{\text{max}})$ and finite mean $\mu (F) = \int_0^{y_{\text{max}}} y \, dF (y)$. The left inverse continuous distribution function or quantile function, showing the income level of an individual that covers position $p \in (0, 1)$ on 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 reminder of the paper we will also equivalently denote with $y (p)$ the quantile function. The average income could be calculated as $\mu (F) = \int_0^1 F^{-1} (p) \, dp$. Consider a set of 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 expressed as

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

(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. 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 evaluation is defined as

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

By defining $v (p) = \delta (1 - p)^{\delta - 1}$ we can rewrite (1) as

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

---

3 See also Aaberge (2000) and Maccheroni et al. (2005)

4 For general details see Lambert (2001).
which is the class of Generalized Gini SEF parameterized by $\delta \geq 1$ introduced by Donaldson and Weymark (1983) and Yitzhaki (1983). The parameter $\delta$ 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 as\footnote{The Gini index of inequality parameterized by $\delta$ is expressed as $G (F) = \frac{1}{\mu(F)} \int_0^1 \left[ 1 - \delta (1 - p)^{\delta - 1} \right] F^{-1} \, dp$, which becomes the standard Gini coefficient for $\delta = 2$.}

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

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

### 3.1 Weighting functions

#### 3.1.1 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 the weight as

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

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

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

With a "non-traditional" interpretation of the absolute Gini index, inequality could be measured by considering the difference between the incomes with 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 $F^{-1} \left( \frac{1}{2} + t \right) - F^{-1} \left( \frac{1}{2} - t \right)$ and weights it with the weight $2t$. That is

$$\mu (F) G (F) = \int_{1/2}^1 2 \left[ \frac{1}{2} - p \right] F^{-1} (p) \, dp - \int_0^{1/2} 2 \left[ \frac{1}{2} - p \right] F^{-1} (p) \, dp.$$
The weights attached to the income differences increase as the position of the individuals move 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 on the same side with respect to the median, also reduce inequality because 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 that 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.

3.1.2 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 is reduced 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 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). The associated SEF is rank dependent with a weighting function that can be formalized as:

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

(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 on different positions $p$ around the median. For $\beta = 1$ and $\delta = 2$ the weights $v_P(p)$ are linear and increasing,

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

(5)

We focus primarily on this weighting function as it constitute 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

![Polarization weighting function.](image)

It is linearly increasing both below and above the median and exhibits a jump at the median, with higher weights below the median and lower above the median.

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

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

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

\[ \mu (F) P (F) = - \int_{0}^{1/2} 2pF^{-1} (p) \, dp + \int_{1/2}^{1} 2(1 - p)F^{-1} (p) \, dp. \]  

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

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

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

\[ w_P (p) = \begin{cases} -[1 - 2(\frac{1}{2} - p)] & \text{if } p \leq \frac{1}{2} \\ [1 - 2(p - \frac{1}{2})] & \text{if } p \geq \frac{1}{2} \end{cases}. \]
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 $= 1/2$ to the extreme positions 0 and 1. For instance, for the incomes that are either $t$ positions above the median and $t$ positions below the median, the index considers the difference between these incomes $F^{-1} \left( \frac{1}{2} + t \right) - F^{-1} \left( \frac{1}{2} - t \right)$ and weights it with the weight $1 - 2t$. That is

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

The weights attached to the income differences decreases linearly as the position of the individuals move 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.

An elementary normative implication of the polarization based welfare weighting function is that in order to maximize the welfare, redistribution should be from 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 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.

### 4 Solution for piecewise linear taxation problems

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

$$
W = \int_{0}^{1} v (p) \left[ y (p) - T (y (p)) \right] \, dp,
$$

where $y (p)$ denotes the quantile function or the inverse of the income distribution. Let $y (p_1) = y_1$ and $y (p_2) = y_2$ 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 $T (x)$,
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 \( T(x) \) defined as follows
\[
T(x) = \begin{cases} 
  t_1 y & \text{if } y \leq y_1 \\
  t_1 y_1 + t_2 (y - y_1) & \text{if } y_1 < y \leq y_2 \\
  t_1 y_1 + t_2 (y_2 - y_1) + t_3 (y - y_2) & \text{if } y > y_2
\end{cases}
\]
or
\[
T(x) = t_1 y + (t_2 - t_1) \cdot \max \{ y - y_1, 0 \} + (t_3 - t_2) \cdot \max \{ y - y_2, 0 \}.
\]

In our analysis we consider situations where the gross incomes are unequally distributed across individuals. The social optimization problem requires to maximize \( W \) 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 group of individuals exhibit the same net income. These distributions could substantially differ depending on whether the social objective is concerned about reducing inequality or about reducing polarization.

4.1 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).

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 of this problem is as follows
\[
\max_{t_1,t_2,t_3,y_1,y_2} \mathcal{L} = W + \lambda \left[ \bar{G} - \int_0^1 T(x(p)) \, dp \right],
\]

with \( t_i \in [0,1] \), \( y_1 < y_2 \). The associated Kuhn-Tucker first order conditions are:

either
\[
\frac{\partial \mathcal{L}}{\partial t_i} \bigg|_{t_i=0} \leq 0,
\]

\footnote{See also the review in Lambert (2001).}
or
\[ \frac{\partial L}{\partial t_i} \bigg|_{t_i \in (0,1)} = 0, \]
or
\[ \frac{\partial L}{\partial t_i} \bigg|_{t_i = 1} \geq 0 \]
for \( i = 1, 2, 3; \)

\[ \frac{\partial L}{\partial y_1} = 0, \]
\[ \frac{\partial L}{\partial y_2} = 0 \]

with \( y_2 > y_1 > 0, \) and
\[ \frac{\partial L}{\partial \lambda} = 0. \]

We provide here a sketch of the proof of the optimization result that is illustrated in details in the Appendix for inequality sensitive SEF. We provide instead the full arguments for the derivation of the result for polarization sensitive SEFs.

As shown in the Appendix, the derivatives of the Lagrangian function are:

\[ \frac{\partial L}{\partial t_i} = - \int_0^1 v(p) h_i(p) \, dp - \lambda \left[ \int_0^1 h_i(p) \, dp \right] \]  \hspace{1cm} (9)

for \( i = 1, 2, 3, \) where

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

the associated cdfs are denoted with \( H_i. \)

The partial derivatives are also

\[ \frac{\partial L}{\partial y_1} = [t_2 - t_1] \left[ 1 - V(p_1) + (1 - p_1)\lambda \right] \]  \hspace{1cm} (10)
\[ \frac{\partial L}{\partial y_2} = [t_3 - t_2] \left[ 1 - V(p_2) + (1 - p_2)\lambda \right] \]  \hspace{1cm} (11)
and
\[
\frac{\partial L}{\partial \lambda} = \mathcal{C} - \frac{3}{\int_0^1 T(x(p)) \, dp} = \mathcal{C} - \sum_{i=1}^{3} t_i \int_0^1 h_i(p) \, dp. \tag{12}
\]

Recall that each SEF can be decomposed into an abbreviated evaluation where the average of a distribution is multiplied by 1 minus a measure of the degree of dispersion of a distribution quantified by a linear index. That is \( W(F) = \mu(F) [1 - I(F)], \) in our case \( I(F) \) could be the for instance the Gini index or a polarization index as those illustrated in the previous section. It follows that
\[
\frac{\partial L}{\partial t_i} = \mu(H_i) \cdot [1 - I(H_i)] - \lambda \cdot \mu(H_i) = -\mu(H_i) \cdot [1 - I(H_i) + \lambda].
\]

Moreover, denote with \( \phi_i(p) \) the quantile function at position \( p \) of distribution of \( \Phi_i \) 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 the individual in positions \( p \geq p_i \), that is
\[
\phi_i(p) = \begin{cases} 
0 \quad & \text{if } p < p_i \\
z \quad & \text{if } p \geq p_i
\end{cases}.
\]

Note that \( \mu(\Phi_i) = z \cdot (1 - p_i) \). It follows that:
\[
\frac{\partial L}{\partial y_1} = [t_2 - t_1] [1 - V(p_1) + (1 - p_1)\lambda] \nonumber
\]
\[
= [t_2 - t_1] [\mu(\Phi_1) \cdot [1 - I(\Phi_1)] + \mu(\Phi_1) \lambda] \tag{13}
\]
\[
\frac{\partial L}{\partial y_2} = [t_3 - t_2] [1 - V(p_2) + (1 - p_2)\lambda] \nonumber
\]
\[
= [t_3 - t_2] [\mu(\Phi_2) \cdot [1 - I(\Phi_2)] + \mu(\Phi_2) \lambda] \tag{14}
\]

and
\[
\frac{\partial L}{\partial \lambda} = \mathcal{C} - \sum_{i=1}^{3} t_i \cdot \mu(H_i). \tag{15}
\]

To summarize: if we let \( \frac{\partial L}{\partial y_i} = 0 \), then either \( t_{i+1} = t_i \) holds or \( \lambda = -[1 - I(\Phi_i)] \).

4.1.1 Inequality concerns.

If we consider SEFs where \( v(p) \) is decreasing as is the case for the Gini based SEF and in general for all SEF that are sensitive to inequality reductions through rank preserving progressive transfers from richer to poorer individuals, then \( I(\Phi_1) < I(\Phi_2) \). It then follows that either (i) \( [t_3 = t_2 = t_1 = t] \) or (ii) \( \lambda = -[1 - I(\Phi_1)] \) and \( [t_3 = t_2 = \tau] \).

The case (i) is not consistent with the solution because according to the revenue constraint we should obtain \( t = \sum_{i=1}^{3} \mu(H_i) / \mathcal{C} \in (0, 1) \). In this case it should be
\[
\frac{\partial L}{\partial t_i} = -\mu(H_i) \cdot [1 - I(H_i) + \lambda] = 0
\]
for all $i = 1, 2, 3$. Given that $I(H_i)$ could be different for all $i$, then $\lambda = I(\Phi_1) - 1$ could not hold for all $i$.

The solution associated to case (ii) then should hold. It then follows that, given that $I(H_i)$ could be different for all $i$, then $= 1$ could not hold for all $i$.

The solution associated to case (ii) then should hold. It then follows that, given that $I(H_i)$ could be different for all $i$, then $= 1$ could not hold for all $i$.

It is possible to prove that $I(H_3) > I(H_2) > I(H_1)$, for any SEF where $v(p)$ is decreasing and there is positive density both below $y_1$ and in between $y_1$ and $y_2$, it then follows that $\frac{\partial C}{\partial t_1} < 0$, and $\frac{\partial C}{\partial t_2} > 0$, $\frac{\partial C}{\partial t_3} > 0$ then $t_1 = 0$, $t_3 = t_2 = \tau = 1$, where $y_1$ and $y_2$ are set such that $\overline{G} = \sum_{i=2}^{3} \mu(H_i)$.

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 $\overline{\tau} \in (0, 1]$ s.t. $\overline{G} \geq \overline{\tau} \cdot \mu(F)$ we can derive the statement highlighted in the next proposition.

Let $W_I$ denote the set of all linear rank-dependent SEFs with decreasing weights $v(p)$. Let $T_r$ denote the set of all piecewise linear taxation schemes with 3 brackets with maximal marginal tax rate $\tau$. The following result holds.

**Proposition 1** The solution of the optimal taxation problem with fixed labour supply for tax schedules in $T_r$ maximizing linear SEFs in $W_I$ is

\[
\begin{align*}
  t_1 &= 0, \\
  t_3 &= t_2 = \tau,
\end{align*}
\]

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

Thus, the optimal taxation problem involves the maximal admissible proportional tax burden in the higher bracket and no taxation for bottom incomes. When $\overline{\tau} = 100\%$ then the solution involves reducing to $y_1$ all incomes that are above this value.

### 4.1.2 Polarization concerns.

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

For these SEFs it is possible to derive $p_1$ and $p_2$ such that $I(\Phi_1) = I(\Phi_2)$. This is the case for instance of 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:

\[
\begin{align*}
  V_P(p) = \begin{cases} 
  p^2 + p & \text{if } p \leq 1/2 \\
  p^2 + 1 - p & \text{if } p > 1/2
\end{cases}.
\end{align*}
\]
We consider first the case (i) that

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

Note that the $\frac{\partial \mathcal{L}}{\partial y_1} = \frac{\partial \mathcal{L}}{\partial y_2} = 0$ 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 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

$$-\lambda = 1 - I(\Phi_1) = 1 - \frac{p_1^2}{1-p_1} = 1 - I(\Phi_2) = p_2$$

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

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 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 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.

We consider first the case (i) where

$$\frac{\partial \mathcal{L}}{\partial y_1} = \frac{\partial \mathcal{L}}{\partial y_2} = 0 \rightarrow \lambda = -1 + I(\Phi_1) = -1 + I(\Phi_2). \quad (16)$$

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

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

$$= -\mu(H_i) \cdot [I(\Phi_2) - I(H_i)],$$

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

Note that for any polarization measure $I(\Phi_2) > I(H_3)$, that is $\frac{\partial \mathcal{L}}{\partial t_3} < 0$, implying that $t_3 = 0$. This result could be shown because the difference between $\Phi_2$ 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 \( \Phi_2 \) all incomes covering these positions are equal. As we have argued, moving from \( H_3 \) to \( \Phi_2 \) increases polarization because increases the identification effect reducing the inequality between the individuals on the same side of the median.

It is possible also to show that \( I(\Phi_1) > I(H_1) \) that is \( \frac{\partial \mathcal{L}}{\partial t_1} < 0 \), implying that \( t_1 = 0 \).

This result could be obtained by properly define distributions \( \Phi_1 \) and \( H_1 \) so that \( \mu(\Phi_1) = \mu(H_1) \). By construction it follows that the distributions cross once for \( p = p_1 \) and for all \( p > p_1 \) with \( p_1 < 1/2 \), incomes are larger in \( \Phi_1 \) 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 \( \Phi_1 \) 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 from \( \Phi_1 \) to \( H_1 \) and given that \( \mu(\Phi_1) = \mu(H_1) \) then \( I(\Phi_1) > I(H_1) \).

To verify the condition related to the sign of \( \frac{\partial \mathcal{L}}{\partial y_2} \), it is possible to combine distributions \( \Phi_1 \) and \( \Phi_2 \) whose linear measures of polarization are the same in order to obtain a new distribution \( \Phi_{12} \) with the same 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 \( I(\Phi_1) = I(\Phi_2) < I(H_2) \), thus \( \frac{\partial \mathcal{L}}{\partial y_2} > 0 \) and therefore \( t_2 = 1 \).

This is the case because by construction \( \Phi_{12} \) 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 closed to it to individuals in the upper tail. Both operations reduce the polarization and thus \( I(H_2) > I(\Phi_{12}) \).

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

Consider now the case \( \text{(ii)} \) where \( t_3 = t_2; t_1 \neq t_2 \) implying that in order to obtain \( \frac{\partial \mathcal{L}}{\partial y_2} = 0 \) necessarily is required that \( \lambda = -1 + I(\Phi_1) \).

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 \( \lambda \) into \( \frac{\partial \mathcal{L}}{\partial t_i} \), we obtain

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

Recall that \( t_3 = t_2 \) implies that the sign of \( I(\Phi_1) - I(H_2) \) should be the same as the sign of \( I(\Phi_1) - I(H_3) \), and this result should hold for any \( p_2 > p_1 \).

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

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 $\overline{G} = \mu (H_1) + \mu (H_2)$ while in the second case $\overline{G} = \mu (H_1)$.

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

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

In fact for (iia) we have that as $\overline{G}$ 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 $\Phi_1$ changes we have that also $-\lambda$ increases. Given the construction of $\Phi_1$ this will not be the case if $p_1 > 1/2$.

For (iiib) 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 $\Phi_1$, we can either have $p_1 < 1/2$, or $p_1 > 1/2$, and therefore both (iia) and (iiib) are admissible cases.

Suppose we take $p_1 < 1/2$.

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

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

Going back now to the case where $I(\Phi_1) - I(H_2) = I(\Phi_1) - I(H_3) = 0$. 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 increase, then $p_1$ changes and accordingly also $\Phi_1$ changes, then $I(\Phi_1)$ is modified and as $H_2$ and $H_3$ are not affected then the sign of $I(\Phi_1) - I(H_2)$ and $I(\Phi_1) - I(H_3)$ changes leading either to $t_3 = t_2 = 1$ or $t_3 = t_2 = 0$. Thus, the solutions where tax rates do not take the extreme values are admissible only for a limited cases or specific revenue values.

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.

Before summarizing the results we make the following remark that is motivated by the fact that cases (iia) and (iiib) holds only if the revenue requirement is sufficiently high. In fact for case (iia) we have $p_1 < 1/2$, and $t_3 = t_2 = 1$ and $t_1 = 0$, and for
(iiib) we have $p_1 > 1/2$ and $t_3 = t_2 = 0$ and $t_1 = 1$. 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 and $H^-$ 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 $H^+$ denote the distribution whose quantile function is

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

with associated averages $\mu(H^-)$ and $\mu(H^+)$ such that by construction their sum coincides with the overall gross income average $\mu(H^-) + \mu(H^+) = \mu(F)$. For many real world distributions $\mu(H^-) > \mu(H^+)$

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

We can now summarize the results in the next proposition.

**Proposition 3** The solution of the optimal taxation problem with fixed labour supply for tax schedules in $\mathcal{I}_{\bar{\tau}}$ 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 $\bar{G} = \bar{\mu}(H_2)$ with

$$t_1 = t_3 = 0, \quad t_2 = \bar{\tau},$$

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

(ii) If $\bar{G} > \bar{\tau} \mu(H^+)$ solution (i) should be compared with $p_1 < 1/2$, where $\bar{G} = \bar{\tau} [\mu(H_2) + \mu(H_3)]$ with

$$t_1 = 0, \quad t_2 = t_3 = \bar{\tau}.$$

(iii) If $\bar{G} > \bar{\tau} \mu(H^-)$ solution (i) should be compared with $p_1 > 1/2$, where $\bar{G} = \bar{\tau} \mu(H_1)$ with

$$t_1 = \bar{\tau}, \quad t_2 = t_3 = 0.$$

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

The proposition highlights the fact that under standard revenue requirements the marginal tax rate is maximal within the central bracket that includes the median income, while for very large revenue requirement maximal marginal tax rates are applied in the tail brackets however note that solution (iiib) involve also 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.
5 Simulation results

In this section we provide numerical results to show the differences in terms of tax rates between inequality and polarization concerns. We consider a government which maximizes a rank-dependent social evaluation function defined over individuals’ disposable income subject to a revenue requirement constraint. There is no redistribution of the collected revenues. Individuals are endowed with a productivity level formalized by the exogenous wage. To simplify we assume that labor supply is fixed and equal to one, so individuals’ gross labor incomes and wages coincide. Following Apps et al. (2014) wages are simulated from a Pareto distribution truncated below and above. In particular, we simulate a Pareto distribution with mean 48.07 and Pareto index 1.4, while the lowest and highest income are equal to 20 and 327 respectively, where the highest limit correspond to the 98th percentile of the income distribution.\footnote{This distribution correspond to the case 1a in Apps et al. (2014).} The median of this distribution is 32.37.

As in Apps et al. (2014) we construct the wage distribution as follows: first, we take 1000000 random draws from a Pareto distribution with the given parameters. Second, we truncate the distribution and the values ranked in increasing order are arranged into 1000 equally sized blocks. Finally, for each of these blocks we compute the mean and the vector of these 1000 means represent the wages distribution used for the simulation.

The tax schedule is piecewise linear, with three income brackets and three marginal tax rates. We define the income brackets as a fraction of the median income ($M$), while the three marginal tax rates are bounded between zero and 50%. For each non-welfarist objectives we consider four level of required per capita revenue required ($G$), expressed as a fraction of the mean income ($\mu$). The next table show the results of the tax schedule for the inequality reduction. The last column gives the level of the social welfare ($W$).

<table>
<thead>
<tr>
<th>$G$</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.1 \times \mu$</td>
<td>0</td>
<td>0.15</td>
<td>0.49</td>
<td>64.05</td>
<td>66.64</td>
<td>29.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.98 $\times M$</td>
<td>2.06 $\times M$</td>
<td>82%</td>
</tr>
<tr>
<td>$0.15 \times \mu$</td>
<td>0</td>
<td>0.3</td>
<td>0.49</td>
<td>44.67</td>
<td>45.32</td>
<td>29.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.38 $\times M$</td>
<td>1.4 $\times M$</td>
<td>68.9%</td>
</tr>
<tr>
<td>$0.2 \times \mu$</td>
<td>0</td>
<td>0.35</td>
<td>0.49</td>
<td>32.37</td>
<td>33.02</td>
<td>28.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 $\times M$</td>
<td>1.02 $\times M$</td>
<td>50%</td>
</tr>
<tr>
<td>$0.25 \times \mu$</td>
<td>0</td>
<td>0.5</td>
<td>0.49</td>
<td>24.60</td>
<td>24.60</td>
<td>26.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.76 $\times M$</td>
<td>0.76 $\times M$</td>
<td>25.6%</td>
</tr>
</tbody>
</table>
For each income threshold $y_i$ with $i = 1, 2$, we provide three values: the monetary value (first row), the fraction in terms of the median income (second row) and the fraction of population with wage lower than that threshold (third row).

For polarization reduction results are reported in the following table:

<table>
<thead>
<tr>
<th>$G$</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.1 \times \mu$</td>
<td>0 0.5 0.01</td>
<td>29.12</td>
<td>0.9 $\times M$</td>
<td>55.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41.7%</td>
<td></td>
<td>39.74</td>
</tr>
<tr>
<td>$0.15 \times \mu$</td>
<td>0 0.5 0.04</td>
<td>64.72</td>
<td>0.8 $\times M$</td>
<td>2 $\times M$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30.9%</td>
<td></td>
<td>37.85</td>
</tr>
<tr>
<td>$0.2 \times \mu$</td>
<td>0 0.5 0.18</td>
<td>90.60</td>
<td>0.8 $\times M$</td>
<td>2.8 $\times M$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30.9%</td>
<td></td>
<td>35.69</td>
</tr>
<tr>
<td>$0.25 \times \mu$</td>
<td>0 0.02 0.49</td>
<td>24.27</td>
<td>0.7 $\times M$</td>
<td>0.75 $\times M$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16%</td>
<td></td>
<td>33.47</td>
</tr>
</tbody>
</table>

As we can see the two tax systems differ qualitatively. When we consider the reduction of income inequality there is no taxation for the bottom of the distribution, while individuals above the first income brackets are taxed. In particular in the highest bracket tax burden is very close to the maximal admissible. When the required revenues increase more individual are taxed, the first income threshold goes down. In other words, to reduce income inequality the required revenue should be collected from individuals below a given threshold. However, we have to note that the second income bracket is very small, this results depends on the structure of the grid used for the simulation algorithm. In particular, we define the grid for the tax rate with a stepsize of 1%, while for income thresholds the stepsize is equal to 0.2 times the median. Except for the first row the difference between the two income thresholds is just one step, this means that qualitatively numerical results confirm the theoretical proposition about inequality concerns.

When we move to polarization concern results change. In this case the optimal tax schedule is such that the segment of the income distribution within the two income thresholds is taxed at the maximal admissible tax rate. There is a no taxation area in the bottom of the distribution, while the top is taxed at a tax rate lower than that one levied on the second bracket. When the required revenues increase the solution requires to move the two thresholds toward the ends of the distribution, compare the first two rows of the second table. More specifically, when the required revenue increases from 10% to 15% of the mean income, the fraction of individuals within the second brackets widens from 36% to 52%. It is also interesting to highlight two aspects. First, for all the required revenues the optimal tax schedule requires to tax the median of the distribution. This means that in order to reduce polarization all
incomes between the two income thresholds are pushed down near the first income threshold, in this way it is possible to create a middle class. Second, when the government wants to collect a high amount of tax revenues the optimal tax schedule becomes quite similar to that one obtained for inequality reduction case, check the fourth row of both tables.

6 Conclusions

In this paper we have computed the optimal tax income schedule within a non-welfarist framework. We have considered a piecewise linear income tax system and we have derived the optimal tax schedule for the reduction of income inequality and polarization. Results have revealed that the two redistributive objectives lead to two different patterns for tax rates. In particular, the reduction of inequality requires a heavy taxation for the top of the distribution and no taxation for the bottom. While polarization concerns involve the maximal admissible taxation on the middle of the distribution and no taxation for the ends. We also use numerical simulations to quantitatively assess these differences. Numerical results substantially confirm the theoretical results, moreover they show that when the government wants to collect a high amount of tax revenues the two tax schedules become very similar.

Appendix

Recall the SEF constrained optimization problem where

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

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

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

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

$$\frac{\partial T(x)}{\partial t_1} = \min \{ x, y_1 \} ,$$

$$\frac{\partial T(x)}{\partial t_2} = \begin{cases} 0 & \text{if } x \leq y_1 \\ x - y_1 & \text{if } y_1 < x \leq y_2 \\ y_2 - y_1 & \text{if } x > y_2 \end{cases}$$
and
\[
\frac{\partial T(x)}{\partial t_3} = \max \{x - y_2, 0\}.
\]

Hence the partial derivatives with respect the three tax rates \(t_i\) are respectively
\[
\frac{\partial L}{\partial t_1} = -\int_0^{p_1} v(p) x(p) \, dp - \int_{p_1}^{1} v(p) y_1 \, dp - \lambda \left[ \int_0^{p_1} x(p) \, dp + \int_{p_1}^{1} y_1 \, dp \right]
\]
(17)
\[
\frac{\partial L}{\partial t_2} = -\int_{p_1}^{p_2} v(p) [x(p) - y_1] \, dp - \int_{p_2}^{1} v(p) [y_2 - y_1] \, dp
\]
- \(\lambda \left[ \int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^{1} (y_2 - y_1) \, dp \right]
\]
(18)
or, after rearranging,
\[
\frac{\partial L}{\partial t_2} = -\int_{p_1}^{1} v(p) \min \{x(p), y_2\} \, dp + \int_{p_1}^{1} v(p) y_1 \, dp - \lambda \left[ \int_{p_1}^{1} \min \{x(p), y_2\} \, dp - \int_{p_1}^{1} y_1 \, dp \right],
\]
and
\[
\frac{\partial L}{\partial t_3} = -\int_{p_2}^{1} v(p) [x(p) - y_2] \, dp - \lambda \int_{p_2}^{1} [x(p) - y_2] \, dp.
\]
(19)
The two FOCs with respect the income thresholds \(y_1\) and \(y_2\) are:
\[
\frac{\partial L}{\partial y_1} = -\int_0^{1} v(p) \frac{\partial T(x)}{\partial y_1} \, dp - \lambda \left[ \int_0^{1} T(x) \, dp \right] = 0
\]
and
\[
\frac{\partial L}{\partial y_2} = -\int_0^{1} v(p) \frac{\partial T(x)}{\partial y_2} \, dp - \lambda \left[ \int_0^{1} T(x) \, dp \right] = 0
\]
where the derivatives of the tax function with respect to the income threshold are respectively
\[
\frac{\partial T(x)}{\partial y_1} = \begin{cases} 
0 & \text{if } x \leq y_1 \\
t_1 - t_2 & \text{if } x > y_1 
\end{cases}
\]
and
\[
\frac{\partial T(x)}{\partial y_2} = \begin{cases} 
0 & \text{if } x \leq y_2 \\
t_2 - t_3 & \text{if } x > y_2 
\end{cases}
\]
The two associated FOCs can be then rewritten as
\[
\frac{\partial L}{\partial y_1} = -\int_{p_1}^{1} v(p) [t_1 - t_2] \, dp - \lambda \int_{p_1}^{1} (t_1 - t_2) \, dp = 0
\]
(20)
and
\[
\frac{\partial L}{\partial y_2} = -\int_{p_2}^{1} v(p) [t_2 - t_3] \, dp - \lambda \left[ \int_{p_2}^{1} (t_2 - t_3) \, dp \right] = 0.
\]
(21)
The FOC with respect the Lagrangian multiplier is

\[
\frac{\partial L}{\partial \lambda} = \mathcal{G} - \int_0^1 T(x(p)) \, dp = 0.
\] (22)

In order to identify the solution of the problem we derive three values of \( \lambda_i^* \) which set (17), (18) and (19) equal to zero. By comparing these three values we will show that at most only for one tax rate it is possible to set \( \frac{\partial L}{\partial \lambda_i} = 0 \). The subscript of \( \lambda_i^* \) denotes the associated tax rate, we have therefore

\[
\lambda_1^* = -\frac{\int_0^{p_1} v(p) x(p) \, dp + \int_{p_1}^{p_2} v(p) y_1 \, dp}{\int_0^{p_1} x(p) \, dp + \int_{p_1}^{p_2} y_1 \, dp},
\]

\[
\lambda_2^* = -\frac{\int_{p_1}^{p_2} v(p) [x(p) - y_1] \, dp + \int_{p_2}^{1} v(p) [y_2 - y_1] \, dp}{\int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^{1} (y_2 - y_1) \, dp},
\]

and

\[
\lambda_3^* = -\frac{\int_{p_2}^{1} v(p) [x(p) - y_2] \, dp}{\int_{p_2}^{1} [x(p) - y_2] \, dp}.
\]

We show that each \( \lambda_i^* \) takes non-positive values. Moreover, we can rewrite the three values of \( \lambda_i^* \) as follows

\[
\lambda_1^* = -\int_0^1 v(p) g_1(p) \, dp
\]

\[
\lambda_2^* = -\int_0^1 v(p) g_2(p) \, dp
\]

\[
\lambda_3^* = -\int_0^1 v(p) g_3(p) \, dp
\]

where

\[
g_1(p) = \min \{ x(p), y_1 \} \frac{\int_0^{p_2} x(p) \, dp + \int_{p_2}^{1} y_1 \, dp}{\int_{p_1}^{p_2} x(p) \, dp + \int_{p_1}^{p_2} y_1 \, dp}
\]

\[
g_2(p) = \max \{ x(p) - y_1, 0 \} \frac{\int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^{1} (y_2 - y_1) \, dp}{\int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^{1} (y_2 - y_1) \, dp}
\]

\[
g_3(p) = \max \{ x(p) - y_2, 0 \} \frac{\int_{p_2}^{1} [x(p) - y_2] \, dp}{\int_{p_2}^{1} [x(p) - y_2] \, dp}
\]
knowing that \( \int_0^1 v(p)\,dp = 1 \), and verifying that by construction \( \int_0^1 g_i(p)\,dp = 1 \), see the next figure.

Integrating by parts we have that

\[
\lambda_i^* = - \left( \left[ v(p) \int_0^p g_i(t)\,dt \right]_0^1 - \int_0^1 \left( v'(p) \int_0^p g_i(t)\,dt \right)\,dp \right),
\]

that is

\[
\lambda_i^* = -v(1) + \int_0^1 \left( v'(p) \int_0^p g_i(t)\,dt \right)\,dp
\]

which we can rewrite as

\[
\lambda_i^* = \int_0^1 v'(p) G_i(p)\,dp - v(1),
\]

where \( G_i(p) = \int_0^p g_i(t)\,dt \). In order to rank the three different \( \lambda_i^* \)'s we have to sign the following difference

\[
\lambda_i^* - \lambda_j^* = \int_0^1 v'(p) [G_i(p) - G_j(p)]\,dp
\]

for all \( i,j \in \{1,2,3\} \) with \( i < j \). In general if \( g_i(p) \) and \( g_j(p) \) do not coincide and
If \( i < j \), there exists an interval \([p_A^*, p_B^*]\) where \( p_A^* \leq p_B^* \) such that:

\[
\begin{align*}
g_i (p) &> g_j (p) \quad \forall \quad p < p_A^* \\
g_i (p) &< g_j (p) \quad \forall \quad p > p_B^* \\
g_i (p) &= g_j (p) \quad \forall \quad p \in [p_A^*, p_B^*]
\end{align*}
\]

In our case the functions \( g_i (p) \) for \( i \in \{1, 2, 3\} \) exhibit two properties. First, they are non-decreasing in \( p \). Second, \( g_i (p) \) and \( g_j (p) \) for \( i < j \) are single crossing curves, that is:

\[
\begin{align*}
g_i (p) &\geq g_j (p) \quad \forall \quad p \leq p^* \\
g_i (p) &\leq g_j (p) \quad \forall \quad p \geq p^*
\end{align*}
\]

for all \( i, j \in \{1, 2, 3\} \), and \( i < j \). Then, it follows that \( G_i (p) > G_j (p) \quad \forall \quad p \in (0, 1) \), therefore the following relationship holds

\[
G_1 (p) \geq G_2 (p) \geq G_3 (p) \quad \forall p \in [0, 1].
\]

See next figure

![Graph showing relationship between G1, G2, and G3]

implying that

\[
|\lambda_1^*| \geq |\lambda_2^*| \geq |\lambda_3^*|.
\]

This relationship is true for all the weighting function decreasing in \( p \) \( \left( v' (p) < 0 \right) \), including the Gini weighting function which is a particular case of decreasing and linear weighting functions. Because the Lagrangian multipliers \( \lambda_i^* \) are a measure of the marginal impact on the SEF of an increase in a tax rate, and given the previous

---

\(^8\)We have to note that the two functions \( g_i (p) \) and \( g_j (p) \) can coincide when income distribution is discrete and there is a group of incomes located exactly at \( p_1 \) and another one above \( p_2 \). In other words, functions \( g_i (p) \) and \( g_j (p) \) share same traits if there are no incomes lower than \( y_1 \) and between \( y_1 \) and \( y_2 \).
relationship among them, it follows that with weighting function decreasing in $p$, the tax system will be as follows $0 \leq t_1 \leq t_2 \leq t_3 \leq 1$.\footnote{If we consider the Gini weighting function $v(p) = 2(1-p)$, we note that $v(0) = 2$ and $v(1) = 0$, while $V(0) = 0$ and $V(1) = 1$, then the generic $\lambda^*_i$ is $\lambda^*_i = \int_0^1 \left( v'(p) \int_0^p g_i(t) \, dt \right) \, dp = -2 \int_0^1 \int_0^p g_i(t) \, dt \, dp < 0$ because $v'(p) = -2 < 0$. By replacing $G_i(p) = \int_0^p g_i(t) \, dt$ we obtain $\lambda^*_i = -2 \int_0^1 G_i(p) \, dp.$}

Given that at most for only one tax rate it is possible to have $\frac{\partial C}{\partial t_i} = 0$, we consider the three cases when $\frac{\partial C}{\partial t_i} = 0$, and then we check the sign of the two remaining $\frac{\partial C}{\partial t_j}$ with $i, j \in \{1, 2, 3\}$, and $i \neq j$. In all the following cases we consider the Gini weighting function.

**Case I:** $\frac{\partial C}{\partial t_1} = 0 \implies t_1 \in (0, 1)$

We assume that the FOC with respect to $t_1$ is equal to zero and hence that there is an interior solution for $t_1$. Then, we derive the value of $\lambda^*_1$ which sets the FOC equal to zero and we replace this value into the other two FOCs with respect to $t_2$ and $t_3$. From the sign of these two equations we can derive restrictions on the values of $t_2$ and $t_3$. The level of $\lambda^*_1$ such that $\frac{\partial C}{\partial t_1} = 0$ is

$$\lambda^*_1 = -\frac{\int_{p_1}^{p_2} v(p) x(p) \, dp + \int_{p_1}^1 v(p) y_1 \, dp}{\int_{p_1}^{p_2} x(p) \, dp + \int_{p_1}^1 y_1 \, dp} = \int_0^1 v'(p) G_1(p) \, dp,$$

by substituting this value into (18) we obtain

$$\frac{\partial L}{\partial t_2} = -\int_{p_1}^{p_2} v(p) [x(p) - y_1] \, dp - \int_{p_2}^1 v(p) [y_2 - y_1] \, dp
- \left( \int_0^1 v'(p) G_1(p) \, dp \right) \left[ \int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^1 (y_2 - y_1) \, dp \right]$$

which we can rewrite as

$$\frac{\partial L}{\partial t_2} = \frac{-\int_{p_1}^{p_2} v(p) [x(p) - y_1] \, dp + \int_{p_2}^1 v(p) [y_2 - y_1] \, dp}{\int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^1 (y_2 - y_1) \, dp} = \frac{-\int_{p_1}^{p_2} v(p) [x(p) - y_1] \, dp + \int_{p_2}^1 v(p) [y_2 - y_1] \, dp}{\int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^1 (y_2 - y_1) \, dp}$$

$$- \int_0^1 v'(p) G_1(p) \, dp.$$
From the previous section we know that the first term on the right hand side is 
\[ \int_{0}^{1} v'(p) G_2(p) \, dp \]. Then, we can conclude that given that \( \frac{\partial C}{\partial t_2} \) is proportional to
\[ \int_{0}^{1} v'(p) G_2(p) \, dp - \int_{0}^{1} v'(p) G_1(p) \, dp = - \int_{0}^{1} v'(p) [G_1(p) - G_2(p)] \, dp, \]
given that \( G_1(p) \geq G_2(p) \) [with strict inequality > holding for an interval of values of \( p \)] and \( v'(p) < 0 \), we have that the previous equation is positive, that is \( \frac{\partial C}{\partial t_2} > 0 \), hence \( t_2 \) has to be equal to 1. When we replace \( \lambda^*_1 \) into (6!!) we similarly obtain that \( \frac{\partial C}{\partial t_3} \) is proportional to
\[ \frac{- \int_{p_2}^{1} v(p) [x(p) - y_2] \, dp}{\int_{p_2}^{1} [x(p) - y_2] \, dp} - \int_{0}^{1} v'(p) G_1(p) \, dp \]
where the first term corresponds to \( \int_{0}^{1} v'(p) G_3(p) \, dp \) and given that \( G_1(p) \geq G_3(p) \) [with strict inequality > holding for an interval of values of \( p \)] it follows that \( \frac{\partial C}{\partial t_3} > 0 \), thereby requiring \( t_3 = 1 \). Now we replace the three tax rates \( t_1 \in (0, 1) \) and \( t_2 = t_3 = 1 \) into the two FOCs with respect to the two income thresholds and we obtain from (7!!)
\[ \lambda = - \frac{1 - V(p_1)}{1 - p_1}. \]
The next step is to check if this value sets equation (4!!) equal to zero, that is if
\[ - \int_{0}^{p_1} v(p) x(p) \, dp - \int_{p_1}^{1} v(p) y_1 \, dp + \left( \frac{1 - V(p_1)}{1 - p_1} \right) \left[ \int_{0}^{p_1} x(p) \, dp + \int_{p_1}^{1} y_1 \, dp \right] = 0 \]
which reduces to
\[ - \int_{0}^{p_1} v(p) x(p) \, dp + \left( \frac{1 - V(p_1)}{1 - p_1} \right) \int_{0}^{p_1} x(p) \, dp = 0 \]
and
\[ \left[ 1 - V(p_1) \right] \left[ - \int_{0}^{p_1} v(p) x(p) \, dp + \frac{\int_{p_1}^{1} v(p) \, dp}{\int_{p_1}^{1} x(p) \, dp} \right] = 0 \]
which can not be true because the two terms into the square brackets are positive and negative respectively. So given this contradiction we can exclude the case \( \frac{\partial C}{\partial t_1} = 0 \mapsto t_1 \in (0, 1) \).

**Case II:** \( \frac{\partial C}{\partial t_2} = 0 \mapsto t_2 \in (0, 1) \)

In this case, we are assuming that the FOC with respect to \( t_2 \) is equal to zero and then that \( t_2 \in (0, 1) \). Therefore, from (5??) we derive the value of \( \lambda \) satisfying our
assumption, that is
\[ \hat{\lambda}_2 = - \frac{\int_{p_1}^{p_2} v(p) [x(p) - y_1] dp + \int_{p_1}^{p_2} v(p) [y_2 - y_1] dp}{\int_{p_1}^{p_2} [x(p) - y_1] dp + \int_{p_2}^{p_1} (y_2 - y_1) dp} = \int_0^1 v'(p) G_2(p). \]

By replacing this value into (4??) we obtain that \( \frac{\partial c}{\partial t_1} \) is proportional to
\[ - \frac{\int_{p_1}^{p_1} v(p) x(p) dp + \int_{p_1}^{p_1} v(p) y_1 dp}{\int_{p_1}^{p_1} x(p) dp + \int_{p_1}^{p_1} y_1 dp} - \int_0^1 v'(p) G_2(p) \]
where the first term is \( \lambda_1 \), hence we have
\[ \int_0^1 v'(p) G_1(p) - \int_0^1 v'(p) G_2(p) < 0 \]
given that \( G_1(p) \geq G_2(p) \) and \( v'(p) < 0 \), then \( t_1 = 0 \). Now, by replacing \( \hat{\lambda}_2 \) into (6??) we have that \( \frac{\partial c}{\partial t_2} \) is proportional to
\[ - \frac{\int_{p_2}^{p_2} v(p) [x(p) - y_2] dp}{\int_{p_2}^{p_2} [x(p) - y_2] dp} - \int_0^1 v'(p) G_2(p) \]
and
\[ \int_0^1 v'(p) G_3(p) - \int_0^1 v'(p) G_2(p) > 0 \]
given that \( G_2(p) \geq G_3(p) \) and \( v'(p) < 0 \), then \( t_3 = 1 \). Then, we replace the three tax rates \( t_1 = 0, t_2 \in (0, 1) \) and \( t_3 = 1 \) into the two FOCs with respect to the income thresholds and we obtain from (7)
\[ t_2 \left[ \int_{p_1}^{p_1} v(p) dp + \lambda \int_{p_2}^{p_2} dp \right] = 0 \]
which is true if
\[ \lambda = - \frac{1 - V'(p_1)}{1 - p_1} < 0 \]
and from (8)
\[ (1 - t_2) \left[ \int_{p_2}^{p_2} v(p) + \lambda \int_{p_2}^{p_2} dp \right] = 0 \]
requiring that
\[ \lambda = - \frac{1 - V'(p_2)}{1 - p_2} < 0 \]
however given \( p_1 < p_2 \) the two values of \( \lambda \) are different, hence we can exclude the triplet \( (t_1 = 0, t_2 \in (0, 1) \) and \( t_3 = 1 \) from the solution of the tax problem.
Case III: $\frac{\partial C}{\partial t_3} = 0 \implies t_3 \in (0, 1)$

If the FOC with respect to $t_3$ is equal to zero, this tax rate is a value between zero and one. Then, from (6) we have that the value of $\lambda$ which sets this condition equal to zero is

$$\tilde{\lambda}_3 = -\frac{\int_{p_2}^1 v(p) [x(p) - y_2] dp}{\int_{p_2}^1 [x(p) - y_2] dp} = \int_0^1 v'(p) G_3(p)$$

by replacing this value into (4) and (5) we have that both these two equations are negative, therefore the tax system is such that $t_1 = t_2 = 0$ and $t_3 \in (0, 1)$. From (8) we have that

$$t_3 \left[ \int_{p_2}^1 v(p) dp + \lambda \int_{p_2}^1 dp \right] = 0$$

which is true if the term into the square brackets is equal to zero, hence if

$$\lambda = -\frac{1 - V(p_2)}{1 - p_2} < 0$$

then, we have to check if with this value of $\lambda$ equation (6) is equal to zero,

$$-\int_{p_2}^1 v(p) [x(p) - y_2] dp + \frac{1 - V(p_2)}{1 - p_2} \int_{p_2}^1 [x(p) - y_2] dp = 0$$

and

$$[1 - V(p_2)] \left[ -\frac{\int_{p_2}^1 v(p) x(p) dp}{\int_{p_2}^1 v(p) dp} + \frac{\int_{p_2}^1 x(p) dp}{\int_{p_2}^1 dp} \right] = 0$$

which is true if one of the two terms into square brackets is equal to zero. Specifically, if $V(p_2) = 1$ it means that there is only one group, hence no income is taxed with $t_3$, then given that $t_1 = t_2 = 0$, the government budget constraint does not hold, so we can conclude that this term can not be equal to zero. However, with a weighting function where weights are decreasing ($v'(p) < 0$) the term into the second square brackets is always positive, so the FOC with respect to $t_3$ can not be equal to zero.

Therefore, we can conclude that there are no interior solutions for the optimal tax system, so we have to look for a corner solution. Given that we are considering a tax system with three tax rates, we have to consider eight possible triplets of $t_1, t_2$ and $t_3$. From this set we can remove the case where all the tax rates are equal to zero because the budget constraint is not satisfied. Then, for each possible solution we replace the three values of the tax rates into the two FOCs with respect to the income threshold, and then, if there are no contradictions, we derive the two optimal income thresholds. Finally, among the subset of the possible solutions we choose the one which maximizes social welfare.

We start by considering the case where only the top of the income distribution is taxed at a confiscatory tax rate, while the remaining part is not taxed.
Case A: \( t_1 = t_2 = 0 \) and \( t_3 = 1 \)

If we replace this triplet of tax rates into the two FOCs with respect to the income thresholds we obtain from (7) that \( p_1 \) and hence \( y_1 \) could be whatever, while from (8) we obtain

\[
\tilde{\lambda} = -\frac{1 - V(p_2)}{1 - p_2} < 0.
\]

Now, given this value of \( \tilde{\lambda} \) we have to check if the FOCs with respect to the three tax rates are coherent with the tax system considered, specifically we require that the (4) and (5) are negative, while (6) has to be positive. By replacing this \( \tilde{\lambda} \) into (6) we have

\[
- \int_{p_2}^{1} v(p) [x(p) - y_2] dp + \frac{1 - V(p_2)}{1 - p_2} \int_{p_2}^{1} [x(p) - y_2] dp
\]

and

\[
[1 - V(p_2)] \left[ - \int_{p_2}^{1} v(p) \frac{x(p) dp}{1 - V(p_2)} + \frac{\int_{p_2}^{1} x(p) dp}{1 - p_2} \right] > 0
\]

which is positive because the weighting function \( v(p) \) is decreasing in \( p \), while \( x(p) \) is increasing. Then, by replacing \( \tilde{\lambda} \) into (4) we obtain

\[
- \int_{0}^{p_1} v(p) x(p) dp - \int_{p_1}^{1} v(p) y_1 dp + \frac{1 - V(p_2)}{1 - p_2} \left[ \int_{0}^{p_1} x(p) dp + \int_{p_1}^{1} y_1 dp \right]
\]

which we can rewrite as

\[
- \int_{0}^{p_1} v(p) x(p) dp - (1 - V(p_1)) y_1 + \frac{1 - V(p_2)}{1 - p_2} \left[ \int_{0}^{p_1} x(p) dp + y_1 \left( \frac{1 - p_1}{1 - p_2} \right) (1 - V(p_2)) \right]
\]

and

\[
[1 - V(p_2)] \begin{cases} 
\left[ \frac{\int_{0}^{p_1} v(p) x(p) dp}{\int_{p_2}^{1} v(p) dp} + \frac{\int_{0}^{p_1} x(p) dp}{\int_{p_2}^{1} dp} \right] < 0 \\
\left[ \frac{1 - V(p_1)}{1 - V(p_2)} - \frac{1 - p_1}{1 - p_2} \right] > 1 \\
\left[ \frac{y_1}{1 - V(p_2)} \right] > 1 
\end{cases}
\]
which is absolutely negative if the term into the second square brackets is positive, that is if
\[
\frac{1 - V(p_1)}{1 - p_1} > \frac{1 - V(p_2)}{1 - p_2}
\]
which is true given that \( p_2 > p_1 \).

As to equation (5) we have
\[
- \int_{p_1}^{p_2} v(p) [x(p) - y_1] \, dp - \int_{p_2}^{1} v(p) [y_2 - y_1] \, dp
+ \left[ \frac{1 - V(p_2)}{1 - p_2} \right] \left[ \int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^{1} (y_2 - y_1) \, dp \right]
\]
and
\[
[1 - V(p_2)] \left[ \frac{- \int_{p_1}^{p_2} v(p) x(p) \, dp}{\int_{p_2}^{1} v(p) \, dp} + \frac{\int_{p_1}^{p_2} x(p) \, dp}{\int_{p_2}^{1} dp} \right]
+ y_1 \left[ \frac{(V(p_2) - V(p_1)) \, dp}{>0} - \frac{1}{1 - p_2} (p_2 - p_1) \right]_{>0}
\]
which has to be negative. This condition holds only if the term into the second square brackets is negative, because both the other two are positive.\(^{10}\) Finally, if the signs of the three FOCs with respect to the tax rates are those required, then from the government budget constraint we derive the level of \( p_2 \) such that
\[
\int_{p_2}^{1} (x(p) - y_2) \, dp = \bar{G}
\]
and
\[
\int_{p_2}^{1} x(p) \, dp - (1 - p_2) y_2 = \bar{G}.
\]

**Case B: \( t_1 = 0 \) and \( t_2 = t_3 = 1 \)**

If we replace this triplet into (7) and (8) we obtain from (7) that
\[
t_2 \left[ \int_{p_1}^{1} v(p) \, dp + \lambda \int_{p_1}^{1} dp \right] = 0
\]

\(^{10}\) The term into the third square brackets is positive, that is \( \frac{V_2 - V_1}{1 - p_2} > \frac{p_2 - p_1}{1 - p_2} \). Given the Gini weighting function \( v(p) = 2(1 - p) \), we have that \( V(p) = \int_{0}^{p} v(p) \, dp = 1 - (1 - p)^2 \), then the previous condition reduces to \( \frac{(1 - p_1)^2 - (1 - p_2)^2}{(1 - p_1)^2} > \frac{p_2 - p_1}{1 - p_2} \), which can be simplified as \( 1 - p_2 < \frac{p_2(2 - p_2) - p_1(2 - p_1)}{(p_2 - p_1)} \) and \( p_1 (p_2 - p_1) < (p_2 - p_1) \), which is true given that \( p_1 < p_2 < 1 \).
which is true if
\[ \lambda^{**} = \frac{1 - V(p_1)}{1 - p_1} < 0 \]

while the only requirement for \( p_2 \) is that it is greater than \( p_1 \). Now, by replacing \( \lambda^{**} \) into (5) we have

\[
\begin{align*}
&- \int_{p_1}^{p_2} v(p) [x(p) - y_1] \, dp - \int_{p_2}^{1} v(p) [y_2 - y_1] \, dp \\
&+ \left[ \frac{1 - V(p_1)}{1 - p_1} \right] \left[ \int_{p_1}^{p_2} [x(p) - y_1] \, dp + \int_{p_2}^{1} (y_2 - y_1) \, dp \right]
\end{align*}
\]

which we can rewrite as

\[
[1 - V(p_1)] \left[ - \frac{\int_{p_1}^{p_2} v(p) [x(p) - y_1] \, dp}{\int_{p_1}^{1} v(p) \, dp} + \frac{\int_{p_1}^{p_2} [x(p) - y_1] \, dp}{\int_{p_1}^{1} \, dp} \right]
- (y_2 - y_1) \left[ 1 - V(p_2) - \left( \frac{1 - p_2}{1 - p_1} \right) (1 - V(p_1)) \right]
\]

this equation has to be positive and to this regard we can say that the term into the third square brackets is negative if \( \frac{1 - V(p_1)}{1 - p_1} > \frac{1 - V(p_2)}{1 - p_2} \). From (6) we obtain

\[
[1 - V(p_1)] \left[ - \frac{\int_{p_1}^{1} v(p) [x(p) - y_2] \, dp}{\int_{p_1}^{1} v(p) \, dp} + \frac{\int_{p_1}^{p_2} [x(p) - y_2] \, dp}{\int_{p_1}^{1} \, dp} \right] > 0.
\]

Finally, by replacing \( \lambda^{**} \) into (4) we have

\[
- \int_{0}^{p_1} v(p) \, x(p) \, dp - \int_{p_1}^{1} v(p) \, y_1 \, dp + \left( \frac{1 - V(p_1)}{1 - p_1} \right) \left[ \int_{0}^{p_1} x(p) \, dp + \int_{p_1}^{1} y_1 \, dp \right]
\]

and

\[
[1 - V(p_1)] \left[ - \frac{\int_{0}^{p_1} v(p) \, x(p) \, dp}{\int_{p_1}^{1} v(p) \, dp} + \frac{\int_{0}^{p_1} x(p) \, dp}{\int_{p_1}^{1} \, dp} \right] < 0
\]

which is coherent with \( t_1 = 0 \). Then, from the government budget constraint we derive the optimal \( p_1 \) which solves the tax problem,

\[
\int_{p_1}^{1} (x(p) - y_1) \, dp = \mathcal{G}
\]

and

\[
\int_{p_1}^{1} x(p) \, dp - (1 - p_1) y_1 = \mathcal{G}.
\]
Case C: $t_1 = t_3 = 0$ and $t_2 = 1$

In this case we consider a tax system where the ends of the income distribution are no taxed, while the middle is taxed with $t_2 = 1$. By replacing this tax system into (7) and (8) we obtain respectively

$$t_2 \left[ \int_{p_1}^{1} v(p) \, dp + \lambda \int_{p_1}^{1} dp \right] = 0$$

requiring

$$\bar{\lambda} = -\frac{1 - V(p_1)}{1 - p_1}$$

and

$$-t_2 \left[ \int_{p_2}^{1} v(p) \, dp + \lambda \int_{p_2}^{1} dp \right] = 0$$

which is true if

$$\lambda = -\frac{1 - V(p_2)}{1 - p_2}.$$ 

However, given that $p_1 < p_2$ the two values $\bar{\lambda}$ and $\lambda$ are different, so given this contradiction we have to exclude this triplet.

Case D: $t_1 = 1$ and $t_2 = t_3 = 0$

This tax system is such that only the bottom of the income distribution, that is all incomes below $y_1$ are taxed. From (7) we have

$$-t_1 \left[ \int_{p_1}^{1} v(p) \, dp + \lambda \int_{p_1}^{1} dp \right] = 0$$

which requires

$$\lambda = -\frac{1 - V(p_1)}{1 - p_1}.$$ 

By replacing this value into (4) we obtain

$$- \int_{0}^{p_1} v(p) x(p) \, dp - \int_{p_1}^{1} v(p) y_1 \, dp + \left[ \frac{1 - V(p_1)}{1 - p_1} \right] \left[ \int_{0}^{p_1} x(p) \, dp + \int_{p_1}^{1} y_1 \, dp \right]$$

which is equivalent to

$$\left[ 1 - V(p_1) \right] \left[ -\frac{\int_{0}^{p_1} v(p) x(p) \, dp}{\int_{p_1}^{1} v(p) \, dp} + \frac{\int_{0}^{p_1} x(p) \, dp}{\int_{p_1}^{1} dp} \right] < 0$$

which is negative, then according to the Kuhn-Tucker conditions $t_1$ can not be equal to one, it follows that we can exclude this case.
**Case E:** \( t_1 = t_2 = 1 \) and \( t_3 = 0 \)

This tax system is equivalent to the previous one, where all incomes below a given threshold are taxed. From (8) we obtain

\[
-t_2 \left[ \int_{p_2}^{1} v(p) \, dp + \lambda \int_{p_2}^{1} \, dp \right] = 0
\]

and

\[
\lambda = -\frac{1 - V(p_2)}{1 - p_2}.
\]

When we replace the value of \( \lambda \) into (6) we obtain

\[
-\int_{p_2}^{1} v(p) [x(p) - y_2] \, dp + \frac{1 - V(p_2)}{1 - p_2} \int_{p_2}^{1} [x(p) - y_2] \, dp
\]

which is the same condition that we derived in case A. Then, given that we know that this equation is positive, it follows that \( t_3 \) cannot be equal to zero.

**Case F:** \( t_1 = t_3 = 1 \) and \( t_2 = 0 \)

The last triplet considered corresponds to a tax system where only the ends of the income distribution are taxed, while on the middle there is no taxation. In this case from (7) and (8) we have respectively

\[
t_1 \left[ \int_{p_1}^{1} v(p) \, dp + \lambda \int_{p_1}^{1} \, dp \right] = 0
\]

\[
t_3 \left[ \int_{p_2}^{1} v(p) \, dp + \lambda \int_{p_2}^{1} \, dp \right] = 0
\]

which are satisfied with two different values of \( \lambda \). This condition leads to same contradiction which we find for the case C. Therefore, from this analysis we can conclude that for weighting function decreasing in \( p \), in order to satisfy the budget constraint, the optimal tax has to be such that only incomes above a given threshold are taxed at a confiscatory tax rate. The income threshold is obtained from the budget constraint equation.

**References**


