UNIVERSITY OF VERONA

Department of Economics

GRADUATE SCHOOL OF ECONOMICS AND MANAGEMENT (GSEM)

Doctoral Program in Economics and Finance

Cycle XXVIII October 2012 – September 2015

INEQUALITY, POLARIZATION AND REDISTRIBUTIVE POLICIES

S.S.D. *SECS-P/03*

Coordinator:

Prof. *Giam Pietro CIPRIANI*

Supervisor:

Prof. *Claudio ZOLI*

Doctoral Student:

Vincenzo PRETE

Acknowledgements

I sincerely thank all the people have helped and supported me over these last four years. Without their help I would never have been able to complete this thesis.

First and foremost I would like to express my sincere deepest gratitude to my supervisor Prof. Claudio Zoli, whose suggestions, comments, encouragement and help have been crucial for this work. Thank you for everything.

Also I am really indebted and grateful to Prof. Alessandro Sommacal for his helpful suggestions and valuable programming assistance. Thank you for the time, efforts and contributions to this work.

A very special thank goes to Prof. Rolf Aaberge for his insightful and helpful suggestions and comments. Thanks to Prof. Ugo Colombino for his useful comments.

I thank Prof. Eugenio Peluso and Francesco Andreoli for their useful comments and discussions.

I am grateful to Prof. Vittoria Levati for having supported and encouraged me.

Then, it has been a pleasure working with my colleagues, who were always willing to discuss and share their knowledge with me. Thanks to Gianni, Alain, Riccardo, Chiara and Ivan.

Thanks also to Matjia for his support.

Finally, my heartfelt thanks to my parents and to Marica for her continuous support, understanding and encouragement.

Contents

Bibliography 156

Introduction

This dissertation focuses on optimal income taxation theory and analyses how the optimal income tax schedule changes when non-welfarist objectives expressed in terms of *inequality and* polarization reduction are pursued. The structure of the thesis consists of four chapters:

- Chapter 1: Optimal income taxation and non-welfarist objectives. Theoretical vs practical perspectives.
- Chapter 2: Optimal non-welfarist income taxation for inequality and polarization reduction.
- Chapter 3: Optimal redistribution with non-welfarist objectives.
- Chapter 4: Optimal non-welfarist taxation with non-constant labour supply elasticity.

Over the last two decades many countries have experienced a remarkable increase in income inequality and polarization. On the other hand, policy makers have continuously modified the tax systems towards more simplified schemes with few income brackets and lower top tax rates. Recent literature has addressed the question if the increasing trends of income inequality can be explained by these changes.

In this work we investigate how the optimal tax system should be designed in order to achieve inequality and polarization reduction objectives. To this end, in line with recent works by Kanbur et al. (2006) and Saez and Stantcheva (2016) we adopt a non-welfarist approach, moreover we focus on piecewise linear tax systems. By choosing the non-welfarist approach we recognize that redistributive objectives are crucial per se to the determination of the optimal tax schedule, and not necessarily because of the shape of agents' utility functions. In fact, to this regard Kanbur et al. (2006) claim that governments should not evaluate social welfare taking into account only individuals' utility. Individuals preferences indeed, can be manipulated or agents' behaviors could not be socially optimal.

Conventionally the main difference between welfarism and non-welfarism is that in the latter, the argument of the government's social welfare function is different from individuals' utility.

To this end, we focus on incomes as the most appropriate variable to investigate when the government objectives are the reduction of inequality, poverty or polarization.

We formalize the non-welfarist objectives by assuming that the government maximizes a rank-dependent social evaluation function defined over individuals' net incomes, subject to a budget constraint. Then, the evaluation of the income distribution can be summarized by the mean income of the distribution and a linear index of dispersion dependent on the choice of the weighting function. More specifically, we consider two weighting functions which allow to formalize redistributive objectives expressed in terms of changes in the Gini index of incomes in case of inequality considerations. Then, by appropriate modifications of the positional weights it is possible, within the same evaluation model, to shift towards evaluation concerned with the polarization of incomes.

We consider piecewise linear tax systems that represent the most commonly internationally adopted scheme and the easiest way to identify changes in the tax schedule when the government's concerns move from inequality to polarization reduction. The results we obtain make explicit the interlink between the redistributive objective and the theoretical optimal shape of the tax schemes.

Chapter 1 presents a survey of the main results of the optimal taxation literature, distinguishing between the welfarist and the non-welfarist approach.

Chapter 2 represents the core of this dissertation and it aims at investigating the effect of different redistributive objectives on the shape of the optimal tax schedule.¹ In particular, we consider a piecewise linear tax system with three income brackets and two different regime of the tax rates, i.e. convex $(t_1 \le t_2 \le t_3)$ and non-convex $(t_1 \le t_3 \le t_2)$. The goal of the chapter is to identify the socially desirable mechanism collecting a given amount of taxes, given a specific

¹A version of Chapter 2 is also available as: Prete, Sommacal and Zoli (2016), "Optimal Non-Welfarist Income Taxation for Inequality and Polarization Reduction".

non-welfarist redistributive objective, i.e. inequality and polarization reduction.

The optimal tax system is derived both with fixed and with variable labour supply. The most interesting result we obtain is that the redistributive objectives matter as the optimal tax schedule substantially changes depending on whether the government is inequality or polarization sensitive. In particular, with Öxed labour supply (which represents the benchmark case) the optimal tax system reducing income inequality requires a no-tax area until a given threshold and the maximal admissible taxation above that threshold, which is set in order to satisfy the revenue requirement. In other words, our results suggest that to reduce income inequality the solution is to reduce the income distance between incomes within the tax-area and between these incomes and those in the no-tax area. As to polarization reduction, the optimal tax system requires to tax with the maximum admissible tax rate all incomes belonging to a central interval, which includes also the median income. Marginal tax rates within the two external brackets are set equal to zero. Therefore, the solution to reduce polarization requires to reduce the distance between the incomes in the central bracket, in order to create a sort of less disperse middle class. Note that all incomes in the highest bracket are taxed according to a lump-sum taxation, keeping their absolute dispersion unchanged.

By introducing labour supply elasticity, which is assumed to be constant throughout the entire income distribution, the results are qualitatively unaffected. That is, the optimal tax system reducing inequality is convex, unless when labour supply elasticity is high. While the optimal tax system for the reduction of polarization is non-convex with reduced marginal tax rate for the upper income bracket.

Chapters 3 and 4 integrate the analysis of Chapter 2. In particular, the main novelty of Chapter 3 is the introduction of the possibility to use lump-sum transfers (taxation and subsidy). In this chapter we consider a set of piecewise tax systems with two and three income brackets and two regimes of the tax rates, i.e. convex and non-convex. Here, the results we obtain are completely different with respect to those described in Chapter 2 where redistribution is not allowed. More specifically, when labour supply is fixed the solution is independent of the government's redistributive objective. The optimal tax system is based on a proportional taxation with the highest admissible tax rate. The collected amount is then equally redistributed, eventually keeping the share of income needed to cover the revenue requirement.

When we introduce labour supply elasticity the scenario becomes less obvious. In particular, the non-welfarist objectives become again crucial as the optimal tax system changes when the government's focus shifts from inequality to polarization considerations. More specifically, to reduce inequality the optimal two brackets tax system requires a no taxation area below a given threshold, above which taxation is proportional. The threshold and the tax rate are respectively increasing and decreasing in the level of elasticity. As to polarization reduction, the optimal two brackets tax system requires a proportional taxation for all incomes below a given threshold and zero marginal tax rate for all incomes above. Both the threshold and the tax rate are decreasing in the level of labour supply elasticity, moreover when elasticity is high the optimal tax system reducing polarization is based only on lump-sum taxation.

With three income brackets, the optimal tax system for inequality reduction is convex unless when labour supply elasticity is high or the initial level of income inequality is low. For the reduction of polarization the optimal tax system is always non-convex. In both cases the marginal tax rates are decreasing in the level of labour supply elasticity.

The interesting aspect of the results of Chapter 3 is that the design of the optimal tax system is independent of the revenue requirement, and the sign of the lump-sum transfer depends on the difference between the collected amount and the required revenue. Moreover the lump-sum transfer is positive (subside) or negative (tax) depending on the combination of the level of labour supply elasticity and the index of gross incomes dispersion. We also find that lump-sum taxation is more likely to be used to reduce polarization.

Last, in Chapter 4 we extend the analysis of Chapter 2 introducing the hypothesis that labour supply elasticity is non-constant. In line with Aaberge et al. (2013) we assume that labour supply elasticity is decreasing in the level of individuals' wage. The results we obtain here, are qualitatively in line with those presented in Chapter 2. The optimal tax system reducing inequality and polarization are respectively convex and non-convex. The differences with respect to Chapter 2 are related to the magnitude of the marginal tax rates, which are lower for the income percentiles exhibiting higher elasticity. In this last chapter, we also consider a third weighting function taking into account both inequality and polarization concerns. Specifically, this weighting function is obtained as a linear combination of the two weighting functions used for inequality and polarization concerns. Then, the shape of the optimal tax system associated with this particular weighting function depends on the level of pre-tax dispersion. In particular, for high (low) level of pre-tax inequality and polarization the optimal tax system is convex (non-convex).

Chapter 1

Optimal Income Taxation and Non-Welfarist Objectives. Theoretical vs Practical Perspectives

In this work we investigate if normative recommendations of the optimal tax theory represent a useful guide to understand reforms in the personal income tax systems. In other words, we check if theory and practice move along the same directions. This analysis shows that some theoretical results can be easily identifiable in the tax changes of the last twenty-five years. This is the case of the attempts to reduce the complexity and to improve the efficiency of the tax systems. However, other trends of the tax reforms, like the reduction of the top tax rates and of the progressivity of the tax systems, contrast with theoretical prescriptions suggesting, instead, more progressive taxation and higher tax rates when the level of income inequality is high.

1.1 Introduction

Evidence shows that OECD countries collect a remarkable fraction of their GDP in taxes. Over the last Öfty years, the tax to GDP ratio has been slowly increasing in most of the advanced economies, exhibiting an average value of 34.4% in 2014.¹

In addition, the total tax burden is not equally distributed among the different income sources. More specifically, as noted by Piketty and Saez (2012), the tax structure is such that three quarters of taxes fall on labour income, with the personal income tax (PIT) collecting about twenty-Öve percent of the total tax revenues.

	1965	1975	1985	1995	2005	2010	2013
Personal income tax	26	30	30	26	24	24	25
Corporate income tax	9	8	8	8	10	8	8
Social security contributions	18	22	22	25	25	26	26
(employee)	(6)	(7)	(7)	(9)	(9)	(9)	(10)
(employer)	(10)	(14)	(13)	(14)	(14)	(15)	(15)
Payroll taxes							
Property taxes	8	6	٢		6	ς	6
General consumption taxes	12	13	16	19	20	20	20
Specific consumption taxes	24	18	16	13	11	11	10
Other taxes	n	2	2	3	3	3	3
Total	100	100	100	100	100	100	100

Tax structure in OECD area

Source: OECD (2015), "Revenue Statistics: Comparative tables", OECD Tax Statistics (database).

The magnitude of these figures suggests that the tax system has a relevant impact on the welfare of the society. To this regard, the main lesson from the optimal tax theory is that a "good" tax system should achieve its economic and redistributive objectives, by minimizing both efficiency and welfare losses. In other words, when designing their tax system, governments have to accept a compromise between the need to collect the required revenue to finance public services and to carry out redistributive policies, and the need to limit as much as possible the distortions on individuals' decisions.

This normative prescription has its foundation in the seminal work by Mirrlees (1971), which represents the first attempt to explicitly introduce efficiency concerns in the design of

 1 See OECD (2015).

the optimal income taxation. Since then both academic researchers and policy makers have devoted a notable effort in the attempt to identify the optimal tax system, namely, the efficient and equitable distribution of the tax burden across a population of heterogeneous individuals, who differ in terms of their abilities and or incomes. The relevance of this (re)search is not only confirmed by the presence of several theoretical models but also by the continuous reforms of the tax systems. To this regard, indeed, Peter et al. (2009) considering a sample of 189 countries, argue that from 1981 to 2005 on average each year about 50% of the countries modify the personal income tax schedule, either by readjusting the statutory tax rates or by changing the number of income brackets. Moreover compared to low and middle income countries, high income economies tend to engage more in the identification of the optimal tax structure.²

After a more careful analysis, it is possible to identify two well defined traits of these changes, that is a reduced level of progressivity and a higher simplicity of the PIT schedule. In particular, the former is the result of the decline of the top tax rates financed by a shift of the tax burden over the middle of the income distribution. The latter, instead, is due to the reduction of the number of tax brackets and the use of less sophisticated allowances and tax formula.³ Hence, actual personal income tax systems appear less progressive and complex than those in the 80s.

However, are the evolution of tax reforms and theoretical literature following the same patterns?

More generally, the spirit of this question is to investigate how divergent tax policy reforms are from theoretical prescriptions, that is if policy makers are looking for the optimal tax system by taking into account normative recommendations. To this regard, since Mirrlees (1971) all theoretical models have formalized the equity-efficiency trade-off in a tax formula which is a function of three elements: the level of inequality in the wage or income distribution, the government's attitude toward inequality and the extent to which individuals react to taxation. In other words, given the dispersion level of individuals' income distribution, a government might want to reduce this inequality through the tax system and redistributing from high income agents to low income ones. However, this redistributive policy may discourage individuals⁷ labour supply decisions and then leads to efficiency losses, whose size depends on the individuals

 2 See Peter et al (2009) Table 1.

 3 See Peter et al. (2009) Table 5 and Table 6.

responsiveness to tax changes. Hence, inequality reduction entails a cost in terms of economic efficiency and welfare losses. From this perspective the choice of a particular tax schedule can be seen as the result of the combination between government's objectives and initial circumstances. Then, the previous question can be reformulated as follows:

 Can tax changes represent the government's attempt to achieve some redistributive objectives? Are the changes in the level of inequality (and polarization) driving the variations in the PIT schedule? Does individuals responsiveness to taxation shape changes in the tax rates?

The remainder of this chapter provides an answer to those questions and it is structured as follows: first, we review the main theoretical results and their normative prescriptions. Then, we analyze how changes in PIT schedule can be related to the theoretical literature. Second, we describe the role of objectives and circumstances in shaping the optimal tax formula and the redistributive policies. More specifically, with regard to the objectives, we compare the standard welfarist and the non-welfarist approach and we highlight how the optimal tax schedule changes in both cases. Then, we see toward which approach the tax changes are moving. Finally, with respect to circumstances, we analyze if changes in PIT can be reconciled with the trend of inequality and polarization, or if there have been variations in the individuals labour supply elasticity which can explain the pattern of tax changes.

1.2 The (re)-search of the optimal tax system. Theory vs practice

As it has been largely acknowledged Mirrlees' (1971) paper represents the origin of the modern analysis of optimal income taxation. By adopting this framework, researchers developed several models to identify the optimal tax system which achieves equity and minimizes efficiency losses. A shared element among all these models is the way in which the optimal tax problem is formalized. More specifically, the optimal tax system is the result of an optimization problem faced by the government, which maximizes a social welfare function, defined over individuals' utility, subject to a budget constraint and taking into account individualsíreactions to taxation. The way in which the utility of different individuals are weighted represents the government's concerns for equity. While the individuals' reactions to taxation are measured by the level of labour supply elasticity. Hence, when inequality is low social welfare increases. However, reducing inequality is costly in terms of efficiency losses.

The main qualitative conclusions of the theoretical literature can be summarized into three points: Örst, marginal tax rates are non negative and lower than 100 percent. Second, the marginal tax rate on the individual with the highest income is zero (Sadka (1976) and Seade (1977)). Third, if the income distribution is bounded from below and there is no bunching of individuals at the bottom, then the marginal tax rate for the lowest income is also zero (Seade (1977)). In addition to these qualitative recommendations, some other works also provide quantitative prescriptions about the marginal tax rates profile. In particular, Mirrlees' (1971) computations suggest that tax schedule is roughly linear, with low and decreasing marginal tax rates. Atkinson (1973) by adopting a different specification of the social welfare function, confirm the regressive trait of the tax schedule, however marginal tax rates are higher than those obtained by Mirrlees. Thereafter, Stern (1976), Tuomala (1984) and Kanbur and Tuomala (1994) show that the magnitude of the results obtained by Mirrlees and Atkinson depends on the assumption about individuals' labour supply elasticity, while the regressive pattern is due to the hypothesis about the level of initial dispersion within individuals' income or ability distribution. Specifically, as regard to the individuals' reaction to taxation, by considering lower value of labour supply elasticity, Stern (1976) and Tuomala (1984) find higher marginal tax rates than Mirrlees (1971) and Atkinson (1973), even if they are still decreasing in income.⁴ Kanbur and Tuomala (1994), instead, prove that a progressive tax system, with increasing marginal tax rates, is obtained when there is high dispersion within individuals' ability. Afterward, Saez (2001) concludes that the optimal tax schedule exhibits marginal tax rates increasing in income, with the top tax rates ranging between 50 and 80%.

Hence, the theoretical prescriptions tend to be unanimous as regard the fact that tax system should be progressive when inequality is high. However, there is a substantial disagreement about the treatment of the top incomes. In particular, from one side Mirrlees et al. (2011) argue that top tax rates should be reduced or at least not increased, while from the other side Piketty et al. (2014) claim the opposite, i.e. that top tax rates should be raised.

⁴Marginal tax rates in Tuomala (1984) range from 60 percent at the bottom to 25 percent at the 95th percentile.

In light of the state of art of the theoretical literature, when looking at the salient trends of the reform in PIT, we observe a general tendency to reduce top tax rates and to Önance this cut by shifting the tax burden towards the middle of income distribution. As reported by Peter et al (2009) this downward trend is common within non OECD economies as well. Their investigation reveals that, in the last two decades, the fraction of countries with top tax rates higher than 40 percent sharply declined from 71% to 17%.⁵

This scenario seems to be in line with the position of Mirrlees et al. (2011) about the reduction of top tax rates. However, the fact that these cuts reduce the progressivity of the tax systems allows us to claim that tax changes are moving in an opposite direction than the theoretical prescriptions.

Another peculiarity of the tax changes of the last twenty-five years is the attempt to simplify the tax schedules. To this regard, policies reforms appear in line with some theoretical prescriptions. In particular, theory recognizes that a good tax system has to be simple in order to avoid additional costs on the society and to limit evasion and avoidance opportunities. Slemrod and Sorum (1984) estimate these costs for the U.S. as 5-7% of the total amount of collected taxes. According to Peter et al. (2009) there are several aspects which contribute to the overall level of complexity of the PIT. Among those, in order to analyze the trend of the tax reforms, they choose the use of non-standard allowances, surtaxes, complex tax formula and multiple brackets schedule as indices of the complexity of the personal income tax system. What they find is that all these indicators have declined due to the tax changes occurred over the last two decades, leading to reduction of the overall level of complexity of the PIT.

The reduction of the number of brackets and the decline in progressivity could suggest that policy makers are moving toward a flat tax system. However, to this regard there are no convergent opinions in the literature. More specifically, from one side the proponents of this kind of tax schedule emphasize its advantages in terms of high simplicity which implies low compliance and administration costs, reduced tax evasion and strong labour supply incentives. On the other side, instead, the opponents claim that the argument of lower compliance costs does not make sense, at least for developed economies. Moreover, these countries are characterized by a large middle class which will be disadvantaged by the introduction of a flat tax schedule

⁵ See Peter et al. (2009) Table 2.

which shifts the tax burden from the top to the middle of the income distribution. In addition there is no convincing evidence that a flat tax system increases labour supply incentives and the tax revenue. In particular the magnitude of this effect strictly depends on the parameters of the tax system.⁶

1.3 The role of objectives. Welfarism vs Non-Welfarism

One lesson from the optimal taxation literature is that a crucial element of the optimal tax formula is the way in which the government evaluates the welfare of different individuals. This choice reflects the particular objective pursued by the social planner and its attitude toward redistribution. In this section we start by describing two different approaches used in the theoretical literature to identify government's objectives, then we analyze if the changes in the PIT schedule during the last two decades can be reconciled with one of these two approaches.

The standard approach, known as welfarist, characterized optimal taxation literature since its foundation. Under this approach, an utilitarian government is aimed at maximizing a social welfare function defined over individuals' utility. Then, in order to redistribute from individuals with high income (and/or ability) to individuals with low income (and/or ability), this welfarist government uses a non-linear (Mirrlees (1971)) or linear (Sheshinski (1972)) tax schedule. Theoretical literature has criticized welfarist approach from different perspectives. Sen (1985), for example, argued that utility is only one aspect of individuals' welfare, as there are several other dimensions that should be considered in the design and in the evaluation of tax policies. In other words, by focusing only on the individuals utility other relevant information are ignored, i.e. individuals differ in term of their functioning and capabilities. Then, even if resources were equally allocated across individuals, their utility could not be the same, due to the heterogeneity in terms of functioning and capabilities.

Starting from Senís critique Kanbur et al. (1994) develop an alternative approach known as non-welfarism. More specifically, they propose a framework to analyze how the optimal tax schedule changes when the government is aimed at achieving an objective different from the maximization of the sum of individuals' utility.

⁶ See Paulus and Peichl (2009) and the references therein.

It could be happen, indeed, that individuals preferences are manipulated or that individuals behaviors are not socially optimal. Hence, in all these situations there is room for a government's intervention through the tax system. To this regard O'Donoghue and Rabin (2003) and Schroyen (2005) are two examples of this corrective (paternalistic) taxation. In particular, the formers derive an optimal tax schedule to reduce the consumption of harmful good, while the latter designs a tax system which aims to promote the consumption of some merit goods.

However, it has to be noted that there is no consensus on the dividing line between the welfarist and the non-welfarist approach. In both cases the optimal tax problem is formalized as a constrained maximization of a social welfare function, given the revenue requirement and taking into account individuals' reactions. The difference between the two approaches is the argument of the social welfare function. More specifically, a government is said to be non welfarist when its social welfare function is not defined over individuals utilities.⁷

To this regard, the non-welfarist approach to evaluate social welfare by focusing on individuals' income instead of their utility, seems to be supported by the fact that social indicators measuring inequality, poverty and polarization are based on incomes and not on utilities. Kanbur, Keen and Tuomala (1994) study how the optimal tax formula changes when the government's objective is the reduction of poverty. In this case, the optimal tax problem is formalized as the minimization of a poverty index, defined over individuals' income. The optimal tax formula reducing poverty qualitatively differs from the welfarist optimal tax formula. However, in order to quantify this difference in terms of marginal tax rates, they provide some numerical simulations, which show that non-welfarist tax system envisages higher marginal tax rates than the welfarist one, even if both approaches show a decreasing marginal tax rates.

Another example of non-welfarism is the use of a social welfare function where individuals incomes are weighted according to their position in the income distribution. For example, in a rank dependent social evaluation function consistent with the Gini index, the weights attached to individuals incomes, ranked in ascending order, are linearly decreasing and bounded above by two and below by zero. Aaberge and Colombino (2013) adopt a rank dependent social welfare function in their micro econometric optimal tax model. Their results confirm that different social objectives lead to different tax systems. Specifically, the more egalitarian the

⁷ See Kanbur, Pirtilla and Tuomala (2006).

social welfare function, the more progressive the tax schedule. Recently Saez and Stantcheva (2016) propose an approach to evaluate tax reforms, where the weights attached to individuals incomes taking into account concerns of fairness.

To summarize, the welfarist approach tends to focus more on efficiency concerns, while the non-welfarist one considers the possibility that the social-planner could pursue different social objectives. As the examples for poverty reduction have shown, these differences have a remarkable impact on the design of the optimal tax schedule. Hence, changes in objectives could lead to different tax schedules. However, when observing the trend of PIT changes it is not a easy task to identify a specific redistributive objective pursued by policy makers. In particular, the reduction of the overall progressivity of the tax systems does not appear consistent with a non-welfarist objective of inequality reduction. This is confirmed by the upward trend of inequality measures as well. Hence, one can conclude that government' objectives might be the improvement of the efficiency and the simplicity of tax schedule. In addition, the reduction of top tax rates seems to move toward one of the most famous welfarist result, i.e. zero marginal tax rate on the top of distribution.

1.4 The role of circumstances

1.4.1 Elasticity

The size of the tax to GDP ratio suggests that the effect of taxation on individuals welfare is not negligible. Individuals labour supply decisions are one channel through which this effect materializes and the labour supply elasticity measures the magnitude of individuals responsiveness to taxation. Therefore, it might be interesting to understand if PIT changes reflect in some way variations in the individuals' responsiveness to taxation. To this regard the main lesson from theoretical models is that the more responsive are the individuals, the lower the tax rates.⁸ Researchers have devoted a considerable effort to estimate individuals' reaction to

⁸As noted before, one reason for the low tax rates obtained by Mirrlees (1971) and Atkinson (1973) is their assumption on the elasticity between consumption and labour. In particular, they consider an elasticity level equal to one and constant throughout the wage distribution. Stern (1976) and Tuomala (1984) consider respectively a level of elasticity equal to 0.4 and 0.5 and obtain higher tax rates.

taxation.⁹ However, it has to be considered that these reactions can materialize along several dimensions. For example, individuals may change their labour supply or vary the composition of their income, by preferring income sources which are subject to lower taxation.¹⁰

Then, available estimates are quite different and their heterogeneity is due to the estimation procedure, the characteristics of the tax system, the country and the period analyzed. To this regard, Creedy (2009) suggests that elasticity changes both over time and across countries which are different in terms of tax systems and regulations. In any case, it is possible to outline an overview of the main results of the literature and then try to understand if tax changes are consistent with theoretical prescriptions.

More specifically, with regard to taxable income elasticity there is a broad consensus about a range of values between 0.1 and 0.4. In addition, Saez, Slemrod and Giertz (2009) find that the elasticity of taxable income is increasing in income. If we consider, instead, individuals reactions in term of labor supply decisions (hours of work) the picture is that elasticity is inversely related to income.¹¹ Moreover, males appear less responsive than females and this difference is more evident with respect to married females or females with children. Moreover, among males those with low and middle education levels are more sensitive than those with high education. In addition Bargain and Peichl (2016) argue that there has been a fall in the labours supply elasticity since 1980.

However heterogeneity in terms of individuals' reaction to taxation, the variety of the available estimates and the complexity of tax systems make difficult to relate the continuous tax changes to some pattern of the elasticities.

1.4.2 Inequality (and Polarization)

Since the seminal work by Mirrlees (1971), optimal taxation models recognize that the level of inequality plays an important role in determining the level of tax rates. More precisely, the

⁹ See Giertz (2004), Meghir and Philips (2008), Saez, Slemrod and Giertz (2009) and Bargain and Peichl (2016) for an overview of the literature relating labour supply, taxes and taxable income elasticity.

 10 Meghir and Phillips (2008) for example suggest that another possible dimension is related to the human capital accumulation.

 11 ¹¹ According to Aaberge and Colombino (2013) this is due to the fact that individuals with high income, on average participate more and work more hours. Then, if their wage decreases they have to change less their labor supply with respect to individuals with low income.

higher the level of inequality, the lower the social welfare and the more progressive the tax system. Then, by looking at the next figure we can observes that there has been an upward trend in the level of inequality, as shown by the increase in the Gini index in most of the OECD economies. The fact that economies became more unequal during the last two decades is also confirmed by the evolution of two other inequality measures: the Palma ratio and the ratio between the income shares of the richest 10% and the poorest 10%. Both indicators increased over this period.

Trend in Gini indexes in OECD countries

Note: Dashed line Gini disposable income, post taxes and transfers. Solid line: Gini market income, before taxes and transfers. Some elaboration on OECD data.

Trend in Palma ratio and other income share ratio

Note: In the right vertical axis, Palma ratio (green solid line). In the left vertical axis: share of all income received by the top quintile divided by the share of the first (red dashed line); ratio of the upperbound value of 9th decile to that of the upperbound value of 1stdecile (blue dashed line). Source: own elaboration on OECD data.

Another interesting aspect concerning income distribution is the level of polarization, which is to some extent related to the dimension of the middle class. The next figure shows the trends of polarization indices.

Trend in Polarization indexes in OECD countries

Note Dashed line: Polarization DER (Duclos Esteban Ray) index (alpha=1), solid line: Polarization Wolfson index. Source: own elaboration on LIS database

Given these trends and the normative prescription of optimal tax theory one should expect that tax changes move toward more progressive tax systems. However, evidence suggests the opposite, thus personal income taxation become less progressive and top tax rates decline. Put in these terms, theory and practice seem to move along two different patterns.

1.5 Conclusion

In this chapter we investigate if theoretical prescriptions of the optimal taxation theory and tax reforms occurred in the last twenty-Öve years move along the same direction. In addition, by considering tax policies as the result of objectives and circumstances, we have analyzed if theoretical implications on the relevance of those two crucial elements, can help to understand the changes in the PIT schedule. Our analysis shows that some theoretical results can be recognizable in tax policy changes, for example the tendency to improve simplicity and efficiency of the PIT system goes through this direction. However, there are also some trends in the tax reforms which appear not consistent with the theory. More specifically, the reduction of the top tax rates and the decline in the level of progressivity contrast with the theoretical framework developed since Mirrlees (1971), which claims a positive relationship between progressivity and inequality. Indeed, evidence shows that while during this period inequality has been rising, the tax burden has been shifted from the top to the middle of the income distribution. This is also not consistent with the non-welfarist objective of inequality reduction. Last, it is difficult to use theoretical prescriptions about the role of elasticity to explain tax changes.

Chapter 2

Optimal Non-Welfarist Income Taxation for Inequality and Polarization Reduction

We adopt a non-welfarist approach to investigate the effect of different redistributive objectives on the shape of the optimal tax schedule. We consider inequality and income polarization reduction objectives and we identify socially desirable three brackets piecewise linear tax systems that allow to collect a given amount of per-capita revenue. The optimal tax problem is formalized as the maximization of families of rank-dependent social evaluation functions defined over individuals' net income. These functions allow to incorporate within the same social evaluation model concerns for inequality and for polarization reduction.

Both with fixed and with variable labour supply the optimal tax schemes substantially differ as the focus moves from the reduction of inequality to the one of polarization. In the case of inequality concerns the optimal tax system is mainly convex exhibiting increasing marginal tax rates unless when labour supply elasticities are higher. While in case of polarization concerns the optimal tax scheme is non-convex with reduced marginal tax rate for the upper income bracket.

2.1 Introduction

Over the last two decades many countries have experienced a remarkable increase in income inequality and income polarization.¹ At the same time, policy makers have continuously modified the tax systems, especially the personal income tax schedule, which on average represents one quarter of the total amount of revenues for OECD countries. More specifically, these changes have been related to the reduction in the number of income brackets and the decline of the top marginal tax rates with a shift of the tax burden from the top to the middle of the income distribution. As a result, tax systems have become more simplified and less progressive over the last twenty five years. 2

Recent literature has addressed the question of the optimal level of taxation for top incomes and analyzed if the upward trend in inequality can be related to the reduction of the top tax rates. However, divergent views about these aspects have been developed.³

In this paper we focus on the entire income distribution and 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' income instead of their utility. Individuals' preferences do not play a direct role into the social welfare, but they still play a role in the design of the optimal tax system in that they shape the individualsí reaction to taxation in terms of consumption and labour supply decisions We believe that the focus on income is the most appropriate choice when the government objective is to reduce inequality, poverty or polarization whose indicators are defined in terms of individuals' income.

As to the formalization of the redistributive objective, we assume that the non-welfarist government maximizes a rank-dependent social evaluation function defined over individuals

¹See OECD (2015) and Brzezinski (2013) for a description of the trend of income inequality and polarization respectively.

² See Peter et al. (2009) for an overview of the changes emerged in the personal income taxation schedule from 1981 to 2005 in 189 countries, where it is shown that on average each year about 50% of the countries modify the income tax schedule. See also Piketty and Saez (2013) and OECD (2015) for a description of the tax systems and their evolution.

³Piketty, Saez and Stantcheva (2014) argue that top tax rates should be raised, while on the other hand, Mirrlees et al. (2011) suggest that top tax rates should not be increased.

income, given a revenue requirement constraint. According to this evaluation model incomes are aggregated linearly and are weighted according to their position in the income ranking. The specific form of the weighting function captures the government's non-welfarist objective. This class of social evaluation function allows to formalize redistributive objectives that could be quantified in the case of the concern for inequality in terms of changes in the Gini index of incomes. Moreover, by suitable modifications of the positional weighting function it is possible to move within the same social evaluation model from evaluations concerned on inequality to those focussing on the polarization of the incomes.

In our analysis we focus on piecewise linear tax systems not only because these tax schemes represent those most commonly adopted, but also because they allow to make explicit the changes in tax schedule driven by the government's shift from a concern for inequality to a concern for polarization reduction. In fact, as we will show, for the benchmark case with fixed labour supply a scheme with three brackets is sufficiently flexible to highlight the distinctive features of the inequality reducing optimal taxation compared to the one that aims at reducing income polarization.

Our unified approaches that focus on inequality and polarization reduction objectives extend existing literature on non-welfarist taxation that has focussed on poverty alleviation (Kanbur, Keen and Tuomala, 1994). We identify the socially preferred three brackets linear taxation schemes that collect a given amount of taxes, given a specific redistributive objective both for the case of fixed labour supply and when labour supply elasticity is constant. The interesting aspect of our work is that both with fixed and with variable labour supply the optimal tax schemes substantially differ as the focus moves from the reduction of inequality to the one of polarization. In the case of inequality concerns the optimal tax system is mainly convex exhibiting increasing marginal tax rates unless when labour supply elasticities are high. While in case of polarization concerns the optimal tax scheme is non-convex with reduced marginal tax rate for the upper income bracket.

The remainder of the paper proceeds as follows: the next subsection briefly reviews the literature on optimal income taxation, comparing the welfarist and non-welfarist approaches. Section 2 presents the linear rank-dependent social evaluation function and describes the two different weighting functions adopted in the paper to formalize the non-welfarist redistributive objective. Section 3 formalizes the optimal tax problem faced by the non-welfarist government. The solution of this problem is presented in Section 4 under the assumption of exogenous fixed labour supply. Section 5 describes first the agents' optimization problem, then it presents numerical results about the optimal tax schedule reducing income inequality and polarization with fixed and elastic labor supply. Section 6 concludes.

2.1.1 Optimal Welfarist and Non-Welfarist Income Taxation

Optimal Welfarist Income Taxation

An optimal tax system is the result of a constrained optimization problem, where the government maximizes a social welfare function, subject to a set of constraints. These constraints are related to the amount of required revenues and to individuals' reaction to taxation. A celebrated result of the seminal Mirrlees' (1971) model of non-linear income taxation is that marginal tax rates should be non-negative and lower than 100%. In addition, when income distribution is bounded the optimal marginal tax rate is zero both on the highest income (Sadka, 1976 and Seade, 1977) and on the lowest one (Seade, 1977). Mirrlees also provided some numerical computations for the optimal tax system that suggest that the tax schedule is quite linear and regressive. By adopting a different specification of the social welfare function, Atkinson (1973) confirms the regressive pattern of tax rates which are however higher than those obtained in Mirrlees (1971). Thereafter, Stern (1976), Tuomala (1984), and Kanbur and Tuomala (1994) show that such results derive from the assumption of low levels of inequality and high labour supply elasticity. More specifically, by considering lower values of labour supply elasticity Stern (1976) and Tuomala (1984) obtain higher marginal tax rates than Mirrlees (1971) and Atkinson (1973). While Kanbur and Tuomala (1994) derive a progressive tax system by considering an high level of inequality within the distribution of individuals abilities.

The complexity of the non-linear tax model has led the theoretical literature to formulate the optimal tax problem in a simpler way, by considering the case of linear taxation. Sheshinski (1972) identifies the optimal marginal tax rate and the level of lump-sum subsidy which maximize an utilitarian social welfare function given the budget constraint. Tuomala (1985) then provides a simplified formula for the optimal linear income tax, where the equity-efficiency trade-off is easily identifiable. Specifically, the higher is the elasticity of labour supply, the

higher are the efficiency costs of taxation and the lower is the optimal tax rate. Likewise, the optimal tax rate decreases with higher levels of inequality within the distribution of individualsí wage.

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 varying between brackets but constant within. Given this consideration, the theoretical literature has developed models of piecewise linear optimal taxation starting from the works by Sheshinski (1989) and Slemrod et al. (1994). More specifically, 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) challenge Sheshinskiís 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 convex and non-convex, and by using numerical simulations they analyze the conditions under which each system is optimal. They find essentially convex systems unless when labour elasticities are high. Also Aaberge et al. (2013) Önd that the optimal piecewise tax system is convex with increasing marginal tax rates. Moreover, by comparing their results with the current tax system they show that, the optimal tax system requires lower marginal rates on low incomes and higher marginal tax rates on the top of the distribution.

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 an utilitarian social welfare function that weights all individuals' utilities equally and maximizes their sum independently 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 little significance given by the welfarism to the role of differences in redistributive objectives has motivated the development of an alternative approach by Kanbur et al. (1994), which takes into account the possibility that governments can pursue objectives different from the maximization of the sum of individuals' utility such as the reduction of social indicators like inequality, poverty and polarization. Consequently, the two approaches lead to two different profiles of the optimal tax formula, which is, under the non-welfarist approach, strictly related to the specific governmentís objective. For example, Kanbur et al. (1994) derive the optimal income tax schedule reducing poverty measured in terms of a family of additively separable poverty index. Moreover, they 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 socially optimal. Therefore, in these situations there is room for a government's corrective intervention, who uses taxes and transfers to correct individuals' behaviors. 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. Recently Saez and Stantcheva (2016) propose an alternative approach to evaluate tax reforms according to a social evaluation model where the weights attached to individuals' net income allow to take into account different concepts of justice.

The main indication arising from theoretical literature is that a government is "non-welfarist" when evaluates social welfare by using a criterion different from individuals' preferences. 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 differently than individuals do. Conventionally, following Kanbur et al. (1994) a non-welfarist government is one that goes beyond utility and maximizes a social welfare function which is not defined over individuals' utility. In our case the objective function will be defined over the net income distribution and will take a rank-dependent formalization.

2.2 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 $v(p) \ge 0$ 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

$$
W_v(F) = \int_0^1 v(p) F^{-1}(p) dp
$$
\n(2.1)

where $v(p) \geq 0$ is the weight attached to the income of individual ranked p. The normative basis for this evaluation function has 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 $(2015).⁴$ 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

⁴See also Aaberge (2000), Aaberge et al. (2013) and Maccheroni et al. (2005).

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 evaluation⁵ 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 (2.1) 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 as⁶

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

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.

2.2.1 Weighting functions

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

 $\overline{\text{5}}$ For general details see Lambert (2001).

⁶The single parameter family of relative Gini index of inequality parameterized by δ is expressed as $G(F)$ = $\frac{1}{\mu(F)} \int_0^1 \left[1 - \delta(1-p)^{\delta-1}\right] F^{-1}(p) dp$, which becomes the standard Gini coefficient for $\delta = 2$.

poorer to richer individuals. Alternatively we can write these weights as

$$
v_G(p) = \begin{cases} 1 - \left[-2\left(\frac{1}{2} - p\right) \right] \text{ if } p \le \frac{1}{2} \\ 1 - 2\left(p - \frac{1}{2}\right) \text{ if } p \ge \frac{1}{2} \end{cases} \tag{2.2}
$$

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

$$
w_G(p) = \begin{cases} -2\left(\frac{1}{2} - p\right) & \text{if } p \le \frac{1}{2} \\ 2\left(p - \frac{1}{2}\right) & \text{if } p \ge \frac{1}{2} \end{cases} \tag{2.3}
$$

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

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 $(2013).$ ⁷ The associated SEF is rank-dependent with a weighting function that can be formalized as:

$$
v_{P_{(\beta,\delta)}}\left(p\right) = \begin{cases} 1 + \beta \left(2p\right)^{\delta - 1} & \text{if } p \le \frac{1}{2} \\ 1 - \beta \left(2 - 2p\right)^{\delta - 1} & \text{if } p \ge \frac{1}{2} \end{cases} \tag{2.4}
$$

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,

$$
v_P(p) = \begin{cases} 2p+1 \text{ if } p \le \frac{1}{2} \\ 2p-1 \text{ if } p \ge \frac{1}{2} \end{cases}.
$$
 (2.5)

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 (2.5) is illustrated in the following figure.

The weighting function for the Polarization based SEF.

⁷An alternative approach to the construction of polarization sensitive SEFs is presented in Rodriguez (2015).
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.

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

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

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

$$
\mu(F) P(F) = -\int_0^{1/2} 2p F^{-1}(p) dp + \int_{1/2}^1 2(1-p) F^{-1}(p) dp.
$$
 (2.6)

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

$$
v_P(p) = \begin{cases} 1 - \{-[1 - 2(\frac{1}{2} - p)]\} & \text{if } p \le \frac{1}{2} \\ 1 - [1 - 2(p - \frac{1}{2})] & \text{if } p \ge \frac{1}{2} \end{cases} . \tag{2.7}
$$

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

$$
w_P(p) = \begin{cases} -[1 - 2(\frac{1}{2} - p)] \text{ if } p \le \frac{1}{2} \\ [1 - 2(p - \frac{1}{2})] \text{ if } p \ge \frac{1}{2} \end{cases}.
$$
 (2.8)

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.⁸

An elementary normative implication of the polarization based welfare weighting function is that in order to maximize the welfare, redistribution should be from the individuals above the median to those below. However, when tax schedules are set over few brackets 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.

2.3 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

$$
W_v = \int_0^1 v(p) \left[y(p) - T(y(p)) \right] dp, \tag{2.9}
$$

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

⁸The 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.

budget constraint is

$$
\int_0^1 T\left(y\left(p\right)\right) dp = \overline{G}
$$

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

$$
T(y) := \begin{cases} t_1y & \text{if } y \le y_1 \\ t_1y_1 + t_2(y - y_1) & \text{if } y_1 < y \le y_2 \\ t_1y_1 + t_2(y_2 - y_1) + t_3(y - y_2) & \text{if } y > y_2 \end{cases}
$$
(2.10)

or in equivalent terms

$$
T(y) := t_1y + (t_2 - t_1) \cdot \max\{y - y_1, 0\} + (t_3 - t_2) \cdot \max\{y - y_2, 0\}.
$$

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.

2.4 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) .⁹ 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

⁹See also the review in Lambert (2001).

constraint that collects the per-capita value \overline{G} .

The constrained optimization Lagrangian function for this problem is as follows

$$
\max_{t_1, t_2, t_3, y_1, y_2} \mathcal{L} = W_v + \lambda \left[\overline{G} - \int_0^1 T(y(p)) \, dp \right],\tag{2.11}
$$

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

The SEF W_v is presented in (2.9). 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.

2.4.1 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 $\bar{\tau} \in (0, 1]$.

The family of SEFs considered is denoted by \mathcal{W}_I that represents the set of all linear rankdependent 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 (2.2) 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 $\overline{\tau}$ is s.t. $\overline{G} \leq \overline{\tau} \cdot \mu(F)$ 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

$$
t_1 = 0,
$$

$$
t_3 = t_2 = \overline{\tau},
$$

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 5.

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 $t_3 = \overline{\tau}$, $t_1 = t_2 = 0$ 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 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 $\bar{\tau}\geq \frac{T(y)-T(y')}{y-y'}$ $\frac{y-y}{y-y'} \ge 0$ 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 alterative post-tax distributions [see Marshall et al. 2011]. That is, let T^0 denote the optimal tax function then the Lorenz curve of the post tax income distribution is obtained as $L_{T^0}(p) = \frac{1}{\mu_{T^0}}$ $\int_0^p \left[y(q) - T^0(y(q)) \right] dq$ where $\mu_{T^0} = \int_0^1 \left[y(q) - T^0(y(q)) \right] dq$ 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 $L_{T^0}(p) = \frac{1}{\mu_{T^0}}$ $\int_0^p \left[y(q) - T^0(y(q)) \right] dq \ge L_T(p) = \frac{1}{\mu_T} \int_0^p \left[y(q) - T(y(q)) \right] dq$ for all $T \in T_{\overline{\tau}}$ and all $p \in [0, 1]$. Recalling that all the alternative tax schemes should guarantee the same revenue, the condition could be simplified as $\int_0^p [y(q) - T^0(y(q))] dq \ge \int_0^p [y(q) - T(y(q))] dq$, that is after simplifying for $y(q)$ we obtain

$$
\int_{0}^{p} T^{0} (y(q)) dq \leq \int_{0}^{p} T (y(q)) dq
$$
\n(2.12)

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

Recall that by construction (i) $T^0(y(p)) = 0$ for all $p \le p_1$, and that (ii) $\bar{\tau} = \frac{T^0(y) - T^0(y')}{y - y'}$ $\frac{y-y}{y-y'} \geq$

 $T(y)-T(y')$ $\frac{y-x(y)}{y-y'}$ 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 (2.12) holds for all $T \in T_{\overline{\tau}}$ 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.

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

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 $t_2 = \overline{\tau}$. 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 $t_1 = \overline{\tau}$, 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 $\mathcal{T}_{\overline{\tau}}$ maximizing linear SEFs in \mathcal{W}_P is:

(*i*) $p_1 < 1/2 < p_2$ where $\frac{1-V_P(p_1)}{1-p_1} = \frac{1-V_P(p_2)}{1-p_2}$ $\frac{\Gamma(\nu_P(p_2))}{1-p_2}$ and such that the revenue constraint is satisfied with

$$
t_1 = t_3 = 0,
$$

$$
t_2 = \overline{\tau},
$$

if $\overline{G} \le \min\{G^+, G^-\}.$

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

$$
t_1 = 0,
$$

$$
t_2 = t_3 = \overline{\tau}
$$

where p_1 [and so also y_1] is such that the revenue constraint is satisfied (iib) If $\overline{G} > G^-$ solution (i) should be compared with $p_1 > 1/2$, and

$$
t_1 = \bar{\tau},
$$

$$
t_2 = t_3 = 0,
$$

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

(iii) If \overline{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 7.

The proposition highlights the fact that under standard revenue requirements $\overline{G} \le \min\{G^+,G^-\}$ 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 $\overline{G} > G^+$ and $\overline{G} > G^-$ 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.

2.5 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.

2.5.1 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 $\phi : [0, L] \to \mathbb{R}$ 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

$$
\phi(l) = k \cdot l^{\alpha} \tag{2.13}
$$

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 $y = w l$. Agents are subject to taxation $T(y) \geq 0$ formalized by (2.10), 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

$$
x := y - T(y) = \begin{cases} (1 - t_1)y & \text{if } y \in Y_1 \equiv [0, y_1) \\ (t_2 - t_1)y_1 + (1 - t_2)y & \text{if } y \in Y_2 \equiv [y_1, y_2) \\ (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)y & \text{if } y \in Y_3 \equiv [y_2, \infty) \end{cases}
$$
(2.14)

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

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

when $y^* \in Y_i \setminus y_{i-1}$, where the function $\phi'^{-1}(.)$ by construction is positive and strictly increasing.

Given the definition of $y = w_l$, one obtains also the associated optimal labour supply

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

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

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

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

$$
y^* = w \cdot \left[\frac{(1 - t_i)w}{k\alpha} \right]^{\frac{1}{\alpha - 1}} = w^{\frac{\alpha}{\alpha - 1}} \left[\frac{(1 - t_i)}{k\alpha} \right]^{\frac{1}{\alpha - 1}} \tag{2.15}
$$

$$
l^* = \left[\frac{(1 - t_i)w}{k\alpha} \right]^{\frac{1}{\alpha - 1}}
$$

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 $\frac{1}{(\alpha-1)}$.¹⁰In this paper we will consider as a reference distribution the gross income distribution in absence of taxation. Then, by setting $t_i = 0$ from (2.15) 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

$$
y(p) := w(p)^{\frac{\alpha}{\alpha - 1}} \left[\frac{1}{k\alpha} \right]^{\frac{1}{\alpha - 1}} = w(p)^{1 + \varepsilon} \left[\frac{\varepsilon}{k(\varepsilon + 1)} \right]^{\varepsilon}
$$
(2.16)

represents the gross income of the individual covering position p under the assumption of notaxation, 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 .

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

Case A, is such that $t_1 \le t_2 \le t_3$, while case B considers the configuration where $t_1 \le t_3 < t_2$.

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.

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

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

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

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

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

$$
y_{t}(p) = \begin{cases} y(p) (1-t_{1})^{\varepsilon} & if y(p) < \frac{y_{1}}{(1-t_{1})^{\varepsilon}} \\ y_{1} & if \frac{y_{1}}{(1-t_{1})^{\varepsilon}} \leq y(p) \leq \frac{y_{1}}{(1-t_{2})^{\varepsilon}} \\ y(p) (1-t_{2})^{\varepsilon} & if \frac{y_{1}}{(1-t_{2})^{\varepsilon}} < y(p) \leq \frac{y_{2}}{(1-t_{2})^{\varepsilon}} \\ y_{2} & if \frac{y_{2}}{(1-t_{2})^{\varepsilon}} \leq y(p) \leq \frac{y_{2}}{(1-t_{3})^{\varepsilon}} \\ y(p) (1-t_{3})^{\varepsilon} & if y(p) > \frac{y_{2}}{(1-t_{3})^{\varepsilon}} \end{cases}
$$
(2.17)

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

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 $y \in Y_3 \backslash y_2$, while all incomes below are such that $y \in Y_2 \backslash y_1$. Appendix B illustrates the derivation of \hat{y} that is

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

with $t_2 > t_3$, while if $t_2 \rightarrow t_3$ then $\hat{y} = \frac{y_2}{(1-t_2)^2}$ $\frac{y_2}{(1-t_2)^{\epsilon}}$. It follows that

$$
y_{t}(p) = \begin{cases} y(p) (1-t_{1})^{\varepsilon} & if y(p) < \frac{y_{1}}{(1-t_{1})^{\varepsilon}} \\ y_{1} & if \frac{y_{1}}{(1-t_{1})^{\varepsilon}} \leq y(p) < \frac{y_{1}}{(1-t_{2})^{\varepsilon}} \\ y(p) (1-t_{2})^{\varepsilon} & if \frac{y_{1}}{(1-t_{2})^{\varepsilon}} \leq y(p) \leq \widehat{y} \\ y(p) (1-t_{3})^{\varepsilon} & if y(p) > \widehat{y} \end{cases}
$$
(2.18)

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.

2.5.2 Numerical results

The optimal taxation problem described in the previous section is solved numerically.¹¹ To this end, we need to assign a value to the parameters α and k of the utility function (2.13) and to specify the distribution ξ_w of individual wages and the exogenous revenue requirement \overline{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,

¹¹We use a grid search method. More specifically, we define the grids for t_1, t_2, y_1 and y_2 , with $t_1 \le t_2$ 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.

and accordingly we set α respectively equal to 11, 6 and 3^{12} . 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.¹³ 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 $327.¹⁴$

Finally we consider different values of the exogenous revenue requirement \overline{G} , namely we alternatively set \overline{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 $t_1 \le t_2 \le t_3$, while the non-convex tax regime considers the configuration where $t_1 \le t_3 \le t_2$. 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

 12 The 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).

¹³Given 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 (2.15) and set $t = 0$ and $k = 1/\alpha$).

¹⁴The 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) (case (1.a)), we assume L=20, $\alpha = 1.4$ and we truncate the distribution at the 98^{th} percentile which corresponds to a value of 327.

Table 1. Optimal tax systems with fixed labor supply

Panel A: Gini Social Evaluation Function.

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

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

\overline{G}	t_{1}	t_2	t_3	y_1	y_2	W_P		
$0.1 \times \mu$	0	50%	$\boldsymbol{0}$	27.5	52.00	39.75		
				0.37	0.75			
$0.15 \times \mu$	θ	50%			$\boldsymbol{0}$	26.0	71.50	37.81
				0.31	0.85			
$0.2 \times \mu$	$\overline{0}$	50%	$\boldsymbol{0}$	24.25	106.50	35.62		
						0.25	0.92	
$0.25 \times \mu$	0	50%	$\boldsymbol{0}$	22.5	171.50	33.40		
				0.15	0.97			

Panel B: Polarization Social Evaluation Function.

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

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

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 (\overline{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.

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

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

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

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

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 low¹⁵, 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

 15 See the Gini weighting function in figure 1.

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.

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

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

\overline{G}	ε	t_{1}	t_2	t_3	y_1	y_2	W
$0.1 \times \mu$	0.1	$\overline{0}$	$\boldsymbol{0}$	24.53%	30.00	30.00	38.30
					0.46	0.46	
$0.15 \times \mu$	0.1	1%	1%	36.53%	20.00	30.00	35.89
					$\overline{0}$	0.48	
$0.2 \times \mu$	0.1	9%		9% 37.81%	20.00	30.00	33.42
					0.01	0.48	
$0.25 \times \mu$	0.1	18%	18%	37.59%	30.00	30.00	30.94
					0.48	0.48	
$0.1 \times \mu$	$\rm 0.2$	10.52%	10.52%	10.52%			37.63
$0.15 \times \mu$	$0.2\,$	15.51%	15.51\%	15.51\%			34.99
$0.2 \times \mu$	$\rm 0.2$	20.96\%	20.96%	20.96%			32.30
$0.25 \times \mu$	$\rm 0.2$	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 4A. Optimal convex tax-system for polarization based SEF.

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

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

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 \overline{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.

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

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

2.6 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 Öxed 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 Öxed 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

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

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 $i = 1, 2, 3, \frac{\partial \mathcal{L}}{\partial y_i}$ 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

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

$$
\frac{\partial T(y)}{\partial t_2} = \begin{cases}\n0 & \text{if } y \leq y_1 \\
y - y_1 & \text{if } y_1 < y \leq y_2 \\
y_2 - y_1 & \text{if } y > y_2\n\end{cases}
$$

and

$$
\frac{\partial T\left(y\right)}{\partial t_{3}}=\max\left\{ y-y_{2},\,0\right\}.
$$

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

$$
\frac{\partial \mathcal{L}}{\partial t_1} = -\int_0^{p_1} v(p) \, y(p) \, dp - \int_{p_1}^1 v(p) \, y_1 dp - \lambda \left[\int_0^{p_1} y(p) \, dp + \int_{p_1}^1 y_1 dp \right],\tag{2.19}
$$

$$
\frac{\partial \mathcal{L}}{\partial t_2} = -\int_{p_1}^{p_2} v(p) \left[y(p) - y_1 \right] dp - \int_{p_2}^{1} v(p) \left[y_2 - y_1 \right] dp
$$
\n
$$
-\lambda \left[\int_{p_1}^{p_2} \left[y(p) - y_1 \right] dp + \int_{p_2}^{1} \left(y_2 - y_1 \right) dp \right]
$$
\n(2.20)

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

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

and

$$
\frac{\partial \mathcal{L}}{\partial t_3} = -\int_{p_2}^1 v(p) \left[y(p) - y_2 \right] dp - \lambda \int_{p_2}^1 \left[y(p) - y_2 \right] dp. \tag{2.21}
$$

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

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

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

$$
\frac{\partial \mathcal{L}}{\partial y_1} = -\int_{p_1}^1 v(p) \left[t_1 - t_2 \right] dp - \lambda \int_{p_1}^1 (t_1 - t_2) \, dp = 0 \tag{2.22}
$$

and

$$
\frac{\partial \mathcal{L}}{\partial y_2} = -\int_{p_2}^1 v(p) \left[t_2 - t_3 \right] dp - \lambda \left[\int_{p_2}^1 (t_2 - t_3) \, dp \right] = 0. \tag{2.23}
$$

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

$$
\frac{\partial \mathcal{L}}{\partial \lambda} = \overline{G} - \int_0^1 T(y(p)) \, dp = 0. \tag{2.24}
$$

:

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

$$
\frac{\partial \mathcal{L}}{\partial y_1} = 0, \n\frac{\partial \mathcal{L}}{\partial y_2} = 0
$$

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 (2.11) are:

$$
\frac{\partial \mathcal{L}}{\partial t_i} = -\int_0^1 v(p) \, h_i(p) \, dp - \lambda \left[\int_0^1 h_i(p) \, dp \right] \tag{2.25}
$$

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

$$
h_1(p) : = \begin{cases} y(p) & \text{if } p < p_1 \\ y_1 & \text{if } p \ge p_1 \end{cases} ;
$$

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

$$
h_3(p) : = \begin{cases} 0 & \text{if } p < p_2 \\ y(p) - y_2 & \text{if } p \ge p_2 \end{cases} .
$$

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

$$
\frac{\partial \mathcal{L}}{\partial y_1} = [t_2 - t_1] [1 - V(p_1) + (1 - p_1)\lambda]
$$
\n(2.26)

$$
\frac{\partial \mathcal{L}}{\partial y_2} = [t_3 - t_2] \left[1 - V (p_2) + (1 - p_2) \lambda \right]
$$
 (2.27)

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

$$
\frac{\partial \mathcal{L}}{\partial \lambda} = \overline{G} - \int_0^1 T(y(p)) \, dp = \overline{G} - \sum_{i=1}^3 t_i \int_0^1 h_i(p) \, dp. \tag{2.28}
$$

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 $D_v(.)$ 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

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

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

$$
\frac{\partial \mathcal{L}}{\partial y_1} = [t_2 - t_1] [1 - V (p_1) + (1 - p_1) \lambda] \n= [t_2 - t_1] [\mu (\Phi_1) \cdot [1 - D_v (\Phi_1)] + \mu (\Phi_1) \lambda],
$$
\n(2.29)
\n
$$
\frac{\partial \mathcal{L}}{\partial y_2} = [t_3 - t_2] [1 - V (p_2) + (1 - p_2) \lambda] \n= [t_3 - t_2] [\mu (\Phi_2) \cdot [1 - D_v (\Phi_2)] + \mu (\Phi_2) \lambda],
$$
\n(2.30)

:

and

$$
\frac{\partial \mathcal{L}}{\partial \lambda} = \overline{G} - \sum_{i=1}^{3} t_i \cdot \mu(H_i).
$$
 (2.31)

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

Remark 4 The partial derivatives of the Lagrangian optimization problem in (2.11) are:

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

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

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

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

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

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 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 W_I **.** Consider the results in Remark 4. If we consider SEFs where $v(p)$ is decreasing as is the case for the Gini based SEF and in general for all SEFs that are sensitive to inequality reductions through rank preserving progressive transfers from richer to poorer individuals, then $D_v(\Phi_1) < D_v(\Phi_2)$ [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 $t = \sum_{i=1}^{3} \mu(H_i) / \overline{G} \in (0,1)$. In this case it should be

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

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 $\overline{G} = \sum_{i=2}^{3} \mu(H_i)$.

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 $\overline{G} = \mu(H_2) + \mu(H_3)$. 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 5 (1A) A solution of the optimal taxation problem with fixed labour supply for tax schedules in $\mathcal{T}_{\bar{\tau}}$ maximizing linear SEFs in \mathcal{W}_I is

$$
t_1 = 0,
$$

$$
t_3 = t_2 = \overline{\tau},
$$

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

Polarization concerns.

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

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

For these SEFs it is possible to derive p_1 and p_2 such that $D_v(\Phi_1) = D_v(\Phi_2)$. This is the case for instance for the SEF whose weights are represented in (2:5): For these measures it is possible to derive the associated $V(p)$ and compute $\frac{1-V(p)}{1-p}$. They are respectively:

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

with

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

:

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 $-\lambda = \frac{1 - V_P(p_1)}{1 - p_1}$ $\frac{N_P(p_1)}{1-p_1} = \frac{1-N_P(p_2)}{1-p_2}$ $\frac{-V_P(p_2)}{1-p_2}$. 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 $-\lambda = \frac{1 - V_P(p_1)}{1 - p_1}$ $\frac{N_P(p_1)}{1-p_1} = \frac{1-N_P(p_2)}{1-p_2}$ $\frac{-v_P(p_2)}{1-p_2}$ for $-\lambda > 1/2$.

In this case

$$
-\lambda = 1 - D_v(\Phi_1) = 1 - \frac{p_1^2}{1 - p_1}
$$

$$
= 1 - D_v(\Phi_2) = p_2
$$

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 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 $p \le 1/2$, 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 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

$$
\frac{\partial \mathcal{L}}{\partial y_1} = \frac{\partial \mathcal{L}}{\partial y_2} = 0 \to \lambda = -1 + D_v(\Phi_1) = -1 + D_v(\Phi_2). \tag{2.32}
$$

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

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

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

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 $\overline{G} = \mu(H_2)$.

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 $\overline{G} = \mu(H_2)$ 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 (2.32) and the revenue constraint that is such that $\lambda = -1 + D_v(\Phi_1) = -1 + D_v(\Phi_2)$ and $\overline{G} = \mu(H_2)$. 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 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 (2.22) and (2.23) 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 $G = \overline{G}$, we have that $dG = 0$ and so

$$
(1 - p1)dy1 = (1 - p2)dy2.
$$
\n(2.33)

Analogously the differential of the SEF is

$$
dW_v = [1 - V(p_1)] dy_1 - [1 - V(p_2)] dy_2.
$$
 (2.34)

Substituting for dy_2 from (2.33) we obtain

$$
dW_v = (1 - p_1) \cdot \left[\frac{1 - V(p_1)}{1 - p_1} - \frac{1 - V(p_2)}{1 - p_2} \right] dy_1.
$$
 (2.35)

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 $\frac{1-V(p_1)}{1-p_1} > \frac{1-V(p_2)}{1-p_2}$ $\frac{-v(p_2)}{1-p_2}$ 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 $\frac{1-V(p_1)}{1-p_1} < \frac{1-V(p_2)}{1-p_2}$ $\frac{-v(p_2)}{1-p_2}$ leading to $dW_v < 0$ according to (2.35). 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 $\overline{G} = \mu(H_1) + \mu(H_2)$, while in the second case it is required that $\overline{G} = \mu(H_1)$.

As \overline{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 \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 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 $\overline{\tau} \in (0,1]$. Let $y(1/2) = y_M$ denote the median income. Then, let H^- denote the distribution whose quantile function is

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

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

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

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

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

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

We can now summarize the results in the next proposition.

Proposition 7 (3A) The solution of the optimal taxation problem with fixed labour supply for tax schedules in $\mathcal{T}_{\bar{\tau}}$ 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 $\overline{G} = \overline{\tau}\mu(H_2)$ with

$$
t_1 = t_3 = 0,
$$

$$
t_2 = \overline{\tau},
$$

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

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

$$
t_1 = 0,
$$

$$
t_2 = t_3 = \overline{\tau}
$$

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

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

$$
t_1 = \bar{\tau},
$$

$$
t_2 = t_3 =
$$

 θ ,

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

(iii) If \overline{G} > max $\{\overline{\tau}\mu(H^+), \overline{\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 Örst 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{w} in the wage distribution such that all wages above \hat{w} are such that the associated $y \in Y_3 \setminus y_2$, while for all wages in $\left(\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t)^{\alpha}} \right] \right)$ $(1-t_2)$ $\Big]^{\frac{1}{\alpha}}\, ;\widehat{w}$ $\overline{ }$ the associated gross income is such that $y \in Y_2 \backslash y_1$.

For all $w > y_1^{\alpha-1}$ $\int k\alpha$ $(1-t_2)$ $\int_{0}^{\frac{1}{\alpha}}$ the optimal gross income is $y^* > y_1$. If $t_2 > t_3$, the conditions in (2.15) could identify two potential levels of incomes one in $Y_2 \ Y_1$ and one in $Y_3 \ Y_2$ 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 (2.14) 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

$$
U_2 = U(x_2^*, l_2^*) = x_2^* - k \cdot l_2^{*\alpha} = (t_2 - t_1)y_1 + (1 - t_2)wl_2^* - kl_2^{*\alpha},
$$

\n
$$
U_3 = U(x_3^*, l_3^*) = x_3^* - k \cdot l_3^{*\alpha} = (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)wl_3^* - kl_3^{*\alpha}.
$$

It then follows that $l^* = l_2^*$ when $w > y_1^{\alpha-1}$ $\int k\alpha$ $(1-t_2)$ $\int_{0}^{\frac{1}{\alpha}}$ if and only if $U_2 \ge U_3$, otherwise we have $l^* = l_3^*$.

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

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

which can be simplified as

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

After substituting for l_i^* one obtains

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

$$
-(1-t_3)w\left[\frac{(1-t_3)w}{k\alpha}\right]^{\frac{1}{\alpha-1}} + k\left[\frac{(1-t_3)w}{k\alpha}\right]^{\frac{\alpha}{\alpha-1}}
$$

$$
\geq (t_3-t_2)y_2,
$$

that is

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

leading to

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

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

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

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

:

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

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

$$
\frac{\alpha}{(\alpha - 1)} \frac{(t_2 - t_3)}{\left(1 - t_3\right)^{\frac{\alpha}{\alpha - 1}} - (1 - t_2)^{\frac{\alpha}{\alpha - 1}}}\n<\n\frac{1}{(1 - t_2)^{\frac{1}{\alpha - 1}}},\n\frac{\alpha}{(\alpha - 1)}(t_2 - t_3) < \left(\frac{1 - t_3}{1 - t_2}\right)^{\frac{1}{\alpha - 1}}(1 - t_3) - (1 - t_2),\n\frac{\alpha}{(\alpha - 1)} \frac{(t_2 - t_3)}{(1 - t_2)} <\n\left(\frac{1 - t_3}{1 - t_2}\right)^{\frac{1}{\alpha - 1}}\frac{(1 - t_3)}{(1 - t_2)} - 1,\n1 + \frac{\alpha}{(\alpha - 1)} \frac{(t_2 - t_3)}{(1 - t_2)} <\n\left(\frac{1 - t_3}{1 - t_2}\right)^{\frac{\alpha}{\alpha - 1}}.
$$

Let $\delta = \frac{t_2-t_3}{1-t_2}$ $\frac{t_2-t_3}{1-t_2} > 0, \frac{1-t_3}{1-t_2}$ $\frac{1-t_3}{1-t_2} = 1 + \delta$ 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 \hat{w} converges to $\left[y_2^{\alpha-1} \frac{k\alpha}{(1-t_1)^{1+\alpha}} \right]$ $(1-t_2)$ $\int_{0}^{\frac{1}{\alpha}}$ from below.

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

Remark 8 If
$$
t_2 > t_3
$$
, (i) $\hat{w} < \left[y_2^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}}$, (ii) \hat{w} is increasing in t_3 , and (iii) $\lim_{t_3 \to t_2^-} \hat{w} = \left[y_2^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}}$.

It could however be possible that $\widehat{w} < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t)} \right]$ $(1-t_2)$ $\int_{0}^{\frac{1}{\alpha}}$, that is the threshold \hat{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 $Y_2 \backslash y_1$. All gross incomes are therefore in the non adjacent intervals $Y_1 \y_0$ and $Y_3 \y_2$. Given that $t_1 \leq t_3$ then in accordance with case A for all $w < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t)} \right]$ $(1-t_1)$ $\int_{0}^{\frac{1}{\alpha}}$ we have $l^* = l_1^* = \left[\frac{(1-t_1)w}{k\alpha} \right]_{0}^{\frac{1}{\alpha-1}}$ and $y^* = y_1^* = w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_1)}{k\alpha} \right]_{0}^{\frac{1}{\alpha-1}}$ with $y_1^* \in Y_1 \backslash y_0$.

If $\widehat{w} < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t)} \right]$ $(1-t_2)$ $\int_{0}^{\frac{1}{\alpha}}$ then for all wages where $w \geq \hat{w}$ 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 $Y_3 \ Y_2$, 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 $\hat{w} < \left[y_1^{\alpha-1}\frac{k\alpha}{(1-t)}\right]$ $(1-t_2)$ $\int_{0}^{\frac{1}{\alpha}}$ and $w \in$ $\int \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t)} \right]$ $(1-t_1)$ $\Big]^{\frac{1}{\alpha}}$; \widehat{w} \setminus :

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

$$
U_1 = U((1 - t_1) y_1, y_1/w) = (1 - t_1) y_1 - k \cdot (y_1/w)^{\alpha},
$$

\n
$$
U_3 = U(x_3^*, l_3^*) = (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)wl_3^* - kl_3^{*\alpha}.
$$

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

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

The condition can be simplified into

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

that is

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

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

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

Case B.1. Let $\hat{w} := k^{\frac{1}{\alpha}} (\alpha - 1)^{\frac{1}{\alpha}} \frac{\alpha}{(\alpha - 1)!}$ $\frac{\alpha}{(\alpha-1)}\left[\frac{(t_2-t_3)y_2}{\left((1-t_3)^{\frac{\alpha}{\alpha-1}}-(1-t_2)^{\frac{\alpha}{\alpha-1}}\right)}\right]$ $\frac{\alpha-1}{\alpha}$, and assume that $\hat{w} \geq$ $\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t)} \right]$ $(1-t_2)$ $\left.\begin{array}{l} \frac{1}{\alpha} \\ \end{array} \right.$ It follows that

$$
y^* = \begin{cases} w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_1)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} \\ y_1 & \text{if } w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_1)} \right]^{\frac{1}{\alpha}} ; \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} \right) \\ w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_2)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w \in \left[\left[y_1^{\alpha-1} \frac{k\alpha}{(1-t_2)} \right]^{\frac{1}{\alpha}} ; \widehat{w} \right] \\ w^{\frac{\alpha}{\alpha-1}} \left[\frac{(1-t_3)}{k\alpha} \right]^{\frac{1}{\alpha-1}} & \text{if } w > \widehat{w} \end{cases} \tag{2.36}
$$

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

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

:

:

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

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

By using the definition of the gross income in (2.16) 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 $\hat{y} \alpha^{\frac{1}{\alpha-1}} = \frac{\hat{w}^{\frac{\alpha}{\alpha-1}}}{k^{\frac{1}{\alpha-1}}}$ $\frac{w^{\alpha-1}}{k^{\alpha-1}}$ that is after substituting

$$
\widehat{y} = \frac{\alpha}{\alpha - 1} \left[\frac{(t_2 - t_3)y_2}{\left((1 - t_3)^{\frac{\alpha}{\alpha - 1}} - (1 - t_2)^{\frac{\alpha}{\alpha - 1}} \right)} \right]
$$
\n
$$
= (1 + \varepsilon) \left[\frac{(t_2 - t_3)y_2}{\left((1 - t_3)^{(1 + \varepsilon)} - (1 - t_2)^{(1 + \varepsilon)} \right)} \right]
$$

It then follows that the post tax gross income distribution is

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

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

Case B.2. Suppose that $\widehat{w} < \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t)^{\alpha}} \right]$ $(1-t_2)$ $\int_{0}^{\frac{1}{\alpha}}$. Let \tilde{w} denote the solution of

$$
(1-t_2)y_1 + (t_2-t_3)y_2 = k \cdot y_1^{\alpha} \cdot \tilde{w}^{-\alpha} + (1-t_3)^{\frac{\alpha}{\alpha-1}} \cdot \tilde{w}^{\frac{\alpha}{\alpha-1}} \left(\frac{1}{k\alpha}\right)^{\frac{1}{\alpha-1}} \left(\frac{\alpha-1}{\alpha}\right)
$$
(2.37)

such that $\tilde{w} \in$ $\int \left[y_1^{\alpha-1} \frac{k\alpha}{(1-t)} \right]$ $(1-t_1)$ $\left] ^{\frac{1}{\alpha}},\widehat{w}\right.$ $\overline{ }$: 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 i 1

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

:

At the same time, by using (2.16) and substituting in the implicit definition of \tilde{w} we have that \widetilde{y} is the solution of:

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

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

Then the post tax gross income distribution is

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

Chapter 3

Optimal Redistribution with Non-Welfarist Objectives

We consider optimal income taxation redistributive schemes under non-welfarist objectives. We derive theoretical and simulation results that highlight the different impact of redistributive income policies under income inequality reducing objectives compared to polarization reducing ones. The analyzed mechanism considers piecewise linear income taxation schemes supplemented with lump sum transfers (taxation or subsidy). The sign of these transfers is determined by the combination of the level of gross income dispersion and the value of labour supply elasticity.

With two income brackets the optimal tax system reducing inequality exhibits a proportional taxation with a no-tax area. While in case of polarization concerns the optimal tax system is proportional with zero top marginal tax rate.

With three income brackets the optimal tax system for inequality concerns is convex unless when the level of initial dispersion is low or the labour supply elasticity is high. As to polarization reduction the optimal tax system is non-convex with the maximal admissible tax rate within the central bracket and zero in the two external ones.

3.1 Introduction

Since its foundation (Mirrlees 1971), the optimal taxation theory has investigated the shape of the optimal tax schedule under different possible tax configurations. The optimal tax system is obtained through the maximization of a social welfare function subject to a set of constraints, dealing with the amount of required revenue and the agents' reactions to taxation. These exercises consider a social welfare function based over individuals' utility and the focus is on the efficiency costs of taxation.

Typically, little importance is given to the effect of differences in the redistributive objectives on the shape of the optimal tax formula. Moreover, the focus on individuals' utility seems to be not the most appropriate, since utility is only one aspect of individuals' welfare and policy makers, instead of maximizing the sum of individuals' utility, could aim to reduce the level of social indicators like inequality, poverty or polarization, which are defined in terms of income and not utility.

Hence, the traditional welfarist approach is not the proper way to investigate the effect of different redistributive objectives on the shape of the optimal tax system. Therefore, in order to provide a justification for income redistribution, we adopt an alternative approach, proposed by Kanbur et al. (1994) and known as non-welfarist.

In particular, in this paper we consider a set of piecewise linear taxation schemes supplemented with lump-sum transfers (taxation or subsidies) and we analyze the impact of these schemes in terms of a rank-dependent social evaluation function with different distributive objectives. More specifically, in line with the literature on income inequality measurement we will consider two families of rank-dependent evaluation functions defined over net incomes, that could incorporate either concerns for inequality reduction or concerns for income polarization reduction.

Our results supplement those presented in Chapter 2 whose aim was to highlight the shape of optimal taxation schemes when redistribution is not allowed. The focus there was only on the socially desirable mechanism that guarantees to collect a given level of per-capita revenue. Here, the mechanisms are integrated with lump-sum taxation and subsidies that provide the basic tools for redistribution within the piecewise linear taxation schemes.

The results we obtain are completely different from those derived when redistribution is not

allowed. A first intuition could be obtained when considering fixed labour supply. In particular, in this case the optimal tax system requires to tax all incomes with the highest admissible tax rate and then redistribute them equally, eventually keeping the share of income necessary to cover the revenue requirement. This conceivable result holds irrespective of whether the social evaluation function is inequality or polarization sensitive and represents a remarkable difference with respect to the case without redistribution, where there exist two clear and welldefined patterns of the tax rates as the focus shifts from inequality to polarization concerns. More specifically, with fixed labour supply and no redistribution, in case of inequality concerns the socially desirable mechanism collecting a given revenue requirement exhibits an income threshold with no taxation below and the maximal admissible tax rate above. While the optimal tax system associated with a polarization sensitive social evaluation function is such that the marginal tax rate is the maximal admissible within a central interval including also the median income, and zero outside.

When we introduce labour supply elasticity the scenario becomes less obvious. In this case, indeed, the different redistributive objectives play a relevant role in determining the shape of the optimal tax system, and we analyze this role comparing different tax regimes.

In particular, in the case of a two brackets piecewise linear tax system we have that for the inequality based social evaluation functions the optimal tax system requires a no-taxation area below a given threshold, and proportional taxation above. The tax rate is decreasing in the level of labour supply elasticity, while the exemption area is increasing. As to polarization reduction, the optimal tax system is based on a proportional taxation for all incomes below a given threshold, above which the marginal tax rate is set equal to zero. Both the tax rate and the income threshold are decreasing in the level of elasticity.

With three income brackets, the optimal tax system reducing inequality is convex with increasing marginal tax rates, unless when the level of initial gross income dispersion is not very high or when labour supply elasticity is high. For polarization concerns the optimal tax system is mainly non-convex. In both cases, proportional taxation is supplemented with a lump-sum transfer, whose sign is determined by the combination of the level of labour supply elasticity and the index of initial gross income dispersion.

Generally the lump-sum taxation is more likely to dominate proportional taxation for po-

larization sensitive social evaluation functions than for the inequality ones.

The remainder of the paper proceeds as follows: the next section describes the agentsí optimization problem and introduces the linear rank-dependent social evaluation function. In Section 3 we formalize the optimal redistributive tax problem and derive the socially desirable tax schedule considering alternative configurations with different tax rates regimes. Section 4 concludes.

3.2 Setting

In this section, first we formalize the agents' optimization problem under the assumption of a piecewise linear three brackets tax system. As demonstrated in Chapter 2, this tax schedule represents the easiest way to highlight the differences in terms of marginal tax rates driven by the governmentís shift from inequality to polarization concerns. Then, we present the linear rank-dependent social evaluation function and show how the policy maker weights individuals net incomes according to its specific non-welfarist objective.

3.2.1 The agents' optimization problem

For most derivations we consider agents endowed with quasi-linear preferences between consumption and leisure exhibiting constant labour supply wage elasticity. In particular, agents make labour supply decisions based on the constrained optimization of the following function

$$
U(x,l) = x - \phi(l),\tag{3.1}
$$

where $x \in \mathbb{R}$ denotes the net disposable income/consumption and $l \in [0, L]$ is the labour supply. The function $\phi : [0, L] \to \mathbb{R}$ 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

$$
\phi(l) = k \cdot l^{\alpha} \tag{3.2}
$$

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 $y = w l$. Agents are subject to taxation formalized by the tax schedule $T(y) \geq 0$, that leads to the net disposable income, considered in their utility function, obtained as $x = y - T(y)$.

The tax schedule is piecewise linear, with three income brackets identified by two gross income thresholds $y_1 < y_2$ and three marginal tax rates $t_1, t_2, t_3 \in [0, 1]$. Formally

$$
T(y) := \begin{cases} t_1y & \text{if } y \in [0, y_1) \\ t_1y_1 + t_2(y - y_1) & \text{if } y \in [y_1, y_2) \\ t_1y_1 + t_2(y_2 - y_1) + t_3(y - y_2) & \text{if } y \ge y_2 \end{cases}
$$

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

$$
x := y - T(y) = \begin{cases} (1 - t_1)y & \text{if } y \in Y_1 \equiv [0, y_1) \\ (t_2 - t_1)y_1 + (1 - t_2)y & \text{if } y \in Y_2 \equiv [y_1, y_2) \\ (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)y & \text{if } y \in Y_3 \equiv [y_2, \infty) \end{cases}
$$
(3.3)

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 gross income levels 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

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

when $y^* \in Y_i \setminus y_{i-1}$, where the function $\phi'^{-1}(.)$ by construction is positive and strictly increasing.

Given the definition of $y = w_l$, one obtains also the associated optimal labour supply

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

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

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

In order to simplify the exposition, and in line with the results obtained in Chapter 2, we consider a piecewise tax system with three income brackets and three marginal tax rates. Depending on what case is considered with respect to the value of t_3 compared to $t_1 \leq t_2$, we could either have (if $t_1 \le t_2 \le t_3$) that some agents experience the same gross income coinciding with one of the thresholds y_1 and y_2 , or (as under the case where $t_1 \le t_3 < t_2$) that this could happen for y_1 while around y_2 the map of y^* w.r.t. w is discontinuous, but still increasing.

When necessary we will consider in details these issues when $\phi(l) = k \cdot l^{\alpha}$ with $\alpha > 1$.

Recall that in this case the condition $MRS_{yx} = (1 - t_i)$ requires that

$$
y^* = w \cdot \left[\frac{(1 - t_i)w}{k\alpha} \right]^{\frac{1}{\alpha - 1}} = w^{\frac{\alpha}{\alpha - 1}} \left[\frac{(1 - t_i)}{k\alpha} \right]^{\frac{1}{\alpha - 1}}
$$
(3.4)

$$
l^* = \left[\frac{(1 - t_i)w}{k\alpha} \right]^{\frac{1}{\alpha - 1}}
$$

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. w is constant and equals $\frac{1}{\alpha-1}$.

In this paper we will consider a simplified exposition of the problem taking as reference distribution the gross income distribution in absence of taxation. That is we consider $t_i = 0$ and derive $y^* = w^{\frac{\alpha}{\alpha-1}} \left[\frac{1}{k\alpha}\right]^{\frac{1}{\alpha-1}}$ 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 the gross wages ranked in non-decreasing order. It then follows that the following monotonically increasing transformation of the wage

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

denotes the gross income of this individual under the assumption of no-taxation, with the

associated labour supply $l(p) = \left\lceil \frac{w(p)}{k\alpha} \right\rceil^{\frac{1}{\alpha-1}}$.

We will analyze the redistributive schemes taking as reference the gross income distribution formalized by the quantile (or inverse distribution) function $y(p)$.

According to (3.4) when considering a linear taxation scheme with a unique marginal tax rate t the associate gross income distribution will generate the following quantile distribution

$$
y_t(p) = w^{\frac{\alpha}{\alpha - 1}} \left[\frac{1}{k\alpha} \right]^{\frac{1}{\alpha - 1}} [(1 - t)]^{\frac{1}{\alpha - 1}} = y(p)(1 - t)^{\frac{1}{\alpha - 1}}
$$
(3.5)
= $y(p)(1 - t)^{\varepsilon}$

for $p \in [0, 1]$.

As special cases we will get that when labour supply is not elastic $(\varepsilon = 0) y_t(p) = y(p) = w(p)$ with $l(p) = 1$. While, if $\alpha = 2$ then $\varepsilon = 1$ and $y_t(p) = y(p)(1-t)$ with $y(p) := w(p)^2 \frac{1}{2k}$ and $l(p) = w(p) \frac{1}{2l}$ $\frac{1}{2k}$.

3.2.2 Social evaluations

The redistributive schemes are evaluated according to rank-dependent social evaluation functions defined over the distribution of the net incomes, in line with the exposition presented in Chapter 2.

Let $F(y)$ denote the cumulative income distribution function with quantile function $y(p)$ inf $\{y : F(y) \geq p\}$. The rank-dependent social evaluation function [SEF] (see Yaari, 1987, 1988) and Weymark, 1982) aggregates incomes weighted according to weights $v (p) \geq 0$ for $p \in [0, 1]$ that depend on the individuals' position $p \in [0, 1]$ in the income ranking, it is expressed as

$$
W_v(F) = \int_0^1 v(p) y(p) dp,
$$
\n(3.6)

where $\int_0^1 v(p) dp = 1.$

The positional welfare weights could formalize different distributional concerns for the social evaluation. In fact, as argued in Chapter 2, we could consider two families of weights that represent inequality concerns or polarization concerns.

As special cases of these two families of weights we can consider those where $v(p) := v_G(p)$

with

$$
v_G(p) = \begin{cases} 1 - \left[-2\left(\frac{1}{2} - p\right) \right] \text{ if } p \le \frac{1}{2} \\ 1 - 2\left(p - \frac{1}{2}\right) \text{ if } p \ge \frac{1}{2} \end{cases},\tag{3.7}
$$

that formalize inequality reducing concerns. These weights could be compared to those formalizing polarization concerns that are represented by the function

$$
v_P(p) = \begin{cases} 2p+1 \text{ if } p \le \frac{1}{2} \\ 2p-1 \text{ if } p \ge \frac{1}{2} \end{cases}.
$$
 (3.8)

For both weighting functions we can derive the cumulative weights obtained as

$$
V(p) := \int_0^p v(t) dt.
$$

The weight $V(p)/p$ denotes the average weight for the income of the individuals covering the poorer p quantiles of the population, while the weight $\left[1 - V(p)\right] / \left[1 - p\right]$ considers the average social weight for the income of those in the upper $1-p$ proportion of the population. Both for $v_G(p)$ and $v_P(p)$ we have that $V(p)$ is increasing with $V(p) > p$ and $[1 - V(p)] < [1 - p]$ for $p \in (0, 1)$. However $V_G(p)$ is concave, while $V_P(p)$ is convex in the interval $p \in (0, 1/2)$ and in the interval $p \in (1/2, 1)$. In fact $V_G(p) := 1 - (1 - p)^2$ and therefore

$$
\frac{V_G(p)}{p} := \frac{1 - (1 - p)^2}{p} = 2 - p, \text{ and } \frac{1 - V_G(p)}{1 - p} = 1 - p;
$$

it follows that $V_G(p)/p$ is decreasing and linear w.r.t. p and $\frac{1-V_G(p)}{1-p}$ is increasing and linear w.r.t. $(1 - p)$ for all $p \in (0, 1)$.

While for $V_P(p)$ we have that

$$
V_P(p) := \begin{cases} p^2 + p & \text{if } p \le \frac{1}{2} \\ (1-p)^2 + p & \text{if } p \ge \frac{1}{2} \end{cases}
$$

therefore

$$
V_P(p)/p := \begin{cases} p+1 & \text{if } p \le \frac{1}{2} \\ p+1/p-1 & \text{if } p \ge \frac{1}{2} \end{cases}, \text{ and } \frac{1-V_P(p)}{1-p} := \begin{cases} 2+p-\frac{1}{1-p} & \text{if } p \le \frac{1}{2} \\ p & \text{if } p \ge \frac{1}{2} \end{cases}.
$$

¹For graphical representations of the weighting functions and further illustrations and discussions the reader is referred to Chapter 2.

Thus we have that $V_P(p)/p$ is increasing w.r.t. p for $p \in (0, 1/2)$ and then decreases in the interval $p \in (1/2, 1)$. On the other hand $\frac{1-V_P(p)}{1-p}$ is decreasing and linear w.r.t. $(1-p)$ for all $p \in (1/2, 1)$, and increasing and concave w.r.t. $(1 - p)$ for all $p \in (0, 1/2)$.

Under both approaches the social evaluation can be summarized by the mean income of the distribution $\mu(F)$ and a linear index of dispersion $D_v(F)$ dependent on the choice of the weighting function v . The SEF could then be decomposed as

$$
W_{v}(F) = \mu(F) [1 - D_{v}(F)].
$$

3.3 Solutions for optimal piecewise redistributive linear taxation

In this section we formalize the optimal redistributive tax problem faced by a non-welfarist government. The SEF is a general rank-dependent function W defined over net incomes, with generic non-negative positional weights $v(p)$ with

$$
W = \int_0^1 v(p) \left[y_T(p) - T \left(y_T(p) \right) \right] dp,
$$

where $y_T(p)$ denotes the quantile function or the inverse of the post-tax gross income distribution and $T(y_T(p))$ is the tax return associated with the post-tax gross income $y_T(p)$. The taxation scheme could also involve positive income transfers to individuals such as a generalized subsidy.

We will first provide the theoretical results for the case of pure redistributive linear taxation with lump-sum taxation and subsidies and then we move to consider redistributive taxation schemes that guarantee a required level of per-capita revenue.

3.3.1 Optimal linear taxation

In this subsection we consider the purely redistributive taxation in which the amount of tax revenue collected through a proportional taxation with marginal tax rate t is redistributed through a lump-sum subsidy S .

Hence, the taxation scheme is

$$
T\left(y\right) = -S + ty.
$$

Under this scheme the post-tax gross income satisfies $y_T(p) = (1-t)^{\epsilon} y(p)$. While the net post-tax income is $y_T(p) - T(y_T(p)) = (1 - t)^{\varepsilon + 1} y(p) + S.$

Let F_T denote the post-tax net-income distribution. If the total amount of collected revenue is redistributed as lump-sum subsidy the social welfare $W_v(F_T)$ becomes

$$
W_v(F_T) = \int_0^1 \left[(1-t)^{\varepsilon+1} y(p) + S \right] v(p) \, dp = \int_0^1 (1-t)^{\varepsilon+1} y(p) \, v(p) \, dp + S,
$$

where the per-capita subsidy is $S = \int_0^1 t (1-t)^{\varepsilon} y(p) dp$.

If we replace the definition of the lump-sum subsidy in the previous expression and given that $W_v(F) = \int_0^1 y(p) v(p) dp = \mu(1 - D_v(F))$ where F denotes the income distribution under no-taxation with average $\mu = \int_0^1 y(p) dp$, we have that the social welfare with lump-sum redistribution is

$$
W_v(F) = \mu (1 - D_v(F))(1 - t)^{\varepsilon + 1} + \mu t (1 - t)^{\varepsilon}.
$$

By rearranging the terms we have that

$$
W_v(F) = \mu \left(1 - t\right)^{\varepsilon} - \mu D_v(F) \left(1 - t\right)^{\varepsilon + 1}.
$$

The $F.O.C.$ w.r.t. t is

$$
\frac{\partial W_v}{\partial t} = -\varepsilon (1-t)^{\varepsilon-1} + D_v(F) (1+\varepsilon) (1-t)^{\varepsilon} = 0.
$$

Then, the optimal tax rate does not depend on the level of average income that is

$$
t^* = 1 - \left(\frac{\varepsilon}{1+\varepsilon}\right) \frac{1}{D_v(F)}.
$$

In line with our intuition the proportional level of redistribution t^* is decreasing with the labour supply elasticity ε and is increasing in the level of dispersion $D_v(F)$.

Moreover, it is important to note the implications of the constraints on t^* on the level of labour supply elasticity ε and on the index of initial gross income dispersion D_v . In particular, when ε increases, in order to obtain an optimal tax rate bounded between 0 and 0.5 the level of initial dispersion has to be very high. For example, for very high level of labour supply elasticity $\varepsilon = 1$, the optimal tax rate falls within the admissible range if $D_v \in \left[\frac{1}{2}\right]$ $\frac{1}{2}$, 1]. When $\varepsilon = 0.5, D_v \in \left[\frac{1}{3}\right]$ $\frac{1}{3}$, $\frac{2}{3}$ $\frac{2}{3}$, while for $\varepsilon = 0.2$ then $D_v \in \left[\frac{1}{6}\right]$ $\frac{1}{6}, \frac{1}{3}$ $\frac{1}{3}$. Zero labour supply elasticity leads to a completely confiscatory tax rate of $t^* = 1$.

Moreover, by considering the weights v_G and v_P we can derive the optimal tax rate when the dispersion is formalized making use of respectively an inequality index I or a polarization index P. For a given gross income distribution under no-taxation F the two indices reach different values and therefore also the associated optimal level of taxation/redistribution t^* could differ. The following remark holds

Remark 9 For any F, $I(F) \ge P(F)$, that is

$$
t_I^* = 1 - \left(\frac{\varepsilon}{1+\varepsilon}\right) \frac{1}{I(F)} \ge t_P^* = 1 - \left(\frac{\varepsilon}{1+\varepsilon}\right) \frac{1}{P(F)}.
$$

The claim that $I(F) \geq P(F)$ could be proved directly by considering that the positional weights associated with the inequality index are decreasing for all $p \in [0, 1]$, while the weights associated with the polarization index are increasing for all $p \in [0, 1/2)$ and for all $p \in (1/2, 1]$ with the same average weight for both indices in the two intervals $[0, 1/2)$ and $(1/2, 1]$. Given that incomes are non-decreasing in p , then the social welfare will be higher for the polarization sensitive SEF. It follows that, for a given average income level the polarization index is (strictly) smaller than the inequality index. The two indices coincide in value only if all the incomes below the median are equal as well as those that are above the median.

Note that in the above remark we use the generic index of dispersion $I(F)$ because that result holds for any weighting function whose weights are decreasing in the individuals' positions, not necessarily in a linear way as for the weighting function considered for the Gini SEF.

Redistributive taxation and revenue constraints

Here we consider the case where the tax function denoted by $T(y)$ should generate a nonnegative per-capita amount of revenue G that is

$$
\int_0^1 T(y(p)) dp = G.
$$

Under this assumption we first investigate whether such a budget constraint could be satisfied with a lump-sum taxation instead of relying on proportional taxation.

Is there any room for lump-sum taxation? In this subsection we compare proportional tax scheme with lump-sum taxation and we derive the value of labour supply elasticity such that a lump-sum tax leads to a higher level of social welfare than a proportional taxation. We focus only on the proportional tax case and then, on a SEF sensitive to the dispersion reduction formalized through the index D_v , that could lead to a formalization of the inequality index I or of the polarization index P.

Let W^L the social welfare associated to a lump-sum tax regime, which is equal to

$$
W_v^L = W_v - G,\t\t(3.9)
$$

where W_v is the level of social welfare with no taxation and G is the amount of collected revenue, which is equal to $\mu t (1 - t)^{\epsilon}$ in the proportional tax case. The formula in (3.9) is derived considering that individuals' preferences do not exhibit income effects, therefore any lump-sum tax does not affect the labour supply, and the fact that the SEF is linear in incomes with average positional weight equal to 1.

Then, given the definition of G and by using the abbreviated form of the SEF we can rewrite (3.9) as

$$
W_v^L = \mu \left(1 - D_v \right) - \mu t \left(1 - t \right)^{\varepsilon},
$$

where D_v denotes the dispersion index calculated on the before tax income distribution. The

social welfare associated to a proportional tax system is

$$
W_v^P = W_v (1-t)^{\varepsilon+1},
$$

which we can rewrite as

$$
W_v^P = \mu \left(1 - D_v\right) \left(1 - t\right)^{\varepsilon + 1}.
$$

Then, a lump-sum taxation is preferred to a proportional tax system when

$$
(1 - D_v) - t (1 - t)^{\varepsilon} > (1 - D_v) (1 - t)^{\varepsilon + 1}
$$

:

Thus, when

$$
(1-D_v)\left[1-(1-t)^{\varepsilon+1}\right] > t\left(1-t\right)^{\varepsilon},
$$

which implies that

$$
(1 - D_v) > \frac{t (1 - t)^{\varepsilon}}{1 - (1 - t)^{\varepsilon + 1}},
$$

where the term on the right hand side is equal to $\frac{1}{1+\varepsilon}$ when t tends to zero. Hence, the next remark follows

Remark 10 The lump-sum taxation leads to a higher level of social welfare than a proportional tax system when

$$
\varepsilon > \frac{D_v}{1 - D_v}.
$$

As argued in the earlier section $I(F) \ge P(F)$, it then follows that $\frac{I(F)}{1-I(F)} \ge \frac{P(F)}{1-P(F)}$ $\frac{F(F)}{1-P(F)}$, thus lump-sum taxation is more likely to dominate proportional taxation for polarization sensitive evaluations than for inequality sensitive ones. In particular, the grey area illustrated in Figure 1 shows all the combinations of labour supply elasticity ε and dispersion level D_v such that the lump-sum taxation is socially preferred to a proportional scheme.

Fig.1. Combinations of ε and D_v leading to a lump-sum taxation.

Redistributive taxation with demo-grant

We consider here a more general taxation scheme with lump-sum transfers, linear taxation and a revenue constraint. Let

$$
W_v = \int v(p) y(p) (1-t)^{\varepsilon+1} dp - a \tag{3.10}
$$

denote the SEF expressed in terms of net incomes, where a is a demo-grant, which can be negative (positive) in case of lump-sum subsidy (lump-sum taxation). The government budget constraint is

$$
G = a + t \int y(p) (1 - t)^{\varepsilon} dp.
$$

A non-welfarist government who wants to collect a given revenue amount G maximizes (3.10) w.r.t. t given the constraint that the revenue equals G .

When $G = 0$ we obtain the optimal solution for the linear income tax case where $t^* =$ $1-\left(\frac{\varepsilon}{1+}\right)$ $1+\varepsilon$ $\begin{pmatrix} 1 \end{pmatrix}$ $\frac{1}{D_v(F)}$. However, in general if $G > 0$ the optimization problem requires to maximize the same objective function. That is the social decision requires that

$$
\max_{t} W(t, G) = \int v(p) y(p) (1-t)^{\varepsilon+1} dp + t \int y(p) (1-t)^{\varepsilon} dp - G.
$$

That can be rewritten as

$$
\max_{t} W(t, G) = \mu (1 - D_v) (1 - t)^{\varepsilon + 1} + t\mu (1 - t)^{\varepsilon} - G.
$$

Taking the $F.O.C.$ w.r.t. t

$$
\frac{\partial W(t, G)}{\partial t} = -\mu (1 - D_v) (1 - t)^{\varepsilon} (\varepsilon + 1) + \mu (1 - t)^{\varepsilon} - t\mu (1 - t)^{\varepsilon - 1} \varepsilon = 0.
$$

Dividing by $\mu (1-t)^{\varepsilon-1}$ one obtains $(1-t^*) = t^* \varepsilon + (1-D_v) (1-t^*) (\varepsilon+1)$ implying that

$$
(1-t^*) = \frac{\varepsilon}{(\varepsilon+1)} \frac{1}{D_v},
$$

for $t^* \in (0,1)$. So that we have

$$
t^* = 0 \text{ if } D_v \le \frac{\varepsilon}{(\varepsilon + 1)} \to a = G,
$$

leading to a lump-sum taxation, otherwise

$$
t^* \in (0,1) \text{ if } D_v > \frac{\varepsilon}{(\varepsilon+1)}.
$$

By using the optimal solution for the proportional tax case we can rewrite the budget constraint as $G = a + t^* (1 - t^*)^{\varepsilon} \int y(p) dp$ that is

$$
G = \left(1 - \frac{\varepsilon}{\varepsilon + 1} \frac{1}{D_v}\right) \left(\frac{\varepsilon}{\varepsilon + 1}\right)^{\varepsilon} \frac{1}{(D_v)^{\varepsilon}} \mu + a
$$

$$
a = G - \mu \cdot \varepsilon^{\varepsilon} \left(\frac{1}{(\varepsilon + 1) D_v}\right)^{\varepsilon + 1} (D_v - \varepsilon (1 - D_v)).
$$

Then, we obtain $a \leq 0$, that is we have a lump-sum subsidy if

$$
\frac{G}{\mu} \leq \varepsilon^{\varepsilon} \left(\frac{1}{(\varepsilon + 1) D_y} \right)^{\varepsilon + 1} (D_v - \varepsilon (1 - D_v))
$$
\n
$$
= \left(\frac{\varepsilon}{(\varepsilon + 1) D_v} \right)^{\varepsilon + 1} \left(\frac{(\varepsilon + 1) D_v}{\varepsilon} - 1 \right)
$$
\n
$$
= \left(\frac{\varepsilon}{(\varepsilon + 1) D_v} \right)^{\varepsilon} - \left(\frac{\varepsilon}{(\varepsilon + 1) D_v} \right)^{\varepsilon + 1}.
$$

Note that the term $\frac{G}{\mu}$ denotes the tax revenue expressed as a proportion of the average income computed in the case of no-taxation. Figure 2 illustrates the combinations of dispersion D_v and elasticity ε levels that identify the threshold where $a = 0$ for the three cases of percentage revenue $\frac{G}{\mu}$ equal respectively to 0%, 20% and 50%. These values are associated respectively with the bottom, the central and the top dashed curves in the graph. The values associated with a positive subsidy are those above the reference revenue curve. For the combinations below the curve the optimal taxation scheme involves the use of a lump-sum taxation. Note moreover that for all the combinations of ε and D_v that are below the bottom threshold line the optimal tax system envisages only a lump-sum taxation, while for all the combinations above the bottom threshold line we have a positive value of t^* . That is the tax scheme is such that the proportional taxation supplements the lump-sum transfer. For instance, in case of a revenue requirement $\frac{G}{\mu} = 0.2$ we have that all the combinations of values of ε and D_v that are below the bottom threshold line are associated with a lump-sum taxation. For the combinations that are comprised in between the two bottom threshold lines we have that a lump-sum taxation is combined with a proportional taxation, while for those coinciding with the central threshold line the revenue is obtained solely with proportional taxation. Finally, for the values above the central threshold the revenue requirement is covered combining proportional taxation with a lump-sum subsidy.

Fig.2. Combinations of ε and D_v affecting the sign of the lump-sum transfer.

3.3.2 Optimal two brackets redistributive taxation

In this section we move to a two brackets piecewise linear income tax scheme. We consider two different regimes which depend on the ranking of the two marginal tax rates. In particular we start with the case of increasing marginal tax rates with a no tax area for all incomes lower than a given threshold. Then, we focus on a tax schedule exhibiting decreasing marginal tax rates, where all incomes lower than a given threshold are subject to a proportional taxation, while for those above the marginal tax rate is set equal to zero. Hence, these incomes pay a lump-sum tax. For both regimes, we formalize the government's optimization problem, then we provide a quantitative illustration of the optimal solution by using numerical simulations.

Two brackets taxation with no tax area

In this subsection we consider a piecewise linear tax schedule $T(y)$ with a tax exemption area for gross incomes y lower than y_1 and a proportional taxation at rate t for incomes above y_1 that is:

$$
T(y) := \begin{cases} 0 & \text{if } y \in [0, y_1) \\ t(y - y_1) & \text{if } y \ge y_1 \end{cases}
$$

Let $y(p)$ denote the gross income quantile function with no taxation. Under the assumptions on the shape of $T(y)$ and on agents' preferences in (3.1) the associated post taxation gross income quantile function $y_t(p)$ is:

$$
y_t(p) := \begin{cases} y(p) & \text{if } y(p) < y_1 \\ y_1 & \text{if } y_1 \le y(p) < \frac{y_1}{(1-t)^{\varepsilon}} \\ y(p)(1-t)^{\varepsilon} & \text{if } \frac{y_1}{(1-t)^{\varepsilon}} \le y(p) \end{cases} \tag{3.11}
$$

:

or alternatively

$$
y_t(p) := \begin{cases} y(p) & \text{if } y(p) < y_1 \\ \max\{y(p)(1-t)^{\varepsilon}; y_1\} & \text{if } y_1 \le y(p) \end{cases} . \tag{3.12}
$$

Define $p_1^L := \inf\{p : y(p) = y_1\}$ and $p_1^H := \sup\{p : y(p)(1-t)^{\varepsilon} = y_1\}$. Note that p_1^H depends on t for a given ε . We have that for all individuals in the positions included in the interval $[p_1^L,$ p_1^H the gross income is y_1 .

Then the associated SEF based on net incomes is

$$
W_v = \int_0^{p_1^L} v(p) y(p) dp + y_1 \int_{p_1^L}^{p_1^H} v(p) dp + \int_{p_1^H}^{1} v(p) [y_1 + (1-t) (y (p) (1-t)^\varepsilon - y_1)] dp - a,
$$

with the revenue constraint

$$
G = a + t \int_{p_1^H}^1 [y(p) (1-t)^{\varepsilon} - y_1] dp.
$$

It then follows that the tax schedule optimization problem requires to derive

$$
\max_{t,y_1} \tilde{W}_v = \int_0^{p_1^L} v(p) y(p) dp + y_1 \int_{p_1^L}^{p_1^H} v(p) dp
$$
\n
$$
+ \int_{p_1^H}^1 v(p) \left[y(p) (1-t)^{\varepsilon+1} + ty_1 \right] dp
$$
\n
$$
+ t \int_{p_1^H}^1 \left[y(p) (1-t)^{\varepsilon} - y_1 \right] dp - G.
$$
\n(3.13)

The F.O.Cs. for the optimal level of t requires to compute $\frac{\partial \tilde{W}_v}{\partial t}$ and $\frac{\partial \tilde{W}_v}{\partial y_1}$. To simplify the exposition we assume that the cumulative distribution function is increasing with at most a finite number of discontinuities such that the r.h.s. and the l.h.s. of $\frac{\partial p_1^H}{\partial t}$, $\frac{\partial p_1^L}{\partial t}$ and of $\frac{\partial p_1^L}{\partial y_1}$, $\frac{\partial p_1^H}{\partial y_1}$ exist. Let $\mu_1 := \int_{p_1^H}^1 y(p) dp$ and $\mu_1(1-D_{v1}) := \int_{p_1^H}^1 v(p) y(p) dp$, that is we compute the average income and the abbreviated SEF for the distribution where for all $p < p_1^H$ all incomes are set equal to zero and coincide with $y(p)$ for $p \geq p_1^H$. We can then derive (the detailed calculations are illustrated in Appendix A1):

$$
\frac{\partial \tilde{W}_v}{\partial t} = (1-t)^{\varepsilon} (\varepsilon + 1) \mu_1 D_{v1} - [V(p_1^H) - p_1^H] y_1 - \mu_1 (1-t)^{\varepsilon - 1} \varepsilon,\n\frac{\partial \tilde{W}_v}{\partial y_1} = V (p_1^H) - V (p_1^L) + t(1 - V (p_1^H)) - t(1 - p_1^H).
$$

It then follows that the F.O.C. w.r.t. the tax rate $t \in (0,1)$ requires that

$$
\frac{\partial \tilde{W}_v}{\partial t} = 0 \to (1-t)^{\varepsilon} (\varepsilon + 1) D_{v1} - (1-t)^{\varepsilon-1} \varepsilon = \frac{\left[V(p_1^H) - p_1^H\right] y_1}{\mu_1},
$$

implying that

$$
(1-t)^{\varepsilon-1} [(1-t) (\varepsilon+1) D_{v1} - \varepsilon] = \frac{[V(p_1^H) - p_1^H] y_1}{\mu_1}.
$$
 (F.O.C. t)

While the F.O.C. for an internal solution w.r.t. y_1 requires to set $\frac{\partial \tilde{W}_v}{\partial y_1} = 0$, it then follows that

$$
\frac{\partial \tilde{W}_v}{\partial y_1} = 0 \to V(p_1^H) - V(p_1^L) = t \left[V(p_1^H) - p_1^H \right].
$$

Recall that by construction $V(p) > p$ for all $p \in (0, 1)$, it follows that

$$
t = \frac{V(p_1^H) - V(p_1^L)}{V(p_1^H) - p_1^H}.
$$
 (F.O.C. y_1)

Combining the two $F.O.Cs$ we have

$$
(1-t)^{\varepsilon-1} [(1-t) (\varepsilon+1) D_{v1} - \varepsilon] = [V(p_1^H) - p_1^H] \frac{y_1}{\mu_1},
$$

$$
1-t = \frac{V(p_1^L) - p_1^H}{V(p_1^H) - p_1^H}.
$$

A special cases. Before moving to the general solution it is useful to highlight the solution for the special case where labour supply is fixed, that is when $\varepsilon = 0$.

In this case $p_1^H = p_1^L = p_1$ and the partial derivatives w.r.t. t and y_1 of the objective function are

$$
\frac{\partial \tilde{W}_v}{\partial t} = \int_{p_1}^1 v(p) [y_1 - y(p)] dp + \int_{p_1}^1 [y(p) - y_1] dp
$$

=
$$
- \int_{p_1}^1 [v(p) - 1][y(p) - y_1] dp > 0
$$

=
$$
\mu_1 D_{v1} - [V(p_1) - p_1] y_1 > 0,
$$

for all $p_1 \in [0, 1)$, and

$$
\frac{\partial \tilde{W}_v}{\partial y_1} = t \int_{p_1}^1 \left[v(p) - 1 \right] dp = -t \left[V(p_1) - p_1 \right] < 0,
$$

for all $p_1 \in (0,1)$ and for all $t > 0$. As a result t^* equals the maximal admissible value of the marginal tax rate and $y_1 = 0$. So irrespective of whether the weighting function is inequality or polarization sensitive the optimal redistributive policy is to tax at 100% all incomes and redistribute them equally eventually keeping the share of income necessary to cover the revenue requirement G .

The general solution. Recall that by construction $\mu_1 D_{\nu1} > [V(p_1) - p_1] y_1$ and $V(p_1) - p_1 >$ 0 for all $p_1 \in (0, 1)$. Note also that for any p_1 by construction $D_{v1} > D_v$ both in terms of inequality and polarization evaluations, with $P_{v1} \leq I_{v1}$.

Moreover recall that the partial derivatives of the objective function are

$$
\frac{\partial \tilde{W}_v}{\partial t} = -\frac{\varepsilon \mu_1}{(1-t)^{1-\varepsilon}} + (1-t)^{\varepsilon} (\varepsilon + 1) \mu_1 D_{v1} - [V(p_1^H) - p_1^H] y_1
$$

$$
\frac{\partial \tilde{W}_v}{\partial y_1} = V(p_1^H) - V(p_1^L) - t [V(p_1^H) - p_1^H].
$$

Note that $\frac{\partial \tilde{W}_v}{\partial t}$ is decreasing in t for $\varepsilon \leq 1$, also verify that for $t = 0$ we have that

$$
\left. \frac{\partial \tilde{W}_v}{\partial t} \right|_{t=0} = \mu_1 D_{v1} - \varepsilon \mu_1 (1 - D_{v1}) - \left[V(p_1^H) - p_1^H \right] y_1,
$$

while for high values of t we obtain that $\frac{\partial \tilde{W}_v}{\partial t} < 0$. These conditions guarantee that for sufficiently small values of ε there exists a value of $t \in (0,1)$ s.t. $\frac{\partial \tilde{W}_v}{\partial t} = 0$. Moreover, for small values of D_{v1} and sufficiently large values of ε we could have that $\frac{\partial \tilde{W}_v}{\partial t}$ $\Big|_{t=0}$ < 0. If this is the case then $t^* = 0$. Recalling that $P_{v1} \leq I_{v1}$, it follows that for polarization evaluations with sufficiently high levels of ε the optimal marginal tax rate t^* equals 0.

Taking into account the partial derivative $\frac{\partial \tilde{W}_v}{\partial y_1}$, then when $t^* \in (0,1)$ we have that the optimal threshold should satisfy the F.O.C. where $\frac{\partial \tilde{W}_v}{\partial y_1} = 0$. The case where $y_1 = 0$ is ruled out as ε increases because in this case both t^* decreases and $V(p_1^H) - V(p_1^L)$ increases for a given y_1 .

Illustrative simulation results. A quantitative illustration of the optimal two brackets linear piecewise tax system with a tax exemption area at the bottom of the income distribution is presented in Tables 1 and 2 for Gini and Polarization SEFs respectively.²

Recall that the simulations are based on a Pareto distribution of gross incomes under no taxation which is bounded between 20 and 327, whose mean is equal to 48.04, while the level

²The simulations are generated according to the model presented in Chapter 2 which is summarized in Appendix B at the end of this chapter.

of Inequality and Polarization are equal to 0.37 and 0.11 respectively³.

$G\$	ε	$t\,$	y_1	\boldsymbol{a}	G	ε	$t\,$	y_1	\boldsymbol{a}
			21					21	
$0.1 \times \mu$	0.01	97%	11.3%	19.92	$0.2 \times \mu$	0.01	97%	11.3%	$15.12\,$
			4.5%					4.5%	
			25					25	
$0.1 \times \mu$	0.1	80%	42.4%	9.47	$0.2 \times \mu$	0.1	80%	42.4\%	4.67
			14.9%					14.9%	
			29					29	
$0.1 \times \mu$	$0.2\,$	66%	57.2%	4.60	$0.2 \times \mu$	$0.2\,$	66%	57.2%	-0.20
			15.7%					15.7%	
			36					36	
$0.1 \times \mu$	$0.4\,$	47%	70.6%	0.30	$0.2 \times \mu$	$0.4\,$	47\%	70.6%	-4.51
			13.3%					13.3%	

Table 1. Optimal tax system: Gini based SEF.

Note. The first two values reported in the column y_1 express the threshold in terms

of the income level and the associated percentile in the income distribution.

The third value represents the fraction of population with a gross income equal to y_1 .

 In order to illustrate the impact of the combinations of the initial gross income dispersion and the level of labour supply elasticity we consider alternative gross income distributions, with different levels of dispersion (inequality and polarization) which are obtained by changing the bounds of the Pareto distribution. The results of the simulations with these alternative distributions are reported in Appendix B at the end of this chapter.

$G\$	ε	$t\,$	y_1	\boldsymbol{a}	G	ε	t_{\rm}	y_1	\boldsymbol{a}
			23					23	
$0.1 \times \mu$	0.01	92%	21.1%	17.50	$0.2 \times \mu$	0.01	92%	21.1%	12.69
			2.9%					2.9%	
			29					29	
$0.1 \times \mu$	0.1	31%	44.5%	1.34	$0.2 \times \mu$	0.1	31\%	44.5%	-3.47
			4%					4%	
			31					31	
$0.1 \times \mu$	0.15	6%	47.5%	-3.62	$0.2 \times \mu$	0.15	6%	47.5%	-8.42
			0.6%					0.6%	
$0.1 \times \mu$	$0.2\,$	θ	$\overline{0}$	-4.80	$0.2 \times \mu$	0.2	$\overline{0}$	θ	-9.61

Table 2. Optimal tax system: Polarization based SEF.

In both Tables, the left (right) panel reports the simulations when the revenue requirement is equal to 10% (20%) of the average income. As anticipated in our analysis the design of the tax scheme, that is the optimal choice of y_1 and t , is independent of the revenue requirement G. This aspect is clear from the optimization problem in (3.13) where G does not affect the first order conditions for optimization. The the use of lump-sum taxation or subsidies depends on the difference between the revenue generated by the optimal taxation scheme compared to the revenue requirement. When this difference is positive (negative) the proportional taxation is supplemented with a lump-sum subsidy (tax).

More specifically, for the Gini based SEFs Table 1 shows that the threshold y_1 is increasing in the level of labour supply elasticity while the marginal tax rate t decreases.

In case of polarization sensitive SEFs (Table 2) the simulations generate higher values for y_1 and lower values for t for a given level of elasticity, (compare rows 1 and 2 of Tables 1 and 2). However, as elasticity increases the marginal tax rate sharply decreases (compare first three rows of Table 2) and for elasticity level $\varepsilon \geq 0.2$ the optimal marginal tax rate is 0 and the optimal taxation scheme reducing polarization is only based on lump-sum taxation.

Two brackets taxation with zero marginal tax rate at the top.

This subsection focuses on a two brackets income tax scheme with decreasing marginal tax rates. More specifically, gross incomes lower than y_1 are subject to a proportional taxation with marginal tax rate τ , while the marginal tax rate above this threshold is set equal to zero. That is, gross incomes greater than y_1 are subject to a lump-sum taxation. The tax function $\mathcal{T}(y)$ is:

$$
\mathcal{T}(y) := \begin{cases} \tau y & \text{if } y \in (0, y_1] \\ \tau y_1 & \text{if } y > y_1 \end{cases}
$$

:

Let $y(p)$ denote the gross income quantile function with no taxation. Under the assumption of the tax schedule $\mathcal{T}(y)$ the associated post taxation gross income quantile function $y_{\tau}(p)$ is

$$
y_{\tau}(p) := \begin{cases} y(p)(1-\tau)^{\varepsilon} & \text{if } y(p) < \widehat{y} \\ y(p) & \text{if } y(p) \ge \widehat{y} \end{cases},
$$

which is discontinuous at the income threshold \hat{y} defined as⁴

$$
\widehat{y} = (1 + \varepsilon) \left[\frac{\tau y_1}{1 - (1 - \tau)^{1 + \varepsilon}} \right].
$$

For $\tau > 0$ and $\varepsilon > 0$ the income threshold \hat{y} is always greater than y_1 , while if $\tau \to 0$ or $\varepsilon \to 0$ then $\hat{y} \to y_1$. Moreover, if $y(p) < \hat{y}$ then $y_\tau(p) < y_1$. With this tax system if $\tau > 0$ there are no gross incomes at the income threshold y_1 . Let $\hat{p} := \sup\{p : y(p) = \hat{y}\}\$ denote the position associated with the income threshold that separates the income recipients with gross incomes lower than y_1 from those with gross incomes above y_1 and therefore subject to 0 marginal tax rate. The associated SEF based on net incomes is

$$
W_v = \int_0^{\widehat{p}} v(p) y(p) (1 - \tau)^{\varepsilon + 1} dp + \int_{\widehat{p}}^1 v(p) [y(p) - \tau y_1] dp - a
$$

⁴This condition and all related to the agents' optimization problem are special cases of those discussed in details in Appendix B of the previous chapter for the three brackets tax function case.

with the revenue constraint

$$
G = a + \tau \int_0^{\widehat{p}} y(p) (1 - \tau)^{\epsilon} dp + \tau \int_{\widehat{p}}^1 y_1 dp.
$$

Then, substituting for a from the revenue constraint to the SEF, the social optimization problem becomes

$$
\max_{\tau, y_1} \hat{W}_v = \int_0^{\hat{p}} v(p) y(p) (1 - \tau)^{\varepsilon + 1} dp + \int_{\hat{p}}^1 v(p) [y(p) - \tau y_1] dp \n+ \tau \int_0^{\hat{p}} y(p) (1 - \tau)^{\varepsilon} dp + \tau \int_{\hat{p}}^1 y_1 dp - G.
$$
\n(3.14)

The F.O.Cs. for the optimal level of τ requires to compute $\frac{\partial \hat{W}_v}{\partial \tau}$ and $\frac{\partial \hat{W}_v}{\partial y_1}$.

To simplify the exposition we assume that the cumulative distribution function is increasing with at most a finite number of discontinuities such that the r.h.s. and the l.h.s. of $\frac{\partial \hat{p}}{\partial \tau}$ exist.

We can then derive (the detailed calculations are illustrated in Appendix A2):

$$
\frac{\partial \hat{W}_v}{\partial \tau} = -\frac{\partial \hat{p}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \tau} \tau y_1 \left[v \left(\hat{p} \right) \varepsilon + \frac{1 - \left(1 + \varepsilon \tau \right) \left(1 - \tau \right)^{\varepsilon}}{1 - \left(1 - \tau \right)^{1 + \varepsilon}} \right] \tag{3.15}
$$

$$
+[1-(1+\varepsilon)(1-\tau)^{\varepsilon}]\int_0^{\widehat{p}}\left[v\left(p\right)-1\right]y(p)dp\qquad \qquad (3.16)
$$

$$
-\varepsilon (1-\tau)^{\varepsilon-1} \int_0^{\widehat{p}} y(p) \, dp \tag{3.17}
$$

$$
+\int_{0}^{\widehat{p}}\left[v\left(p\right)-1\right]\left[y_{1}-y\left(p\right)\right]dp, \tag{3.18}
$$

and

$$
\frac{\partial \hat{W}_v}{\partial y_1} = -\frac{\partial \hat{p}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial y_1} \tau y_1 \left[v \left(\hat{p} \right) \varepsilon + \frac{1 - (1 + \varepsilon \tau) (1 - \tau)^{\varepsilon}}{1 - (1 - \tau)^{1 + \varepsilon}} \right]
$$
(3.19)

$$
+\tau\left(V\left(\widehat{p}\right)-\widehat{p}\right).\tag{3.20}
$$

Both partial derivatives are presented in order to decompose the effect related to the existence of a positive elasticity of labour supply and the effect holding with fixed labour supply. This latter effect is formalized in the last term of both partial derivatives.

A special cases. As also done for the previous taxation scheme we start by highlight the solution for the special case where labour supply is fixed, that is when $\varepsilon = 0$. In this case

$$
\frac{\partial \hat{W}_v}{\partial \tau} = \int_0^{\widehat{p}} \left[v(p) - 1 \right] \left[y_1 - y(p) \right] dp > 0 \tag{3.21}
$$

$$
\frac{\partial \hat{W}_v}{\partial y_1} = \tau (V(\hat{p}) - \hat{p}) > 0. \tag{3.22}
$$

Both partial derivatives are positive for all and for all y_1 and for all $\tau > 0$. As a result τ^* equals the maximal admissible value of the marginal tax rate and $y_1 = y^{\text{max}}$ that is the area with zero marginal tax rate at the top is eliminated and all individuals are subject to the maximal marginal tax rate. This result is irrespective of whether the weighting function is inequality or polarization sensitive. As for the previous tax scheme the optimal redistributive policy is to tax at 100% all incomes and redistribute them equally eventually keeping the share of income necessary to cover the revenue requirement G .

The general solution. If $\varepsilon > 0$ we obtain mitigating effects on the sign of the partial derivatives deriving from the distributive and distortionary welfare effects of taxation on labour supply.

We first consider the first term appearing on both partial derivatives that we denote by $v(\widehat{p})\,\varepsilon\,+\,f(\varepsilon,\tau)$ with $f(\varepsilon,\tau) := \frac{1-(1+\varepsilon\tau)(1-\tau)^{\varepsilon}}{1-(1-\tau)^{1+\varepsilon}}$ $\frac{(1+\epsilon^{\gamma})(1-\gamma)}{1-(1-\tau)^{1+\epsilon}}$. For $\epsilon > 0$, the term $f(\epsilon, \tau)$ is positive and increasing in τ for $\tau \in (0, 1)$, with $\lim_{\tau \to 0} f(\varepsilon, \tau) = 0$ and $f(\varepsilon, 1) = 1$, moreover it is also increasing in ε with $f(0, \tau) = 0$ and $\lim_{\varepsilon \to 0} f(\varepsilon, \tau) = 1$. Considering that $\frac{\partial \widehat{p}}{\partial \widehat{y}} \geq 0$, that $\frac{\partial \widehat{y}}{\partial \tau} = \widehat{y}$ $\left[\frac{1-(1+\tau \varepsilon)(1-\tau)^{\varepsilon}}{1-\tau} \right]$ $\tau [1-(1-\tau)^{1+\varepsilon}]$ $\Big] > 0$, that $\frac{\partial \widehat{y}}{\partial y_1} = \frac{(1+\varepsilon)\tau}{1-(1-\tau)^3}$ $\frac{(1+\varepsilon)\tau}{1-(1-\tau)^{1+\varepsilon}} = \frac{\hat{y}}{y_1} > 0$ and that $v(\hat{p})\varepsilon > 0$ for $\hat{p} \in [0,1)$ irrespective of whether we are considering the polarization or the inequality sensitive representation, we have that the first term in (3.15) and in (3.19) is non-positive. Moreover, the effect is increasing in $\varepsilon > 0$ because all the terms are increasing in ε .

If we consider first the F.O.C. $\frac{\partial \hat{W}_v}{\partial y_1} = 0$, we have that for an internal optimal value for y_1 it is necessary that

$$
\frac{\partial \widehat{p}}{\partial \widehat{y}} \frac{\partial \widehat{y}}{\partial y_1} y_1 \left[v \left(\widehat{p} \right) \varepsilon + \frac{1 - (1 + \varepsilon \tau) \left(1 - \tau \right)^{\varepsilon}}{1 - (1 - \tau)^{1 + \varepsilon}} \right] = V \left(\widehat{p} \right) - \widehat{p}.
$$
\n(3.23)
As $V(\hat{p}) - \hat{p} \to 0$ for $\hat{p} \to 1$, then for sufficiently high values of $\varepsilon > 0$ the F.O.C. should be satisfied for $y_1 > 0$. Moreover, considering that $v_P(\hat{p}) > v_G(\hat{p})$ for $\hat{p} \in (3/4, 1)$ and that $V_p(\hat{p}) - \hat{p} = (1 - \hat{p})^2 < V_G(\hat{p}) - \hat{p} = \hat{p}(1 - \hat{p})$ for $\hat{p} \in (1/2, 1)$ then for a given tax rate τ the optimal level of y_1 in the upper part of the distribution should be lower in case of concerns for polarization. It follows, in line with Mirrlees (1971) result that

Remark 11 When $\varepsilon > 0$, the optimal taxation system requires that at least the top incomes are taxed at a 0 marginal tax rate. For sufficiently low levels of elasticity the threshold y_1 , is lower in the case of polarization sensitive evaluations than for inequality sensitive evaluations.

By combining with the F.O.C. related to $\frac{\partial \hat{W}_v}{\partial \tau}$ one obtains

$$
\left[V\left(\widehat{p}\right) - \widehat{p}\right] \frac{\frac{\partial \widehat{y}}{\partial \tau}}{\frac{\partial \widehat{y}}{\partial y_1}} \tau = \left[1 - (1 + \varepsilon)(1 - \tau)^{\varepsilon}\right] \int_0^{\widehat{p}} \left[v\left(p\right) - 1\right] y(p) dp \tag{3.24}
$$

$$
-\varepsilon (1-\tau)^{\varepsilon-1} \int_0^{\widehat{p}} y(p) \, dp \tag{3.25}
$$

$$
+\int_{0}^{\widehat{p}}\left[v\left(p\right)-1\right]\left[y_{1}-y\left(p\right)\right]dp, \tag{3.26}
$$

where $\frac{\frac{\partial \hat{y}}{\partial \tau}}{\frac{\partial \hat{y}}{\partial y_1}}$ $=\frac{\widehat{y}\frac{1-(1+\tau \varepsilon)(1-\tau)^{\varepsilon}}{\tau [1-(1-\tau)^{1+\varepsilon}]}}{\widehat{y}}$ $\tau[1-(1-\tau)^{1+\varepsilon}]$ $\frac{\widehat{y}}{y_1}$ $= y_1 \frac{1-(1+\tau \varepsilon)(1-\tau)^{\varepsilon}}{\tau [1-(1-\tau)^{1+\varepsilon}]}$ $\frac{(- (1+\tau \varepsilon)(1-\tau)^{\varepsilon}}{\tau [1-(1-\tau)^{1+\varepsilon}]}, \text{ that is } \frac{\frac{\partial \hat{y}}{\partial \tau}}{\frac{\partial \hat{y}}{\partial y_1}}$ $= f(\varepsilon, \tau) \frac{y_1}{\tau}$ $\frac{y_1}{\tau}$. It follows that the associated internal optimal level of τ should satisfy

$$
\left[V\left(\widehat{p}\right) - \widehat{p}\right]f(\varepsilon,\tau)y_1 = \left[1 - (1+\varepsilon)\left(1-\tau\right)^{\varepsilon}\right]\int_0^{\widehat{p}} \left[v\left(p\right) - 1\right]y(p)dp \tag{3.27}
$$

$$
-\varepsilon (1-\tau)^{\varepsilon-1} \int_0^{\widehat{p}} y(p) \, dp \tag{3.28}
$$

$$
+\int_{0}^{\widehat{p}}\left[v\left(p\right)-1\right]\left[y_{1}-y\left(p\right)\right]dp.\tag{3.29}
$$

This condition can be rearranged as follows

$$
\left[V\left(\widehat{p}\right) - \widehat{p}\right]f(\varepsilon,\tau)y_1 = -\varepsilon(1-\tau)^{\varepsilon}\int_0^{\widehat{p}} \left[v\left(p\right) - 1\right]y(p)dp \tag{3.30}
$$

$$
-\varepsilon (1-\tau)^{\varepsilon-1} \int_0^{\widehat{p}} y(p) \, dp \tag{3.31}
$$

$$
+\int_{0}^{\widehat{p}} \left[v\left(p\right) -1 \right] \left[y_{1} - y\left(p\right) (1-\tau)^{\varepsilon} \right] dp, \tag{3.32}
$$

where the term on the l.h.s is always positive, while the last term on the r.h.s. is also always positive by construction. The derivation of the optimal level of $\tau \in (0,1)$ could be obtained through simulations. We illustrate them in the next subsection by showing that polarization SEFs exhibit lower level of τ for any $\varepsilon > 0$.

Illustrative simulation results. Tables 3 and 4 provide a quantitative illustration of the optimal tax system with zero marginal tax rate at the top for Gini and Polarization SEF respectively.

G	ε	τ	y_1	\boldsymbol{a}	$G\,$	ε	τ	y_1	\boldsymbol{a}
$0.1 \times \mu$	0.01	97%	318	40.02	$0.2 \times \mu$	0.01	97%	318	35.40
			99.9%					99.9%	
$0.1 \times \mu$	0.1	76%	284	26.91	$0.2 \times \mu$	0.1	76%	284	22.11
			99.6%					99.6%	
$0.1 \times \mu$	$0.2\,$	56%	280	18.09	$0.2 \times \mu$	$0.2\,$	56%	280	13.29
			99.5%					99.5%	
$0.1 \times \mu$	$0.5\,$	10%	312	-0.25	$0.2 \times \mu$	$0.5\,$	10%	312	-5.05
			99.8%					99.8%	
$0.1 \times \mu$	0.6	θ	$\overline{0}$	-4.80	$0.2 \times \mu$	0.6	$\overline{0}$	θ	-9.61

Table 3. Optimal tax system: Gini based SEF.

$\mathcal G$	ε	τ	y_1	\boldsymbol{a}	G	ε	τ	y_1	\boldsymbol{a}
$0.1 \times \mu$ 0.01		93%	128	35.68	\parallel 0.2 \times μ 0.01		93%	128	30.87
			94.4%					94.4%	
$\vert 0.1 \times \mu \vert 0.1$		42%	68		10.86 $\ $ 0.2 $\times \mu$	0.1	42%	68	6.05
			83.7%					83.7%	
$\pm 0.1 \times \mu$	$0.2\,$	θ	θ	-4.80	0.2 $\times \mu$	$0.2\,$	$\overline{0}$	$\overline{0}$	-9.61

Table 4. Optimal tax system: Polarization based SEF.

Analogously to the case of no taxation at the bottom of the income distribution, the choice of the optimal tax system with zero marginal tax rate at the top turns out to be independent of the revenue requirement G (compare the left and right panels of Tables 3 and 4). For the Gini sensitive SEF the optimal system requires to tax almost the entire gross income distribution with a marginal tax rate which is decreasing as elasticity increases. The marginal tax rate is set equal to zero only for the extreme right tail of the distribution. This finding echoes the celebrated Mirrleesí (1971) result of no taxation at the top of the income distribution. For levels of elasticity $\varepsilon \geq 0.6$ the optimal tax system consists of a lump-sum taxation.

When the focus shifts to polarization concerns the optimal tax system envisages lower values for τ and y_1 for each elasticity level (compare the first two rows of Tables 3 and 4). Then, the optimal tax system is based on lump-sum taxation for elasticity level $\varepsilon \ge 0.2$.

Therefore, a higher social welfare is associated to a tax system with a no taxation area (for Gini based SEFs) and with zero top marginal tax rate (for polarization based SEFs).

3.3.3 Optimal three brackets redistributive taxation

In this section we move to a piecewise linear tax system with three income brackets identified by two income thresholds $y_1 \lt y_2$ and three marginal tax rates $t_1, t_2, t_3 \in [0, 1)$. In line with Chapter 2, we focus only on tax schedule where $t_1 \leq t_2$, and then we assume two possible tax rates regimes, i.e. convex and non-convex, depending on the ranking of t_2 and t_3 . In particular, the convex regime is such $t_1 \leq t_2 \leq t_3$, while non-convex case deals with configurations where $t_1 \le t_3 \le t_2.$

We formalize the optimal tax problem faced by the non-welfarist government, then we provide numerical simulations to illustrate the optimal tax system for both regimes under the case of SEFs sensitive to inequality or to polarization reduction.

Three brackets convex tax system

Under the convex regime marginal tax rates are increasing and the tax function $T(y)$ is

$$
T(y) := \begin{cases} t_1y & \text{if } y \in [0, y_1) \\ t_1y_1 + t_2(y - y_1) & \text{if } y \in [y_1, y_2) \\ t_1y_1 + t_2(y_2 - y_1) + t_3(y - y_2) & \text{if } y \ge y_2 \end{cases}
$$

Then, from the agents' optimization problem described in Section 3.2.1 we obtain that the associated post tax gross income quantile function $y_t(p)$ is

$$
y_{t}(p) = \begin{cases} y(p) (1-t_{1})^{\varepsilon} & if y(p) < \frac{y_{1}}{(1-t_{1})^{\varepsilon}} \\ y_{1} & if \frac{y_{1}}{(1-t_{1})^{\varepsilon}} \leq y(p) \leq \frac{y_{1}}{(1-t_{2})^{\varepsilon}} \\ y(p) (1-t_{2})^{\varepsilon} & if \frac{y_{1}}{(1-t_{2})^{\varepsilon}} < y(p) \leq \frac{y_{2}}{(1-t_{2})^{\varepsilon}} \\ y_{2} & if \frac{y_{2}}{(1-t_{2})^{\varepsilon}} \leq y(p) \leq \frac{y_{2}}{(1-t_{3})^{\varepsilon}} \\ y(p) (1-t_{3})^{\varepsilon} & if y(p) > \frac{y_{2}}{(1-t_{3})^{\varepsilon}} \end{cases}
$$

:

With this tax regime some agents experience a gross income equal to one of the thresholds y_1 or y_2 . In particular, define $p_1^L := \inf\{p : y(p)(1-t_1)^{\varepsilon} = y_1\}, p_1^H := \sup\{p : y(p)(1-t_2)^{\varepsilon} = y_1\},$ $p_2^L := \inf\{p : y(p)(1-t_2)^{\varepsilon} = y_2\},\$ and $p_2^H := \sup\{p : y(p)(1-t_3)^{\varepsilon} = y_2\},\$ we have that all individuals covering the positions included in the interval $\left[p_1^L, p_1^H\right]$ have a gross income equal to y_1 , while for all individuals in the positions $p \in [p_2^L, p_2^H]$ the gross income is y_2 .

The SEF based on net incomes is

$$
W_v = \int_0^{p_1^L} v(p) y(p) (1 - t_1)^{\varepsilon + 1} dp + y_1 (1 - t_1) \int_{p_1^L}^{p_1^H} v(p) dp
$$

+
$$
\int_{p_1^H}^{p_2^L} v(p) [y(p) (1 - t_2)^{\varepsilon + 1} + y_1 (t_2 - t_1)] dp
$$

+
$$
[y_2 (1 - t_2) + y_1 (t_2 - t_1)] \int_{p_2^L}^{p_2^H} v(p) dp
$$

+
$$
\int_{p_2^H}^{1} v(p) [y(p) (1 - t_3)^{\varepsilon + 1} + y_2 (t_3 - t_2) + y_1 (t_2 - t_1)] dp - a,
$$

with the revenue constraint

$$
G = a + t_1 \int_0^{p_1^L} y(p)(1 - t_1)^{\varepsilon} dp + t_1 \int_{p_1^L}^1 y_1 dp
$$

+ $t_2 \int_{p_1^H}^{p_2^L} [y(p)(1 - t_2)^{\varepsilon} - y_1] dp + t_2 \int_{p_2^L}^1 (y_2 - y_1) dp$
+ $t_3 \int_{p_2^H}^1 [y(p)(1 - t_3)^{\varepsilon} - y_2] dp.$

We solve the social optimization problem numerically and report in Table 5 the results with fixed labour supply for Gini and Polarization based SEFs respectively. Note that we assume that the marginal tax rates can not exceed an upper limit which we set equal to 50%.

			Gini based SEF			Polarization based SEF						
G	t_1	t_2	t_3	y_1	y_2	\boldsymbol{a}	t_1	t_2	t_3	y_1	y_1	\boldsymbol{a}
$0.10 \times \mu$	θ	$\overline{0}$	50%	20	20	9.22	$\overline{0}$	$\overline{0}$	50%	$20\,$	20	9.22
				θ	$\overline{0}$					$\overline{0}$	θ	
$0.15 \times \mu$	$\overline{0}$	$\overline{0}$	50%	20	20	6.82	$\overline{0}$	$\overline{0}$	50%	20	20	6.82
				θ	$\overline{0}$					θ	$\overline{0}$	
$0.20 \times \mu$	θ	θ	50%	20	20	4.41	$\overline{0}$	$\overline{0}$	50%	20	20	4.41
				θ	θ					$\overline{0}$	$\overline{0}$	
$0.25 \times \mu$	θ	$\overline{0}$	50%	20	20	2.01	$\overline{0}$	$\overline{0}$	50%	20	20	2.01
				$\overline{0}$	θ					$\overline{0}$	θ	

Table 5. Optimal convex tax system with fixed labour supply.

As shown in Table 5 with fixed labour supply and redistribution the results are completely different from those obtained in Chapter 2 when redistribution was not allowed. In particular, here the optimal tax system turns out to be independent of the redistributive objectives and requires to tax all incomes at the highest admissible tax rate. Then tax revenues net of the income share covering the revenue requirement, are equally redistributed.

Recall that with no redistribution and fixed labour supply the optimal tax system exhibits two clear and different patterns of the marginal tax rates. For Gini based SEF there is an income threshold above which taxation is the maximal admissible and zero below. In the case of polarization SEF, the optimal tax system requires to tax, at the maximum admissible tax rate, all incomes within the central interval. Marginal tax rates outside such interval are set equal to zero.

Tables 6 and 7 reports some selected results of the simulations for positive values of labour supply elasticity. Simulations results for other level of revenue requirements and different values of labour supply elasticity are reported in Tables 15A and B for Gini SEFs and in Table 18 for polarization SEFs.

G	ε	t_{1}	t_2	t_3	$y_1\,$	y_2	\boldsymbol{a}
$0.1 \times \mu$	0.1	$\boldsymbol{0}$	45%	50%	24	25	6.02
					29%	34%	
$0.2 \times \mu$	0.1	$\overline{0}$	45%	50%	24	25	1.22
					29%	34\%	
	0.2	$\overline{0}$	38%	50%	28	29	3.62
$0.1 \times \mu$					46%	52%	
	0.2	$\boldsymbol{0}$	38%	50%	$28\,$	29	-1.18
$0.2 \times \mu$					46%	52%	
			35%	48%	35	41	0.34
$0.1 \times \mu$	0.4	$\overline{0}$			65%	76%	
$0.2 \times \mu$	0.4	$\overline{0}$	35%	48%	35	41	-4.46
					65%	76%	

Table 6. Optimal convex tax system: Gini based SEF.

$G\$	ε	t_1	t_2	t_3	y_1	y_2	\boldsymbol{a}
$0.1 \times \mu$	0.1	$\boldsymbol{0}$	14\%	31\%	$28\,$	29	1.41
					40\%	44%	
$0.2 \times \mu$	0.1	$\boldsymbol{0}$	14%	31\%	28	29	-3.39
					40\%	44\%	
$0.1 \times \mu$	0.15	$\boldsymbol{0}$	$\boldsymbol{0}$	6%	20	31	-3.62
					$\boldsymbol{0}$	47%	
$0.2 \times \mu$	0.15	$\boldsymbol{0}$	$\boldsymbol{0}$	6%	20	31	-8.43
					$\boldsymbol{0}$	47\%	
	0.2	$\overline{0}$	$\boldsymbol{0}$	$\overline{0}$	20	$20\,$	-4.80
$0.1 \times \mu$					$\boldsymbol{0}$	$\boldsymbol{0}$	
$0.2 \times \mu$	0.2	θ	θ	θ	20	20	-9.61
					$\boldsymbol{0}$	$\boldsymbol{0}$	

Table 7. Optimal convex tax system: Polarization based SEF.

Similarly to the case with two income brackets analyzed in the previous section, with redistribution the optimal tax system is independent of the revenue requirement for a given value of labour supply elasticity $\varepsilon > 0$. When the revenue generated by the optimal tax system is greater (lower) than the revenue requirement the proportional tax system is supplemented with a lump-sum subsidy (tax). For the Gini based SEF the optimal tax system envisages a no tax area for all incomes within the first bracket $(t_1 = 0)$. The threshold y_1 identifying this exemption area is increasing in the level of labour supply elasticity. Marginal tax rates within the other two intervals $(t_2 \text{ and } t_3)$ are positive and decreasing with the level of labour supply elasticity.

When the focus shift to polarization concerns the optimal tax system exhibits lower tax rates and tends to be based on lump-sum taxation for elasticity level $\varepsilon \geq 0.2$. Recall that the simulations in Chapter 2 show that, under the assumption of no redistribution and with $\varepsilon \geq 0.2$, the optimal convex tax system reducing polarization requires a proportional taxation with a marginal tax rate increasing in the amount of required revenue.

Three brackets non-convex tax system

When we assume a non-convex regime of the tax rates we have the following post tax gross income quantile function $y_t(p)$

$$
y_{t}(p) = \begin{cases} y(p) (1-t_{1})^{\varepsilon} & \text{if } y(p) < \frac{y_{1}}{(1-t_{1})^{\varepsilon}} \\ y_{1} & \text{if } \frac{y_{1}}{(1-t_{1})^{\varepsilon}} \leq y(p) < \frac{y_{1}}{(1-t_{2})^{\varepsilon}} \\ y(p) (1-t_{2})^{\varepsilon} & \text{if } \frac{y_{1}}{(1-t_{2})^{\varepsilon}} \leq y(p) \leq \widehat{y} \\ y(p) (1-t_{3})^{\varepsilon} & \text{if } y(p) > \widehat{y} \end{cases}
$$

;

where gross incomes are the same for all incomes that are in the first bracket and at the first threshold, while the result changes for incomes within the second and the third bracket. In particular, differently from the convex regime, there are no agents experiencing a gross income equal to the second income threshold y_2 , where the gross income distribution exhibits a discontinuity. Hence, there exists a threshold level \hat{y} such that all incomes lower than \hat{y} fall in the second bracket, while all incomes above \hat{y} belong to third bracket. This threshold is

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

and it is derived in Appendix B of Chapter 2. Define $p_1^L := \inf\{p : y(p)(1 - t_1)^{\varepsilon} = y_1\},\$ $p_1^H := \sup\{p : y(p)(1-t_2)^{\varepsilon} = y_1\}$ and $\widehat{p} := \sup\{p : y(p) = \widehat{y}\}\$, we have that the SEF based on net incomes is

$$
W_v = \int_0^{p_1^L} v(p) y(p) (1 - t_1)^{\varepsilon + 1} dp + y_1 (1 - t_1) \int_{p_1^L}^{p_1^H} v(p) dp
$$

+
$$
\int_{p_1^H}^{\widehat{p}} v(p) [y(p) (1 - t_2)^{\varepsilon + 1} + y_1 (t_2 - t_1)] dp
$$

+
$$
\int_{\widehat{p}}^1 v(p) [y(p) (1 - t_3)^{\varepsilon + 1} + y_2 (t_3 - t_2) + y_1 (t_2 - t_1)] dp - a,
$$

with the revenue constraint

$$
G = a + t_1 \int_0^{p_1^L} y(p)(1 - t_1)^{\varepsilon} dp + t_1 \int_{p_1^L}^1 y_1 dp
$$

+ $t_2 \int_{p_1^H}^{\widehat{p}} [y(p)(1 - t_2)^{\varepsilon} - y_1] dp + t_2 \int_{\widehat{p}}^1 (y_2 - y_1) dp$
+ $t_3 \int_{\widehat{p}}^1 [y(p)(1 - t_3)^{\varepsilon} - y_2] dp.$

The social optimization problem is solved numerically for values of elasticity $\varepsilon \geq 0$. Table 8 reports the simulations results related to case of fixed labour supply, while Tables 9 and 10 illustrate the optimal non-convex tax system for inequality and polarization sensitive SEFs respectively.

			Gini based SEF				Polarization based SEF					
G	t_1	t_2	t_3	y_1	y_2	\boldsymbol{a}	t_{1}	t_2	t_3	y_1	y_1	\boldsymbol{a}
$0.10 \times \mu$		50% Ω	50%	20	20	9.22	$\overline{0}$	50%	50%	20	20	9.22
				θ	θ					Ω	$\overline{0}$	
$0.15 \times \mu$	θ	50%	50%	20	20	6.82	$\overline{0}$	50%	50%	20	20	6.82
				θ	$\overline{0}$					Ω	$\overline{0}$	
$0.20 \times \mu$	Ω	50%	50%	20	20	4.41	$\overline{0}$	50%	50%	20	20	4.41
				θ	$\overline{0}$					Ω	θ	
$0.25 \times \mu$	θ	50%	50%	20	20	2.01	$\overline{0}$	50%	50%	20	20	2.01
				θ	$\overline{0}$					θ	θ	

Table 8. Optimal non-convex tax system with fixed labour supply.

With fixed labour supply the optimal tax non-convex tax system coincides with the convex one and it is independent of the distributive objective. All incomes are taxed with the highest admissible tax rate and then the amount of revenue exceeding the revenue requirement is equally redistributed.

$G\$	ε	t_{1}	t_2	t_3	y_1	$y_2\,$	\boldsymbol{a}
$0.1 \times \mu$	0.1	θ	50%	50%	24	24	6.05
					30%	30%	
$0.2 \times \mu$	0.1	$\overline{0}$	50%	50%	24	24	1.25
					30%	30%	
$0.1 \times \mu$	0.2	$\overline{0}$	50%	35%	28	240	3.69
					49\%	99%	
$0.2 \times \mu$	0.2	$\overline{0}$	50%	35%	28	240	-1.11
					49%	99%	
$0.1 \times \mu$	0.4	θ	49%	19%	36	210	0.47
					71%	99%	
$0.2 \times \mu$	0.4	$\boldsymbol{0}$	49%	19%	36	210	-4.33
					71%	99%	

Table 9. Optimal non-convex tax system: Gini based SEF.

Table 9 shows that the optimal non-convex tax system reducing inequality always require a no tax area which is increasing as elasticity increases (see column y_1). Then, unless when elasticity is low ($\varepsilon \leq 0.1$), the optimal value of t_3 is lower than t_2 and this difference becomes sizeable as the labour supply elasticity increases. This result is in line with the case of two income brackets (see Table 3) and with the simulations in Chapter 2. However, differently from these last ones,

when redistribution is allowed, it is always optimal to set $t_1 = 0$.

G	ε	t_1	t_2	t_3	y_1	y_2	\boldsymbol{a}
$0.1 \times \mu$	0.1	θ	50%	0%	29	66	0.42
					47%	83%	
$0.2 \times \mu$	0.1	$\overline{0}$	50%	0%	29	66	-4.38
					47%	83%	
$0.1 \times \mu$	0.15	$\overline{0}$	50%	0%	30	54	-1.07
					52%	77%	
$0.2 \times \mu$	0.15	$\overline{0}$	50%	0%	30	54	-4.38
					52%	77%	
$0.1 \times \mu$	0.2	$\overline{0}$	46%	$\boldsymbol{0}$	30	48	-2.03
					53%	72%	
$0.2 \times \mu$	0.2	$\overline{0}$	46%	$\overline{0}$	30	48	-6.84
					53%	72%	

Table 10. Optimal non-convex tax system: Polarization based SEF.

As to polarization SEFs, the optimal tax system is such that taxation is the maximal admissible within the central interval, whose size is reducing in the level of labour supply elasticity. Moreover, differently from the case with no redistribution (Chapter 2), the optimal marginal tax rates outside the central interval are always equal to zero and lump-sum taxation supplements proportional tax in order to cover the revenue requirement.

To summarize the optimal three brackets tax system reducing income inequality is convex unless when the level of labour supply elasticity is high, while the optimal tax schedule to reduce polarization is non-convex. Last, we derive the optimal three brackets tax system by considering gross wage distributions with different levels of pre-tax inequality and polarization. The simulations results are reported in Tables 16 and 17 for Gini SEFs and in Tables 19 and 20 for polarization SEFs. The results we obtain show that when the level of initial inequality is not high the optimal three brackets tax system for Gini based SEF is non-convex (compare Tables 16 and 17 in Appendix B). While for the polarization SEFs the optimal three brackets tax system is always non-convex (see Tables 19 and 20 in Appendix B).

3.4 Concluding remarks

In this paper we derive the optimal income taxation redistributive scheme under two nonwelfarist objectives, i.e. inequality or polarization reduction. We consider a set of alternative piecewise linear tax schemes integrated with lump-sum taxation and subsidies. More specifically, we analyze tax systems with two and three income brackets, as well as different tax rates regimes depending on their ranking. These different configurations are evaluated according to a rankdependent social evaluation function defined over net incomes, which formalizes the specific government's non-welfarist objective.

The interesting aspect of this work is related to the introduction of a redistributive mechanism that leads to results contrasting those obtained in Chapter 2. In that case, indeed, redistribution was not allowed and the focus was on the socially desirable mechanism collecting a given revenue requirement.

There are four main interesting results. First, intuitively with fixed labour supply the optimal tax system is based on a proportional taxation with the highest admissible tax rate. The collected amount, net of the income share covering the revenue requirement, is equally redistributed. This result is independent of the non-welfarist distributive objective.

Second, with positive values of labour supply elasticity the non-welfarist objective becomes a crucial determinant of the shape of the optimal tax system. In particular, simulations results show that the optimal two brackets system reducing inequality requires a no taxation area below a given threshold and proportional taxation above. As elasticity increases, the threshold and the tax rate increases and decreases respectively. In case of polarization reduction, the optimal two brackets tax system requires a proportional taxation for all incomes below a given threshold a zero marginal tax rate for all incomes above. Both the threshold and the tax rate are decreasing in the level of labour supply elasticity, and for value of $\varepsilon \geq 0.2$ the optimal tax system is based only on lump-sum taxation.

With three income brackets we obtain that the optimal tax system for Gini based SEF is convex unless when labour supply elasticity is high. Moreover, the optimal tax system reducing inequality is convex when the initial level of gross income dispersion is high. As to polarization reduction the optimal tax system is always non-convex. In both cases the marginal tax rates are decreasing in the level of labour supply elasticity.

Third, the design of the optimal tax system is independent of the revenue requirement, and the sign of the lump-sum transfer depends on the difference between the collected amount and the required revenue. Moreover the sign of the lump-sum transfer is affected by the combination of the level of labour supply elasticity and the index of gross income dispersion.

Last, it is more likely that the optimal tax system is based only on lump-sum tax with polarization sensitive SEFs.

Appendix

Appendix A

In Appendix A we illustrate the main calculations related to the optimization problems behind the derivations of the optimal two brackets redistributive taxation schemes.

A.1 The calculations for two brackets with no-tax area.

The partial derivatives w.r.t. the choice variables for the optimization problem in (3.13) are respectively $\frac{\partial \tilde{W}_v}{\partial t}$ and $\frac{\partial \tilde{W}_v}{\partial y_1}$. We derive them here in details. We first consider

$$
\frac{\partial \tilde{W}_v}{\partial t} = y_1 v (p_1^H) \frac{\partial p_1^H}{\partial t} - v (p_1^H) \left[y (p_1^H) (1 - t)^{\varepsilon + 1} + ty_1 \right] \frac{\partial p_1^H}{\partial t} \n- \int_{p_1^H}^1 v (p) y (p) (\varepsilon + 1) (1 - t)^{\varepsilon} dp + \int_{p_1^H}^1 v (p) y_1 dp - t \left[y (p_1^H) (1 - t)^{\varepsilon} - y_1 \right] \frac{\partial p_1^H}{\partial t} \n+ \int_{p_1^H}^1 \left[y (p) (1 - t)^{\varepsilon} - y_1 \right] dp - t \int_{p_1^H}^1 y (p) \varepsilon (1 - t)^{\varepsilon - 1} dp \n= y_1 v (p_1^H) \frac{\partial p_1^H}{\partial t} - v (p_1^H) \left[y (p_1^H) (1 - t)^{\varepsilon} (1 - t) + ty_1 \right] \frac{\partial p_1^H}{\partial t} \n- (\varepsilon + 1) \int_{p_1^H}^1 v (p) y (p) (1 - t)^{\varepsilon} dp + \int_{p_1^H}^1 v (p) y_1 dp - t \left[y (p_1^H) (1 - t)^{\varepsilon} - y_1 \right] \frac{\partial p_1^H}{\partial t} \n+ \int_{p_1^H}^1 y (p) (1 - t)^{\varepsilon} dp - \int_{p_1^H}^1 y_1 dp - t \int_{p_1^H}^1 y (p) \varepsilon (1 - t)^{\varepsilon - 1} dp.
$$

Recall that $y(p_1^H)(1-t)^\varepsilon = y_1$ we then obtain

$$
\frac{\partial \tilde{W}_v}{\partial t} = y_1 v (p_1^H) \frac{\partial p_1^H}{\partial t} - v (p_1^H) y_1 \frac{\partial p_1^H}{\partial t} - (\varepsilon + 1) (1 - t)^{\varepsilon} \int_{p_1^H}^1 v (p) y (p) dp + \int_{p_1^H}^1 v (p) y_1 dp + (1 - t)^{\varepsilon} \int_{p_1^H}^1 y (p) dp - \int_{p_1^H}^1 y_1 dp - t\varepsilon (1 - t)^{\varepsilon - 1} \int_{p_1^H}^1 y (p) dp.
$$

That is we obtain that

$$
\frac{\partial \tilde{W}_v}{\partial t} = -(\varepsilon + 1) (1 - t)^{\varepsilon} \int_{p_1^H}^1 v(p) y(p) dp + \int_{p_1^H}^1 [v(p) - 1] y_1 dp
$$

$$
+ (1 - t)^{\varepsilon} \left[1 - \frac{t}{1 - t} \varepsilon \right] \int_{p_1^H}^1 y(p) dp.
$$

Let $\int_{p_1^H}^1 y(p) dp = \mu_1$ and $\int_{p_1^H}^1 v(p) y(p) dp = \mu_1 (1 - D_{v1})$, that is we compute the average income and the abbreviated SEF for the distribution where for all $p < p_1^H$ all incomes are set equal to 0 and coincide with $y(p)$ for $p \geq p_1^H$. We then obtain

$$
\frac{\partial \tilde{W}_v}{\partial t} = -(1-t)^{\varepsilon} (\varepsilon + 1) \mu_1 (1 - D_{v1}) - [V(p_1^H) - p_1^H] y_1 + (1-t)^{\varepsilon} \left[1 - \frac{t}{1-t} \varepsilon \right] \mu_1
$$

= $(1-t)^{\varepsilon} (\varepsilon + 1) \mu_1 D_{v1} - [V(p_1^H) - p_1^H] y_1 + (1-t)^{\varepsilon} \left[1 - \frac{t}{1-t} \varepsilon - (\varepsilon + 1) \right] \mu_1$
= $(1-t)^{\varepsilon} (\varepsilon + 1) \mu_1 D_{v1} - [V(p_1^H) - p_1^H] y_1 - \mu_1 (1-t)^{\varepsilon - 1} \varepsilon.$

In order to derive the $F.O.C.$ w.r.t. the income threshold level y_1 we compute

$$
\frac{\partial \tilde{W}_v}{\partial y_1} = \frac{\partial p_1^L}{\partial y_1} v (p_1^L) y (p_1^L) + \int_{p_1^L}^{p_1^H} v (p) dp - y_1 \frac{\partial p_1^L}{\partial y_1} v (p_1^L) + y_1 \frac{\partial p_1^H}{\partial y_1} v (p_1^H) \n- \frac{\partial p_1^H}{\partial y_1} v (p_1^H) \left[y (p_1^H) (1-t)^{\varepsilon+1} + ty_1 \right] + \int_{p_1^H}^1 v (p) t dp \n- \frac{\partial p_1^H}{\partial y_1} t \left[y (p_1^H) (1-t)^{\varepsilon} - y_1 \right] - t \int_{p_1^H}^1 dp.
$$

Recall that $y(p_1^H) (1-t)^{\varepsilon} = y_1$ and $y(p_1^L) = y_1$, it follows that

$$
\frac{\partial \tilde{W}_v}{\partial y_1} = \int_{p_1^L}^{p_1^H} v(p) \, dp + y_1 \frac{\partial p_1^H}{\partial y_1} v(p_1^H) - \frac{\partial p_1^H}{\partial y_1} v(p_1^H) [y_1 (1-t) + ty_1] \n+ t \int_{p_1^H}^{1} v(p) \, dp - t \int_{p_1^H}^{1} dp \n= \int_{p_1^L}^{p_1^H} v(p) \, dp + t \int_{p_1^H}^{1} v(p) \, dp - t \int_{p_1^H}^{1} dp \n= V (p_1^H) - V (p_1^L) + t(1 - V (p_1^H)) - t(1 - p_1^H) \n= V (p_1^H) - V (p_1^L) - t [V (p_1^H) - p_1^H].
$$

A.2 The calculations for two brackets with no marginal taxation at the top.

The partial derivatives w.r.t. the choice variables for the optimization problem in (3.14) are respectively

$$
\frac{\partial \hat{W}_v}{\partial \tau} = \left[v \left(\hat{p} \right) \hat{y} (1 - \tau)^{\varepsilon + 1} \right] \frac{\partial \hat{p}}{\partial \tau} - (1 + \varepsilon) (1 - \tau)^{\varepsilon} \int_0^{\hat{p}} v \left(p \right) y(p) dp \tag{3.33}
$$

$$
-v \left(\hat{p} \right) \left[\hat{y} - \tau y_1 \right] \frac{\partial \hat{p}}{\partial \tau} - y_1 \int_{\hat{p}}^1 v \left(p \right) dp + \tau \hat{y} (1 - \tau)^{\varepsilon} \frac{\partial \hat{p}}{\partial \tau}
$$

$$
+ (1 - \tau)^{\varepsilon} \int_0^{\hat{p}} y \left(p \right) dp - \tau \varepsilon (1 - \tau)^{\varepsilon - 1} \int_0^{\hat{p}} y \left(p \right) dp + y_1 (1 - \hat{p}) - \tau y_1 \frac{\partial \hat{p}}{\partial \tau}
$$

for the tax rate τ , and

$$
\frac{\partial \hat{W}_v}{\partial y_1} = \left[v \left(\hat{p} \right) y \left(\hat{p} \right) \left(1 - \tau \right)^{\varepsilon + 1} \right] \frac{\partial \hat{p}}{\partial y_1} - v \left(\hat{p} \right) \left[y \left(\hat{p} \right) - \tau y_1 \right] \frac{\partial \hat{p}}{\partial y_1} - \tau \int_{\hat{p}}^1 v \left(\hat{p} \right) dp \quad (3.34)
$$

$$
+ \tau y \left(\hat{p} \right) \left(1 - \tau \right)^{\varepsilon} \frac{\partial \hat{p}}{\partial y_1} + \tau \left(1 - \hat{p} \right) - \tau y_1 \frac{\partial \hat{p}}{\partial y_1}.
$$

for the income threshold level $y_{\rm 1}$

We now simplify and readjust these equations, starting from $\frac{\partial \hat{W}_v}{\partial \tau}$.

Recalling that $y(\hat{p}) = \hat{y}$ and readjusting (3.33) we obtain

$$
\frac{\partial \hat{W}_v}{\partial \tau} = \left[(1 - \tau)^{\varepsilon} - \tau \varepsilon (1 - \tau)^{\varepsilon - 1} \right] \int_0^{\widehat{p}} y(p) dp + y_1 (V(\widehat{p}) - \widehat{p}) \n- (1 + \varepsilon) (1 - \tau)^{\varepsilon} \int_0^{\widehat{p}} v(p) y(p) dp \n+ \frac{\partial \widehat{p}}{\partial \tau} \left[v(\widehat{p}) \widehat{y}(1 - \tau)^{\varepsilon + 1} - v(\widehat{p}) [\widehat{y} - \tau y_1] + \tau \widehat{y}(1 - \tau)^{\varepsilon} - \tau y_1 \right].
$$
\n(3.35)

Recalling that

$$
\widehat{y} := (1 + \varepsilon) \left[\frac{\tau y_1}{1 - (1 - \tau)^{1 + \varepsilon}} \right],\tag{3.36}
$$

and substituting we get

$$
\frac{\partial \hat{W}_v}{\partial \tau} = (1 - \tau)^{\varepsilon} \left[1 - \frac{\tau}{(1 - \tau)} \varepsilon \right] \int_0^{\widehat{p}} y(p) dp + y_1 (V(\widehat{p}) - \widehat{p}) \n- (1 + \varepsilon) (1 - \tau)^{\varepsilon} \int_0^{\widehat{p}} v(p) y(p) dp \n+ \frac{\partial \widehat{p}}{\partial \tau} [-v(\widehat{p}) (1 + \varepsilon) \tau y_1 + v(\widehat{p}) \tau y_1 + \tau \widehat{y} (1 - \tau)^{\varepsilon} - \tau y_1].
$$

$$
\frac{\partial \hat{W}_v}{\partial \tau} = \frac{\partial \hat{p}}{\partial \tau} \left[-v(\hat{p}) \varepsilon \tau y_1 + \tau (1 - \tau)^{\varepsilon} (1 + \varepsilon) \left[\frac{\tau y_1}{1 - (1 - \tau)^{1 + \varepsilon}} \right] - \tau y_1 \right] \n+ (1 - \tau)^{\varepsilon} \left[1 - \frac{\tau}{(1 - \tau)^{\varepsilon}} \right] \int_0^{\hat{p}} y(p) dp + y_1 (V(\hat{p}) - \hat{p}) \n- (1 + \varepsilon) (1 - \tau)^{\varepsilon} \int_0^{\hat{p}} v(p) y(p) dp,
$$
\n(3.37)

where the term in the first square bracket could be simplified as $-v(\widehat{p}) \varepsilon + \frac{\tau(1-\tau)^{\varepsilon}(1+\varepsilon)-1+(1-\tau)^{1+\varepsilon}}{1-(1-\tau)^{1+\varepsilon}}$ $\frac{(1+\epsilon)-1+(1-\gamma)^{-1}}{1-(1-\tau)^{1+\epsilon}}=$ $-v(\widehat{p}) \varepsilon - \frac{1-(1+\varepsilon\tau)(1-\tau)^{\varepsilon}}{1-(1-\tau)^{1+\varepsilon}}$ $\frac{1-(1+\varepsilon\tau)(1-\tau)}{1-(1-\tau)^{1+\varepsilon}}$. It follows that

$$
\frac{\partial \hat{W}_v}{\partial \tau} = -\frac{\partial \hat{p}}{\partial \tau} \tau y_1 \left[\frac{1 - (1 + \varepsilon \tau) (1 - \tau)^{\varepsilon}}{1 - (1 - \tau)^{1 + \varepsilon}} + v(\hat{p}) \varepsilon \right] \n+ (1 - \tau)^{\varepsilon} \left[1 - \frac{\tau}{(1 - \tau)^{\varepsilon}} \right] \int_0^{\hat{p}} y(p) dp + y_1 (V(\hat{p}) - \hat{p}) \n- (1 + \varepsilon) (1 - \tau)^{\varepsilon} \int_0^{\hat{p}} v(p) y(p) dp.
$$
\n(3.38)

The terms in (3.38) can be further readjusted by adding and subtracting $\int_0^{\hat{p}} [y(p) - y_1] dp \int_0^{\widehat{p}} v(p) [y(p) - y_1] dp$, while in the first term one can consider that $\frac{\partial \widehat{p}}{\partial \tau} := \frac{\partial \widehat{p}}{\partial \widehat{y}}$ $\partial \widehat{y}$ $\frac{\partial \widehat{y}}{\partial \tau}$ where $\frac{\partial \widehat{y}}{\partial \tau}$ could be computed by considering the definition of \hat{y} in (3.36) so that after simplifications one obtains

$$
\frac{\partial \hat{W}_v}{\partial \tau} = -\frac{\partial \hat{p}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \tau} \tau y_1 \left[v(\hat{p}) \varepsilon + \frac{1 - (1 + \varepsilon \tau) (1 - \tau)^{\varepsilon}}{1 - (1 - \tau)^{1 + \varepsilon}} \right] \n+ (1 - \tau)^{\varepsilon} \left[1 - \frac{\tau}{(1 - \tau)} \varepsilon \right] \int_0^{\hat{p}} y(p) dp + y_1 (V(\hat{p}) - \hat{p}) \n- (1 + \varepsilon) (1 - \tau)^{\varepsilon} \int_0^{\hat{p}} v(p) y(p) dp \n- \int_0^{\hat{p}} [y(p) - y_1] dp + \int_0^{\hat{p}} v(p) [y(p) - y_1] dp \n+ \int_0^{\hat{p}} [v(p) - 1] [y_1 - y(p)] dp
$$
\n(3.39)

that leads to the following partition of the components

$$
\frac{\partial \hat{W}_v}{\partial \tau} = -\frac{\partial \hat{p}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \tau} \tau y_1 \left[v(\hat{p}) \varepsilon + \frac{1 - (1 + \varepsilon \tau) (1 - \tau)^{\varepsilon}}{1 - (1 - \tau)^{1 + \varepsilon}} \right] \n+ [1 - (1 + \varepsilon) (1 - \tau)^{\varepsilon}] \int_0^{\hat{p}} [v(p) - 1] y(p) dp \n- \varepsilon (1 - \tau)^{\varepsilon - 1} \int_0^{\hat{p}} y(p) dp \n+ \int_0^{\hat{p}} [v(p) - 1] [y_1 - y(p)] dp.
$$
\n(3.40)

We now consider $\frac{\partial \hat{W}_v}{\partial y_1}$, recalling that $y(\hat{p}) = \hat{y}$ and readjusting (3.34) we obtain

$$
\frac{\partial \hat{W}_v}{\partial y_1} = \left[v \left(\hat{p} \right) \hat{y} \left(1 - \tau \right)^{\varepsilon + 1} \right] \frac{\partial \hat{p}}{\partial y_1} - v \left(\hat{p} \right) \left[\hat{y} - \tau y_1 \right] \frac{\partial \hat{p}}{\partial y_1} + \tau \hat{y} \left(1 - \tau \right)^{\varepsilon} \frac{\partial \hat{p}}{\partial y_1} + \tau \left(V \left(\hat{p} \right) - \hat{p} \right) - \tau y_1 \frac{\partial \hat{p}}{\partial y_1}.
$$

After rearranging the terms we get

$$
\frac{\partial \hat{W}_v}{\partial y_1} = \left[\hat{y} (1 - \tau)^{\varepsilon} \tau [1 - v(\hat{p})] + \hat{y} [1 - v(\hat{p})] - \hat{y} [1 - (1 - \tau)^{\varepsilon} v(\hat{p})] - \tau y_1 [1 - v(\hat{p})]] \frac{\partial \hat{p}}{\partial y_1} + \tau (V(\hat{p}) - \hat{p}).
$$

Note that the first term that multiplies $\frac{\partial \hat{p}}{\partial y_1}$ has already been simplified in the above calculations for $\frac{\partial \hat{W}_v}{\partial \tau}$, moreover we can consider that $\frac{\partial \hat{p}}{\partial y_1} := \frac{\partial \hat{p}}{\partial \hat{y}}$ $\partial \widehat{y}$ $\frac{\partial \hat{y}}{\partial y_1}$ where $\frac{\partial \hat{y}}{\partial y_1}$ could be computed by considering the definition of \hat{y} in (3.36) so that after substitutions we obtain

$$
\frac{\partial \hat{W}_v}{\partial y_1} = -\frac{\partial \hat{p}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial y_1} \tau y_1 \left[v(\hat{p}) \varepsilon + \frac{1 - (1 + \varepsilon \tau) (1 - \tau)^{\varepsilon}}{1 - (1 - \tau)^{1 + \varepsilon}} \right] + \tau (V(\hat{p}) - \hat{p}).
$$

Appendix B

In this appendix we recall the main aspects of the simulation procedure which is described in Chapter 2 and then we provide additional tables whose results supplement those presented in this chapter.

All simulations are based on a Pareto distribution of 1000 individuals gross wages. This distribution is truncated both above and below and it is generated following Apps et al. (2014). In particular, as in Apps et al. (2014) we assume that the lower bound is equal to 20, while the upper bound corresponds to the 98^{th} percentile⁵. This distribution is equivalent to the case $(1.a)$ in Apps et al. (2014) and the associated levels of inequality and polarization are 0.37 and 0.11 respectively.

In order to investigate the impact of the dispersion level on the optimal tax system, we consider two alternative distributions characterized by different levels of inequality and polarization. These distributions are obtained by changing the bounds of the first distribution, keeping the mean constant. In particular, the second distribution is bounded between 17 and 1009, with pre-tax inequality and polarization equal to 0.45 and 0.10. The third distribution is bounded between 25 and 142, while the level of inequality and polarization are 0.26 and 0.18. The graph below illustrates these three distributions.

⁵Recall that the cdf of a Pareto distributed variable x is $F(x) = 1 - \left(\frac{L}{x}\right)^{\alpha}$, where L is the scale parameter, representing the lower bound of the distribution, while the parameter α is the Pareto index. Given $L = 20$ and $\alpha = 1.4$, it follows that the value associated to the upper bound is equal to 327.

Fig. 3. Simulated gross wage distributions.

Recall that the agents' utility function adopted in the paper leads to a labour supply elasticity which is constant throughout the entire distribution. Moreover, for a given wage distribution, changes in the level of labour supply elasticity have two effects: first, they impact on the gross income distribution in absence of taxation; second, they determine the individuals' reaction to taxation. In this work, we want to focus only on the second effect, hence when the labour supply elasticity changes, we keep constant the gross income distribution in absence of taxation. To this end, we need an appropriate rescaling of the wage distribution, where each element is raised to the power of $(1 + \varepsilon)$ (see equation (3.4) and set $t = 0$, $k = 1/\alpha$ and $\alpha = \frac{\varepsilon + 1}{\varepsilon}$ $\frac{\pm 1}{\varepsilon}$).

The simulations are performed by using a grid search method. More specifically, we define the grids for the three tax rates and for the two thresholds. Then, for each combination of these tax parameters we compute the value of the demo-grant which keeps the government budget balanced and then we compute the value of social evaluation function. Last, we select the combination of policy parameters associated with the highest social welfare.

The following tables report the simulations results for different tax systems, with two and

 $% \left\langle \cdot ,\cdot \right\rangle _{0}$ three income brackets, under different regimes of the tax rates.

Simulations results for two brackets tax system

$I = 0.45.$			No-tax area					Zero top marginal tax rate		Opt.
G	ε	y_1	\boldsymbol{t}	\boldsymbol{a}	\cal{SW}	y_1	τ	\boldsymbol{a}	\cal{SW}	Syst.
$0.10 \times \mu$	0.01	18 12.6% 4.8%	98%	21.43	39.51	900 99.9%	98%	40.46	40.97	ZeroTop
$0.10 \times \mu$	0.1	22 46.2% 15.7%	84%	11.39	32.40	850 99.9%	80%	27.95	32.43	ZeroTop
$0.10 \times \mu$	0.15	23 52% 17.3%	77%	8.90	30.46	800 99.9%	72%	23.79	29.87	No-tax
$0.10 \times \mu$	0.20	25 59.8% 17.8%	73%	6.83	29.02	800 99.9%	64%	20.28	27.99	No-tax
$0.10 \times \mu$	0.40	32 74.9% 15.9%	58%	2.19	25.93	800 99.9%	37%	9.98	23.74	No -tax
$0.10 \times \mu$	0.60	37 80.9% 14.2%	48%	0.13	24.44	850 99.9%	18%	2.88	22.01	No-tax

Table 11. Optimal tax system: Gini based SEF. Distribution 2.

$I = 0.26$.			No-tax area					Zero top marginal tax rate		Opt.
G	ε	$y_{\rm 1}$	\boldsymbol{t}	\boldsymbol{a}	\cal{SW}	$y_1\,$	τ	\boldsymbol{a}	SW	Syst.
$0.10 \times \mu$	0.01	26 10.4% 4.4%	96%	14.97	41.25	137 99.5%	96%	$39.85\,$	41.22	No-tax
$0.10 \times \mu$	0.1	$31\,$ 41.8% 13.2%	72%	4.86	35.25	127 98.4%	66%	23.72	34.49	No-tax
$0.10 \times \mu$	0.15	33 48.9% 13.5%	62%	2.51	$33.90\,$	128 98.5%	51%	17.26	32.82	$\operatorname{No}\text{-}\operatorname{tax}$
$0.10 \times \mu$	0.20	$35\,$ 54.5% 13.2%	54%	0.89	33.02	129 98.6%	64%	20.28	27.99	No -tax
$0.10 \times \mu$	0.40	44 69.6% 9.6%	32%	-2.49	31.47	25 $\boldsymbol{0}$	$\overline{0}$	-4.80	30.62	No-tax
$0.10 \times \mu$	0.60	$52\,$ 77.5% 7.1%	21\%	-3.68	30.98	25 $\overline{0}$	$\overline{0}$	-4.80	30.62	No -tax

Table 12. Optimal tax system: Gini based SEF. Distribution 3:

$P = 0.10$			No-tax area					Zero top marginal tax rate		Opt.
G	ε	y_1	$t\,$	\boldsymbol{a}	\cal{SW}	$y_1\,$	τ	\boldsymbol{a}	\cal{SW}	Syst.
$0.10 \times \mu$	0.01	$20\,$ 23.2% 2.7%	92%	18.81	41.81	158 95.9%	93%	33.09	41.84	ZeroTop
$0.10 \times \mu$	0.1	24 41.7% 3.2%	32\%	2.40	38.63	63 84.4%	44%	9.65	38.80	ZeroTop
$0.10 \times \mu$	0.15	27 48.9% 1%	10%	-2.64	38.31	59 82.9	20%	1.76	38.37	ZeroTop
$0.10 \times \mu$	0.20	17 $\overline{0}$	θ	-4.80	38.28	17 θ	θ	-4.80	38.28	Lump-sum

Table 13. Optimal tax system: Polarization based SEF. Distribution 2

Table 14. Optimal tax system: Gini based SEF. Distribution 3

$P = 0.18$			No-tax area					Zero top marginal tax rate		Opt.
G	ε	y_1	$t\,$	\boldsymbol{a}	\cal{SW}	$y_1\,$	τ	\boldsymbol{a}	\cal{SW}	Syst.
$0.10 \times \mu$	0.01	29 23.5% 2.8%	91%	11.99	41.75	93 92.2%	92%	37.02	41.79	ZeroTop
$0.10 \times \mu$	0.1	35 44.5% 3.2%	30%	-0.56	38.68	66 81.6%	31%	8.21	38.67	ZeroTop
$0.10 \times \mu$	0.15	38 49.5% 0.8%	7%	-3.87	38.44	25 $\boldsymbol{0}$	$\overline{0}$	-4.80	38.43	No-tax
$0.10 \times \mu$	0.20	25 $\overline{0}$	$\overline{0}$	-4.80	38.43	25 $\overline{0}$	$\overline{0}$	-4.80	38.43	Lump-sum

Simulations results for three brackets tax system

Table 15A.

Distribution 1: Wages from 20 to 326, mean: 48.04. Inequality (Gini) before taxes: 0.37.														
Gini based SEF					Convex tax system					Non-convex tax system				Optimal System
G	S	ł,	ł2	43	弍	呉	a	ł,	Ł2	ł3	弍	呉	a	
	0.25	\circ	38%	50%	0.53 8	0.61 32	2.61	0	50%	21%	0.57 ඝ	0.99 240	2.74	Non-convex
$0.10 \times \mu$	0.30	0	35%	50%	0.56 51	0.67 55	1.86	0	50%	23%	0.63 32	0.99 240	1.89	Non-convex
	0.40	\bullet	35%	48%	0.65 35	0.76 ₽	0.34	0	49%	19%	0.71 86	0.99 210	0.47	Non-convex
	0.25	0	38%	50%	0.53 8	0.61 32	0.21	\circ	50%	21%	0.57 8	6.99 240	0.34	Non-convex
$0.15 \times \mu$	0.30	0	35%	50%	0.56 51	0.67 8	-0.54	\circ	50%	23%	0.63 32	0.99 240	-51	Non-convex
	0.40	0	35%	48%	0.65 35	0.76 41	-2.06	0	49%	19%	0.71 96	0.99 210	-1.93	Non-convex
	0.25	0	38%	50%	0.53 8	0.61 32	-2.19	0	50%	21%	0.57 8	0.99 240	-2.06	Non-convex
$0.20 \times \mu$	0.30	\circ	35%	50%	0.56 ಹ	0.67 55	-2.94	\circ	50%	23%	0.63 32	0.99 240	-2.91	Non-convex
	0.40	\circ	35%	48%	0.65 ౘ	0.76 4	-4.46	\circ	49%	19%	0.71 36	0.99 210	4.33	Non-convex
	0.25	0	38%	50%	0.53 8	0.61 32	4.59	0	50%	21%	0.57 8	6.99 240	$\frac{4.46}{4}$	Non-convex
$0.25 \times \mu$	0.30	\bullet	35%	50%	0.56 51	0.67 8	-5.34	\circ	50%	23%	0.63 32	0.99 240	-5.31	Non-convex
	0.40	0	35%	48%	0.65 35	0.76 Ч	-6.86	0	49%	19%	$\overline{5}$ 96	0.99 210	-6.74	Non-convex

Table 15B.

Table 16.

	Optimal System		Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex
		a	3.21	2.65	1.75	0.77	0.81	0.25	-0.66	-1.63	-1.59	-2.15	-3.06	4.04	-3.99	4.55	-5.46	-6.44
		S,	0.98 120	0.98 20	0.98 120	0.98 120	0.98 120	0.98 20	0.98 120	0.98 120	0.98 120	0.98 120	0.98 120	0.98 120	0.98 120	0.98 120	0.98 120	0.98 120
	Non-convex tax system	5,	0.33 80	0.37 51	0.45 33	0.53 8	0.33 30	0.37 51	0.45 33	0.53 35	0.33 30	0.37 51	0.45 33	0.53 35	0.33 30	0.37 51	0.45 33	0.53 35
		43	%R	29%	21%	15%	%ŧ	29%	21%	15%	94%	29%	21%	15%	94%	29%	21%	15%
		Ł,	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%
		七	0	0	0	0	0	\circ	\circ	0	\circ	\circ	0	\circ	0	\circ	0	\circ
		a	3.15	2.56	1.86	0.58	0.75	0.16	-0.54	-1.81	-1.65	-2.23	-2.94	-4.21	4.06	4.64	-5.34	-6.62
to 142, mean: 48.02. Inequality(Gini) before taxes: 0.26.		S,	0.36 51	0.41 32	0.45 33	0.59 38	0.36 51	0.41 32	0.45 33	0.59 38	0.36 51	0.41 32	0.45 33	0.59 38	0.36 51	0.41 32	0.45 33	0.59 38
	Convex tax system	5,	0.31 30	0.35 51	0.37 32	0.50 35	0.31 30	0.35 51	0.37 32	0.50 35	0.31 8	0.35 51	0.37 32	0.50 35	0.31 80	0.35 ನ	0.37 32	0.50 35
		43	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%
		₽	41%	38%	25%	39%	41%	38%	25%	39%	41%	38%	25%	39%	41%	38%	25%	39%
		4	0	\circ	0	\circ	0	0	\circ	0	0	\circ	0	\circ	\circ	\bullet	0	0
		ω	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20
Distribution 3: Wages from 25	Gini based SEF	Ü			$0.1 \times \mu$				$0.15 \times \mu$				$0.20 \times \mu$				$0.25 \times \mu$	

Table 17.

	Optimal System		Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex
		a	0.42	-0.41	-1.07	-2.03	-1.98	-2.81	-3.48	-4.43	4.38	-5.21	-5.88	-6.84	-6.78	-7.62	-8.28	-9.24
		呉	0.83 8	0.80 60	0.77 54	0.72 $\frac{48}{5}$	0.83 8	0.80 60	0.77 54	0.72 48	0.83 8	0.80 8	0.77 54	0.72 $\frac{48}{5}$	0.83 8	0.80 60	0.77 54	0.72 \$
		Z,	0.47 29	0.50 80	0.52 80	0.53 30	0.47 29	0.50 80	0.52 80	0.53 8	0.47 29	0.50 30	0.52 30	0.53 30	0.47 29	0.50 80	0.52 30	0.53 80
	Non-convex tax system	43	\circ	\circ	0	0	0	\bullet	\bullet	0	\bullet	\circ	\circ	\circ	\circ	\bullet	\bullet	\circ
		Ł,	50%	50%	50%	46%	50%	50%	50%	46%	50%	50%	50%	46%	50%	50%	50%	46%
		42	\circ	\circ	\circ	\circ	0	\circ	\circ	0	\circ	\bullet	\circ	\circ	\circ	\bullet	0	\circ
		a	1.41	-0.88	-3.62	4.80	-0.98	-3.28	-6.02	-7.21	-3.39	-5.68	-8.43	-9.61	-5.79	-8.08	-10.83	-12.01
Distribution 1: Wages from 20 to 326, mean: 48.04. Polarization before taxes: 0.11.		Z,	0.44 29	0.46 80°	0.47 51	20 \circ	0.44 29	0.46 80	0.47 51	ನಿ ಂ	0.44 29	0.46 30	0.47 51	$\overline{20}$ \circ	0.44 29	0.46 80	0.47 55	$\frac{1}{20}$
		5,	0.40 28	$\overline{20}$ \circ	20 \circ	20 \circ	0.40 28	20 \circ	$\overline{20}$ \circ	$\frac{1}{20}$	0.40 28	$\overline{20}$ \circ	$\overline{20}$ \circ	$\overline{20}$ \circ	0.40 28	20 \circ	$\overline{20}$ \circ	$\frac{1}{20}$
	Convex tax system	43	31%	20%	6%	\circ	31%	20%	6%	\circ	31%	20%	6%	\circ	31%	20%	6%	\circ
		42	14%	\circ	0	0	14%	\circ	\circ	0	14%	\circ	\circ	\circ	14%	\bullet	0	\bullet
		42	\circ	0	0	0	0	\circ	0	0	\circ	0	0	0	0	\circ	0	\circ
		دیا	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20
	Polarization based SEF	O			$0.1 \times \mu$				$0.15 \times \mu$				$0.20 \times \mu$				$0.25 \times \mu$	

Table 18.

	Optimal System		Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex	Non-convex
		a	-9.52	-1.24	-1.81	-2.76	-2.92	-3.64	4.22	-5.16	-5.32	-6.04	-6.62	-7.56	-7.72	-8.44	-9.02	-9.96
		Ż,	0.79 62	0.76 58	0.72 54	0.69 51	0.79 \mathcal{S}	0.76 58	0.72 54	0.69 51	0.79 62	0.76 58	0.72 54	0.69 51	0.79 62	0.76 58	0.72 54	0.69 51
		弍	0.47 35	0.51 36	0.52 36	0.55 76	0.47 35	0.51 36	0.52 96	0.55 22	0.47 35	0.51 36	0.52 86	0.55 57	0.47 35	0.51 36	0.52 36	0.55 75
	Non-convex tax system	ß	0	0	\circ	0	0	0	\circ	0	0	\circ	\circ	\circ	\circ	\bullet	\circ	\circ
		ł2	50%	50%	49%	42%	50%	50%	49%	42%	50%	50%	49%	42%	50%	50%	49%	42%
		42	\circ	\circ	\circ	\circ	\circ	0	\circ	\circ	\circ	\circ	\circ	\circ	\circ	\circ	\circ	\circ
		a	-0.57	-2.13	-3.87	4.80	-2.97	4.53	-6.27	-7.20	-5.37	-6.93	-8.67	-9.60	-7.77	-9.34	-11.07	-12.00
to 142, mean: 48.02. Polarization before taxes: 0.18.		島	0.44 35	0.49 75	0.49 38	25 \circ	0.44 35	64.0 75	64.0 38	25 \circ	0.44 35	64.0 75	0.49 38	25 \circ	0.44 35	0.49 57	64.0 38	25
		弍	25 \circ	25 \circ	25 \circ	25 \circ	25 \circ	25 \circ	25 \circ	25	25 \circ	25 \circ	25 \circ	25 \circ	25 \circ	25	25	25
	Convex tax system	43	30%	20%	汊	0	30%	20%	汊	\circ	30%	20%	汊	\circ	30%	20%	汊	\circ
		42	0	0	\circ	\circ	\bullet	0	0	0	0	\circ	\circ	0	\circ	\circ	\bullet	\circ
		42	\circ	\circ	\circ	\circ	\circ	\circ	\circ	\circ	\bullet	\circ	\circ	\circ	\circ	\circ	\circ	\circ
		ω	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20	0.10	0.12	0.15	0.20
Distribution 3: Wages from 25	Polarization based SEF	O			$0.1 \times \mu$				$0.15 \times \mu$				$0.20 \times \mu$				$0.25 \times \mu$	

Table 20.

Chapter 4

Optimal Non-Welfarist Taxation with Non-Constant Labour Supply **Elasticity**

In this work we derive the socially desirable non-welfarist tax scheme under the assumption of non-constant labour supply elasticity. We consider a class of social evaluation functions that allows to formalize redistributive objectives related to income inequality and polarization.

The focus is on the optimal tax system collecting a given amount of per-capita revenue when redistribution is not allowed. We consider a three brackets piecewise tax system with two possible different tax rates regimes, i.e. convex where $t_1 \leq t_2 \leq t_3$ and non-convex where $t_1 \le t_3 \le t_2.$

In the case of inequality concerns, the optimal tax system is convex with a no-tax area which vanishes when the revenue requirement is high. For polarization concerns the optimal tax schedule is non-convex.

Last, for social evaluation functions taking into account both inequality and polarization considerations, the convex tax system is more likely to dominate the non-convex configuration for high levels of inequality and polarization.

4.1 Introduction

A typical optimal taxation model which aims at deriving the shape of the optimal tax schedule, is formalized as a constrained maximization exercise. That is, a policy maker maximizes a social welfare function subject to a budget constraint and taking into account agents' reactions to taxation. Therefore, this exercise represents the formalization of the trade-off between equity and efficiency which is at the core of the optimal taxation theory. In other words, the need to satisfy the revenue requirement and the distributive objectives has to be balanced with the efficiency losses associated with the taxation.

Tuomala (1985) provides a simplified formula for the optimal linear tax which makes explicit the equity-efficiency trade-off. In particular, the more responsive are the agents to income taxation, the higher are the efficiency costs and the lower the tax rate. At the same time, the optimal tax rate is increasing in the level of inequality within the wage distribution. Another formulation of the optimal linear tax, where this trade-off is easily identifiable, is the covariance rule provided by Dixit and Sadmo (1977) and Atkinson and Stiglitz (1980).

Therefore, as demonstrated by Ebert and Moyes (2003 and 2007), the design of the redistributive policy can not ignore agentsí behavioral responses to taxation. However, given the complexity of the tax systems, agents can react along different potential dimensions. For instance, agents may decide to change their labour supply or to stop working. The first decision is known as intensive margin, while the second is called extensive margin.¹

Then, assessing labour supply responses to taxation is a crucial element in the design of the optimal tax policy. Typically, these reactions are estimated within structural models and measured by the wage labour supply elasticity. The picture arising from the empirical literature is that these reactions differ according to the gender, the marital status and the socio-economic background of agents, moreover there is no a broad consensus about the size of these reactions.²

In this work we take into account that agents' reactions to taxation can differ. In particular, we consider the case where wage labour supply elasticity is decreasing in the level of agents'

¹Behavioral responses to taxation can involve the possibility to evade tax or the decision to shift income into sources taxed with lower tax rate. However, in this work we do not consider these dimensions and we focus only on reactions in terms of labour supply changes.

² See Bargain and Peichl (2016), Bargain et al. (2014), Meghir and Phillips (2008), Saez et. al (2009) for a review of the literature on labour supply responses to taxation.

wage. This scenario is consistent with the findings of many studies (see Aaberge et al. 1995, 2002 and Aaberge and Colombino 2013). Then, we aim at deriving the optimal non-welfarist tax system reducing income inequality and income polarization. More specifically we want to see how the assumption of non-constant labour supply elasticity changes the results obtained in Chapter 2, where the focus was on the identification of the socially desirable mechanism collecting a given level of per-capita revenue, under the assumption of constant labour supply elasticity.

The introduction of non-constant elasticity represents the first extension with respect to Chapter 2. A second novelty is the consideration of a third weighting function taking into account at the same time both inequality and polarization concerns. In particular, this weighting function is obtained as a combination of the two weighting functions adopted for inequality and polarization considerations.

As in Chapter 2, here we consider a piecewise linear tax system with three income brackets and no redistribution of the collected revenue.

The results we obtain are qualitatively in line with those derived in Chapter 2. More specifically, the optimal tax system reducing income inequality is convex with increasing marginal tax rates. As to polarization reduction the optimal tax system is non-convex, while for the third weighting function combining inequality and polarization considerations, the optimal tax system depends on the levels of pre-tax inequality and polarization. In particular, when inequality and polarization before tax are high the optimal tax system is convex.

The differences with respect to the results obtained in Chapter 2 are related to the features of the labour supply elasticity profile. In particular, since we consider a labour supply elasticity decreasing in the level of wage, we obtain that the tax rates in the first bracket are always lower than those derived in Chapter 2, where elasticity is constant and lower. Moreover, here the median income is always within the Örst interval and taxed with lower tax rates than in Chapter 2.

The rest of the paper is organized as follows. In Section 2 we briefly introduce the social evaluation model used to evaluate the tax systems. The optimization problems faced by the nonwelfarist policy-maker and by each agent are described in Section 3. The numerical solutions of the optimal tax systems are presented in Section 4, while Section 5 concludes.

4.2 The social evaluation model

In this section we recall the social evaluation model adopted to evaluate alternative taxation policies. In line with the exposition of Chapter 2 we consider a class of social evaluation functions defined over net incomes, that allows to formalize redistributive objectives related to inequality and polarization concerns. In particular, when taking into account inequality considerations the redistributive objective could be quantified in terms of changes in the Gini index of incomes. Then, an appropriate modification of the positional weighting function allows to move, within the same social evaluation model, from inequality considerations to those focusing on the polarization of incomes.

Let $F(y)$ denote the cumulative income distribution function with quantile function $y(p)$ $\inf \{y : F(y) \ge p\}$ and mean $\mu(F) = \int_0^1 y(p) dp$. Let $v(p) \ge 0$ be a set of positional weights depending on the individuals' position $p \in [0, 1]$ in the income ranking, such that $\int_0^1 v(p) dp = 1$ and $V(p) = \int_0^p v(t) dt$.

The rank-dependent social evaluation function [SEF] (see Yaari, 1987, 1988 and Weymark, 1982) aggregates incomes weighted according to the individuals' position in the income ranking and it is formalized as

$$
W_v(F) = \int_0^1 v(p) y(p) dp.
$$
 (4.1)

The specific non-welfarist distributive objective is formalized by the particular form of the weighting function $v(p)$. As argued in Chapter 2, we consider two families of weights taking into account inequality or polarization considerations. More specifically, when the focus is on inequality concerns the weights are linearly decreasing in the individuals' positions, moving from poorer to richer individuals. Then, the weighting function $v(p)$ can be formalized as

$$
v(p) := v_G(p) = \begin{cases} 1 - \left[-2\left(\frac{1}{2} - p\right) \right] \text{ if } p \le \frac{1}{2} \\ 1 - 2\left(p - \frac{1}{2}\right) \text{ if } p \ge \frac{1}{2} \end{cases} \tag{4.2}
$$

In the case of polarization considerations the weighting function is expressed as

$$
v(p) := v_P(p) = \begin{cases} 2p + 1 & \text{if } p \le \frac{1}{2} \\ 2p - 1 & \text{if } p \ge \frac{1}{2} \end{cases}
$$
(4.3)

where weights are linearly increasing both below and above the median. Moreover, the weighting

function $v_P(p)$ exhibits a jump at the median position $p = 1/2$, and weights below the median are higher than those below.

Under both approaches the social evaluation can be summarized by the mean income of the distribution $\mu(F)$ and a linear index of dispersion $D_v(F)$ dependent on the choice of the weighting function v . The SEF could then be decomposed as

$$
W_v(F) = \mu(F) [1 - D_v(F)]. \qquad (4.4)
$$

However, in this work we also consider a third weighting function which is obtained as a linear combination of (4.2) and (4.3) and it is formalized as

$$
v(p) := v_{GP}(p) = \frac{1}{2}v_G(p) + \frac{1}{2}v_P(p) = \begin{cases} 1.5 \text{ if } p \le \frac{1}{2} \\ 0.5 \text{ if } p \ge \frac{1}{2} \end{cases} (4.5)
$$

In other terms, according to (4.5) the weight attached to each net income is equal to the average between the weights attached by the weighting functions (4.2) and (4.3). Then, all incomes below (above) the median are weighted with the same weight equal to 1.5 (0.5). The graph below illustrates the three weighting functions considered in this work.

Fig.1. Weighting functions.

4.3 Theoretical framework

In this section, first we formalize the optimal tax problem faced by a non-welfarist government aimed at identifying the socially optimal three brackets piecewise tax system collecting a given per-capita amount of revenue. Then, we present the agentsí optimization problem. To this regard, differently from Chapter 2 we adopt a specification of the agents' utility function leading to a wage labour supply elasticity which is decreasing in the level of individuals' wage.

4.3.1 The government's optimization problem

A non-welfarist government maximizes a linear rank-dependent SEF defined over individuals⁷ net incomes with generic non-negative positional weights $v(p)$

$$
W_v = \int_0^1 v(p) \left[y(p) - T(y(p)) \right] dp, \tag{4.6}
$$
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

$$
\int_{0}^{1} T\left(y\left(p\right)\right) dp = G,\tag{4.7}
$$

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

$$
T(y) := \begin{cases} t_1y - Z & \text{if } y \le y_1 \\ t_1y_1 + t_2(y - y_1) - Z & \text{if } y_1 < y \le y_2 \\ t_1y_1 + t_2(y_2 - y_1) + t_3(y - y_2) - Z & \text{if } y > y_2 \end{cases}
$$
(4.8)

where Z is a demo-grant which can be negative (positive) in case of lump-sum taxation (subsidy). In line with Chapter 2 we assume that Z is equal to zero. Moreover, we focus only on tax schedules where $t_1 \leq t_2$, and assume two possible regimes (convex and non-convex) of tax rates depending on the ranking of t_2 and t_3 . Convex system is such that $t_1 \le t_2 \le t_3$, while non-convex case considers configurations where $t_1 \leq t_3 \leq t_2$.

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 $y_1 < y_2$. The associated Lagrangian function of the government's optimization problem is

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

where λ denotes the Lagrangian multiplier, measuring the welfare cost of an increase in the revenue requirement.

The optimal tax system is such that groups of individuals could exhibit the same net income. However, the interesting aspect of our formulation is that the optimal tax system substantially differs depending on whether the social objective is concerned about reducing inequality or with reducing polarization.

4.3.2 The agents' optimization problem

In this work we consider agents endowed with preferences between consumption and leisure formalized by the following utility function described in Onrubia et al. (2005)

$$
U(x,l) = x^{\rho} + \alpha (L-l)^{\rho}, \qquad (4.9)
$$

with $\alpha > 0$ and $0 < \rho < 1$, where $x \in \mathbb{R}$ denotes the net disposable income/consumption and $l \in [0, L]$ denotes the labour supply, which can not exceed a maximum level equal to L. 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 .

Each agent is characterized by a productivity level formalized by the exogenous wage $w > 0$. We assume that agents in the economy earn only a gross labour income income $y = w \geq 0$.

With no-taxation $y = x$, hence the derivative of the utility function (4.9) w.r.t. the labour supply l is

$$
\frac{\partial U(x,l)}{\partial l} = \rho w^{\rho} l^{\rho-1} - \alpha \rho (L-l)^{\rho-1}.
$$

Then, setting this derivative equal to zero, one obtains that the optimal level of pre-tax labour supply is

$$
l^* = \frac{L\left(\frac{w}{\alpha}\right)^{\sigma}}{w + \left(\frac{w}{\alpha}\right)^{\sigma}} = \frac{L}{\frac{\alpha^{\sigma}}{w^{\sigma-1}} + 1},\tag{4.10}
$$

where $\sigma = \frac{1}{(1-\rho)} > 1$ is the constant elasticity of substitution between consumption and leisure. The next graph illustrates the pre-tax individuals' labour supply for three different values of σ , i.e. 3, 2 and 1.5, which are associated with the highest, the central and the bottom curve respectively.

Fig. 2. Pre-tax optimal labour supply.

As Figure 2 shows, labour supply is increasing in the level of wage, that is individuals with higher wage work longer hours, moreover as the level of σ increases the labour supply curve becomes higher and each individual increases his labour supply. Given (4.10) it follows that the income in absence of taxation is

$$
y^* = w l^* = \frac{L w^{\sigma}}{\alpha^{\sigma} + w^{\sigma - 1}}.
$$
\n(4.11)

From (4.10) one obtains that the wage labour supply elasticity is

$$
\varepsilon_{l,w} = \frac{\partial l}{\partial w} \frac{w}{l} = \frac{(\sigma - 1) \alpha^{\sigma}}{\alpha^{\sigma} + w^{\sigma - 1}},\tag{4.12}
$$

which is positive for $\sigma > 1$. Moreover, the labour supply elasticity is decreasing in the level of individuals' wage, to this regard one can verify that the derivative of (4.12) w.r.t. the wage is negative, that is

$$
\frac{\partial \varepsilon}{\partial w} = -\frac{(\sigma - 1) w^{\sigma - 2}}{(\alpha^{\sigma} + w^{\sigma - 1})^2} < 0.
$$

Figure 3 illustrates the pattern of the wage labour supply elasticity for values of σ and α equal to 2 and 4.5 respectively.

Fig. 3. Labour supply wage elasticity.

Agents are subject to a taxation formalized by (4.8) that leads to the net disposable income, considered in their utility function, obtained as $x = y - T(y)$. Hence, one can equivalently re-express the problem in the space (x, y) for each agent. In this case the utility function (4.9) can be rewritten as

$$
u(x, y) = U(x, y/w) = x^{\rho} + \alpha (L - y/w)^{\rho}
$$
\n(4.13)

and the relation between x and y is

$$
x := y - T(y) = \begin{cases} (1 - t_1)y & \text{if } y \in Y_1 \equiv [0, y_1) \\ (t_2 - t_1)y_1 + (1 - t_2)y & \text{if } y \in Y_2 \equiv [y_1, y_2) \\ (t_2 - t_1)y_1 + (t_3 - t_2)y_2 + (1 - t_3)y & \text{if } y \in Y_3 \equiv [y_2, \infty) \end{cases}
$$
(4.14)

Where Y_i denotes the income set associated with the i^{th} income bracket, while the set $Y \setminus y_{i-1}$ will denote the set Y_i net of its lower element y_{i-1} , where $y_0 = 0$. Recall that we rule out the possibility to redistribute the collected revenue, so that the demo-grant Z is set equal to zero. Therefore, net incomes can be written in the general form as

$$
x = y(1 - t_i) + \sum_{i} (t_i - t_{i-1}) y_{i-1}.
$$
\n(4.15)

The marginal rate of substitution between gross income y and net income x is

$$
MRS_{yx} = \frac{\frac{\alpha}{w} \left(L - \frac{y}{w} \right)^{\rho - 1}}{x^{\rho - 1}},\tag{4.16}
$$

which is decreasing in the level of wage w, therefore in the space (x, y) the indifference curves are flatter the higher is the individual's wage.

From the agents' optimization problem it follows that for gross income levels that do not coincide with the thresholds $y_1 < y_2$ it should hold that $MRS_{yx} = (1 - t_i)$ when $y \in Y_i$. Given the definition of $y = w l$ one obtains that the above optimality condition requires $y_{i-1} < y^* < y_i$, that is

$$
y_{i-1} < \frac{Lw^{\sigma} - \left(\frac{\alpha}{1-t_i}\right)^{\sigma} \left(\sum_i y_{i-1} \left(t_i - t_{i-1}\right)\right)}{\frac{\alpha^{\sigma}}{\left(1-t_i\right)^{\sigma-1}} + w^{\sigma-1}} < y_i,\tag{4.17}
$$

with the associated post-tax labour supply

$$
l_{t}^{*} = \frac{L\left(\frac{w(1-t_{i})}{\alpha}\right)^{\sigma} - \sum_{i} y_{i-1} (t_{i} - t_{i-1})}{w(1-t_{i}) + \left(\frac{w(1-t_{i})}{\alpha}\right)^{\sigma}}
$$
(4.18)

for $i \in \{1, 2, 3\}$ where $y_3 = +\infty$, $y_0 = 0$. Then, the three sets of values can be expressed in terms of intervals of wages such that

$$
0 < \frac{Lw^{\sigma}}{\frac{\alpha^{\sigma}}{(1-t_1)^{\sigma-1}} + w^{\sigma-1}} < y_1
$$

$$
y_1 < \frac{Lw^{\sigma} - \left(\frac{\alpha}{1-t_2}\right)^{\sigma} (y_1 (t_2 - t_1))}{\frac{\alpha^{\sigma}}{(1-t_2)^{\sigma-1}} + w^{\sigma-1}} < y_2
$$

$$
y_2 < \frac{Lw^{\sigma} - \left(\frac{\alpha}{1-t_3}\right)^{\sigma} [y_2 (t_3 - t_2) + y_1 (t_2 - t_1)]}{\frac{\alpha^{\sigma}}{(1-t_3)^{\sigma-1}} + w^{\sigma-1}} < y_3.
$$

Depending on what tax regime is considered with respect to the value of t_3 compared to $t_1 \leq t_2$, we could have that under the convex regime some agents have a gross income equal to one of the thresholds y_1 and y_2 , or that if the tax system is non-convex, some agents could have a gross income coinciding with y_1 , while around y_2 the map of y^* w.r.t. w exhibits a jump.

In particular if $t_1 \le t_2 \le t_3$ the gross income distribution is as follows

$$
y^* = \begin{cases} \n\frac{Lw^{\sigma}}{\frac{\alpha^{\sigma}}{(1-t_1)^{\sigma-1}} + w^{\sigma-1}} & \text{if } w^{\sigma} \left(L - \frac{y_1}{w} \right) < Q_1 \\
y_1 & \text{if } Q_1 \le w^{\sigma} \left(L - \frac{y_1}{w} \right) \le Q_2 \\
\frac{Lw^{\sigma} - \left(\frac{\alpha}{1-t_2} \right)^{\sigma} [y_1(t_2 - t_1)]}{\frac{\alpha^{\sigma}}{(1-t_2)^{\sigma-1}} + w^{\sigma-1}} & \text{if } w^{\sigma} \left(L - \frac{y_1}{w} \right) > Q_2 \text{ and } w^{\sigma} \left(L - \frac{y_2}{w} \right) < Q_3 \\
y_2 & \text{if } Q_3 \le w^{\sigma} \left(L - \frac{y_2}{w} \right) \le Q_4 \\
\frac{Lw^{\sigma} - \left(\frac{\alpha}{1-t_3} \right)^{\sigma} [y_2(t_3 - t_2) + y_1(t_2 - t_1)]}{\frac{\alpha^{\sigma}}{(1-t_3)^{\sigma-1}} + w^{\sigma-1}} & \text{if } w^{\sigma} \left(L - \frac{y_2}{w} \right) > Q_4\n\end{cases} \tag{4.19}
$$

where the thresholds $\mathcal{Q}_j,$ for $j=1,2,3,4$ are defined as follows

$$
Q_1 = \left(\frac{\alpha}{1 - t_1}\right)^{\sigma} \left[y_1 (1 - t_1)\right],
$$

\n
$$
Q_2 = \left(\frac{\alpha}{1 - t_2}\right)^{\sigma} \left[y_1 (1 - t_1)\right],
$$

\n
$$
Q_3 = \left(\frac{\alpha}{1 - t_2}\right)^{\sigma} \left[y_2 (1 - t_2) + y_1 (t_2 - t_1)\right],
$$

\n
$$
Q_4 = \left(\frac{\alpha}{1 - t_3}\right)^{\sigma} \left[y_2 (1 - t_2) + y_1 (t_2 - t_1)\right].
$$

The labour supply levels associated with the gross income distribution (4.19) are

$$
l^* = \begin{cases}\n\frac{L\left(\frac{w(1-t_1)}{\alpha}\right)^\sigma}{w(1-t_1) + \left(\frac{w(1-t_1)}{\alpha}\right)^\sigma} & \text{if } w^\sigma \left(L - \frac{y_1}{w}\right) < Q_1 \\
y_1/w & \text{if } Q_1 \le w^\sigma \left(L - \frac{y_1}{w}\right) \le Q_2 \\
\frac{L\left(\frac{w(1-t_2)}{\alpha}\right)^\sigma - y_1(t_2 - t_1)}{w(1-t_2) + \left(\frac{w(1-t_2)}{\alpha}\right)^\sigma} & \text{if } w^\sigma \left(L - \frac{y_1}{w}\right) > Q_2 \text{ and } w^\sigma \left(L - \frac{y_2}{w}\right) < Q_3 \quad (4.20) \\
y_2/w & \text{if } Q_3 \le w^\sigma \left(L - \frac{y_2}{w}\right) \le Q_4 \\
\frac{L\left(\frac{w(1-t_3)}{\alpha}\right)^\sigma - y_2(t_3 - t_2) - y_1(t_2 - t_1)}{w(1-t_3) + \left(\frac{w(1-t_3)}{\alpha}\right)^\sigma} & \text{if } w^\sigma \left(L - \frac{y_2}{w}\right) > Q_4\n\end{cases}
$$

The next figure illustrates the optimal combination of (y, x) for five agents with different wage levels under the convex regime of tax rates. Given that the MRS_{yx} is decreasing in w, we have that as the wage increases the indifference curves become flatter.

Fig. 4. Gross & Net incomes under the CONVEX regime.

Under the non-convex regime $(t_1 \le t_3 \le t_2)$ the gross income distribution is

$$
y^* = \begin{cases} \n\frac{Lw^{\sigma}}{(1-t_1)^{\sigma-1}} + w^{\sigma-1} & \text{if } w^{\sigma} \left(L - \frac{y_1}{w} \right) < Q_1 \\
y_1 & \text{if } Q_1 \le w^{\sigma} \left(L - \frac{y_1}{w} \right) \le Q_2 \\
\frac{Lw^{\sigma} - \left(\frac{\alpha}{1-t_2} \right)^{\sigma} [y_1(t_2 - t_1)]}{\frac{\alpha^{\sigma}}{(1-t_2)^{\sigma-1}} + w^{\sigma-1}} & \text{if } w^{\sigma} \left(L - \frac{y_1}{w} \right) > Q_2 \text{ and } w^{\sigma} \left(L - \frac{y_2}{w} \right) \le \widehat{Q} \\
\frac{Lw^{\sigma} - \left(\frac{\alpha}{1-t_3} \right)^{\sigma} [y_2(t_3 - t_2) + y_1(t_2 - t_1)]}{\frac{\alpha^{\sigma}}{(1-t_3)^{\sigma-1}} + w^{\sigma-1}} & \text{if } w^{\sigma} \left(L - \frac{y_2}{w} \right) > \widehat{Q} \n\end{cases} \tag{4.21}
$$

where the gross incomes in the first income bracket and at the first threshold are the same as in the convex regime. The result changes for incomes within the second and the third bracket. More specifically, there are no agents having a gross income equal to y_2 , where the gross income distribution exhibits a discontinuity. Hence, there exists a threshold level \widehat{Q} distinguishing between incomes within the second and the third bracket. We do not derive explicitly this threshold and in the numerical simulations, for all individuals located above y_1 we identify their optimal gross income as the one associated with the larger utility.

The optimal levels of labour supply under the non-convex regime are

$$
l^* = \begin{cases}\n\frac{L\left(\frac{w(1-t_1)}{\alpha}\right)^{\sigma}}{w(1-t_1) + \left(\frac{w(1-t_1)}{\alpha}\right)^{\sigma}} & \text{if } w^{\sigma} \left(L - \frac{y_1}{w}\right) < Q_1 \\
y_1/w & \text{if } Q_1 \leq w^{\sigma} \left(L - \frac{y_1}{w}\right) \leq Q_2 \\
\frac{L\left(\frac{w(1-t_2)}{\alpha}\right)^{\sigma} - y_1(t_2 - t_1)}{w(1-t_2) + \left(\frac{w(1-t_2)}{\alpha}\right)^{\sigma}} & \text{if } w^{\sigma} \left(L - \frac{y_1}{w}\right) > Q_2 \text{ and } w^{\sigma} \left(L - \frac{y_2}{w}\right) \leq \widehat{Q} \tag{4.22} \\
\frac{L\left(\frac{w(1-t_3)}{\alpha}\right)^{\sigma} - y_2(t_3 - t_2) - y_1(t_2 - t_1)}{w(1-t_3) + \left(\frac{w(1-t_3)}{\alpha}\right)^{\sigma}} & \text{if } w^{\sigma} \left(L - \frac{y_2}{w}\right) > \widehat{Q}\n\end{cases}
$$

Figure 5 shows the optimal levels of gross income y and the associated net income x for four agents with different wages under a non-convex regime of tax rates. Recall that higher wages are associated with flatter indifference curves.

Fig. 5. Gross & Net income under the NON-convex regime.

4.4 Simulations results

In this section we present numerical results showing the optimal tax system for each weighting function under the two tax rates regimes (convex and non-convex). In line with Chapter 2, simulations are based on a truncated Pareto distribution of 1000 gross wages. This distribution is truncated both below and above, with the two bounds equal to 20 and 327 respectively. The average gross wage is equal to 48.04. Then, in order to solve the optimal tax problem described in the previous section, we need to assign a value to the parameters L, ρ and α of the utility function (4.9) . We set the maximal level of labour supply L equal to one, while the choice of the other two parameters is crucial to determine the profile of the wage labour supply elasticity (4.12) . In particular, we set ρ and α in such a way that the pattern of the wage labour supply elasticity is broadly consistent with the main Öndings provided by the empirical literature. To this end, recall that empirical estimates (see Bargain and Peichl 2016 and Meghir and Phillips 2008) of labour supply elasticity are higher for the extensive margin (choice between working and not working) than for the intensive margin (choice on how much to work). The estimates suggest also that some demographics groups (married women) are more responsive than others (men). Moreover, labour supply elasticity tends to decline as the individuals' income increases.

The parameters ρ and α we choose lead to a wage labour supply elasticity profile broadly consistent with the estimates obtained in Aaberge and Colombino (2013), although we do not consider different demographics groups.

More specifically we consider two different profiles of the wage labour supply elasticity, which are illustrated in Figure 6. Specifically the solid (dashed) line is obtained setting ρ and α equal to 0.67 (0.5) and 8.712 (4.5) respectively. In both case the average labour supply is equal to 0.64.

Fig. 6. Wage labour supply elasticity Distr. 1A and 1B.

As anticipated the elasticity sharply declines as the wage increases. Recall that we consider a truncated wage distribution with lower bound equal to 20, then the elasticity at the bottom percentile is around 0.5 and 1.3 for the dashed and the solid line respectively. However, from (4.11) one can note that the combination of (ρ, α) has an impact on the gross income distribution in absence of taxation as well. Figure 7 illustrates the gross income distributions associated with the two pairs of (ρ, α) considered above.

Fig. 7. Gross Income Distributions 1A and 1B.

In particular, Distribution $1A(1B)$ is associated with the elasticity profile illustrated by the dashed (solid) line in Figure 6. The average gross income μ and the median income M in absence of taxation are respectively equal to 35.03 and 19.90 for Distribution 1A and 37.01 and 19.84 for Distribution 1B. Moreover, Distribution 1A exhibits lower levels of pre-tax inequality and polarization than Distribution 1B, that is $I_{1A} = 0.47$, $P_{1A} = 0.13$ and $I_{1B} = 0.51$, $P_{1B} = 0.15$.

As in Chapter 2 the simulations are performed by using a grid search method.³ Recall also that the revenue requirement G is defined as a fraction of the average gross income in absence of taxation. More specifically, we consider four different levels of G , i.e. 10% , 15% , 20% and 25% of the average gross income. We also assume that the value of the marginal tax rates can not exceed an upper limit which we set equal to 50%.

³We define the grids for t_1, t_2, y_1 and y_2 , with $t_1 \le t_2$ and $y_1 \le y_2$. Then, for each combination of these policy parameters we compute the value of t_3 which satisfies the budget constraint and then we compute the value of the SEF. Last, we identify the combination associated with the highest value of the SEF.

					Table 1. Optimal tax system for Distribution 1A ($\mu = 35$, $I = 0.47$ and $P = 0.13$).						
$\bar{l} = 0.64$	Convex tax-system						Opt.				
$G\$	t_1	t_2	t_3	\mathfrak{y}_1	\mathcal{y}_2	t_1	t_2	$t_{3}% \rightarrow-\infty$	$y_1\,$	\mathcal{y}_2	Syst.
Gini based SEF											
$0.10 \times \mu$	$\boldsymbol{0}$	30%	46\%	$40\,$	$54\,$	θ	44\%	32%	$42\,$	260	$\mbox{Con.}$
				0.83	0.90				0.87	0.99	
$0.15 \times \mu$	θ	36%	50%	24	$36\,$	$\boldsymbol{0}$	50%	34\%	26	260	Con.
				0.72	0.85				0.80	0.99	
$0.20 \times \mu$	8%	36%	50%	$22\,$	$34\,$	8%	50%	38\%	$24\,$	230	Con.
				0.69	0.83				0.78	0.99	
$0.25 \times \mu$	18%	40\%	50%	24	$36\,$	18%	50%	40\%	26	230	Con.
				0.72	0.84				0.78	$0.99\,$	
Polarization based SEF											
$0.10 \times \mu$	10%	10%	10%	350	$350\,$	7%	28%	8%	18	$32\,$	$\mbox{Non-}C.$
				$\mathbf{1}$	$\mathbf{1}$				0.57	0.73	
$0.15 \times \mu$	16%	16%	16%	300	300	13%	30%	14%	18	32	Non-C.
				$\mathbf{1}$	$\mathbf{1}$				0.57	0.73	
$0.20 \times \mu$	21\%	21\%	21\%	300	300	20%	34\%	20%	18	$30\,$	Non-C.
				$\mathbf{1}$	$\mathbf{1}$				0.58	0.72	
$0.25 \times \mu$	27%	27%	27%	$300\,$	300	26%	38\%	26%	16	28	Non-C.
				$\mathbf{1}$	$\mathbf{1}$				0.54	0.71	
$(1/2Gini + 1/2$ Polarization) based SEF											
$0.10 \times \mu$	$\boldsymbol{0}$	19%	24%	$20\,$	$46\,$	$\overline{0}$	22%	1%	$20\,$	260	Non-C.
				$0.58\,$	$0.83\,$				0.60	0.99	
$0.15 \times \mu$	3%	27%	31%	18	$34\,$	3%	30%	4%	18	260	Non-C.
				$0.58\,$	0.78				0.59	0.99	
$0.20 \times \mu$	12\%	29%	33%	$18\,$	$40\,$	12\%	32%	14\%	$18\,$	230	Non-C.
				$0.57\,$	$0.81\,$				0.58	0.99	
$0.25 \times \mu$	22%	31%	34%	18	34_{151}	22%	34%	22%	18	230	Non-C.
				$0.55\,$	$0.77\,$				$0.57\,$	$0.99\,$	

Tables 1 and 2 report the simulations results for Distribution 1A and 1B respectively.

Note. The two values reported in the columns y_1 and y_2 express the thresholds in

terms of the income level and the associated percentile in the income distribution.

$\bar{l} = 0.64$	$Convex$ tax-system				Non-convex tax-system					Opt.	
G	t_1	t_2	$t_{3}% \sqrt{t_{1}}\cdot\vec{r}_{2}$	$y_1\,$	$\mathcal{Y}2$	t_1	t_2	t_3	$y_1\,$	y_2	Syst.
Gini based SEF											
$0.10 \times \mu$	θ	34%	50%	53 0.86	83 0.93	$\boldsymbol{0}$	48%	46\%	$55\,$ 0.89	89 0.93	Con.
$0.15 \times \mu$	θ	32%	50%	29 0.75	51 0.88	θ	48%	40%	31 0.83	300 0.99	Con.
$0.20 \times \mu$	5%	36%	50%	21 0.70	41 0.86	6%	50%	40%	25 0.80	300 0.99	Con.
$0.25 \times \mu$	16%	38\%	50%	21 0.69	41 0.85	16%	50%	40\%	25 0.79	300 0.99	Con.
Polarization based SEF											
$0.10 \times \mu$	θ	18%	24\%	19 0.58	120 0.94	0%	24%	18%	19 0.61	31 0.72	Non-C.
$0.15 \times \mu$	5%	24\%	28\%	17 0.57	110 0.94	5%	30%	24\%	17 0.61	25 0.68	Non-C.
$0.20 \times \mu$	13%	28%	32%	17 0.58	105 0.93	14%	32%	28\%	17 0.61	$25\,$ 0.69	Non-C.
$0.25 \times \mu$	23%	32%	36%	15 0.55	110 0.94	24\%	36%	32%	15 0.58	23 0.67	Non-C.
$(1/2Gini + 1/2$ Polarization) based SEF											
$0.10 \times \mu$	$\overline{0}$	18%	34%	$29\,$ 0.70	$83\,$ $0.91\,$	$\boldsymbol{0}$	28%	24%	$35\,$ 0.77	$300\,$ $0.99\,$	Con.
$0.15 \times \mu$	$\overline{0}$	26%	40\%	$19\,$ $0.62\,$	$73\,$ 0.90	$\boldsymbol{0}$	32%	14\%	$19\,$ 0.66	260 0.99	Con.
$0.20 \times \mu$	6%	32%	44\%	17 0.62	65 0.89	6%	38%	28\%	17 0.66	$300\,$ $0.99\,$	$\mbox{Con.}$
0.25 \times μ	17%	34\%	46\%	17 $0.61\,$	63 $0.89\,$	17%	40%	30%	17 $0.54\,$	$300\,$ 0.77	Con.

Table 2. Optimal tax system for Distribution 1B ($\mu = 37, I = 0.51$ and $P = 0.15$).

For both distributions the optimal tax system reducing inequality is convex with increasing marginal tax rates. The tax rate in the first income bracket is set equal to zero unless when the revenue requirement is high (0.20% and 0.25%). However, note that Distribution 1A is characterized by a narrow first bracket and higher tax rates in the first two brackets than Distribution 1B (compare top left panel of Tables 1 and 2). This result is related to the fact that Distribution 1A exhibits lower wage labour supply elasticity at the bottom than Distribution 1B (see Figure 6). When the required revenue increases the marginal tax rates in the Örst two brackets increase, while the two income thresholds decrease.

This result is qualitatively in line with the results we obtain in Chapter 2, where the wage labour supply elasticity is assumed to be constant throughout the entire distribution. However, differently from there, here the first income threshold is always greater than the median income. In Chapter 2, indeed, we obtain that for low values of elasticity and high revenue requirements the median income is taxed with a marginal tax rate which is almost equal to the maximal admissible. To this regard, the results reported in Tables 1 and 2 are more consistent with those in Chapter 2 related to the case of high elasticity.

When the government's non-welfarist objective shifts to polarization concerns the optimal tax system, as obtained in Chapter 2, becomes non-convex. This result holds for both distributions (see central panel of Tables 1 and 2). More specifically, the optimal tax system reducing polarization requires a central bracket with a high marginal tax rate. The marginal tax rates increase with the revenue requirement (see central right panel of Tables 1 and 2). Comparing the optimal non-convex tax systems reducing polarization for Distributions 1A and 1B two interesting findings should be noted. First, the central bracket is larger in the case of Distribution 1A (compare columns y_1 and y_2 of central right panel of Tables 1 and 2). Second, marginal tax rates are higher for Distribution 1A, but the interesting aspect is the profile of the two marginal tax rates t_1 and t_3 . In particular, in the case of Distribution 1A we have that $t_1 = t_3$, while for Distribution 1B it happens that $t_1 < t_3$. The reasons for choosing $t_1 < t_3$ is related to the fact that the wage labour supply elasticity of the incomes in the bottom percentiles is higher in Distribution 1B than in Distribution 1A (see Figure 6). The difference in terms of the wage elasticity profile has another consequence. In particular, when we consider the optimal tax system reducing polarization under the convex regime we obtain that for Distribution 1A

this tax system is based on a proportional taxation, while for Distribution 1B the marginal tax rates are increasing with the optimal t_1 always lower than the one obtained for Distribution 1A (compare central left panel of Tables 1A and B). The median income falls within the first income brackets, this result is consistent with the conclusions of Chapter 2, where the median income belongs to the central interval only for low levels of labour supply elasticity.

Last, for SEFs taking into account both inequality and polarization concerns we have that the optimal tax system is non-convex in the case of Distribution 1A and convex for Distribution 1B (compare bottom panel of Tables 1 and 2). Specifically, for Distribution 1A, where the top wage elasticity is greater than Distribution 1B, it is socially desirable to reduce the marginal tax rate on the extreme right tail. The marginal tax rates with this SEF are lower than those obtained with the Gini and Polarization SEFs, except for the levels of t_1 which are comprised between the two values obtained under the Gini and Polarization SEFs. As to the two income thresholds, we have that y_1 is in line with the values obtained in the case of Polarization SEFs, while y_2 coincides with the levels associated to the Gini SEFs.

When we consider Distribution 1B, characterized by a high levels of inequality and polarization and showing higher (lower) level of labour supply elasticity at the bottom (top), the optimal tax system considering inequality and polarization concerns, requires increasing marginal tax rates. Both the optimal marginal tax rates and the income thresholds are comprised in between the levels obtained with the Gini and Polarization SEFs.

4.5 Concluding remarks

In this work we extend the analysis of Chapter 2, where we investigate how the specific government's redistributive objective impacts on the shape of the optimal tax system. As in Chapter 2 we consider a piecewise three brackets linear tax schedule and we focus on the optimal tax scheme collecting a given amount of per-capita revenue under the assumption of no redistribution of the collected income. We analyze two different regimes of the tax rates depending on the ranking of t_2 and t_3 , i.e. convex $t_2 \le t_3$ and non-convex $t_2 \ge t_3$.

With respect to Chapter 2, the novelty of this work is that, here we replace the assumption of constant labour supply elasticity throughout the entire distribution, with a more realistic assumption that elasticity changes according to individuals' characteristics. In particular, in line with the empirical estimates of other works (Aaberge and Colombino 2013), we assume that labour supply elasticity is decreasing in the level of individualsíwage. Moreover, we introduce also a third weighting function, which is obtained as a combination of the two weighting functions expressing concerns for inequality or polarization reduction.

The results we obtain qualitatively confirm those obtained in Chapter 2, that is the optimal tax system changes according to the non-welfarist redistributive objective. In particular the optimal tax system for inequality reduction is convex exhibiting increasing marginal tax rates. With respect to Chapter 2 we find that the first bracket is wider (including always the median income) and taxed with lower marginal tax rate. This result is related to the introduction of non-constant elasticity with high values in the bottom percentiles.

As to polarization SEFs the optimal tax system is non-convex, however the marginal tax rates are lower than those obtained in Chapter 2. Hence, the main implication of the assumption of non-constant elasticity is the reduction of the tax burden for income percentiles exhibiting higher elasticity.

Last, when the weighting function takes into account both inequality and polarization concerns, the shape of the optimal tax system depends on the level of initial dispersion. More specifically, for high level of pre-tax inequality and polarization the optimal tax scheme is convex, while when the levels of inequality and polarization are low the optimal system is nonconvex. In both cases the tax rates and the income thresholds are comprised in between the values associated with the tax system obtained for Gini and polarization SEFs respectively.

Bibliography

- [1] Aaberge, R. (2000). Characterizations of Lorenz curves and income distributions. Social Choice and Welfare, 17 (4): 639-653.
- [2] Aaberge, R. and Atkinson, A.B. (2013). The median as watershed. Discussion Papers No. 749, (2013). Statistics Norway, Research Department.
- [3] Aaberge, R. and Colombino, U. (2013). Using a microeconometric model of household labour supply to design optimal income taxes. Scandinavian Journal of Economics, 115 (2): 449-475.
- [4] 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] Dixit, A. and Sadmo, A. (1977). Some simplified formulae for optimal income taxation. Scandinavian Journal of Economics, 79, 417-423.
- [20] Donaldson, D. and Weymark, J.A. (1983). Ethically flexible Gini indices for income distributions in the continuum. Journal of Economic Theory, 29: 353-358.
- [21] Duclos, J. Y., Esteban, J. and Ray, D. (2004). Polarization: Concepts, Measurement, Estimation. Econometrica, 72: 1737-1772.
- [22] Ebert, U. and Moyes, P. (2003). The difficulty of income redistribution with labour supply. Economics Bullettin, 8 (2): 1-9.
- [23] Ebert, U. and Moyes, P. (2007). Income taxation with labor responses. Journal of Public Economic Theory, 9 (4): 653-682.
- [24] Esteban, J. and Ray, D. (1994). On the Measurement of Polarization. Econometrica, 62: 819-852.
- [25] Fellman, J. (1976). The effect of transformation on Lorenz curves. Econometrica, 44: 823-824.
- [26] Giertz, S.H. (2004). Recent literature on taxable income elasticities. Tax Analysis Division Congressional Budget Office, Washington D.C., paper 2004-16.
- [27] Jakobsson, U. (1976). On the measurement of the degree of progression. Journal of Public Economics, 5: 161-168.
- [28] Kakwani, N. C. (1977). Applications of Lorenz curves in economic analysis. Econometrica, 87: 71-80.
- [29] 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.
- [30] Kanbur, R, Pirtilla, J. and Tuomala, M. (2006). Non-Welfarist Optimal Taxation and Behavioural Public Economics. Journal of Economic Surveys, 20 (5): 849-868.
- [31] Kanbur R, and Tuomala, M. (1994). Inherent Inequality and the Optimal Graduation of Marginal Tax Rates. Scandinavian Journal of Economics 96 (2): 275-282.
- [32] Keen, M., Papapanagos H. and A. Shorrocks (2000): Tax Reform and Progressivity. The Economic Journal, Vol. 110, No. 460, pp. 50-68
- [33] Lambert, P.J. (2001). The Distribution and Redistribution of Income. Manchester University Press.
- [34] Maccheroni, F., Muliere P., Zoli C. (2005). Inverse stochastic orders and generalized Gini functionals. Metron, LXIII, (3): 529-559.
- [35] Mankiw, N. G., Weinzierl, M. and Yagan, D. (2009). Optimal Taxation in Theory and Practice. The Journal of Economic Perspectives, 23 (4): 147-174.
- [36] Marshall, A. W., Olkin, I. and Arnold, B. C. (2011). Inequalities: Theory of Majorization and Its Applications, Springer.
- [37] Meghir, C. and Phillips, D. (2008). Labour Supply and Taxes. Iza Discussion Paper no. 3405.
- [38] Mirrlees, J.A. (1971). An Exploration in the Theory of Optimum Income Taxation. Review of Economic Studies, 38: 175-208.
- [39] Mirrlees, J.A., Adam, S., Besley, T. et al. (2011). Tax by design. Oxford University Press.
- [40] OECD (2011). Revenues Statistics, 1965-2010. OECD, Paris, 2011 Edition.
- [41] OECD (2015). Revenues Statistics, 1965-2014. OECD, Paris, 2015 Edition.
- [42] O'Donoghue, T. and Rabin, M. (2003). Studying optimal paternalism, illustrated by a model of sin taxes. American Economic Review, 93: 186-191.
- [43] Onrubia, J., Salas, R. and Sanz, J.F. (2005). Redistribution and labour supply. Journal of Economic Inequality, (2005) 3: 109-124.
- [44] 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.
- [45] Piketty, T. and Saez, E. (2012). Optimal Labor Income Taxation. NBER Wp series 18521.
- [46] 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.
- [47] Prete, V., Sommacal, A. and Zoli, C. (2016). Optimal Non-Welfarist Income Taxation for Inequality and Polarization Reduction. Working Paper:23 December 2016, Department of Economics, University of Verona.
- [48] 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.
- [49] Sadka, E. (1976). On Income Distribution, Incentive Effects and Optimal Income Taxation. Review of Economic Studies, 43: 261-268.
- [50] Saez, E. (2002). Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses. Quarterly Journal of Economics 117, 1039-1073.
- [51] 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.
- [52] Saez, E. and Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. American Economic Review. 106 (1): 24-45.
- [53] Schroyen, F. (2005). An alternative way to model merit good arguments. Journal of Public Economics, 89: 957-966.
- [54] Seade, J. (1977). On the Shape of Optimal Tax Schedules. Journal of Public Economics, 7: 203-235.
- [55] Sen, A. (1985). Commodities and Capabilities. Amsterdam: North-Holland.
- [56] Sheshinski, E. (1972). The optimal linear income tax. Review of Economic Studies, 39: 297-302.
- [57] Sheshinski, E. (1989). Note on the shape of the optimum income tax schedule. Journal of Public Economics, 40: 201-215.
- [58] Slack, S.E. (2015). Revisiting the optimal linear income tax with categorical transfers. Economics Letters, 134: 73-77.
- [59] Slemrod, J. and Sorum, N. (1984). The Compliance Costs of the U.S. Individual Tax System. National Tax Journal, 37 (4): 461-474.
- [60] Slemrod, J., Yitzhaki, S., Mayshar, J. and Lundholm, M. (1994). The optimal two-bracket linear income tax. Journal of Public Economics, 53: 269-290.
- [61] Stern, N. H. (1976). On the Specification of Models of Optimum Income Taxation. Journal of Public Economics, 6: 123-162.
- [62] Suits, D. (1977): Measurement of Tax Progressivity. American Economic Review, 67: 747- 752.
- [63] Tuomala, M. (1984). On Optimal Income Taxation. Some further numerical results. Journal of Public Economics, 23: 351-366.
- [64] Tuomala, M. (1985). Simplified Formulae for Optimal Linear Income Taxation. Scandinavian Journal of Economics 87 (4): 668-672.
- [65] Tuomala, M. (1990). Optimal Income Tax and Redistribution. Clarendon Press, Oxford.
- [66] Wang, Y. and Tsui, K. (2000). Polarization orderings and new classes of polarization indices. Journal of Public Economic Theory, 2, 349-363.
- [67] Weymark, J.A. (1981). Generalized Gini inequality indices. Mathematical Social Sciences, 1: 409-430.
- [68] Wolfson, M. C. (1994). When inequalities diverge. American Economic Review, 84(2), 353-358.
- [69] Yaari, M.E. (1987). The dual theory of choice under risk. Econometrica 55: 99-115.
- [70] Yaari, M.E. (1988). A controversial proposal concerning inequality measurement. Journal of Economic Theory 44: 381-397.
- [71] Yitzhaki, S. (1983). On an extension of the Gini inequality index. International Economic Review 24: 617-628.