Paper submitted for the EEA conference in Barcelona 2023 – to be complemented with simulation results and proofs

Optimal Taxation and Social Preferences**

Thomas Aronsson and Olof Johansson-Stenman+*

February 2023

Abstract

The present paper analyzes optimal redistributive income taxation in a Mirrleesian framework extended with social preferences at the individual level. We start by examining a general model where the social preference component of the individual utility functions is formulated to encompass almost any form of preferences for other people's consumption, and then continue with four prominent special cases. Two of these reflect self-centered inequality aversion, based on Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), whereas the other two reflect non-self-centered inequality aversion, where people have preferences for a low Gini coefficient and a high minimum consumption level in the society, respectively. We find that social preferences may have a considerable impact on the structure and levels of marginal taxation, including the top income tax levels, and that different types of social preferences have very different implications for optimal taxation.

JEL: D62, D90, H21, H23.

Keywords: Optimal Taxation, Redistribution, Social Preferences, Inequality Aversion.

** The authors would like to thank Thomas Gaube and seminar participants at Colorado State University, University of Wyoming, and the IIPF conference in Tampere 2018 for helpful comments and suggestions. Research grants from The Marianne and Marcus Wallenberg Foundation (MMW 2015.0037) and The Swedish Research Council (ref 2020-02208) are also gratefully acknowledged.

* Address: Department of Economics, Umeå School of Business, Economics and Statistics, Umeå University, SE – 901 87 Umeå, Sweden. E-mail: Thomas.Aronsson@umu.se.

+ Address: Department of Economics, School of Business, Economics and Law, University of Gothenburg, SE – 405 30 Gothenburg, Sweden. E-mail: Olof.Johansson@economics.gu.se.

1. Introduction

Growing empirical evidence suggests that people typically have preferences that go beyond those of the narrowly selfish *Homo Economicus*, i.e., they also have social preferences. The purpose of the present paper is to integrate such preferences, in general as well as more specific forms, into the theory of optimal redistributive taxation.

Most models of optimal income taxation imply that the government prefers lower to higher inequality for a given aggregate gross income. One rational for this is to assume a concave utility function at the individual level, such that low-income individuals have higher marginal utility of consumption than high-income individuals. Another rational is to assume a prioritarian social welfare function in the sense that social welfare is concave in individual utilities (e.g., as in Diamond 1998). In fact, most models allow for both of these mechanisms, i.e., where both the individual utility functions and the social welfare function are (or can be) concave (e.g., Mirrlees 1971, Saez 2001). Sometimes the social objective function is modeled directly in terms of a concave function of individual consumption (e.g., Atkinson 1970), and sometimes, as in Saez and Stantcheva (2016), direct social welfare weights are applied where these weights are inversely related to consumption. In each of these cases, we can say that *the government* is inequality averse. At the same time, *individuals* are in these models almost always assumed *not* to care about inequality, or have social preferences more generally. That is, their utility is typically modelled to depend solely on their own consumption and labor supply/effort, and not on any measure of others' consumption.

The present paper, in contrast, analyzes the implications of social preferences, in the sense that people also care about measures of others' consumption, for optimal redistributive income taxation. Such a model enables us to present a very general characterization of the marginal tax schedule along the whole income distribution, including top income tax rates. Our approach is described in greater detail below.

We begin by presenting a very general model of optimal income taxation and social preferences, where the social preference component of the individual utility functions encompasses almost any form of preferences with respect to other people's consumption. In other words, this model is not restricted to inequality aversion or other types of pro-social preferences. It also encompasses models of social comparisons driven by concerns for social status, and even more generally reflect almost any form of consumption externality. Despite the complexity of the underlying model, the results show that the optimal marginal income tax can be written as a sum of two terms. One is a modified redistributive component (an analogue to the *ABC*-component described in Diamond's (1998) and Saez' (2001) interpretation of the solution to Mirrlees' (1971) optimal tax problem), where the modification arises because externalities affect the social costs and/or benefits of redistribution. The other is the value of the marginal externality that each individual imposes on other people. Note that the latter is type-specific, since the externality that social preferences give rise to is typically non-atmospheric in the sense that the marginal contribution to this externality differs among individuals. We also show that the corrective and redistributive aspects of tax policy interact in important ways, where redistributive elements directly affect the type-specific value of the marginal externality.

While there are many kinds of social preferences, there is extensive evidence suggesting that people tend to be inequality averse, in the sense of having preferences for a more equal distribution of consumption in society. We will therefore also analyze four specific models of inequality aversion, which are thus special cases of our general model. In each case, the theoretical analysis is combined with numerical simulations allowing us to go beyond the policy rules and quantify the importance of social preferences for the optimal marginal tax schedule as well as for the overall redistribution policy.

Two of the models focus on *self-centered* inequality aversion, based on the seminal contributions by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), respectively. By self-centered, we mean that individuals care about the relationships between their own consumption and other people's consumption, rather than about inequality *per se*. Note that the externalities generated here are typically more complex than those following in related models of status consumption, where people impose negative consumption externalities on one another (see Section 2). In the models presented below, increased consumption of a specific individual may either lead to more or less inequality, and can thus lead to a negative or positive externality depending on this individual's position in the distribution of disposable incomes.

In (our continuous-type version of) the model by Fehr and Schmidt (1999), individuals compare their own consumption both upwards and downwards in the distribution of consumption (or disposable incomes) and experience disutility of discrepancies in both dimensions, although possibly to a larger extent in the upward direction. The resulting consumption externality is then *non-atmospheric*. Thus, while each individual will experience an externality that depends on other people's consumption, the contribution to that externality per consumption unit will vary in the population. In particular, additional consumption by an individual whose consumption is above that of an individual *A* will impose a negative externality on *A*, while the externality is positive for an individual whose consumption is below that of *A*. Consequently, by increasing their consumption, individuals impose positive externalities on people with higher consumption and negative externalities on people with lower consumption than themselves. In the first-best, this mechanism induces a monotonically increasing marginal income tax schedule, with a negative marginal income tax rates in the lower end of the income distribution and a positive one at the top, i.e., a progressive marginal tax schedule. Indeed, we are able to present a closed-form solution showing that this marginal income tax rate depends only on the ordinal income rank and on the two parameters of the Fehr and Schmidt model, α and β, reflecting disadvantageous and advantageous inequality aversion, respectively. The second-best optimal taxation schedule is naturally more complex. Here we show that…

The Bolton and Ockenfels (2000) inequality aversion model of is in many ways similar to the Fehr and Schmidt model. In their model each individual's utility depends on the relation between their own consumption and the average of others' consumption in a non-monotonic way, such that each individual prefers to have a consumption level as close as possible to the average, *ceteris paribus*, i.e., for a constant level of their own consumption. This means that each individual's utility will depend on their own consumption as well as the average consumption level, where utility increases in average consumption for individuals whose consumption is above the average, and vice versa. But since the contribution to the average is of course the same per consumption unit for all, the externalities are here *atmospheric*, in contrast to the Fehr and Schmidt model. This, in turn, means that in the first-best, the externality-correcting marginal income tax is here the same for alland simply equal to the sum of all people's marginal willingness to pay for a decreased average consumption, based on a Pigovian logic. However, since all people with a consumption level below the average will have a positive marginal willingness to pay and those above average a negative one, the overall sum, and hence the first-best marginal income tax, is often rather small; indeed, we will present examples where it is exactly zero. Therefore, seemingly very similar models of inequality aversion, such as those presented in Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), may have very different policy implications. In the second-best…

The remaining two models are based on *non-self-centered* inequality aversion, where individuals are concerned with inequality *per se* (and not the relationship between their own consumption and other people's consumption). We take here a broad perspective we abeyond narrow hedonic interpretations, where we can alternatively interpret these models as reduced forms taking instrumental effects of inequality, such as crime and the implications thereof, into account (see the next section).

In one of these models, people prefer to live in a society with lower inequality as measued by the consumption Gini coefficient. Since consumption increases by different individuals will then clearly affect the Gini coefficient differently, the externalities are here non-atmospheric as for the Fehr and Schmidt model. Moreover, the first-best marginal income taxation will then be negative for individuals for which a small consumption increase reduces the Gini coefficient, which is shown to be for a majority of the population up to the income level where a small consumption increase leaves the Gini unchanged. More precisely, this is the case for the bottom income share $(1+G)/2$, where G is the Gini coefficient. Correspondingly, the marginal income tax is positive for the share above this threshold, i.e., for the remaining upper income share $(1-G)/2$. We present algebraic second-best rules for the marginal income tax for several functional form classes. In our numerical simulations we find that inequality aversion typically leads to higher marginal income tax rates for a large range of incomes as well as a more progressive marginal tax structure. This outcome is thus reminiscent of the way Fehr-Schmidt preferences for self-centered inequality aversion affects the marginal tax structure.

As our seond model of non-self-centered inequality aversion, we consider a case of Rawlsian poverty aversion, where people are concerned about the consumption of the poorest group in society, inspired by Charness and Rabin (2002). Assuming that the poorest group in society is unemployed (or work very little), this type of social preference does not give rise to any corrective motive for income taxation. Yet, it tends, nevertheless, to increase the marginal income tax rates for redistributive reasons.

We also present top marginal income tax rates of two kinds: First, we follow convention based on an unbounded ability-distribution, where we present analytic expressions on top marginal tax rates that depend on both the thickness of the upper tail of the ability distribution, as expressed by an assumed constant Pareto-parameter, and expressions that vary depending on the specific type of the inequality aversion. Second, realizing that an unbounded ability distribution is of course impossible in the finite world we live in, we present optimal marginal income tax rates at what we denote *the very top* of the income distribution, i.e. for the single individual with the highest income. Without externalities, this marginal income tax rate would clearly be zero, as shown already by Sadka (1976) and Seade (1977). However, except for the Rawlsian case, we show that inequality aversion would typically make this top marginal income tax rate positive and substantial.

The remainder of the paper is outlined as follows. Section 2 provides a brief literature review followed by the presentation and analysis of the general model of optimal income taxation under social preferences in Section 3. Sections 4 and 5 present the corresponding analyses of social preference models characterized by self-centered and non-self-centered inequality aversion, respectively. Section 6 presents the analysis of top-income marginal tax rates and Section 7 concludes the paper; proofs are presented in the Appendix.

2. Literature Review

There is a large experimental and empirical literature on social preferences in general, and inequality aversion in particular. For experimental work on inequality aversion, see, e.g., Fehr and Schmidt (1999, 2003), Bolton and Ockenfels (2000), Fisman et al. (2007), Bellemare et al. (2008), and Bruhin et al. (2019). The broad message is that people prefer a more equal to unequal allocation, *ceteris paribus*, and that they are willing to trade off some of their own income in order to obtain a more equitable allocation. Part of this literature focuses on the potential context-dependence of preferences for equality, where these preferences largely depend on the perceived fairness, suggesting that some inequalities are perceived as more fair than others (e.g., Cappelen et al. 2013; Almås et al. 2020), and that people's willingness to forsake their own income is related to others' previous actions and associated perceived intentions (e.g., Charness and Rabin 2002). This adds to the perhaps obvious conclusion that it is far from straightforward to generalize quantitative estimates from specific experimental settings to a broader real-life social setting. Alesina and Giuliano (2011) provide a broad overview of the literature on other-regarding preferences.

Regarding the potential instrumental effects of inequality, there is substantial cross-country evidence of a robust positive correlation between the incidence of crime and the extent of income inequality (e.g., Fajnzylber et al. 2002), while there is, not surprisingly, less straightforward to clearly identify causal relationships; see Glaeser et al. (1996) and Kelly (2000). There are also studies on the potential impact of inequality on social capital. For example, Alesina and La Ferrara (2000) find that participation in social activities is lower in more unequal societies, whereas Alesina and La Ferrara (2002) find that trust is lower in areas with a more uneven distribution of income. Some argue that inequality contributes to a society's degree of polarization, which in turn may lead to social tension in general and in extreme cases to problems such as civil war (Esteban and Ray 1994; Collier and Hoeffler 1998; Blattman and Miguel 2010).

The theoretical policy-oriented research based on models where people have motives other than material self-interest is considerably smaller. Yet, there is by now a sizable literature dealing with optimal taxation and public expenditure in economies where people are motivated by their relative consumption (or relative income).[1] This research typically assumes that people derive utility from their own consumption relative to that of referent others, i.e., individuals prefer to consume more than others and dislike consuming less, implying that people impose negative positional consumption externalities on one another. A natural interpretation is that relative consumption indicates social status, even if other interpretations are possible as well. Several of these studies find that positional externalities may motivate much higher marginal tax rates compared with conventional models of optimal taxation.

There are also research on other kinds of externalities and optimal taxation. Piketty et al. (2014) analyze negative effects of rent seeking among top earners, where such activities induce personal enrichment rather than increasing the size of the pie, and show that such behavior can motivate substantially larger marginal top income tax rates; Rothschild and Scheuer (2016) generalize and extend this analysis. Lockwood et al. (2017) analyze implications of varying externalities from different professions, arguing that high-paying professions tend to have negative externalities and low-paying professions positive ones. This implies higher optimal marginal taxes on top incomes.

---

[1] See e.g. Boskin and Sheshinski (1978), Oswald (1983), Frank (1985, 2005, 2008), Tuomala (2015), Corneo and Jeanne (1997), Ljungqvist and Uhlig (2000), Ireland (2001), Dupor and Liu (2003), Abel (2005), Aronsson and Johansson-Stenman (2008, 2010, 2015, 2018, 2021), Alvarez-Cuadrado and Long (2011, 2012), Eckerstorfer and Wendner (2013), and Kanbur and Tuomala (2013).

The theoretical policy-oriented literature allowing for *prosocial* preferences, in contrast, is very small. In fact, despite the extensive empirical and experimental evidence referred to above, social preferences are almost absent in the modern theory of optimal redistributive taxation. An exception is the study by Eckerstorfer and Wendner (2013) examining the joint implication of relative consumption concerns and altruism for optimal commodity taxation.[2] In their model, the altruism component in people's preferences means that each individual's utility depends positively on the average utility level in the economy as a whole.

Nyborg-Sjøstad and Cowell (2022) is most closely related to the present paper. The analytical part of their study is based on a Mirrleesian model of optimal taxation where agents have quasi-linear preferences, and where the measure of equality that people care about is given by the Gini coefficient. Thus, they focus on a specific form of non-self-centered inequality aversion. They show how the resulting inequality externality affects the policy rules for marginal income taxation, and based on numerical simulations that inequality aversion may lead to a more progressive marginal tax structure. They also provide a broad and insightful discussion more generally of consumption externalities induced by preferences for equality.

Simula and Trannoy (2022) also examine a model of optimal nonlinear taxation in a Mirrleesian tradition. Like Nyborg-Sjøstad and Cowell, they largely focus on inequality in terms of the Gini coefficient, although they also consider other versions within the broader S-Gini family, as well as based on the A-family introduced by Aaberge (2000). However, in contrast to Nyborg-Sjøstad and Cowell, as well as most papers on optimal redistributive income taxation (including the present one), they consider a rank-dependent social welfare function (rather than a welfarist one) as the objective function. Their model implies that the social objective function will depend directly on the measure of inequality (e.g., the Gini coefficient), whereas individual utility only depends on the individual's own consumption and labor supply (through an additive specification).[3] The latter means that the individuals are not characterized by social preferences in their study.

---

[2] Dufwenberg et al. (2011) provide a more general theoretical treatment of other-regarding preferences in general equilibrium.

[3] See also Fleurbaey and Maniquet (2018) for a comprehensive and insightful treatment of different social objective functions and optimal income taxation.

Our study differs from, and goes beyond, those of Nyborg-Sjøstad and Cowell (2022) and Simula and Trannoy (2022) in several important ways. First, and foremost, we develop a general model of optimal redistributive taxation under social preferences, which encompasses virtually all possible versions of inequality aversion as well as other kinds of interdependent preferences such as concerns for relative consumption. This enables us to derive a policy rule for marginal income taxation, which is applicable to almost any (atmospheric or non-atmospheric) consumption externality that social preferences may give rise to. To our knowledge, such a framework is novel in the literature on optimal redistributive taxation. Second, we consider a much broader spectrum of specific models of social preferences, all of which are special cases of our general model, including self-centered and non-self-centered inequality aversion, respectively, and a Rawlsian framework where people have preferences for the outcome of the poorest group in society. Among other things, we show that seemingly similar models of inequality aversion can differ significantly in terms of the shape of the optimal marginal tax schedule as well as in terms of the levels of marginal taxation. Finally, in the special case where the relevant measure of inequality is represented by the Gini coefficient, we present results for other and more general settings than in the earlier studies.

3. A General Model of Optimal Income Taxation under Social Preferences

Consider an economy with linear production and competitive markets, implying that ability or marginal productivity reflects a fixed before-tax wage rate per unit of labor, *w*. Let $f(w)$ denote the density function of the ability-distribution. The population is normalized to one for notational convenience such that $\int_0^\infty f(w)dw = 1$. We follow convention in assuming that the single-crossing condition holds, implying that higher ability individuals earn a higher gross income and enjoy more consumption in equilibrium than lower ability individuals. Therefore, $F(w) = \int_0^w f(t)dt$ simultaneously reflects the ability distribution function and the ordinal rank of gross income and consumption.

3.1 Preferences and Individual Behavior

Individuals of any ability-type *w* derive utility from their own consumption, c_w, and labor supply/effort, l_w, as in conventional models. We also assume that individuals care about a social outcome, a key variable in the present paper, which we often interpret as a measure of inequality, although it can be given other interpretations as well. This measure, denoted I_w, typically varies between individuals and depends on the individual's own consumption as well as a type-specific measure of other people's consumption, H_w, such that

$$\mathrm{I}_w = \mathrm{I}(c_w, H_w) , \tag{1}$$

where

$$H_w = \int_0^\infty h_w(c_s) f(s) ds . \tag{2}$$

Thus, the weights attached to other people's consumption are type-specific and given by the function $h_w(\cdot)$ for an individual of type *w*. In addition, although individuals are assumed to care about the consumption distribution in the economy as a whole, they do not care enough to voluntarily give money to others, i.e., there is no charitable giving.[4]

We can then write the utility function as follows:[5]

$$U_w = v(c_w, l_w, \mathrm{I}_w) = v(c_w, l_w, \mathrm{I}(c_w, H_w)) = u(c_w, l_w, H_w) . \tag{3}$$

The function $v(\cdot)$ expresses the preferences in terms of the individual's own consumption and labor supply, respectively, and the measure of inequality described above. This function is increasing in consumption, c, decreasing in labor supply/effort, l, decreasing in the measure of inequality, I, and strictly quasi-concave. $u(\cdot)$ is a convenient reduced form to be used in some of the calculations below. Since the externality represented by H_w is typically non-atmospheric, it follows that a consumption change among type *s*-individuals can affect the utility of an individual of type *w* positively or negatively, depending on whether it leads to increased or decreased inequality, as experienced by individuals of type *w*.

If the preferences are weakly labor separable, a special case often examined in the literature on optimal redistributive taxation, equation (3) can be rewritten to read

$$U_w = v(c_w, l_w, \mathrm{I}_w) = \upsilon(q(c_w, \mathrm{I}_w), l_w) . \tag{4a}$$

[4] For recent research on charitable giving in an optimal taxation framework, see Aronsson et al. (2023a, 2023b).

[5] We follow convention in the literature on optimal taxation by taking the preferences as given. It is not the aim of the present paper to explain why people tend to have certain kinds of social preferences; see, e.g., Alger and Weibull (201x) for an evolutionary approach to social preferences.

With utility function (4a), the marginal rate of substitution between I and c does not depend directly on effort, l. Another frequent special case is the quasi-linear utility function, in which equation (3) can be written as follows:

$$U_w = v(c_w, l_w, \mathrm{I}_w) = V(c_w + g(l_w, \mathrm{I}_w)) \tag{4b}$$

We will return to the special cases of weak labor separability and quasi-linearity below.

For later use, an s-individual's marginal willingness to pay for an individual of type w to decrease her consumption is given by

$$M_{sw} = -\frac{u_H^{(s)}}{u_c^{(s)}} h_{c_w}^{(s)} = -MRS_{H,c}^{(s)} h_{c_w}^{(s)} \tag{5}$$

where a subscript attached to the utility function, the function $h(\cdot)$, or the MRS-function denotes partial derivative. By using $v_c + v_\mathrm{I}\mathrm{I}_c = u_c$, $v_\mathrm{I} I_H = u_H$, we can alternatively write equation (5) as

$$M_{sw} = -\frac{v_\mathrm{I}^{(s)} \mathrm{I}_H^{(s)} h_{c_w}^{(s)}}{v_c^{(s)} + v_\mathrm{I}^{(s)} \mathrm{I}_c^{(s)}} . \tag{6}$$

The aggregate (or mean) marginal willingness to pay among all individuals to avoid the externality generated by an individual of type w then becomes[6]

$$E(M_w) = -\int_0^\infty MRS_{H,c}^{(s)} h_{c_w}^{(s)} ds = -E(MRS_{H,c}) E(h_{c_w}) \left(1 + \kappa_w\right) \tag{7}$$

in which κ_w denotes the normalized covariance between the marginal willingness to pay to avoid the externality generated by type w and the effect of type w's consumption on the measure of inequality, i.e., $\kappa_w = \mathrm{cov}(MRS_{H,c} / E(MRS_{H,c}), h_{c_w} / E(h_{c_w}))$.

The individual budget constraints imply that private consumption equals gross income $y = wl$ minus the income tax

$$y_w - T(y_w) = c_w , \tag{8}$$

where $T(y_w)$ denotes a general, nonlinear tax function (where the tax payment can be either positive or negative).

Individuals are assumed to be atomistic agents in the sense of treating H_w as exogenous, which is a conventional assumption in models with externalities. Each individual of any type w

---

[6] This could, of course, have been expressed through the v-function instead using $v_c + v_\mathrm{I}\mathrm{I}_c = u_c$, $v_\mathrm{I}\mathrm{I}_H = u_H$.

chooses consumption and labor supply subject to the budget constraints implying the following first-order condition:

$$MRS_{l,c}^{(w)} = \frac{v_l^{(w)}}{v_c^{(w)} + v_{\mathrm{I}}^{(w)}\mathrm{I}_c^{(w)}} = \frac{u_l^{(w)}}{u_c^{(w)}} = -w\left(1 - T_y^{(w)}\right) , \tag{9}$$

where $T_y^{(w)}$ denotes the marginal income tax rate facing each individual of ability-type *w*.

3.2 Public Decision-Problem and Optimal Taxation

The government maximizes a generalized utilitarian social welfare function, as in, e.g., Mirrlees (1971) and Saez (2001),

$$W = \int_0^\infty \psi(u_w) f(w) dw , \tag{10}$$

where ψ is weakly concave. The resource constraint for the economy as a whole implies that aggregate production is equal to aggregate consumption

$$\int_0^\infty w l_w f(w) dw = \int_0^\infty c_w f(w) dw . \tag{11}$$

The incentive compatibility constraint, preventing each individual from mimicking the adjacent type with lower productivity (by choosing the labor supply in order to reach the same income as this type), can be written as

$$\frac{dU_w}{dw} = -\frac{l_w u_l^{(w)}}{w} . \tag{12}$$

As this constraint holds for each type, we can use partial integration to derive

$$\int_0^\infty \theta_w \left(\frac{dU_w}{dw} + \frac{u_l(c_w, l_w, H_w))l_w}{w} \right) dw = \int_0^\infty \left(\theta_w \frac{u_l(c_w, l_w, H_w))l_w}{w} - \dot{\theta}_w U_w \right) dw + \theta_w U_w \Big|_{=0}^{=\infty} = 0 , \tag{13}$$

where θ_w is a differentiable multiplier.

The social decision-problem can now be expressed such that utility, U_w, is a state variable while l_w and H_w are control variables. Inverting the function $u(\cdot)$ in equation (3) and solving for c_w gives

$$c_w = k(l_w, H_w, U_w) . \tag{14}$$

The properties of the function $k(\cdot)$, applicable to all types *w*, can be summarized as follows

$$k_U = \frac{1}{u_c}; \;\; k_l = -\frac{u_l}{u_c}; \;\; k_H = -\frac{u_H}{u_c} \;\; . \tag{15}$$

By using the function $k(\cdot)$, the Lagrangean of the social decision-problem can then be written

$$\begin{aligned} L = & \int_0^\infty \psi(U_w)\, f(w)\, dw + \lambda \int_0^\infty \left(w l_w - k(l_w, H_w, U_w) \right) f(w)\, dw \\ & + \int_0^\infty \left(\theta_w \frac{u_l((k(l_w, H_w, U_w), l_w, H_w)) l_w}{w} - \dot{\theta}_w U_w \right) dw \qquad , \\ & + \int_0^\infty \eta_w \left(H_w - \int_0^\infty h_w(k(l_s, H_s, U_s))\, f(s) ds \right) f(w)\, dw \end{aligned} \tag{16}$$

where we have suppressed the term $U_\infty \theta_\infty - U_0 \theta_0$, which is zero by the transversality conditions. λ is the Lagrange multipliers attached to the resource constraint, θ_w the multiplier attached to the incentive compatibility constraint imposed on individuals of type *w*, and η_w is the Lagrange multiplier associated with the type-specific externality, H_w. The social first-order conditions are presented in the Appendix.

Now, let

$$\Gamma_w = \frac{1}{\lambda} \int_0^\infty \eta_s h_s{}'(c_w)\; f(s)\; ds \tag{17}$$

denote society's marginal willingness to pay to avoid the externality generated by the consumption of type *w*-individuals. As such, it will appear in the policy rules for marginal income taxation presented below. Let us also introduce the following short notation:

$$R_s^n = \int_0^\infty \cdots \int_0^\infty \int_0^\infty \int_0^\infty M_{r_n r_{n-1}}\;\; f(r_n)\;\; dr_n \ldots M_{r_2 r_1}\;\; f(r_2)\;\; dr_2 M_{r_1 s}\;\; f(r_1)\;\; dr_1 \tag{18a}$$

so

$$R_s^0 = 1 \;\; , \tag{18b}$$

$$R_s^1 = \int_0^\infty M_{r_1 s}\;\; f(r_1)\;\; dr_1 = E(M_s)\,, \tag{18c}$$

$$R_s^2 = \int_0^\infty \int_0^\infty M_{r_2 r_1}\;\; f(r_2)\;\; dr_2 M_{r_1 s}\;\; f(r_1)\;\; dr_1 = E\left(E(M) M_s \right) \tag{18d}$$

$$\begin{aligned} R_s^3 = & \int_0^\infty \int_0^\infty \int_0^\infty M_{r_3 r_2}\;\; f(r_3)\;\; dr_3 M_{r_2 r_1}\;\; f(r_2)\;\; dr_2 M_{r_1 s}\;\; f(r_1)\;\; dr_1 \\ & = E\left(E\left(E(M) M \right) M_s \right) \ldots \end{aligned} \quad . \tag{18e}$$

Here $E(M_s)$ thus measures the mean (or expected) value of all people's marginal willingness to pay for an individual of type *s* to decrerase her consumption. Since the population size is normalized to one, this also means that $E(M_s)$ reflects the sum of all people's marginal willingness to pay for reduced consumption by an individual of type *s*. $E(M)$ correspondingly denotes the mean of these $E(M_s)$ over all types *s*. Or in probabilistic terms, the expected value of the marginal willingness to pay of a random individual in the population for a decrease in another random individual's consumption.

If the preferences are weakly labor separable (see equation [4a]), the *R*-factors in (18) directly affect the policy rules for marginal income taxation by being part of the value of the marginal externality generated by any type *w*. However, in the general case, where the preferences are not necessarily labor separable, (18) must be modified to capture interaction effects between corrective and redistributive elements in the tax system. To do so, we start by presenting an analogue to the *ABC*-formulation introduced by Diamond (1998), through which the redistributive aspects of marginal taxation can be expressed in terms of estimable behavioral elasticities and the government's preferences for redistribution. Let $\delta_w = \psi'(u_w)u_c^{(w)} / \lambda$ denote the welfare weight the government attaches to individuals of type *w*, and let ς^u and ς^c denote the uncompensated and compensated labor supply elasticity, respectively, with respect to the marginal wage rate derived under a linearized budget constraint. We can then define (for all *w*) the *A*, *B*, and *C* factors introduced by Diamond (1998), but here generalized to allow for preferences that are not quasi-linear

$$A_w = \frac{1+\varsigma_w^u}{\varsigma_w^c} \quad , \tag{19a}$$

$$B_w = \int_w^\infty (1-\delta_s) \exp\left(-\int_w^s \frac{\partial MRS_{lc}^{(m)}}{\partial c} \frac{dy_m}{m} \right) \frac{f(s)}{1-F(w)}\, ds \,, \tag{19b}$$

$$C_w = \frac{1}{\varpi_w} \,, \tag{19c}$$

where

$$\varpi_w = -\frac{\partial(1-F(w))}{\partial w} \frac{w}{1-F(w)}$$

is the Pareto parameter that determines how fast the upper tail of the ability distribution decreases with the gross wage rate. A_w is interpretable as an efficiency mechanism based on behavioral labor supply responses, B_w reflects the urge for redistribution in favor of agents with abilities lower than or equal to *w* (which necessitates tax revenue raised from individuals with abilities higher than *w*), and C_w measures the thickness of the distribution of abilities higher than *w*. In a conventional model without externalities, as in Diamond (1998) or Saez (2001), the optimal marginal income tax rates will simply be given by $T_y^{(w)} / (1-T_y^{(w)}) = A_w B_w C_w$, where the intuition is well-know and well explained elsewhere.

Naturally, this simple rule will not hold in the presence of externalities. Yet, as will be shown, the optimal marginal tax rate can be written as a modified *ABC*-term plus a corrective term measuring the value of the marginal externality that individuals of type *w* impose on other people. The modification of the *ABC*-term refers to the *B*-factor, which takes the following form in our model:

$$\tilde{B}_w = \int_w^{\infty} \left(1-(\delta_s - \Gamma_s)\right) \exp\left(-\int_w^s \frac{\partial MRS_{lc}^{(m)}}{\partial c} \frac{dy_m}{m}\right) \frac{f(s)}{1-F(w)}\, ds \,. \tag{20}$$

Thus, we must deduct the value of the marginal externality, Γ_s, from δ_s in order to obtain the *social* marginal cost of decreased consumption for all individuals affected by the marginal tax increase on type *w*, i.e., for all *s* > *w*. Note that this extra component arises for redistributive reasons; it does hence not reflect the direct externality correction, which will be explored below.

With these preliminaries at our disposal, we are now in the position to modify the *R*-factors described above and then present the policy rule for marginal income taxation. Let

$$\varepsilon_{l(w)}^{H,c} = \frac{\partial MRS_{H,c}^{(w)}}{\partial l} \frac{l_w}{MRS_{H,c}^{(w)}}$$

denote the elasticity of $MRS_{H,c}$ with respect to the labor supply for an individual of type *w*. Under weak labor separability, this elasticity is zero, while it can be either positive or negative in the general case depending on whether effort is complementary with, or substitutable for, the externality. We can then adjust the marginal willingness to pay measure in (5) as follows:

$$\tilde{M}_{r_k r_{k-1}} = M_{r_k r_{k-1}} \left(1+\tilde{B}_{r_k} C_{r_k} \varepsilon_{l(r_k)}^{H,c}\right). \tag{21}$$

Note that equation (21) reflects an interaction effect between the marginal willingness to pay to avoid the externality and the incentive compatibility constraint, since $\tilde{B}_{r_k}$ is directly proportional to the Lagrange multiplier attached to the incentive compatibility constraint for type r_k. The *R*-factors in (18) can then be adjusted correspondingly:

$$\tilde{R}_s^n = \int_0^\infty \dots \int_0^\infty \int_0^\infty \int_0^\infty \tilde{M}_{r_n r_{n-1}} \; f(r_n) \; dr_n \dots \tilde{M}_{r_2 r_1} \; f(r_2) \; dr_2 \tilde{M}_{r_1 s} \; f(r_1) \; dr_1 \tag{22a}$$

so
$$\tilde{R}_s^0 = 1 \; , \tag{22b}$$

$$\tilde{R}_s^1 = \int_0^\infty \tilde{M}_{r_1 s} \; f(r_1) \; dr_1 = E(\tilde{M}_s) \, , \tag{22c}$$

$$\tilde{R}_s^2 = \int_0^\infty \int_0^\infty \tilde{M}_{r_2 r_1} \; f(r_2) \; dr_2 \tilde{M}_{r_1 s} \; f(r_1) \; dr_1 = E\left(E(\tilde{M})\tilde{M}_s\right) \tag{22d}$$

We are now ready to present our main general results:

Proposition 1. *(i) The optimal marginal income tax rate satisfies the following policy rule for any type w supplying labor:*

$$\begin{aligned} \frac{T_y^{(w)}}{1-T_y^{(w)}} &= A_w \tilde{B}_w C_w + \Gamma_w \\ &= A_w \tilde{B}_w C_w + \sum_{i=0}^{\infty} \int_0^\infty \tilde{R}_s^i \tilde{M}_{sw} f(s) \, ds \end{aligned} \; . \tag{23}$$

(ii) If the preferences are weakly labor separable, equation (23) reduces to read

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \sum_{i=0}^{\infty} \int_0^\infty R_s^i M_{sw} f(s) \, ds \, . \tag{24}$$

(iii) If the consumption externality is atmospheric, equation (23) reduces to read

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{E(M)}{1-E(M)}\left(1 - E\left(\tilde{B}_w C_w \varepsilon_l^{H,c} M\right)\right). \tag{25}$$

(iv) If the consumption externality is atmospheric and preferences are weakly labor separable, equation (25) reduces to read

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{E(M)}{1-E(M)} \, . \tag{26}$$

The first line of equation (23) suggests that we can simply add the value of the marginal externality to the *ABC*-term; let be that the *B*-component is different here than in model-economies without externalities. Note also that this “additivity” applies regardless of whether the externality is non-atmospheric. However, the second line implies that the interpretation of this term is far from straightforward in the general case with non-atmospheric externalities.

If the externality is atmospheric, which means that $E(M_w) = E(M)$ for all *w*, we can see from results (iii) and (iv) that more conventional policy rules for marginal taxation under externalities surface, since the corrective component (the final term on the right-hand side) is the same for everybody. In equation (26), which assumes weak labor separability, the value of the marginal externality just reflects the sum of all people's marginal willingness to pay to avoid the externality generated by an individual of type of type *w*.

Equation (25) also shows how the externality interacts with the redistributive tax component in case the preferences are not labor separable. To interpret this interaction effect, suppose that $M > 0$ in which case the interaction effect works to increase (decrease) the marginal income tax if the private marginal willingness to pay to avoid the externality tends to increase (decrease) in the labor supply/effort, such that $\varepsilon_l^{H,c} > 0 (< 0)$ on average. The intuition is, of course, that this adjustment contributes to relax the incentive compatibility constraints by making mimicking less attractive. The adjustment goes in the opposite direction if $M < 0$. Note also that the interaction effect vanishes in the special case of weak labor separability in equation (26), where $MRS_{H,c}$ is independent of effort (such that $\varepsilon_l^{H,c} = 0$). Finally, if the resource allocation is first-best, which coincides with the special case of our model where $\theta_w = \tilde{B}_w = 0$ for all *w*, equation (25) reduces to read $T_y^{(w)} = E(M_w) = E(M)$ for all *w*, which is a conventional Pigouvian tax measuring the sum of all people's marginal willingness to pay to avoid the externality.

Let us now return to the general policy rules in equations (23) and (24), which allow for non-atmospheric externalities, where the value of the marginal externality (the second term on the right-hand side) takes the form of an infinite series. Consider first the somewhat simpler case of weak labor separbility in equation (24), where each addend in this infinite series constitutes a weighted sum of people's marginal willingness to pay to avoid the externality generated by an individual of type *w*. The first addend, with weight factor $R_w^0 = 1$, is given by

$$\int_0^\infty M_{sw} f(s)\, ds = E(M_w),$$

i.e., the unweighted sum of all individual's marginal willingness to pay for a consumption reduction of an individual of type w. The second addend with weight factor R_w^1 becomes

$$\int_0^{\infty}\int_0^{\infty} M_{r_1 s}\ f(r_1)\ dr_1 M_{sw} f(s)\ ds = \int_0^{\infty} E(M_s) M_{sw} f(s)\ ds\ ,$$

where a type *s* individual's marginal willingness to pay to avoid the externality generated by an individual of type *w* is weighted by all people's marginal willingness to pay to avoid the externality generated by a type *s* individual. Similarly, the third addend with weight factor R_w^2 can be written as

$$\int_0^{\infty}\int_0^{\infty}\int_0^{\infty} M_{r_2 r_1}\ f(r_2)\ dr_2 M_{r_1 s}\ f(r_1) dr_1\ M_{sw} f(s)\ ds = \int_0^{\infty}\int_0^{\infty} E(M_{r_1}) M_{r_1 s}\ f(r_1)\ dr_1\ M_{sw} f(s)\ ds$$
$$= \int_0^{\infty} E\left(E(M) M_s\right) M_{sw} f(s)\ ds\ ,$$

which implies that any type r_1-individual's marginal willingness to pay to avoid the externality generated by an individual of type *s* is weighted by other people's marginal willingess to pay to avoid the externality generated by an individual of type r_1. The fourth addend implies a corresponding extension by weighting the integrand of the third addend, and so on. The intuition is that the marginal externalities interact at the social optimum if the externalities are non-atmospheric. More specifically, since the marginal contribution to the externality differs between individuals, and the second-best optimal resource allocation equalizes the social (not the private) marginal utility of consumption among individuals, adjusted for incentive compatibility, it follows that the social marginal benefit of correcting the externality generated by any type *w* depends on the social marginal benefits of correcting the externalties generated by all other individuals. Thus, the corrective tax component implemented for any type *w* may either exceed, or fall short of, the sum of other people's marginal willingness to pay for a type *w* individual to decrease her consumption. This will be described more thoroughly below.

To shed further light on the interpretation of the externality term in equation (24), note that we can rewrite the *R*-factors above using normalized covariances such that, e.g.,

$$\begin{aligned} R_s^2 &= \int_0^{\infty}\int_0^{\infty} M_{r_2 r_1}\ f(r_2)\ dr_2 M_{r_1 s}\ f(r_1)\ dr_1 \\ &= E(M)\ E\left(M_s\right)\left(1+\operatorname{cov}\left(\frac{M}{E(M)}, \frac{M_s}{E\left(M_s\right)}\right)\right), \\ &= E(M) E\left(M_s\right)\left(1+\rho_s\right) \end{aligned} \tag{27}$$

where

$$\rho_s = \text{cov}\left(\frac{M}{E(M)}, \frac{M_s}{E(M_s)}\right)$$

(28)

is the normalized covariance between how much all people are willing to pay for a reduction in consumption of a certain type and how much people of that type are willing to pay for a reduction of the consumption of an individual of type *s*. If individuals are willing to pay more for a reduction in the consumption among the rich (which makes sense), and richer individuals are willing to pay more for a reduction in the consumption of type *s* individuals (which may be the case, and may seem likely due to an income effect), then there is a positive covariance. In general, these covariances will of course vary between types, but in the benchmark case where they do not vary, we are able to present a much simpler version of equation (24).

Corollary 1. *If the preferences are weakly labor separable, and if* $\rho_w = \rho$ *for all w, then*

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{E(M_w)}{1-(1+\rho)E(M)} . \tag{29}$$

To interpret Corollary 1, we start by considering a first-best resource allocation (where $\theta_w = B_w = 0$ for all *w*). By using $\Omega_w = 1/\left(1+E(M_w)-(1+\rho)E(M)\right)$, equation (29) then simplifies to read

$$T_y^{(w)} = \Omega_w E(M_w) . \tag{30}$$

In the special case where $\rho_w = \rho = 0$, we have $T_Y^{(w)} < E(M_w)$ for $E(M_w) > E(M)$ and $T_y^{(w)} > E(M_w)$ for $E(M_w) < E(M)$. As indicated above, the intuition is based on the fact that the externalities are non-atmospheric: in this case, the first-best efficiency condition does not imply that the private marginal utility of consumption (adjusted for social welfare weights) should be the same for each individual, as it would with atmospheric externalities. Instead, the social marginal utility of consumption, i.e., taking the externalities into account, should be the same. This implies, in turn, that the private marginal utility of consumption for an individual who generates negative externalities (typically high-ability individuals) will at optimum be larger than that of an individual who generates positive externalities (or smaller negative externalities).[7] Thus, if $E(M_w) > (<) E(M)$, it follows that $E(M_w)$ overestimates (underestimates) the corrective tax necessary to induce individuals of type w to make the socially desired choice.

Suppose next that $\rho_w = \rho > 0$. This implies that $T_y^{(w)} < E(M_w)$ for $E(M_w) > (1+\rho)E(M)$ and $T_y^{(w)} > E(M_w)$ for $E(M_w) < (1+\rho)E(M)$. While the basic logic is the same as for the case where $\rho = 0$, such that the first-best deviation from a conventional Pigouvian tax is due to that the private marginal utility of consumption (again adjusted for the welfare weights implicit in the function ψ) differs among individuals at the social optimum, the critical levels for when the optimal marginal tax exceeds, or falls short of, a conventional Pigouvian tax have now changed. The intuition is that a positive covariance implies that those who generate large negative externalities and (as explained above) have a high marginal utility of consumption, are also willing to pay more for avoiding the externalities generated by others, and vice versa. Therefore, the modification due to differences in the marginal utility of consumption will be smaller here.

The above reasoning also applies to equation (29), which assumes a second-best optimal resource allocation and labor separable preferences. The only modification here is that the social marginal utility of consumption (which is equalized among individuals at the optimum) must be adjusted to reflect the incentive compatibility constraints. This illustrates the importance of the interaction between the size of the externality caused by a specific individual and the marginal willingness to pay to avoid the externalities generated by other people, and hence the corresponding covariances. The latter insight is, of course, also valid in the case where the covariances are not identical, i.e., underlying equation (24).

Let us now return to the general policy rule given in equation (23). The interpretation of the externality component is similar to that in equation (24), except that externality correction now serves a redistributive purpose as well. Thus, the *R*-factors in equation (24) are replaced by the $\tilde{R}$-factors, defined in equations (22), which imply an interaction effect between the marginal willingness to pay to avoid the externality and the incentive compatibility constraints. As we

[7] Recall that we are in the first best here where distributional concerns are taken care of by individual lump-sum taxes.

explained above, if the marginal willingness to pay to avoid the externality tends to increase (decrease) in work effort, this motivates an upward (downward) adjustment of the externality term compared to the simpler model with labor separability, ceteris paribus, where the marginal willingness to pay to avoid the externality does not depend on the time spent on labor.

4. Optimal Income Taxation under Self-Centered Inequality Aversion

In the previous section, we derived a general policy rule for optimal marginal income taxation when people are inequality averse, or more generally when the utility of each individual depends on the consumption of all individuals. This policy rule was expressed in terms of people's marginal willingness to pay for other people to reduce their consumption. Here we will be able to present much clearer and more easily interpretable results based on the two most famous models of self-centered inequality aversion, suggested by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000).

4.1 Results based on the Fehr and Schmidt model

The Fehr and Schmidt (1999) model is expressed in terms of either two or *n* individuals, but it is straightforward to generalize it to a continuous distribution of individuals, as follows:

$$U_w = u\left(c_w - \beta\int_0^w (c_w - c_s) f(s)ds - \alpha\int_w^\infty (c_s - c_w) f(s)ds, l_w\right) \quad (31)$$

where α reflects disadvantageous and β advantageous inequality aversion, resepctively.[8] It is typically assumed that $\alpha > \beta > 0$, implying that people dislike both advantageous and disadvantageous inequality, but they dislike disdavantageous inequality more.

Note also the close link with the literature on relative consumption. If $\beta = -\alpha$, the utility function changes to read $U_w = u\left(c_w + \alpha(c_w - E(c)), l_w\right)$, i.e., the frequently used difference-comparison formulation (e.g., Aronsson and Johansson-Stenman 2008). In this case, therefore,

[8] For a populations size equal to *N*, this specification implies that an individual's marginal willingness to pay for an individual of a higher type to decrease his/her consumption is equal to α / N, and correspondingly equal to β / N for an individual of a lower type to increase his/her consumption. But since the population is normalized to one, these marginal willingness to pay measures become α and β, respectively.

the Fehr-Schmidt model of inequality aversion reduces to an analytically much simpler form , where the consumption externality is atmospheric .

Despite the fact that the general Fehr-Schmidt specification implies complex non-atmospheric externalities, we are able to present a perhaps not very simple but still interpretable policy rule for marginal income taxation as follows:

Proposition 2. *The optimal marginal income tax rate under Fehr and Schmidt (1999) inequality aversion can be written as*

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{\alpha\left(\exp\left((\alpha+\beta)F(w)\right)-1\right)+\beta\left(\exp\left((\alpha+\beta)F(w)\right)-\exp(\alpha+\beta)\right)}{\alpha+\beta\exp(\alpha+\beta)} . \quad (32)$$

The second term on the right-hand side of equation (32) is the value of the marginal externality and containw two parts, which are proporitional to α and β, respectively. Consider first the bottom of the distribution, where $F(w)=0$. In this case, and if the lowest type supplies labor, the first part of the externality term vanishes, and (32) simplifies to

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w - \frac{\exp(\alpha+\beta)-1}{\alpha+\beta\exp(\alpha+\beta)}\beta < A_w \tilde{B}_w C_w . \quad (33)$$

Individuals at the bottom of the income distribution impose a positive externality on other people (by influencing the advantageous inequality experienced by them), which motivates a corrective subsidy, i.e., the marginal tax implied by (33) falls short of the purely redistributive *ABC* component. Therefore, in the special case where people only dislike disadvantageous inequality, such that $\beta=0$, the whole externality term vanishes and (33) reduces to $T_y^{(w)}/\left(1-T_y^{(w)}\right) = A_w \tilde{B}_w C_w$, since people at the bottom of the income distribution no longer generate externalities. Finally, in the symmetric case where $\beta=\alpha$, (33) implies the following: $T_y^{(w)}/\left(1-T_y^{(w)}\right) = A_w \tilde{B}_w C_w + \left(1-\exp(2\alpha)\right)/\left(1+\exp(2\alpha)\right)$. Thus, the externality term at the bottom decreases monotonically in α (and in the equally large β) starting from zero.

Consider next the top of the income distribution, where $F(w)=1$. This means that the second part of the externality measure vanishes, and equation (32) changes to read

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w\tilde{B}_w C_w + \frac{\exp(\alpha+\beta)-1}{\alpha+\beta\exp(\alpha+\beta)}\alpha > A_w\tilde{B}_w C_w . \tag{34}$$

The disadvantageous inequality that the highest earners impose on other people leads to a negative externality that calls for a corrective tax, i.e., the marginal income tax rate exceeds the purely redistributive *ABC* component. We can also see that the (somewhat unintuitive) special case where people only care about advanteagous inequality, i.e., where $\alpha = 0$, implies that the externality term in (34) vanishes. If the inequality aversion is symmetric with $\beta = \alpha$, (34) reduces to $T_y^{(w)} / \left(1-T_y^{(w)}\right) = A_w\tilde{B}_w C_w + \left(\exp(2\alpha)-1\right)/\left(\exp(2\alpha)+1\right)$. Thus, the externality term at the top increases monotonically in α (and in the equally large β), starting from zero, and approaches $1-\exp(2)/\left(1+\exp(2)\right) \approx 0.79$ when α approaches 1.

Finally, returning to the general equation (32), we can see that the externality term implied by Fehr-Schmidt preferences increases monotonically $F(w)$. This is an intuitive result: the higher an individual's income, the larger will be the number of persons with lower incomes suffering from the negative externality that this individual imposes on them, ceteris paribus. Therefore, if people are inequality averse according to the Fehr-Schmidt model, externality correction will work in the direction of a more progressive marginal income tax schedule.

[Simulation results to be included here]

4.2 Results based on the Bolton and Ockenfels model

In the model suggested by Bolton and Ockenfels (2000), the utility function is given by

$$U_w = u\left(c_w, l_w, \frac{c_w}{E(c)}\right), \tag{35}$$

where $\frac{\partial u}{\partial(c/E(c))} > 0$ for $c < E(c)$, $\frac{\partial u}{\partial(c/E(c))} = 0$ for $c = E(c)$, and $\frac{\partial u}{\partial(c/E(c))} < 0$ for $c > E(c)$.

Given their own consumption and labor supply, an individual prefers the average consumption level to be as close as possible to their own consumption level. The perceived inequality depends on the discrepancy between the individual's own consumption and the average consumption. For Bolton-Ockenfels preferences, therefore, the consumption externality is atmospheric. From the analysis in Section 3, this means that we can write $E(M_w) = E(M)$, and that the policy rule for marginal income taxation is rather straightforward.

Let $F(E(c))$ reflect the ordinal rank of the individual whose consumption is equal to the average consumption, and hence also the share of individuals consuming below the average. Policy rules corresponding to the general utility function (35) as well as for two useful special cases are given in Proposition 3.

Proposition 3. *The policy rule for marginal income taxation under Bolton-Ockenfels (2000) inequality aversion is given by (25), which reduces to (26) if the preferences are labor separable.*

Under the utility specification $U_w = v\left(c_w - \phi\left(\frac{E(c)}{c_w} - 1\right)^2, l_w \right)$, *we obtain*

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{E(c)E\left(1/c^2\right) - E\left(1/c\right)}{1/2 - \phi\left(E(c)E\left(1/c^2\right) - E\left(1/c\right)\right)}\phi, \tag{36a}$$

while utility specification $U_w = v\left(c_w - \phi\left|E(c) - c_w\right|, l_w\right)$ *gives*

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{F(E(c)) - 1/2}{1/2 - \phi\left(F(E(c)) - 1/2\right)}\phi, \tag{36b}$$

where $0 < \phi < 1$.

The first part of the proposition, which is applicable to the general utility function (35), follows because the consumption externality is atmospheric. This fact also implies that the second term on the right-hand side of (25), (26), (36a), and (36b), respectively, is identical for all *w*. As such, this term is interpretable as a standard Pigouvian element of the marginal income tax.

The reason for presenting special case (36a) is that the underlying utility function is discussed explicitly by Bolton and Ockenfels (2000), while the utility function underlying (36b) is interesting in the sense that its linear structure makes it somewhat reminicent of the Fehr-Schmidt specification. For each of these specifications, we can observe that the numerator of the externality term contains two parts with opposite signs. This is because increased consumption for any individual, ceteris paribus, leads to negative externalities on all individuals with consumption levels below the average and positive externalities for all individuals with consumption levels above the average. This suggests that the sum of people's marginal willingness to pay to avoid the externality can be relatively small, since positive and negative terms of this sum tend to cancel out (at least in part). To exemplify, consider equation (36b), where the externality term is positive if, and only if, $F(E(c)) > 1/2$, which is the case if mean income is larger than median income or, in other words, if the second Pearson measure of skewness is positive.

Therefore, an interesting conclusion is that seemingly similar models of inequality aversion have very different policy implications. Whereas the non-atmospheric consumption externality implied by Fehr-Schmidt preferences work in the direction of higher marginal income tax rates for high-income earners and a more progressive marginal tax structure, Bolton-Ockenfels preferences imply an atmospheric externality where the corrective tax element is the same for everybody.

[Simulation results to be included here]

5. Optimal Income Taxation under Non-Self-Centered Inequality Aversion

Although much work on social preferences in behavioral economics has focused on self-centered inequality aversion, one may question this point of departure when studying inequality at the societal level. Instead, individuals may, for a variety of reasons, prefer a more equal consumption distribution to a less equal one regardless of the relationship between their own and other people's consumption. In that case, the inequality aversion is said to be non-self-centered. We will focus on two variants of non-self-centered inequality aversion, where people (i) prefer a more equal to a less equal distribution in terms of the Gini coefficient, the by far most commonly used inequality measure at the social level, or (ii) would like the lowest consumption level in society to be as high as possible.

5.1 Preferences with respect to the Gini coefficient

Let us consider the case where people prefer a low Gini coefficient, *G*, such that $\mathrm{I} = G$ and $U_w = u(c_w, l_w, G)$. Since the Gini coefficient can be written

$$G = \frac{1}{E(c)} \int_0^\infty c_w \left(2F(w) - 1\right) f(w) dw, \tag{37}$$

one can show that the marginal willingness to pay by an individual of type *s* for a decrease in a type *w* individuals consumption is given by

$$M_{sw} = -\frac{\partial G}{\partial c_w} \frac{u_G^{(s)}}{u_c^{(s)}} = \frac{2F(w) - 1 - G}{E(c)} MRS_{G,c}^{(s)}$$

where the first factor reflects how the consumption increase of a type *w* individual affects the Gini coefficient, and the second reflects a type *s* individual's marginal willingness to pay to avoid inequality.

Before we proceed, let us briefly reflect on the well-known fact that consumption changes in the upper end of the distribution tends to have relatively small effects on Gini. Consider, for example, the case where Gini is equal to 0.4, and suppose that the 99.99-percentile consumption level is 1000 times the 99-percentile consumption level. These numbers are roughly consistent with the current disposable income distribution in the US. Let Al's consumption be equal to the 99-percentile consumption level, and let Bob's consumption be 10,000 times larger corresponding to the 99.99 percentile level. Then, naturally, a additional Dollar consumption for Bob will increase the Gini more than an additional Dollar consumption for Al, but only marginally so. Indeed, an additional Dollar to the 10,000 times richer Bob increases Gini by only about 3% more than from an additional Dollar to Al. Clearly, while there are other inequality measures with different properties, we will for space reasons not pursue them here, but this (somewhat unattractive) property of the Gini should be kept in mind.

We are now ready to present both general results with respect to the Gini coefficient and more specific one based on functional form assumptions for the utility function.

Proposition 4. *The policy rule for marginal income taxation under inequality aversion with respect to the Gini coefficient can be written as*

$$\frac{T_y^{(w)}}{1 - T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{2F(w) - 1 - G}{E(c)} \sum_{i=0}^{\infty} \int_0^\infty \tilde{R}_w^i \left(1 + B_s C_s \varepsilon_{l(s)}^{H,c}\right) M_{G,s} f(s)\, ds. \tag{38a}$$

In the case of weak labor separability, equation (38a) simplifies to read

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{2F(w)-1-G}{E(c)} \sum_{i=0}^{\infty} \int_0^{\infty} R_w^i M_{G,s} f(s)\, ds\,. \qquad (38b)$$

Utility specification $U_w = u(c_w - \mu G, l_w)$ *implies*

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{2F(w)-1-G}{E(c)+\mu G}\mu\,. \qquad (38c)$$

$$\begin{aligned} U_w &= u(\ln c_w - \xi G, l_w) \\ &= U'\big(\exp\left(\ln c_w - \xi G\right)\big), l_w) \end{aligned}$$

Utility specification $= U'\big(\exp\left(\ln c_w\right) / \exp\left(\xi G\right)\big), l_w)$ *gives*

$$\begin{aligned} &= U'\big(c_w / \exp\left(\xi G\right)\big), l_w) \\ &= U'\big(c_w \exp\left(-\xi G\right)\big), l_w) \end{aligned}$$

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \left(2F(w)-1-G\right)\xi\,. \qquad (38d)$$

Finally, under utility specification $U_w = u(c_w / G^{\xi}, l_w)$, *we obtain*

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w + \frac{2F(w)-1-G}{G}\xi\,. \qquad (38e)$$

In (38a) and (38b), which are based on the general utility function (37), the externality term is still quite complex, but is in both cases proportional to the factor $\left(2F(w)-1-G\right)/E(c)$, which reflects how increased consumption by a type *w* individual affects the Gini coefficient. Moreover, each policy rule in the proposition implies that the externality term (the second term on the right-hand side) is positive if, and only if, $F(w) > (1+G)/2$. This is because an additional consumption unit for an individual of typ *w* causes a negative externality if, and only if, it leads to an increase in the Gini.

The utility function underlying (38c) is a generalization of the functional form analyzed by Nyborg-Sjøstad and Cowell (2022). This form implies that the wilingness to pay for a reduction in the Gini coefficient is the same for all individuals. However, this property runs counter to empirical evidence, suggesting that the marginal willingness to pay to avoid inequality increases in income (see, e.g., xx). Equation (38d) is even simpler, where the externality term reduces to $\left(1-G\right)\xi$ at the top income (where $F(w)=1$). Finally, (38e) is included since the underlying utility function has been used in experimental research (e.g., Carlsson et al. 2007).

5.2 Rawlsian social preferences

With Rawlsian social preferences utility can be written

$$U_w = u(c_w, l_w, c_{\min}) \tag{39}$$

where $c_{\min}$ constitutes the lowest consumption level in the economy, and $\partial u / \partial c_{\min} > 0$. Clearly, if it is optimal for society that indiduals of the lowest ability type(s) do not work, then $c_{\min}$ constitutes the consumption level among the unemployed, in which case the appearance of social preferences does not lead to any externalities. The intuition is, of course, that unemployed individual cannot influence their consumption space. On the other hand, if it is optimal for society that the type with the lowest consumption level is active in the labor market, then these individuals would impose positive externalities on all other types through their labor supply behavior. To internalize this externality, the government would implement a lower marginal labor income tax at the very bottom of the income distribution than it would otherwise have done, whereas the marginal income taxes implemented for all other individuals (effectively all individuals in the economy except those characterized by the lowest ability) would be governed by a conventional policy rule. Consider Proposition 5.

Proposition 5. *Suppose that individuals with abilities less that or equal to* $\underline{w}$ *share the lowest consumption level. The policy rule for marginal income taxation under Rawlsian social preferences can then be written as*

$$\frac{T_y^{(w)}}{1-T_y^{(w)}} = A_w \tilde{B}_w C_w \quad \textit{for all } w > \underline{w}. \tag{40}$$

Although policy rule (40) takes the same form as in economies without externalities, the levels of marginal taxation, as well as the common for all intercept, will of course depend on the strength of the maximin social preference.

6. Optimal Top Marginal Income Tax Rates

What is the optimal marginal tax treatment of top earners in economies where people have social preferences? We present policy rules for marginal income taxation of the top income for the general model analyzed in Section 3 as well as for the different social preference models examined in Sections 4 and 5. In subsection 6.1, we focus on (a simplified version of) the ability distribution analyzed above where the upper tail is unlimited, as in Diamond (1998), Saez (2001), and several following papers. At the same time, a distribution with an unlimited upper tail is basically equivalent to an infinite population, and in the finite world we happen to live in there must of course be an individual with the highest income, and this income must also be finite. Therefore, in subsection 6.2, we characterize the marginal tax treatment of the individual(s) with the highest ability. The latter corresponds to the case analyzed by Sadka (1976), who derived the famous result that the marginal tax for the highest ability type should be zero. Naturally, due to the appearance of externalities, this result will not apply in our framework, and we show that the optimal marginal tax rate at the very top can be substantial.

6.1 The case when the Upper Tail of the Ability Distribution is Unbounded and Pareto-shaped

We assume that upper tail of the ability distribution is unbounded and given by a Pareto distribution. Following the analyses of top marginal tax rates in Diamond (1998) and Saez (2001), we also simplify by assuming that the underlying utility function is quasi-linear. This implies zero income effects such that the compensated and uncompensated labor supply elasticities coincide, i.e., we may write $\varsigma_w^u = \varsigma_w^c = \varsigma_w$, and also $A_w = (1+\varsigma_w)/\varsigma_w = 1+1/\varsigma_w$. The *B*-factor without income effects takes the following form when ability approaches infinity:

$$B_\infty = \lim_{w\to\infty}\int_w^\infty \left(1-\left(\delta_s - \Gamma_s\right)\right)\frac{f(s)}{1-F(w)}\,ds = \lim_{w\to\infty}\int_w^\infty \left(1+\Gamma_s\right)\frac{f(s)}{1-F(w)}\,ds$$

$$= 1 + \lim_{w\to\infty}\int_w^\infty \Gamma_s \frac{f(s)}{1-F(w)}\,ds = 1+\Gamma_\infty$$

provided that the welfare weight, δ_w, approaches zero. Therefore, we obtain the following simple *ABC*-expression when ability approaches infinity:

$$A_\infty \tilde{B}_\infty C_\infty = \frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty}(1+\Gamma_\infty), \tag{41}$$

where the factor $(1+\Gamma_\infty)$ constitutes the only modification compared to the corresponding *ABC* forumula in Diamond (1998). The interpretation is, of course, that the distortive part of the top marginal income tax rate is modified by the externality: the right-hand side of (41) is scaled down (implying a lower tax distortion) if the extrernality is negative and scaled up (implying a higher tax distortion) if the externality is positive. In Proposition 6, we present results for the general case (albeit based on quasi-linear preferences) as well as for the specific models of inequality aversion examined in Sections 4 and 5.[9]

Proposition 6.

Under social preferences, and if the utility function is quasi-linear, the top marginal income tax rate can be characterized as follows:

$$\frac{T_y^{(\infty)}}{1-T_y^{(\infty)}} = A_\infty \tilde{B}_\infty C_\infty + \Gamma_\infty = \frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty}(1+\Gamma_\infty) + \Gamma_\infty = \frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty} + \left(1+\frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty}\right)\Gamma_\infty . \quad (42)$$

For the specific models of inequality aversion, equation (42) can be written as

Fehr- Schmidt preferences when $U_w = v\left(c_w - \beta\int_0^w (c_w - c_s) f(s)ds - \alpha\int_w^\infty (c_s - c_w) f(s)ds - g(l_w)\right)$*:*

$$\frac{T_\infty^{(w)}}{1-T_\infty^{(w)}} = \frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty} + \left(1+\frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty}\right)\frac{\exp(\alpha+\beta)-1}{\alpha+\beta\exp(\alpha+\beta)}\alpha . \quad (43a)$$

Bolton-Ockenfels preferences when $U_w = u(c_w, z_w, C_w) = v\left(c_w - \phi\left|E(c) - c_w\right| - g(l_w)\right)$*:*

$$\frac{T_\infty^{(w)}}{1-T_\infty^{(w)}} = \frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty} + \left(1+\frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty}\right)\frac{2F(E(c))-1}{1-\phi\left(2F(E(c))-1\right)}\phi . \quad (43b)$$

Inequality aversion with respect to the Gini coefficient when $U_w = u(c_w - \mu G - g(l_w))$ *:*

$$\frac{T_\infty^{(w)}}{1-T_\infty^{(w)}} = \frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty} + \left(1+\frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty}\right)\frac{1-G}{\overline{c}+\mu G}\mu . \quad (43c)$$

Rawlsian inequality aversion when $U_w = u(c_w + h(c_{\min}, l_w))$ *:*

$$\frac{T_\infty^{(w)}}{1-T_\infty^{(w)}} = \frac{1+\varsigma_\infty}{\varsigma_\infty \varpi_\infty} . \quad (43d)$$

Let us next plot these top marginal income tax rates in terms of the strengths of the inequality aversion.

---

[9] It may not be immediately obvious that the Fehr-Schmidt utility specification below is quasi-linear, since private consumption is included also of the integral expression. However, since the utility function can be rewritten as $U_w = V\left(c_w + \frac{\beta}{1+\alpha-\beta}\int_0^w c_s f(s)ds - \frac{\alpha}{1+\alpha-\beta}\int_w^\infty c_s f(s)ds - \frac{g(l_w)}{1+\alpha-\beta}\right) = V\left(c_w + \frac{\beta}{1+\alpha-\beta}\int_0^w c_s f(s)ds - \frac{\alpha}{1+\alpha-\beta}\int_w^\infty c_s f(s)ds - h(l_w)\right)$, it is indeed quasi-linear. Similarly, for the Bolton-Ockenfels specification below we can write $U_w = u(c_w, z_w, C_w | c_w > E(c)) = V\left(c_w + \frac{\phi}{1+\phi}E(c) - h(l_w)\right)$ and $U_w = u(c_w, z_w, C_w | c_w < E(c)) = V\left(c_w - \frac{\phi}{1+\phi}E(c) - h(l_w)\right)$, respectively, implying quasi-linearity.

6.2 The Marginal Tax Rates at the Very Top of the Income Distribution

This case corresponds to the classic case analyzed by Sadka (1976) and Seade (1977), who showed that, in the conventional case without externalities, the optimal marginal income tax rates for the highest ability type equals zero. Note that this case is possible to reconcile with the assumption that the upper tail of the ability distribution is unlimited, and approximately given by a Pareto distribution up to very high ability levels (and corresponding income levels). Yet, a continuous distribution implicitly assumes an infinite population size. In reality, this can of course not be true, implying that there must be a single type with the highest ability in the economy. This is what we hear mean by “the very top” of the ability and income distribution.

Formally, this implies in our case that the *C*-factor in the *ABC*-term of propositions 1-4 approaches zero. There is no point in assuming quasi-linearity here, since the *ABC*-component is always equal to zero (although several specifications below are, nevertheless, quasi-linear). Consider Proposition 7, where the marginal income tax rates at the very top are expressed directly in terms of T_y instead of in terms of $T_y/(1-T_y)$.

Proposition 7. *Under inequality aversion, the marginal income tax rates at the very top can be characterized as follows:*
Fehr-Schmidt preferences:

$$T_{y_{\max}} = \left(1-\exp(-(\alpha+\beta))\right)\frac{\alpha}{\alpha+\beta}. \qquad (44a)$$

Bolton-Ockenfelspreferences when $U_w = u(c_w, z_w, C_w) = v\left(c_w - \phi\left(\frac{E(c)}{c_w}-1\right)^2, l_w\right)$ *:*

$$T_{y_{\max}} = 2\Big(E(c)E\left(1/c^2\right) - E\left(1/c\right)\Big)\phi. \qquad (44b)$$

Bolton-Ockenfelspreferences when $U_w = u(c_w, z_w, C_w) = v\left(c_w - \phi\left|E(c)-c_w\right|, l_w\right)$ *:*

$$T_{y_{\max}} = \left(2F(E(c))-1\right)\phi. \qquad (44c)$$

Inequality aversion with respect to the Gini coefficient when $U_w = u(c_w - \mu G, l_w)$ *:*

$$T_{y_{\max}} = \frac{1-G}{E(c)+\mu}\mu. \qquad (44d)$$

Inequality aversion with respect to the Gini coefficient when $U_w = u(c_w / G^{\beta}, l_w)$ *:*

$$T_{y_{\max}} = \frac{1-G}{G+(1-G)\beta}\beta . \qquad (44e)$$

Inequality aversion with respect to the Gini coefficient when $U_w = u(\ln c_w - \beta G, l_w)$ *:*

$$T_{y_{\max}} = \frac{1-G}{1+\beta(1-G)}\beta . \qquad (44f)$$

Rawlsian inequality aversion:

$$T_{y_{\max}} = 0 . \qquad (44g)$$

Consider two natural special cases of the Fehr-Schmidt model. When $\beta = 0$, such that people are solely motivated by disadvantageous inequity aversion, then clearly $T_{y_{\max}} = 1-\exp(-\alpha)$. Likewise, in the symmetric case, where $\beta = \alpha$, we obtain $T_{y_{\max}} = (1-\exp(-2\alpha))/2$. [to be expanded]

7. Conclusion

As far as we know, this is the first paper to provide a comprehensive chracerization of optimal non-linear income taxation when people have social preferences, where we in particular focus on four different versions of inequality aversion, two self-centered and two non-self-centered.

The take-home message of the paper is twofold: First, empirically and experimentally quantified degrees of inequality aversion have potentially very important implications for optimal income taxation, where simulations reveal substantially more progressive income taxes. The same applies to optimal marginal income tax rates, where we present such tax rates for both the cases where the ability distribution is unbounded, and when there is a finite highest ability level. Second, both the exact nature of the inequality aversion and measures of inequality used matter a great deal for the structure of efficient marginal income taxation.

References

Aaberge, R. (2000). Characterization of Lorenz curves and income distributions. Social Choice and Welfare 17, 639–653.

Abel, A.B., 2005. Optimal Taxation When Consumers Have Endogenous Benchmark Levels of Consumption. Rev. Econ. Stud. 72, 1–19.

Alesina, Alberto and Eliana La Ferrara (2000), 'Participation in heterogeneous communities' Quarterly Journal of Economics, 115 (3), 847−904.

Alesina, Alberto and Eliana La Ferrara (2002), 'Who trusts others?', Journal of Public Economics, 85 (2), 207–234.

Alesina, A. and P. Giuliano (2011). Preferences for redistribution. In Handbook of Social Economics, Volume 1, pp. 93–131. Elsevier.

Almås, I, A. W. Cappelen, B. Tungodden (2020) Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians? Journal of Political Economy, 128(5), 1753-1788.

Alvarez-Cuadrado, F., Long, N.V., 2011. The Relative Income Hypothesis. J. Econ. Dyn. Control. 35, 1489–1501.

Alvarez-Cuadrado, F., Long, N.V., 2012. Envy and Inequality. Scand. J. Econ. 114, 949–973.

Aronsson, T., Johansson-Stenman, O., 2008. When the Joneses' Consumption Hurts: Optimal Public Good Provision and Nonlinear Income Taxation. J. Public Econ. 92, 986–997.

Aronsson, T., Johansson-Stenman, O., 2010. Positional Concerns in an OLG Model: Optimal Labor and Capital Income Taxation. Int. Econ. Rev. 51 1071–1095.

Aronsson, T., Johansson-Stenman, O., 2015. Keeping up with the Joneses, the Smiths and the Tanakas: On International Tax Coordination and Social Comparisons. J. Public Econ. 131, 71–86.

Atkinson, A. B. (1970). On the measurement of inequality. Journal of Economic Theory, 2: 244-263.

Bellemare, Charles, Sabine Kröger, and Arthur Van Soest. 2008. "Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities." Econometrica 76(4):815–839.

Blattman, Christopher, and Edward Miguel. 2010. "Civil War." Journal of Economic Literature 48(1): 3–57.

Bolton, G.E., Ockenfels, A., 2000. ERC: A Theory of Equity, Reciprocity and Competition. Am. Econ. Rev. 90, 166–193.

Boskin, M.J. Sheshinski, E., 1978. Individual Welfare Depends upon Relative Income. Q. J. Econ. 92, 589–601.

Bruhin, A., E. Fehr, and D. Schunk (2019): “The many faces of human sociality: Uncovering the distribution and stability of social preferences,” Journal of the European Economic Association, 17(4), 1025–1069.

Carlsson, F., Daruvala, D. and Johansson-Stenman, O. (2005) Are people inequality averse or just risk averse? Economica, 72, 375–396.

Carlsson, F. and O. Johansson-Stenman 2010. Why Do You Vote and Vote as You Do? Kyklos 63(4):495–516.

Clark, A. E. and D’Ambrosio, C. (2015). Attitudes to Income Inequality: Experimental and Survey Evidence. In Atkinson, A. B. and Bourguignon, F., editors, Handbook of Income Distribution, volume 2 of Handbook of Income Distribution, pages 1147–1208. Elsevier

Collier, Paul, and Anke Hoeffler. 1998. “On the Economic Causes of Civil War.” Oxford Economic Papers. 50: 563-573.

Corneo, G., Jeanne, O., 1997. Conspicuous Consumption, Snobbism and Conformism. J. Public Econ. 66, 55–71.

Dufwenberg, Martin, Paul Heidhues, Georg Kirchsteiger, Frank Riedel & Joel Sobel (2011), “Other-Regarding Preferences in General Equilibrium,” Review of Economic Studies 78, 640-66.

Dupor, B., Liu, W.F., 2003. Jealousy and Overconsumption. Am. Econ. Rev. 93, 423–428.

Eckerstorfer, P., Wendner, R., 2013. Asymmetric and Non-atmospheric Consumption Externalities, and Efficient Consumption Taxation. J. Public Econ. 106, 42–56.

Esteban, Joan-Maria, and Debraj Ray. 1994. “On the Measurement of Polarization.” Econometrica 62(4): 819-852.

Fajnzylber, P., Lederman, D. and Loayza, N. (2002) What Causes Violent Crime? European Economic Review, 46, 1323-1357.

Fehr, D. (2018). Is increasing inequality harmful? Experimental evidence. Games Econ Behav. 107, 123–134.

Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. Q.J. Econ. 114, 817–868.

Fehr, E., Schmidt K. M. (2003). Theories of Fairness and Reciprocity Evidence and Economic Applications. Advances in Economics and Econometrics, Econometric Society Monographs, Eighth World Congress, Vol. 1, pp. 208-257.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits, “Individual Preferences for Giving,” American Economic Review, 2007, 97 (5), 1858–1876

Fleurbaey, M., and F. Maniquet 2018. Optimal Income Taxation Theory and Principles of Fairness. Journal of Economic Literature, 56(3): 1029-79.

Fong, C. 2001. Social Preferences, Self-Interest, and the Demand for Redistribution, J. Public Econ. 82: 225–246.

Frank, R.H., 1985a. The Demand for Unobservable and Other Nonpositional Goods. Am. Econ. Rev. 75, 101–116.

Frank, R.H., 1985b. Choosing the Right Pond: Human Behavior and the Quest for Status. New York, Oxford University Press.

Frank, R.H., 2005. Positional Externalities Cause Large and Preventable Welfare Losses. Am. Econ. Rev. 95, 137–141.

Frank, R.H., 2008. Should Public Policy Respond to Positional Externalities? J. Public Econ. 92, 1777–1786.

Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman. 1996. “Crime and Social Interactions.” Quarterly Journal of Economics 111: 507-548.

Ireland, N.J., 2001. Optimal Income Tax in the Presence of Status Effects. J. Public Econ. 81, 193–212.

Kanbur, R., Tuomala, M., 2014. Relativity, inequality, and optimal nonlinear income taxation. Int. Econ. Rev. 54, 1199–1217.

Kelly, Morgan. 2000. "Inequality and Crime." The Review of Economics and Statistics 82(4): 530-539.

List, J. A. 2011. The Market for Charitable Giving. Journal of Economic Perspectives, 25(2): 157–180.

Ljungqvist, L., Uhlig, H., 2000. Tax Policy and Aggregate Demand Management Under Catching Up with the Joneses. Am. Econ. Rev. 90, 356–366.

Lockwood, B. B., Nathanson, C. G., and Weyl, E. G. (2017). Taxation and the allocation of talent. *Journal of Political Economy*, 125(5):1635–1682

Mueller, D. 2003. Public Choice III. Cambridge: Cambridge University Press.

Oswald, A., 1983. Altruism, Jealousy and the Theory of Optimal Non-Linear Taxation. J. Public Econ. 20, 77–87.

Persson, M., 1995. Why are Taxes so High in Egalitarian Societies? Scand. J. Econ. 97, 569–580.

Piketty, T., E. Saez, and S. Stantcheva (2014) Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy*, 6, 230–71.

Pommerehne, W.W., Weck-Hannemann, H., 1996. Tax rates, tax administration and income tax evasion in Switzerland. Public Choice 88 (1–13), 161–170.

Rothschild, C., and F. Scheuer (2016) Optimal Taxation with Rent-Seeking. *Review of Economic Studies* 83, 1225–62.

Saez, Emmanuel, and Stefanie Stantcheva. 2016. Generalized Social Marginal Welfare Weights for Optimal Tax Theory. American Economic Review, 106 (1): 24-45.

Tuomala, M., 1990. Optimal income tax and redistribution. Clarendon Press, Oxford.

Wendner, R., 2010. Conspicuous Consumption and Generation Replacement in a Model with Perpetual Youth. J. Public Econ. 94, 1093–1107.

Wendner, R., 2014. Ramsey, Pigou, Heterogeneous Agents, and Non-atmospheric Consumption Externalities, J. Public Econ. Theory. 16, 491–521.