Working Paper in Economics No. 785

Inequality Aversion, Externalities, and Pareto-Efficient Income taxation

Thomas Aronsson and Olof Johansson-Stenman

Department of Economics, May 20

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Inequality Aversion, Externalities, and Pareto-Efficient Income Taxation**

Thomas Aronsson and Olof Johansson-Stenman+*

Abstract

This paper analyzes Pareto-efficient marginal income taxation taking into account externalities induced through individual inequality aversion, meaning that people have preferences for equality. In doing so, we distinguish between four different and widely used models of inequality aversion. The results show that empirically and experimentally quantified degrees of inequality aversion have potentially very strong implications for Pareto-efficient marginal income taxation. It also turns out that the type of inequality aversion (self-centered vs. non-self-centered), and the specific measures of inequality used, matter a great deal. For example, based on simulation results mimicking the disposable income distribution in the U.S., the preferences suggested by Fehr and Schmidt (1999) imply monotonically increasing marginal income taxes, with large negative marginal tax rates for low-income individuals and large positive marginal tax rates for high-income ones. In contrast, the in many respects comparable model by Bolton and Ockenfels (2000) implies close to zero marginal income tax rates for all.

Keywords: Pareto-efficient taxation, Inequality aversion, Self-centered inequality aversion, Non-self-centered inequality aversion, Fehr and Schmidt preferences, Bolton and Ockenfels preferences, GINI coefficient, Coefficient of variation.

JEL Classification: D03, D62, H23.

** We are grateful for very useful comments from Martin Dufwenberg, Dirk Engelmann, Björn Gustafsson, Gauthier Lanot, François Salanié, Sven Tengstam, participants at seminars at Toulouse School of Economics, Colorado State University, University of Wyoming, and University of Gothenburg, at a workshop in Delhi University, at the National Tax Association Conference in New Orleans 2018, and at the International Institute of Public Finance conference in Tampere 2018. Research grants from the Bank of Sweden Tercentenary Foundation, the Swedish Council for Working Life and Social Research, and the Swedish Tax Agency (all of them through project number RS10-1319:1), and the Swedish Research Council (ref 2016-02371), are gratefully acknowledged.
* Address: Department of Economics, Umeå School of Business and Economics, Umeå University, SE – 901 87 Umeå, Sweden. E-mail: Thomas.Aronsson@econ.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

There are several reasons for a government to tax its citizens, including redistribution objectives and revenue collection to fund public expenditure. Most models of optimal taxation dealing with income redistribution assume that the government wants to redistribute from the well-off to the not so well-off, since low-income individuals have higher marginal utility of consumption than high-income individuals. We can then say that the government, or the social planner, is inequality averse.

At the same time, *individuals* are not generally assumed to care about inequality *per se* in models dealing with public policy. That is, their utility is typically modelled to depend solely on their own private and public consumption, as well as on their own leisure time, and not on any measure of inequality in society. This is despite the fact that much experimental research suggests that people are inequality averse, in the sense that they prefer a more equitable allocation over an allocation that is in their own narrow material self-interest; see, e.g., Fehr and Schmidt (1999) and Bolton and Ockenfels (2000).[1] In the present paper we will take this experimental evidence seriously and assume that people do not only derive utility from their own consumption and leisure time (as in standard models of optimal taxation) but also prefer a more equal over a less equal distribution of consumption, *ceteris paribus*.

The purpose of the present paper is twofold: First, the paper derives and gives a detailed characterization of the first-best optimal marginal tax policy for different kinds of inequality aversion. In doing so, we distinguish between self-centered inequality aversion (where each individual's aversion to inequality is based on a comparison between their own and other people's consumption) and non-self-centered inequality aversion. We will consider two kinds of self-centered inequality aversion, based on Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), respectively, and two kinds of non-self-centered inequality aversion, where individual utilities depend on the Gini coefficient and the coefficient of variation, respectively.

[1] There is of course also much other empirical evidence for other-regarding behavior, for example with respect to tax compliance (Pommerehne and Weck-Hannemann, 1996, Andreoni et al., 1998;), voting behavior and political preferences (Mueller, 1998; Fong, 2001; Carlsson et al., 2010), and charitable giving (List, 2011; Andreoni and Payne, 2013).

Second, the paper illustrates quantitatively, based on numerical simulations mimicking the disposable income distribution in the U.S. in 2013, how these types of inequality aversion affect the structure of first-best marginal income taxation. In doing so, we start from a realistic distribution of the disposable income and assume that this income distribution is optimal from the perspective of the government. That is, we will assume that the observed income distribution is the result of an optimal tax policy of the government. In turn, the government is assumed to maximize a Paretian (or Bergson-Samuelson) social welfare function where the utility of individuals with different before-tax wage rates are given different weights, which are implicitly defined by the resulting income distribution. By combining the social and private first-order conditions, based on utility functions characterized with different kinds of inequality aversion, we are then finally able to calculate the optimal marginal income tax rates. In general, these rates will vary with the before-tax income levels.

An alternative approach would have been to start with an exogenous ability distribution, together with an ethically motivated social welfare function, and the same utility functions as in the present paper. One could then have derived the socially optimal disposable income distribution, as well as the optimal marginal income tax rates and redistribution policy consistent with this disposable income distribution. We did not pursue this conventional approach for two related reasons. If we were to start with an ethically motivated social welfare function, it would presumably be weakly concave in individual utilities. This, together with concave utility functions in private consumption, implies a very equitable distribution of disposable incomes also without taking equity preferences into account; if anything, low-ability individuals would presumably have higher levels of disposable income compared with high-ability individuals (given complementarity between leisure and consumption). Furthermore, while inequality aversion would affect the optimal allocation in such models, the insights derived from them may nevertheless say little about the policy implications of inequality aversion in economies with the large inequalities we observe in most existing market economies. With the chosen approach, in contrast, we are able to quantitatively analyze a Pareto-efficient income tax structure based on an existing distribution of disposable income or consumption.

There is surprisingly little research on tax policy in economies where people are inequality averse. This stands in sharp contrast to the by now rich literature on various aspects of optimal taxation based on another kind of interdependent utility structure where people instead of caring about inequality have preferences regarding their own relative consumption or relative income. That is, people prefer to have more than others and dislike having less. This literature shows that relative consumption concerns have profound effects on the optimal tax structure by implying much higher marginal labor income and/or commodity tax rates than would be the outcome in standard models (where such concerns are absent), as well as justifies capital income taxation both on efficiency grounds and for redistributive reasons.[2] Although there are important similarities between preferences based on inequality aversion and preferences regarding relative consumption, since people's consumption choices generate externalities in both cases, there are important differences, too. In particular, when people derive utility from their relative consumption, they typically impose negative externalities on one another. When people are inequality averse, on the other hand, the consumption externalities may be either positive or negative, depending on whether an increase in a particular individual's consumption contributes to increase or decrease the inequality that other people care about. As we will see below, the latter also implies that the tax policy implications may differ considerably between different kinds of inequality aversion.

In the present paper, we thus focus on efficiency aspects of inequality aversion, i.e., the tax policy responses that these aspects motivate. This means that we (implicitly) assume that the government can observe individual ability and thus use ability-specific lump-sum taxes for purposes of redistribution. Obviously, we do not propose that governments in reality can implement first-best policies consistent with their social welfare function, or social objectives more generally. Nevertheless, we believe that the approach taken here has important advantages. First, it allows for a detailed characterization of the marginal tax policy incentives caused by inequality aversion *per se* (and the corresponding externalities), since all analyses below presuppose that inequality aversion is the only reason for distorting the labor-leisure choice. Second, since we aim at examining several different measures of inequality, it admits a straightforward comparison of social costs and corrective tax policies between inequality measures. This aim further emphasizes the need for a simple baseline model. Third, it is straightforward to compare our findings with those of many other studies dealing with policy

[2] This literature includes Boskin and Sheshinski (1978), Oswald (1983), Frank (1985a, b, 2005, 2008), Tuomala (1990), Persson (1995), Corneo and Jeanne (1997), Ljungqvist and Uhlig (2000), Ireland (2001), Dupor and Liu (2003), Abel (2005), Aronsson and Johansson-Stenman (2008, 2010, 2015, 2018), Wendner (2010, 2014), Alvarez-Cuadrado and Long (2011, 2012), Eckerstorfer and Wendner (2013), and Kanbur and Tuomala (2013).

responses to externalities, including environmental externalities, which are often analyzed in a first-best setting.

To our knowledge, there are only two other studies dealing with optimal tax policy responses to individual preferences for equality, both of which are recent working papers. Aronsson and Johansson-Stenman (2020) analyze second-best optimal income taxation in an economy where people have social preferences, which include both inequality aversion and poverty aversion (through a preference for the worst-off group in society). Their main contributions are to characterize how the corrective and redistributive roles of taxation interact, and examine how individual preferences for social outcomes impact the redistributive role of the tax system. Nyborg-Sjøstad and Cowell (2020) use a Mirrleesian model of optimal income taxation to examine the implications of a particular form of non-self-centered inequality aversion, where the Gini coefficient represents the measure of inequality that people are concerned with.[3] They find that inequality aversion leads to a more progressive marginal tax structure compared with a conventional economic model without any externalities.

Relative to the other two studies, the contribution of the present paper is twofold. It takes a much broader perspective of inequality aversion by analyzing several kinds of self-centered and non-self-centered inequality aversion (along with associated externalities) in a unified framework. This includes a detailed theoretical characterization of marginal tax policy responses along the whole (continuous) ability distribution, which is important because several of the inequality externalities we examine are non-atmospheric, as well as extensive numerical simulations. In addition, we combine the actual disposable income distribution in a specific country, in our case the U.S., with experimental evidence on individual preferences for equality in order to examine the implications for externality-correcting marginal taxation. As we indicated above, this approach is distinctly different from conventional approaches to optimal taxation and allows us to examine inequality aversion and the policy implications thereof in a framework with substantial inequality, which is arguably realistic.

Section 2 presents a simple model with a continuous ability distribution and derives the choice rule for Pareto-efficient marginal income taxation for a very general measure of consumption inequality. Based on the results in Section 2, we derive efficient marginal tax rates for two

[3] This paper is based on and further extends a very impressive master's thesis by Nyborg-Sjøstad (2019).

different versions of *self-centered inequality aversion* in Section 3, namely the ones proposed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), respectively. As explained above, by "self-centered" we mean measures of inequality that are defined as relations between the individual's own consumption and others' consumption. Despite that the two models are quite similar, their policy implications are surprisingly different, which is particularly clear from the numerical simulations.

Section 4 similarly analyzes efficient taxation in economies with *non-self-centered* inequality aversion, where individuals are inequality averse based on the Gini coefficient and the coefficient of variation, respectively. One may interpret such inequality aversion broadly to also include potential instrumental reasons, including less criminality and better possibilities for policy makers to control the spread of infectious diseases.

Section 5 concludes that experimentally estimated parameters of inequality aversion, if generalized to the overall economy, may indeed motivate substantial marginal income taxes. Yet, it is also demonstrated that the exact nature of the inequality aversion measure has profound implications for the efficient marginal income tax structure. Proofs are presented in the Appendix.

2. Pareto Efficiency and Inequality Aversion

Consider an exogenous and continuous ability distribution $f(w)$, where $f(w) > 0$ for all $w_{\min} < w < w_{\max}$ and ability is measured by the before-tax wage rate, *w*. Let the government maximize a social welfare function

$$W = \int_{w_{\min}}^{w_{\max}} \psi\left(u\right) f(w) dw, \tag{1}$$

where $\psi\left(\cdot\right)$ constitutes an increasing function of individual utility. We do not assume that this function is necessarily concave. In fact, we will subsequently rather assume that it often gives a higher weight to the utility of high-ability individuals. Thus, a natural interpretation is that it reflects the outcome of a political process where different individuals or groups have different bargaining power. Therefore, even though low-ability individuals may dislike the governmental objective function, all individuals will, conditional of this objective function, agree that there are good reasons to obtain a Pareto-efficient allocation. Thus, for any distribution of negotiating power in the economy, all individuals agree that Pareto improvements should be made, and hence that the allocation should be Pareto efficient.

Let us assume that higher ability is always associated with higher private consumption in equilibrium.[4] We also assume that the population size is constant and equal to n, such that

$$\int_{w_{\min}}^{w_{\max}} f(w)dw = n\,.$$

Individual utility depends on own consumption, c, own leisure, z, and a (possibly type-specific) measure of the overall consumption distribution (which we will specify further subsequently), C. We thus assume that each individual cares about the distribution of consumption, but not about the distribution of utility or leisure. We can then write the social objective function as

$$W = \int_{w_{\min}}^{w_{\max}} \psi\left(u(c(w), z(w), C)\right) f(w)dw\,. \tag{2}$$

Individual Behavior

Each individual's utility depends on own private consumption, c, work hours, $l = 1 - z$, and the measure of the overall consumption distribution, C. Individuals treat other people's consumption as exogenous, but may treat C as (partly) endogenous, since C depends also on the individual's own consumption. More specifically, if the inequality aversion is self-centered, such that the individual explicitly compares their own consumption with that of referent others, the individual is assumed to recognize that their own consumption choice influences the perceived inequality. Under non-self-centered inequality aversion, on the other hand, the individual treats C as exogenous. The individual chooses how much to work, and hence consume, in order to maximize utility, $u(c, z, C)$, subject to the budget constraint. For an individual of (exogenous) ability w_k, the budget constraint is given by

$$c_k = w_k l_k - T(w_k l_k) + \tau_k = y_k - T(y_k) + \tau_k\,, \tag{3}$$

[4] Had we instead restricted $\psi(\cdot)$ to be a concave function, reflecting a prioritarian social welfare function, a first-best allocation would of course not generally imply that higher-ability individuals have higher consumption in equilibrium.

where $T(w_k l_k) = T(y_k)$ denotes the individual's income tax payment (positive or negative), $y = wl$ income, and τ_k a type-specific lump-sum transfer. The individual first-order conditions for consumption and number of work hours, and hence leisure, can then be combined as follows:

$$MRS_{cz}(w_k) + \frac{\partial u(c_k, z_k, C_k)}{\partial C_k} \frac{\partial C_k}{\partial c(w_k)} \Big/ \frac{\partial u(c_k, z_k, C_k)}{\partial z_k} = \frac{1}{w_k \left(1 - T'(y_k)\right)}, \tag{4}$$

where $T'(y_k)$ denotes the marginal income tax rate. The first term on the left-hand side is the marginal rate of substitution between the individual's own private consumption and leisure for a given consumption distribution, i.e.,

$$MRS_{cz}(w_k) = \frac{\partial u(c_k, z_k, C_k)}{\partial c_k} \Big/ \frac{\partial u(c_k, z_k, C_k)}{\partial z_k}.$$

The Social Decision Problem

The social optimization problem means choosing private consumption and leisure time (or work hours) for each individual to maximize the social welfare function given in equation (2) subject to a resource constraint for the economy as a whole. In doing so, the social planner also recognizes the relationship between each individual's consumption, *c*, and the measure of the distribution of consumption in the economy as a whole, *C*, which in turn thus in general depends on the consumption of all individuals.

Although we are working with a continuous distribution, we will for clarity formulate this as a Lagrangean optimization problem, rather than applying the maximum principle. The Lagrangean is then formulated in the same way as for a discrete optimization problem and written as follows:

$$L = \int_{w_{\min}}^{w_{\max}} \psi\left(u(c(w), z(w), C)\right) f(w)dw + \lambda \left(\int_{w_{\min}}^{w_{\max}} w(1 - z(w)) f(w)dw - \int_{w_{\min}}^{w_{\max}} c(w) f(w)dw \right). \tag{5}$$

The expression in large parentheses in the second part of (5) thus constitutes the resource constraint, implying that output (or before-tax income) equals private consumption at the aggregate level. Consider again an individual with ability w_k, and corresponding consumption $c_k = c(w_k)$, leisure $z_k = z(w_k)$, and (possibly type-specific) measure of the overall consumption distribution C_k.

The social first-order condition for consumption and leisure, respectively, for an individual of type k can then be written as[5]

$$\begin{aligned}&\psi_k{}'\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial c(w_k)}+\psi_k{}'\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial C_k}\frac{\partial C_k}{\partial c(w_k)}\\&+\int_{w_{\min}}^{w_{\max}}\psi'\frac{\partial u(c(w),z(w),C)}{\partial C}\frac{\partial C}{\partial c(w_k)}f(w)\,dw-\lambda=0\end{aligned}, \tag{6}$$

$$\psi_k{}'\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial z(w_k)}-\lambda w_k=0, \tag{7}$$

where $\psi_k{}'=\dfrac{d\psi}{du(c(w_k),z(w_k),C_k)}$.

We are now ready to characterize the optimal marginal tax policy for the model set out above, in which we have made no assumption about the preferences with respect to inequality aversion (other than that C might be type specific). This general characterization will be useful in later parts of the paper, where the tax policy implications of more specific forms of inequality aversion are addressed.

Let

$$MWTP(c(w),c(w_k))=-\frac{\dfrac{\partial u(c(w),c(w),C)}{\partial C}\dfrac{\partial C}{\partial c(w_k)}}{\dfrac{\partial u(c(w),c(w),C)}{\partial c(w)}} \tag{8}$$

denote the willingness to pay by an individual with consumption level $c(w)$ for an individual with consumption level $c(w_k)$ to decrease their consumption marginally.

We can then derive the following result by combining the private and social first-order conditions in equations (4), (6), and (7):

[5] The social first order conditions should thus be interpreted as marginal effects due to a consumption and leisure increase *for an individual of type k*, and hence not due to a consumption and leisure change *for all individuals of type k* in the continuous distribution (which are zero).

Lemma 1. *Consider a general model of inequality aversion. The Pareto-efficient marginal income tax rate implemented for individuals with gross income* $y(w_k)$ *and consumption* $c(w_k)$ *is then given by*

$$T'(y(w_k)) = \int_{w_{\min}}^{w_{\max}} \frac{1-T'(y(w_k))}{1-T'(y(w))} MWTP(c(w), c(w_k)) f(w) dw, \qquad (9)$$

where $T'(y(w_k)) < 1$.

This tax formula looks almost like a conventional Pigouvian tax, i.e., the sum of all other people's marginal willingness to pay for keeping an individual with gross income $y(w_k)$ and consumption $c(w_k)$ from consuming one additional unit. The only difference is the weight factor $\left(1-T'(y(w_k))\right)/\left(1-T'(y(w))\right)$ attached to the measure of marginal willingness to pay on the right-hand side.

To see the rationale behind this weight factor, consider first the logic behind a conventional Pigouvian tax for an externality-generating good. In that case, the discrepancy between the social and private marginal value, as reflected by the externality-correcting tax, would simply consist of the sum of other people's marginal willingness to pay for the individual not to consume one additional unit of the good. This would have been the case here as well had the first term on the left-hand side of equation (6) been the same for everybody, i.e., if the externality were atmospheric.[6] In general, however, the externality examined here is *non-atmospheric*, meaning that the externality generated by consuming one additional unit will typically differ depending on who consumes it. Therefore, in the present case, the social first-order condition does not imply equalization of the social marginal utility of private consumption among consumers. Thus, $\psi' \partial u(c,z,C)/\partial c + \psi' \partial u(c,z,C)/\partial C\ (\partial C/\partial c)$ is in general not the same for all consumption (and hence ability) levels in optimum. Instead, as revealed from (6), what should be equalized is $\psi' \partial u(c,z,C)/\partial c + \psi' \partial u(c,z,C)/\partial C\ (\partial C/\partial c)$ plus a term that reflects the value of the marginal externality that the individual's consumption imposes on other people. This, in turn, means that the social marginal utility of consumption is larger at the

[6] A similar result would follow if we were to introduce a numeraire good that does not generate externalities. The reason is that a government that maximizes a social welfare function and is able to redistribute without any social cost will equalize the social marginal utility of consumption of the numeraire good among individuals.

optimum for individuals whose consumption generates large negative externalities and vice versa, which explains the weight factor.

Note also that equation (9) can alternatively be written as

$$\frac{T'(y(w_k))}{1-T'(y(w_k))} = \int_{w_{\min}}^{w_{\max}} \frac{MWTP(c(w),c(w_k))}{1-T'(y(w))} f(w)dw .$$

Hence, the ratio of the marginal tax rate to one minus the marginal tax rate (i.e., the part of the additional income that is not taxed away) for an individual of ability w^k equals the sum (measured over all individuals) of the ratio between the marginal willingness to pay and the fraction of the marginal income that is not taxed away. The marginal income tax rate faced by individuals with before-tax wage rate w_k and associated consumption c_k is thus interpretable to depend on other people's marginal willingness to pay measured in terms of their gross income.

An analytically useful special case of equation (9) arises when all marginal income tax rates are low enough, yet not necessarily similar, such that it always holds that

$$\frac{1-T'(y(w_k))}{1-T'(y(w))} \approx 1 ,$$

in which it is possible to obtain an algebraic closed-form solution.[7] In this case, the marginal tax rate faced by an individual of ability w_k and consumption c_k can clearly be approximated as

$$T'(y(w_k)) \approx \int_{w_{\min}}^{w_{\max}} MWTP(c(w),c(w_k)) f(w)dw . \tag{10}$$

The Pareto-efficient marginal income tax rate implemented for any individual would in this case simply equal the sum of all people's marginal willingness to pay for this particular individual to reduce their consumption.

---

[7] Note that equation (10) is not a reduced form, since c_k depends on $T'(w_k l_k)$. Note also that the assumption that all marginal tax rates are low does not mean that their relative size is similar. Instead, since the externalities are generally non-atmospheric, their relative size may vary greatly and some optimal marginal tax rates may be negative while others are positive.

3. Marginal Income Taxation under Self-Centered Inequality Aversion

In the previous section, we derived a general expression for Pareto-efficient marginal taxation when people are inequality averse, or more generally when the utility of each individual depends on the consumption of all individuals. Yet, we have not further explored the determination of the marginal willingness to pay measures *per se*. This is the task of the present section, where we will explore the marginal willingness to pays on the right-hand side of equations (9) and (10) based on the two most famous models of self-centered inequality aversion, namely those suggested by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), and then illustrate how the Pareto-efficient marginal income taxes will vary with the gross income based on a realistic distribution of consumption. We assume that this distribution is the result of a Pareto-efficient income tax policy, including an efficient set of type-specific lump-sum taxes. We can then calculate what the marginal income tax rates must be for a continuum of consumption levels under different assumptions about the structure and magnitude of the inequality aversion.

3.1 The Fehr-Schmidt Model

The model suggested by Fehr and Schmidt (1999) has become something of an industry standard in the context of self-centered inequality aversion. This is presumably due to a combination of a high degree of parsimony, since the model is based on only two parameters, and the model's ability to rather well explain the outcomes of many experimental games.

While the Fehr and Schmidt (1999) model is often used in settings with either two or few individuals, it is straightforward to generalize it to a continuous distribution of individuals. The utility of an individual with consumption $c(w)$ can then be written as

$$\begin{aligned} u(c(w), z(w), C(w)) &= v\big(c(w) - C(w), z(w)\big) \\ &= v\left(c(w) - \frac{\beta}{n} \int_{w_{\min}}^{w} (c(w) - c(\widehat{w})) f(\widehat{w}) d\widehat{w} - \frac{\alpha}{n} \int_{w}^{w_{\max}} (c(\widehat{w}) - c(w)) f(\widehat{w}) d\widehat{w}, z(w) \right). \end{aligned} \tag{11}$$

The parameters $\beta > 0$ and $\alpha > 0$ are interpretable to reflect the strengths of the aversion to inequality that is to the individual's material advantage and disadvantage, respectively. Based on this type of inequality aversion, we can evaluate the marginal willingness to pay measures in the general policy rule for Pareto-efficient marginal income taxation presented in Lemma 1 and immediately obtain the following result:

Proposition 1. *Suppose that the inequality aversion is of the Fehr and Schmidt type. The marginal tax policy can then be characterized as follows:*

i) $$T'(y(w_k)) = \frac{\alpha}{n}\int_{w_{\min}}^{w_k} \frac{1-T'(y(w_k))}{1-T'(y(w))} f(w)dw - \frac{\beta}{n}\int_{w_k}^{w_{\max}} \frac{1-T'(y(w_k))}{1-T'(y(w))} f(w)dw < 1 \quad \forall k \quad (12)$$

ii) $T''(y(w_k)) > 0$ *for* $y_{\min} < y(w_k) < y_{\max}$

iii) $T'(y(w_{\min})) < 0 < T'(y(w_{\max})) < 1$ for $\alpha, \beta > 0$

iv) *If* $f(c(w_{\min})) = f(c(w_{\max})) = 0$, *then* $T''(y(w_{\min})) = T''(y(w_{\max}) = 0$

v) $$\frac{\dfrac{T'(y(w_{\max}))}{1-T'(y(w_{\max}))}}{\dfrac{T'(y(w_{\min}))}{1-T'(y(w_{\min}))}} = -\frac{\alpha}{\beta}$$

vi) *If* $\alpha = \beta > 0$ *then* $T'(y(w_{\max})) < -T'(y(w_{\min}))$

vii) *The inflection point implied by* $T'''(y(w_k)) = 0$ *is obtained to the left of the mode of the ability distribution*

viii) *If* $f(c(w_{\min})) = f(c(w_{\max})) = 0 = f'(c(w_{\min})) = f'(c(w_{\max}))$ *then* $T'''(y(w_{\min})) = T'''(y(w_{\max})) = 0$

ix) $$\lim_{\beta \to 0} T'(y(w_{\min})) = \lim_{\alpha \to 0} T'(y(w_{\max})) = 0$$

Equation (12) is clearly an implicit formulation since the Pareto-efficient marginal income tax implemented for gross income $y(w_k)$ is expressed in terms of the Pareto-efficient marginal income taxes for all consumption levels. Consequently, it is not straightforward to interpret this policy rule in itself. Yet, together with the other properties, a clearer picture can be provided.

(*ii*) and (*iii*) say that the marginal income tax is monotonically increasing in income, where the lowest tax level is negative and the highest positive, while (*iv*) says that the slope of the marginal income tax function approaches zero at both ends, if the ability density function approaches zero at both ends.

(v) presents a relationship between the absolute sizes of the lowest and highest marginal tax rates and the preferences for equality, showing that higher values of β relative to α work to increase the (absolute size of the) negative marginal tax rate for the lowest consumption individuals relative to the highest ones, and vice versa. Yet, the denominators, one minus the marginal tax rate, work to increase the size of the negative marginal tax rate for the lowest consumption individuals relative to the highest ones. Consequently, (vi) implies that $\alpha = \beta$ is a sufficient (but not necessary) condition for the marginal subsidy to the lowest income earners to exceed the marginal tax paid by the highest income earners.

(vii) and ($viii$) relate to the curvature of the marginal income tax curve. More specifically, ($viii$) provides conditions for the curvature to be zero at the end points, whereas (vii) states that the point where the marginal tax curve starts to decrease, i.e., switches from being concave to convex, is to the left of the mode (the highest value) of the ability density function.

(ix) finally states that the lowest and highest marginal tax rates approach zero when β and α approach zero, respectively. This follows intuition since the corresponding externality would then cease to exist.

We will now shed more light on the optimal marginal tax rule in two ways. First we will present the results of the special case given in equation (10), where all marginal tax rates are small enough to imply that the weight factor $\left(1-T'(y(w_k))\right)/\left(1-T'(y(w))\right)$, resulting from non-atmospheric externalities, is close to one. We will then present simulation results based on the general case. Following, e.g., Frank (1985), let $R(c(w))$ be a measure between 0 and 1 reflecting the fraction of the population with lower consumption than $c(w)$. As such, it is a measure of the ordinal rank where 0 reflects the lowest and 1 the highest possible ordinal rank; by assumption, this ordinal rank is then the same for wage (or ability) levels and income levels. We can then obtain a much simpler expression for the optimal marginal tax rule as follows:

Corollary 1. *Suppose that the inequality aversion is of the Fehr and Schmidt type and that all marginal income tax rates are small. The marginal income tax rate implemented for gross income* $y(w_k)$ *can then be written as*

$$T'(y(w_k)) = -\beta + (\alpha + \beta)R(c(w_k)) . \tag{13}$$

Equation (13) implies that the Pareto-efficient marginal income tax rate increases in the consumption rank, and that this relationship is affine. The Pareto-efficient marginal tax for the lowest consumption level (where $R(c(w_k)) = 0$) is given by $T'(y(w_{\min})) = -\beta$, whereas the Pareto-efficient marginal tax for the highest consumption level (where $R(c(w_k)) = 1$) is given by $T'(y(w_{\max})) = \alpha$. Thus, the Pareto-efficient marginal tax rate increases monotonically from $-\beta$ for the individual with the lowest consumption to α for the individual with the highest consumption. This implies that the marginal tax rate is negative when $R(c(w_k) < \beta / (\alpha + \beta)$, zero when $R(c(w_k) = \beta / (\alpha + \beta)$, and positive when $R(c(w_k) > \beta / (\alpha + \beta)$.

It is also straightforward to show the marginal tax rate for the median consumption level as follows: $T'(y(w_{\text{median}})) = (\alpha - \beta) / 2$, implying that the marginal tax level at the median consumption level is strictly positive given that individuals perceive disadvantageous inequality to be worse than advantageous inequality such that $\alpha > \beta > 0$. Note also that while the efficient marginal tax rate increases linearly in the consumption *rank*, it typically increases nonlinearly with the consumption *level*, where the specific pattern depends on the resulting consumption distribution in the population.

To illustrate how the Pareto-efficient marginal tax rates vary with consumption in the general case where the marginal tax rates are not necessarily low, we will make use of numerical simulations, for which we have to make some further assumptions. In particular, the results will depend on the resulting consumption distribution. Let us take the disposable income distribution in the U.S. as a point of departure, where, according to the Luxemburg Income Study, the mean disposable income per (equivalence-scale adjusted) capita was 44,071 USD in 2016 (the latest year available) and the corresponding Gini coefficient was 0.381. For convenience, we will here approximate the actual distribution with a log-normal one, such that mean disposable income and the Gini coefficient equal the above values.[8] Moreover, we will assume that the consumption distribution equals the disposable income distribution. Although the results naturally depend on these distributional assumptions, most qualitative insights remain the same for other realistic distributions. We will use the same distributional assumption throughout this paper, i.e., also for other measures of inequality aversion.

[8] We also have data on the 10th percentile, the median, and the 90th percentile. Our lognormal approximation is reasonably good (for our purposes) also for these values.

We must also make parametric assumptions within the Fehr and Schmidt model of inequality aversion. In accordance with Fehr and Schmidt (1999, p. 844), who based their own judgment on ample experimental evidence, we first assume that $\alpha = 0.85$ and $\beta = 0.315$. These parameter values clearly imply substantial marginal tax rates, suggesting that we cannot rely on equation (13) as a good approximation of the Pareto-efficient marginal tax policy. Indeed, whereas the distribution based on the simplified equation (13) implies a marginal tax range from -0.315 to 0.85, the efficient marginal tax distribution according to equation (12) ranges from approximately -0.6 to about 0.8. Naturally, the case with 25% of the Fehr and Schmidt parameters (i.e., $\alpha = 0.85 \cdot 0.25 = 0.2125$ and $\beta = 0.315 \cdot 0.25 = 0.07875$) implies that the affine function associated with low marginal tax rates provides a somewhat better approximation. The simulation results are presented in Figure 1.

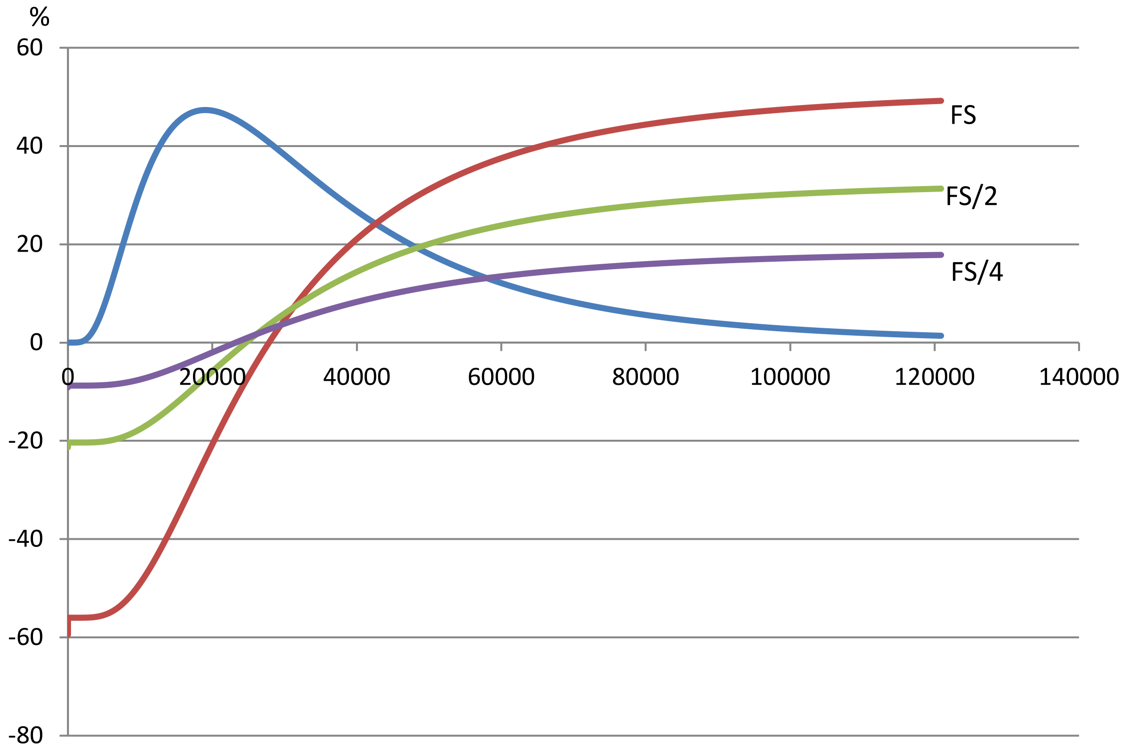

Figure 1. Pareto-efficient marginal income tax rates as a function of disposable income, for a log-normal approximation of the U.S. disposable income distribution in equilibrium, based on the Fehr and Schmidt model of inequity aversion. FS: $\alpha = 0.85$, $\beta = 0.315$; FS/2: $\alpha = 0.425$, $\beta = 0.1575$; FS/4: $\alpha = 0.2125$, $\beta = 0.07875$.

Overall, the Pareto-efficient marginal income tax rates are substantial (recall that we assume that the disposable income distribution is the outcome of an optimal government tax policy, i.e., that the government maximizes eq. [2]). Naturally, the patterns are consistent with the properties stated in Proposition 1: The marginal tax rate increases monotonically, consistent with Proposition 1. The increase is small initially, since there are few individuals with consumption close to zero. The marginal taxes then increase sharply up to a certain consumption (and thus net income) level, before increasing more slowly. Low levels of income should thus be subsidized whereas high levels should be taxed at the margin, in response to inequality aversion. Note also that this qualitative pattern remains the same even if we assume half or a quarter of the values of α and β suggested by Fehr and Schmidt (1999), as can be seen in the figure.

Figure 2 shows the corresponding graphs for the symmetric case where we also make more conservative assumptions: in what we refer to as FS-low, we assume that the number for the β parameter is valid for both α and β. Similarly, in FS-low/2 and FS-low/4, we simply divide this parameter value by 2 and 4, respectively. Obviously the marginal tax rates then become smaller in absolute value. Yet, the qualitative patterns remain similar, although the ratio between the maximum negative marginal tax rate and the maximum positive marginal tax rate becomes larger here, consistent with (v) in Proposition 1.

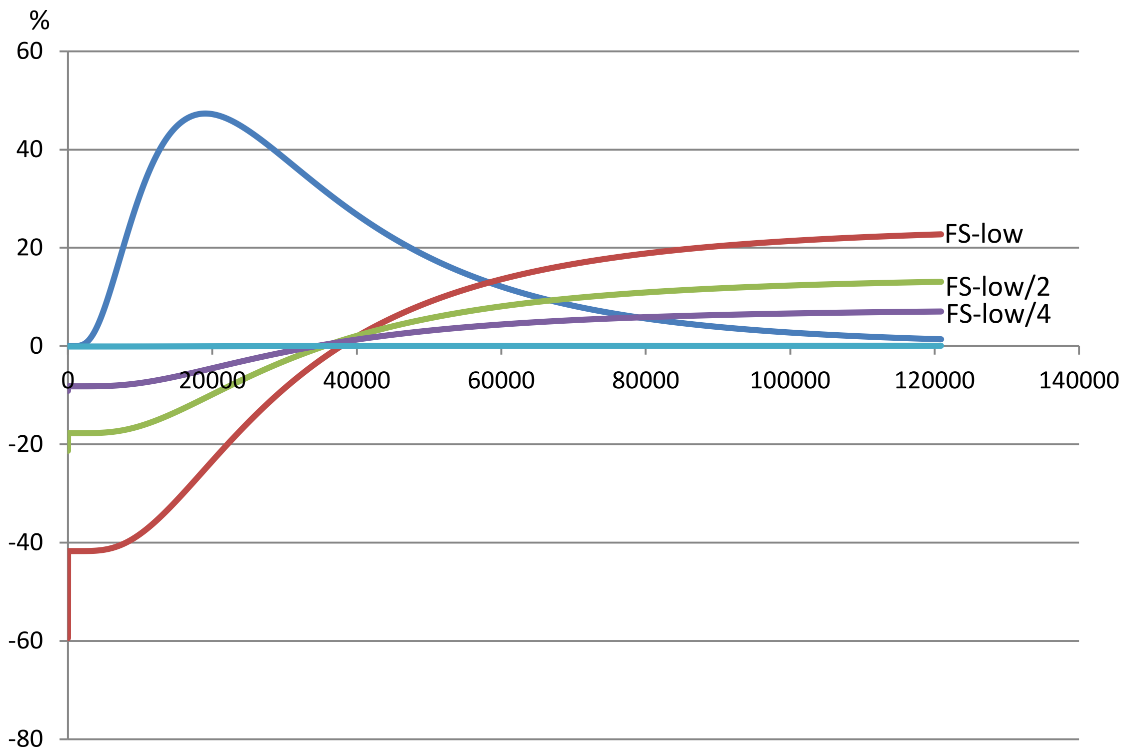

Figure 2. Pareto-efficient marginal income tax rates as a function of disposable income, for a log-normal approximation of the U.S. disposable income distribution in equilibrium, based on a symmetric version of the Fehr and Schmidt model of inequity aversion.

3.2 The Bolton-Ockenfels Model

Bolton and Ockenfels (2000) constitutes the second most often referred to model of self-centered inequality aversion. While also this model is typically used in settings with either two or few individuals (as the Fehr and Schmidt model), the utility function can, of course, be written in the same way in a continuous-type framework. By using $\overline{c}$ to denote the average consumption, the utility function is given as

$$u(c,z,C) = u\left(c, z, \frac{c}{\overline{c}}\right), \tag{14}$$

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

Thus, an individual prefers that the average consumption level is as close as possible to their own consumption level, *ceteris paribus*. Based on equation (14), we observe that

$$\frac{\partial C}{\partial c(w_k)} = -\frac{c(w)}{\overline{c}^2}\frac{\partial \overline{c}}{\partial c(w_k)} = -\frac{c(w)}{\overline{c}^2} f(w_k).$$

Therefore, we can immediately derive the following measure of marginal willingness to pay by using equation (8):

$$MWTP(c(w), c(w_k)) = \frac{c(w)}{\overline{c}^2} \frac{\dfrac{\partial u(c(w), z(w), C)}{\partial C}}{\dfrac{\partial u(c(w), z(w), C)}{\partial c(w)}} = MWTP(c(w)), \tag{15}$$

which is clearly independent of $c(w_k)$. The marginal willingness to pay measure in equation (15) reflects how much an individual with consumption level $c(w)$ is willing to pay for a decrease in *any* individual's consumption. In other words, while an individual's marginal willingness to pay is positive if the average income is higher than the individual's own income, and vice versa, it is independent of which individual the potential consumption change refers to. Consequently, the consumption externality that inequality aversion gives rise to is atmospheric in this case, since each individual only cares about the average consumption in the economy as a whole, in addition to their own consumption and leisure. We can then derive a closed-form solution to the Pareto-efficient tax problem also in the general case, when the marginal tax rates are not low. Lemma 1 and equation (15) imply the following result:

Proposition 2. *Suppose that the inequality aversion is of the Bolton and Ockenfels type. The marginal income tax rate implemented for any individual then takes the following form:*

$$T'(y(w_k)) = \int_{w_{\min}}^{w_{\max}} MWTP(c(w))\, f(w)dw\,, \qquad (16)$$

which is independent of the individual's own gross income.

Equation (16) thus implies that the Pareto-efficient marginal income tax rate is the same for all, irrespective of consumption level. The intuition is as follows: Each individual derives disutility if their consumption deviates from the average consumption in the economy as a whole, *ceteris paribus*. This means that an individual with a consumption level below the mean will prefer that others reduce their consumption. Yet, this individual is indifferent regarding exactly *who* reduces their consumption. Hence, the individual's marginal willingness to pay is the same for a reduction by the rich as for an equally large reduction by the poor. Similarly, an individual above the mean would prefer that others increase their consumption, and they would be willing to pay the same amount to a rich and a poor individual for a given consumption increase. The resulting Pareto-efficient marginal tax rate will then reflect the net effect of such positive and negative marginal willingness to pays.

This can be made more clearly by introducing the short notations

$$\begin{aligned} MWTP^{Below}(c(w)) &\equiv MWTP(c(w))\big| c(w) < \overline{c} \\ &= \frac{c(w)}{n\overline{c}^2} \frac{\partial u(c(w), z(w), C)}{\partial C} \Big/ \frac{\partial u(c(w), z(w), C)}{\partial c(w)}, \\ MWTP^{Above}(c(w)) &\equiv -MWTP(c(w))\big| c(w) > \overline{c} \\ &= -\frac{c(w)}{n\overline{c}^2} \frac{\partial u(c(w), z(w), C)}{\partial C} \Big/ \frac{\partial u(c(w), z(w), C)}{\partial c(w)}. \end{aligned}$$

$MWTP^{Below}(c(w))$ thus reflects the marginal willingness to pay of an individual with consumption $c(w)$ below the mean for a consumption reduction of any individual. Similarly, $MWTP^{Above}(c(w))$ reflects the marginal willingness to pay of an individual with consumption $c(w)$ above the mean for a consumption increase of any individual. Substituting these expressions into equation (16) and letting $w_{\overline{c}}$ reflect the wage level at which the individual consumption equals mean consumption in the economy give

$$T'(y(w_k)) = \int_{w_{\min}}^{w_{\overline{c}}} MWTP^{Below}(c(w))\, f(w)dw - \int_{w_{\overline{c}}}^{w_{\max}} MWTP^{Above}(c(w))\, f(w)dw\,. \qquad (16b)$$

In order to shed more light on the order of magnitude of the Pareto-efficient marginal tax rate, let us consider a more specific and often used quadratic formulation (see, e.g., equation [2] in Bolton and Ockenfels) as follows:

$$u(c,z,C) = v\left(c - \phi\left(\frac{1}{C} - 1\right)^2, z\right) = v\left(c - \phi\left(\frac{\overline{c}}{c} - 1\right)^2, z\right), \tag{17}$$

which clearly reaches its maximum for $\overline{c} = c$. Using this utility function in equation (16) and (16b), one can show that

$$\begin{aligned} T'(y(w_k)) = T' &= 2\phi \int_{w_{\min}}^{w_{\max}} \frac{1}{c(w)}\left(\frac{\overline{c}}{c(w)} - 1\right) f(w)dw \\ &= 2\phi \int_{w_{\min}}^{w_{\overline{c}}} \frac{1}{c(w)}\left(\frac{\overline{c}}{c(w)} - 1\right) f(w)dw - 2\phi \int_{w_{\overline{c}}}^{w_{\max}} \frac{1}{c(w)}\left(1 - \frac{\overline{c}}{c(w)}\right) f(w)dw \end{aligned} . \tag{18}$$

Equation (18) implies (again) that the marginal tax rate is the same for all individuals (the intuition for which we discussed above), and also that it is proportional to a parameter measuring the strength of the aversion to inequality, ϕ.

Let us next use simulations based on the same consumption distribution as in the Fehr and Schmidt model examined above. In order to compare the results from this version of the Bolton and Ockenfels model with the ones from the Fehr and Schmidt model, it makes sense to somehow calibrate the parameters between these models, which can clearly be done in many different ways. We will here consider a simple calibration procedure as follows: Consider two individuals with consumption levels at 0.5 and 1.5 times the mean level, respectively, and let us compare how much they are each willing to pay (or accept) for increasing (decreasing) the consumption level of an individual with a consumption above (below) the average consumption level. Starting with the Fehr and Schmidt model, it is straightforward to show (see equations [A6] and [A8]) that these values equal α and β, respectively. For the Bolton and Ockenfels model, it follows from equation (15) that the low-consumption individual would on the margin be willing to pay

$$\frac{c}{\overline{c}^2} 2\phi\left(\frac{\overline{c}}{c} - 1\right)\frac{\overline{c}^2}{c^2} = \frac{2\phi}{c}\left(\frac{\overline{c}}{c} - 1\right) = \frac{2\phi}{c} = \frac{4\phi}{\overline{c}}$$

for any other individual to decrease their consumption. The calibration thus means that $\alpha = 4\phi / \overline{c}$ such that $\phi = 0.25\alpha\overline{c}$. Similarly for an individual with consumption 50% above the mean, their marginal willingness to pay for an increase in another individual's consumption is given by

$$\frac{2\phi}{c}\left(1-\frac{\overline{c}}{c}\right)=\frac{2}{3}\frac{\phi}{c}=\frac{4}{9}\frac{\phi}{\overline{c}}$$

such that $\phi = 9\beta\overline{c}/4$. Imposing the Fehr and Schmidt parameters used above ($\alpha = 0.85$ and $\beta = 0.315$) would then in the former case imply that $\phi = 0.25 \cdot 0.85\,\overline{c} = 0.2125\,\overline{c}$, while in the latter case we would get that $\phi = 2.25 \cdot 0.315\,\overline{c} = 0.70875\,\overline{c}$. Since there is only one parameter, we will simply take the average of these two values, such that $\phi = 0.46\,\overline{c} = 20273$ USD.

The results from the simulations imply a Pareto-efficient marginal tax rate approximately equal to 0.003%, and it turns out to be very close to zero for all reasonable values of ϕ. Note that the reason for this finding is not that the effects of varying inequality on each individual's utility is small based on the Bolton and Ockenfels model. The reason is instead that both the level and sign of the marginal willingness to pay vary across individuals, as clearly illustrated in equation (18), and that the net effect turns out to be close to zero. Indeed, while this Pareto-efficient marginal tax rate is not generally strictly equal to zero, it will presumably be close to zero in most cases when the aversion to inequality is measured by a general expression that is locally symmetric around $\overline{c}$.

3.2.1 Non-Symmetric Bolton and Ockenfels Inequality Aversion

While Bolton and Ockenfels assumed that utility is twice continuously differentiable in its arguments, they also noted that this assumption is made for mathematical convenience. Let us here keep the continuity assumption but drop the assumption of differentiability at the average consumption level. Thus, we allow for locally asymmetric utility specifications around the mean consumption level, such that individuals may perceive disadvantageous inequality to be worse than advantageous inequality also locally in the neighborhood of the average consumption (as in the model by Fehr and Schmidt). Moreover, also the quadratic term in the formulation of equation (16) was used for analytical convenience (since it implied that utility is maximized for $\overline{c} = c$). Consider therefore the following asymmetric functional forms that are also more closely related to the Fehr and Schmidt model:

$$u(c,z,C) = v\left(c - \phi^{Below}\left(\frac{\overline{c}}{c}-1\right), z\right) \text{ for } c \le \overline{c} \text{ and}$$

$$u(c,z,C) = v\left(c - \phi^{Above}\left(1 - \frac{\overline{c}}{c}\right), z\right) \text{ for } c > \overline{c}\,.$$

Using this asymmetric utility function to calculate the marginal willingness to pay in equation (16), it follows that

$$T'(w_k l_k) = \frac{\phi^{Below}}{\overline{c}} R(\overline{c}) \frac{\overline{c}_{below\,mean}}{\overline{c}} - \frac{\phi^{Above}}{\overline{c}} \left(1 - R(\overline{c})\right) \frac{\overline{c}_{above\,mean}}{\overline{c}}\,. \qquad (16')$$

Thus, in this case we can obtain a closed-form solution. Based on the same distributions as before, we have that $R(\overline{c})$, i.e., the fraction of the population with a consumption lower than mean consumption, equals about 64%. We also have that the mean consumption among those with a consumption below average, $\overline{c}_{below\,mean}$, equals about 25,000 USD, whereas the mean consumption among those who consume above average, $\overline{c}_{above\,mean}$, equals 78,000 USD. Recalling that the overall average consumption, $\overline{c}$, equals 44,071 USD per annum, we obtain that $\frac{\overline{c}_{below\,mean}}{\overline{c}} \approx 0.56$ and $\frac{\overline{c}_{above\,mean}}{\overline{c}} \approx 1.8$. Moreover, following the same calibration procedure as above, the low-consumption individual with a consumption level at 50% of the mean would on the margin be willing to pay $\frac{c}{\overline{c}^2} \phi^{Below} \frac{\overline{c}^2}{c^2} = \frac{\phi^{Below}}{c} = 2\frac{\phi^{Below}}{\overline{c}}$ for another individual to decrease their consumption, such that $\phi^{Below} = 0.5\alpha\overline{c}$. Similarly for an individual with above-average consumption, we get the marginal willingness to pay $\frac{\phi^{Above}}{c} = 1.5\frac{\phi^{Above}}{\overline{c}}$ such that $\phi^{Above} = \frac{2}{3}\beta\overline{c}$. Substituting these calibrated parameters and values into equation (16') above implies

$$T' \approx 0.5\alpha \cdot 0.64 \cdot 0.56 - 0.667\beta \cdot 0.36 \cdot 1.8 = 0.179\alpha - 0.432\beta\,.$$

If we then plug in the parameter values $\alpha = 0.85$, $\beta = 0.315$, we get an optimal marginal tax rate of 1.6%. For half and one quarter of these parameter values, we would then of course get about 0.8% and 0.4%, respectively. Although these marginal tax rates are not negligible, they are clearly much lower than the marginal tax rates implied by the Fehr and Schmidt model.

Overall, the policy implications in terms of Pareto-efficient taxation turn out to be strikingly different between the Fehr-Schmidt and the Bolton-Ockenfels models. This is the case both in terms of structure, where the marginal tax rates increase strongly with the consumption level based on the former model while being constant in the latter, and in terms of levels, which are potentially very high based on the former model and low or negligible based on the latter.

4. Marginal Income Taxation under Non-Self-Centered Inequality Aversion

Although much work on social preferences has focused on self-centered inequality aversion, one may question such a point of departure in a multi-individual society. In particular, an individual may prefer a more equal consumption distribution to a less equal one regardless of the relationship between their own and other people's consumption. For instance, an individual may prefer a society with fewer super rich and super poor persons regardless of their own consumption level, consumption rank, or relative consumption compared with others.

Moreover, there may be instrumental reasons for preferring more equality, including less social tension, less criminality, and, which is an urgent issue when this is written 2020, better possibilities for policy makers to control the spread of infectious diseases. Based on such a perspective, one can think of the models here as reflecting a reduced form of more complex underlying social mechanisms.

In this section, we explore the marginal willingness to pay in equations (9) and (10) based on different models of *non*-self-centered (or general) inequality aversion. This means that the inequality measure is the same for all individuals such that $\partial C_k / \partial c(w_k) = \partial C / \partial c(w_k)$ for all C and C_k.

We consider two measures of non-self-centered inequality, the Gini coefficient and the coefficient of variation, as the basis for studying the optimal tax policy responses to inequality aversion. We will for ease of comparability in each case consider a Cobb-Douglas specification of the individual's preferences over c and C; cf. Carlsson et al. (2005). For each measure of inequality, we can then write the utility function as

$$u(c,z,C) = v\left(c(\eta - C)^{\gamma}, z\right), \tag{19}$$

where $\gamma > 0$ is a parameter reflecting the degree of inequality aversion and η is interpretable as a measure of maximum inequality. Thus, an individual always prefers less to more general inequality, regardless of the relationship between their own and other people's consumption.

4.1 Gini Coefficient

Let us start with the most commonly used inequality measure at the social level, namely the Gini coefficient, *G*, such that $C = G$ in (19), and let us also assume that $\eta = 1$. We can then interpret $\eta - C$ in equation (19) in terms of the difference between the worst case (maximum inequality) and the actual inequality, i.e., as a measure of social equality; $c\gamma / (\eta - C)$ measures the individual's marginal willingness to pay for a decrease in the Gini coefficient. Note also that the Gini coefficient is half of the relative mean absolute consumption difference, which in turn is defined as the ratio of the mean absolute consumption difference, *D*, to the mean consumption. Therefore, $G = 0.5D / \overline{c}$, and hence

$$G = \frac{0.5}{\overline{c}n^2} \int_{w_{\min}}^{w_{\max}} \int_{w_{\min}}^{w_{\max}} \left| c(\widehat{w}) - c(\breve{w}) \right| f(\widehat{w}) f(\breve{w}) d\breve{w} d\widehat{w} .$$

Based on this measure of inequality, we can derive the marginal willingness to pay measures used to form the marginal tax policy rules in equations (9) and (10).

Let us start with the general case where the marginal taxes are not necessarily low, implying the following result:

Proposition 3. *Suppose that the inequality aversion is non-self-centered and based on the Gini coefficient. The marginal income tax policy can then be characterized as follows:*

i) $$T'(y(w_k)) = \gamma \frac{2R(c(w_k)) - G - 1}{n(1-G)\overline{c}} \int_{w_{\min}}^{w_{\max}} c(w) \frac{1 - T'(y(w_k))}{1 - T'(y(w))} f(w) dw < 1 \quad \forall k \quad (20)$$

ii) $T''(y(w_k)) > 0$ *for* $c_{\min} < c(w_k) < c_{\max}$

iii) *If* $f(c(w_{\min})) = f(c(w_{\max})) = 0$*, then* $T''(y(w_{\min})) = T''(y(w_{\max}) = 0$

iv) $T'(y(w_k)) = 0$ *when* $R(c(w_k)) = (1+G)/2$; and $T'(y(w_k)) > (<)\ 0$ *when* $R(c(w_k)) > (<)\ (1+G)/2$

v) $T'(y(w_{\min})) < T'(y(w_{\text{median}})) < 0 < T'(y(w_{\max})) < \gamma$

vi) $$-\frac{\dfrac{T'(y(w_{\min}))}{1 - T'(y(w_{\min}))}}{\dfrac{T'(y(w_{\max}))}{1 - T'(y(w_{\max}))}} = \frac{1+G}{1-G} = 1 + \frac{2G}{1-G} > 1 \text{ implying}$$

$$-\frac{T'(y(w_{\min}))}{T'(y(w_{\max}))} = \frac{1 - T'(y(w_{\min}))}{1 - T'(y(w_{\max}))} \frac{1+G}{1-G} > 1 + \frac{2G}{1-G} > 1 .$$

Again, there is no closed-form algebraic solution in the general case. As in the Fehr-Schmidt case, the marginal tax function is monotonically increasing, starting from a negative value and ending at a positive one. Here, one can also show (*v*) that the marginal tax rate at the median income is negative, while (*vi*) states that the maximum negative marginal income tax tends to be substantially larger than the maximum positive marginal income tax.

Let us next turn to the simplified case where all marginal income tax rates are low, as given in equation (10), where we instead obtain a closed-form solution as follows:

Corollary 2. *If the inequality aversion is based on the Gini coefficient and the marginal income tax rates are low, then*

$$T'(y(w_k)) = \gamma \frac{2R(c(w_k)) - G - 1}{2(1-G)}. \tag{21}$$

Equation (21) is reminiscent of equation (13), i.e., the corresponding marginal tax policy derived under the Fehr and Schmidt type of inequality aversion, and we can observe a monotonically increasing affine relationship between marginal tax rates and consumption (and hence also a monotonic relationship with gross income, by assumption). The marginal tax rate starts from $-0.5\gamma(1+G)/(1-G) < 0$ for the individual with the lowest consumption rank and ends with $0.5\gamma > 0$ for the individual with the highest consumption rank. The intuition is that all individuals would benefit from a more equal consumption distribution, *ceteris paribus*, which can be accomplished through increased consumption in the lower end of the distribution and decreased consumption in the upper end. Since marginal taxation affects the before-tax income via the labor-supply decision, the tendency to supply too much labor in the upper end of the distribution and too little labor in the lower end is counteracted through this marginal tax policy.

Returning to the general case, where we do not assume that the marginal tax rates necessarily are low, let us now consider simulations based on the same consumption distribution as before, with a Gini coefficient of 0.377, and hence a relative mean absolute consumption difference of about 0.75. The results are presented in Figure 2.

As expected from the qualitative analysis above, the Pareto-efficient marginal tax rates vary with the (optimal) consumption level in the same general way as for the Fehr and Schmidt type of inequality aversion, and the increase is also monotonic, consistent with Proposition 3. This means that the non-self-centered inequality aversion discussed here may have tax policy implications qualitatively similar to those associated with self-centered inequality aversion, even if the levels of marginal taxation differ between Figures 1 and 2. One important difference is that the marginal tax rates are negative even for quite high consumption levels in Figure 2.

To provide intuition behind this finding, note first that increased consumption for all individuals below the mean leads to decreased inequality, and hence causes a positive externality. However, a consumption increase for middle-class people will also decrease inequality. In fact, for the Gini coefficient to increase, the initial consumption level must be rather high, such that the rank at which the consumption increases exceeds $(1+G)/2$. That is, if the Gini coefficient reflects perfect equality, such that G=0, then it is sufficient that the consumption level is larger than the median consumption level for the marginal tax to be positive. Yet, for the Gini coefficient used in the simulations (reflecting the U.S. disposable income distribution with G=0.381) the initial consumption must be larger than the consumption level for almost 70% of the population for the externality to be negative and the marginal tax to be positive.

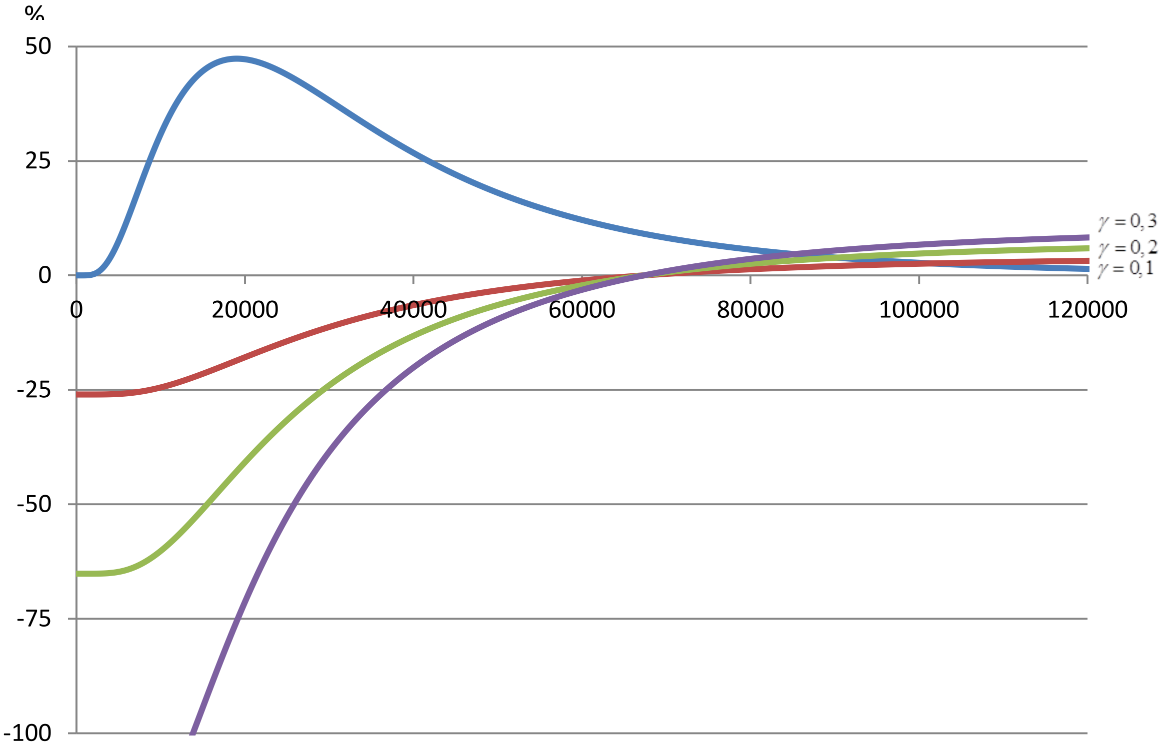

Figure 3. Pareto-efficient marginal income tax rates as a function of disposable income (for a log-normal distribution) in equilibrium, based on non-self-centered inequality aversion where inequality is measured as the Gini coefficient.

4.2 Coefficient of Variation

Consider next another commonly used general inequality measure, namely the coefficient of variation, V, defined as the ratio of the standard deviation of the consumption distribution in the population, σ, to mean consumption, $\bar{c}$, such that $C = V = \sigma / \bar{c}$. Carlsson et al. (2005) analyze and parameterize this measure of inequality based on a questionnaire-experimental approach. They conclude that the mean degree of inequality aversion is such that $\gamma \approx 0.2$ in equation (19).

In the general case, where the marginal taxes are not necessarily low, the utility function in equation (19) and Lemma 1 imply the following result:

Proposition 4. *Suppose that the inequality aversion is non-self-centered and based on the coefficient of variation. Then*

i) $$T'(y(w_k)) = \frac{\gamma}{n\bar{c}} \frac{V}{\eta - V} \left(\frac{c(w_k) - \bar{c}}{\bar{c} V^2} - 1 \right) \int_{w_{\min}}^{w_{\max}} c(w) \frac{1 - T'(y(w_k))}{1 - T'(y(w))} f(w) dw < 1 \ \forall k \qquad (22)$$

ii) $T''(y(w_k)) > 0$

iii) *If* $\dfrac{d^2 c(w_k)}{dy^2(w_k)} < 0$ *then* $T'''(y(w_k)) < 0 \ \forall k$

iv) $T'(y(w_k)) = 0$ *for* $c(w_k) = \left(V^2 + 1\right)\bar{c}$; $T'(y(w_k)) > (<)\ 0$ *for* $c(w_k) > (<) \left(V^2 + 1\right)\bar{c}$

v) $T'((y(\bar{c})) < 0$

vi) *For a sufficiently large* γ *there exists a* $c^* > 0$ *such that* $\lim_{c \to c^*} T'(y(w_k)) = -\infty$

vii) $\lim_{c(w_k) \to \infty} T'(y(w_k)) = 1$.

The monotonic structure of the marginal income tax prevails here too (*ii*), and provided that consumption increases in a concave manner with respect to gross income (a sufficient and not necessary condition) the marginal income tax function is strictly concave in consumption (*iii*). The marginal tax rate is negative at the mean income level (*v*) and must exceed $\left(V^2 + 1\right)$ times mean consumption in order to become positive (*iv*). The basic intuition for why the marginal tax rate is negative for such a large part of the consumption interval is similar to the case where the measure of inequality is based on the Gini coefficient. *vi* implies that for a sufficiently strong inequality aversion, as reflected by γ, the marginal tax rate approaches minus infinity at a certain consumption level. The interpretation is simply that regardless of the weights implied by the social welfare function, it can then never be optimal with consumption levels below this level, and thus, the tax structure must prevent this from happening. *vii* finally implies that the marginal tax rate at the top will approach 100% when consumption (or net income) approaches infinity.

In the simplified case where all marginal income tax rates are small, we obtain a closed-form solution summarized as follows:

Corollary 3. *Suppose that the inequality aversion is non-self-centered and based on the coefficient of variation, and that all marginal income tax rates are low. Then*

$$T'(y(w_k)) = \gamma \frac{V}{\eta - V}\left(\frac{c(w_k) - \overline{c}}{\overline{c}}\frac{1}{V^2} - 1\right). \tag{23}$$

Again we can observe a monotonic positive relationship between the marginal tax rates and consumption, starting from $-\gamma\left(1/V^2+1\right)$ for $c(w_k)=0$. The basic intuition is of course the same as for the marginal tax policy implied by equation (21), where the inequality aversion is based on the Gini coefficient.

The simulation in Figure 4 shows the efficient marginal tax rates for different inequality parameters γ, based on the same distributional assumptions as before for the general case (without assuming small marginal tax rates). We can observe that the Pareto-efficient marginal tax rates vary strongly with the consumption level, that they may become very high at non-extreme consumption levels, and that they can take extreme negative values for low levels of consumption. Indeed, it can be shown that the case with $\gamma = 0.3$ implies that the Pareto-efficient marginal tax approaches minus infinity for a positive consumption level. Despite level differences, however, the general pattern in Figure 4 resembles that in Figure 3.

The difference in pattern compared with the model with Fehr and Schmidt (self-centered) inequality aversion is similar here, and for the same reasons as for the inequality aversion model based on the Gini coefficient. Increased consumption for the fraction of the population satisfying $(c(w_k)-\bar{c})/\bar{c} < V^2$ leads to less inequality and vice versa, meaning that the marginal income tax rate will remain negative up to quite high consumption levels.

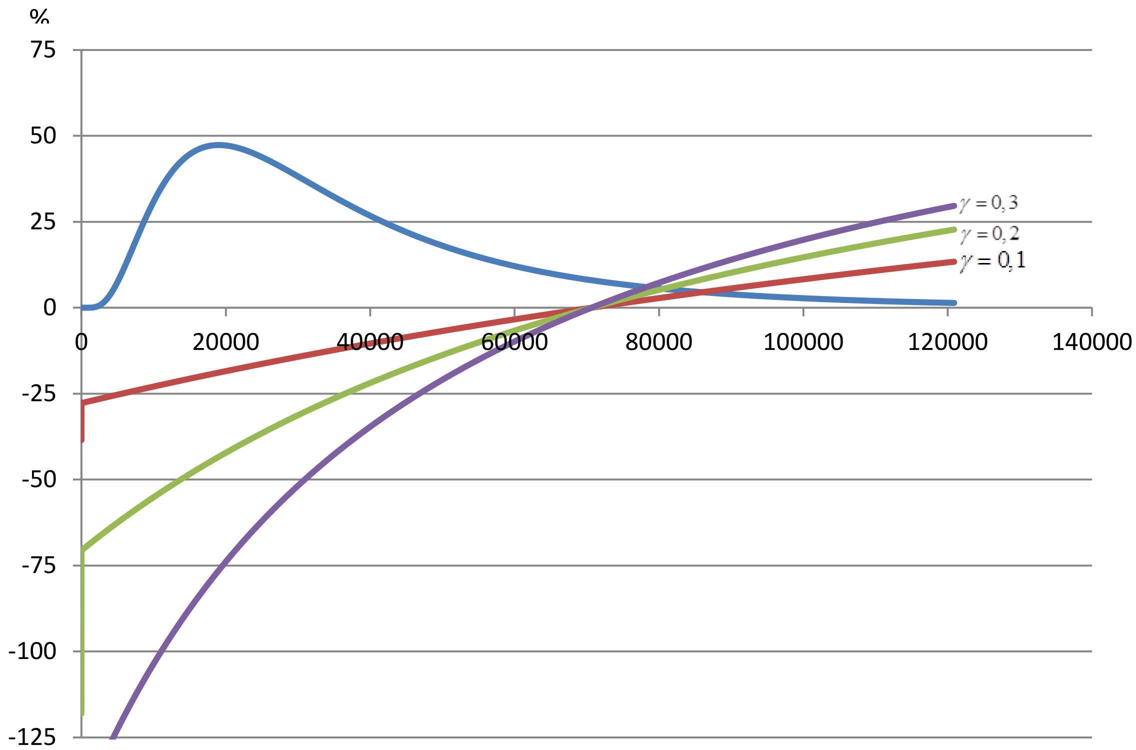

Figure 4. Pareto-efficient marginal income tax rates as a function of disposable income (for a log-normal distribution) in equilibrium, based on non-self-centered inequality aversion where inequality is measured as the coefficient of variation.

5. Conclusions

We started by examining a general model, in which we made no other assumption about the inequality aversion other than that people prefer a more equal distribution of consumption (or disposable income) over a less equal one, *ceteris paribus*. Based on the policy rules for marginal income taxation derived in the context of this general model, we examined the implications of four more specific types of inequality aversion, two self-centered and two non-self-centered. The basic aims were to understand how and why inequality aversion motivates marginal tax wedges in the labor market and how the Pareto-efficient marginal tax rate varies along the distribution of consumption.

The take-home message of the paper is twofold. First, empirically and experimentally quantified degrees of inequality aversion have potentially very important implications for Pareto-efficient marginal income taxation. More specifically, three out of four models of inequality aversion show that the first-best efficient marginal tax rates required to internalize the externalities caused by inequality aversion are both substantial in size and vary substantially with respect to the consumption levels. Moreover, these models imply a progressive marginal tax structure in the sense that low income levels are subsidized at a diminishing marginal rate while high income levels are taxed at an increasing marginal rate.

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. The most striking result comes from comparing the two most cited models of self-centered inequality aversion. Whereas the Fehr and Schmidt (1999) type of inequality aversion implies monotonically increasing marginal income tax rates, with high negative marginal tax rates for low-income individuals and high positive tax rates for high-income individuals, the often considered similar inequality aversion model by Bolton and Ockenfels (2000) implies close to zero marginal income tax rates for all. A crucial underlying reason is that the consumption externality caused by inequality aversion is non-atmospheric in the former case and atmospheric in the latter.

Future research may take different directions. One way to go would be to allow for a broader spectrum of social interaction, where the policy implications of inequality aversion are examined alongside the implications of other (empirically established) forms of social interaction, such as relative consumption concerns and/or social norms.

Appendix

Proof of Lemma 1

By combining (6) and (7) we get

$$\frac{\int_{w_{\min}}^{w_{\max}} \psi' \frac{\partial u(c(w),z(w),C)}{\partial C} \frac{\partial C}{\partial c(w_k)} f(w)\,dw}{\frac{1}{w_k}\psi_k' \frac{\partial u(c(w_k),z(w_k),C_k)}{\partial z(w_k)}} + w_k \frac{\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial C_k} \frac{\partial C_k}{\partial c(w_k)}}{\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial z(w_k)}} + w_k MRS_{cz,k} = 1, \tag{A1}$$

where

$$MRS_{cz,k} = \frac{\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial c(w_k)}}{\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial z(w_k)}}.$$

Next, we make use of the fact that (7) holds for all ability types such that

$$\frac{1}{w_k}\psi_k' \frac{\partial u(c(w_k),z(w_k),C_k)}{\partial z(w_k)} = \lambda = \frac{1}{w}\psi' \frac{\partial u(c(w),z(w),C)}{\partial z(w)}. \tag{A2}$$

Substituting equation (A2) into equation (A1) gives

$$\int_{w_{\min}}^{w_{\max}} w \frac{\frac{\partial u(c(w),z(w),C)}{\partial c(w)}}{\frac{\partial u(c(w),z(w),C)}{\partial z(w)}} \frac{\frac{\partial u(c(w),z(w),C)}{\partial C} \frac{\partial C}{\partial c(w_k)}}{\frac{\partial u(c(w),z(w),C)}{\partial c(w)}} f(w)dw + w_k \frac{\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial C_k} \frac{\partial C_k}{\partial c(w_k)}}{\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial z(w_k)}} + w_k MRS_{cz,k} = 1. \tag{A3}$$

Using equation (4) to derive

$$w_k MRS_{cz}(w_k) + w_k \frac{\partial u(c_k,z_k,C_k)}{\partial C_k} \frac{\partial C_k}{\partial c(w_k)} \Big/ \frac{\partial u(c_k,z_k,C_k)}{\partial z_k} = \frac{1}{1-T'(y_k)}$$

and substituting into equation (A3) gives equation (9).

Finally, for any gross wages w_j and w_k, we can write

$$\frac{1-T'(y(w_j))}{1-T'(y(w_k))}=\frac{\psi_k{}'\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial c(w_k)}+\psi_k{}'\frac{\partial u(c(w_k),z(w_k),C_k)}{\partial C_k}\frac{\partial C_k}{\partial c(w_k)}}{\psi_j{}'\frac{\partial u(c(w_j),z(w_j),C_j)}{\partial c(w_j)}+\psi_j{}'\frac{\partial u(c(w_j),z(w_j),C_j)}{\partial C_j}\frac{\partial C_j}{\partial c(w_j)}}. \quad \text{(A4)}$$

For each type we assume that social welfare increases in utility, and individual utility increases in consumption, implying that the right-hand side of equation (A4) is positive. For the left-hand side to be positive as well, we must have $T'(y(w_k))<1$ for each type *k*.

QED

Equation (10) follows as the special case where $\left(1-T'(y(\tilde{w}))\right)/\left(1-T'((w))\right)\approx 1$.

Proof of Proposition 1

By using equation (11), we can derive $\partial u(c,z,C)/\partial C=-\partial u(c,z,C)/\partial c$, implying

$$MWTP(c(w),c(w_k))=\frac{\partial C}{\partial c(w_k)}. \quad \text{(A5)}$$

For $c(w)<c(w_k)$, it follows that $\partial C/\partial c(w_k)=\frac{\alpha}{n}$ and hence

$$MWTP(c(w),c(w_k))=\frac{\alpha}{n}. \quad \text{(A6)}$$

Thus, all individuals with a consumption level lower than $c(w_k)$ will on the margin be willing to pay the same amount, α/n, per consumption reduction unit for an individual with consumption $c(w_k)$. Similarly, when $c(w)>c(w_k)$, it follows that $\partial C/\partial c(w_k)=-\frac{\beta}{n}$ and

$$MWTP(c(w),c(w_k))=-\frac{\beta}{n}. \quad \text{(A7)}$$

Therefore, all individuals with a consumption level higher than $c(w_k)$ will instead be willing to pay β/n per unit consumption increase for an individual with consumption $c(w_k)$.

Substituting equations (A6) and (A7) into equation (9) directly yields equation (13), and hence part (*i*).

Regarding (*ii*), the monotonicity property, note that we can rewrite equation (12) as

$$
\begin{aligned}
\frac{T'(y(w_k))}{1-T'(y(w_k))} &= \frac{\alpha}{n}\int_{w_{\min}}^{w_k}\frac{f(w)}{1-T'(y(w))}dw - \frac{\beta}{n}\int_{w_k}^{w_{\max}}\frac{f(w)}{1-T'(y(w))}dw \\
&= \frac{\alpha}{n}\int_{y(w_{\min})}^{y(w_k)}\frac{g(y)}{1-T'(y)}dy - \frac{\beta}{n}\int_{y(w_k)}^{y(w_{\max})}\frac{g(y)}{1-T'(y)}dy
\end{aligned}, \tag{A8}
$$

where $g(y)$ is the density function of gross income y in equilibrium. By differentiating both sides of (A8) with respect to $y(w_k)$, we get

$$
\begin{aligned}
&\frac{T''(y(w_k))(1-T'(y(w_k)))+T'(y(w_k))T''(y(w_k))}{(1-T'(y(w_k)))^2} \\
&= \frac{T''(y(w_k))}{(1-T'(y(w_k)))^2} = \frac{\alpha+\beta}{n}\frac{g(y(w_k))}{1-T'(y(w_k))} > 0
\end{aligned},
$$

implying

$$
T''(y(w_k)) = \frac{\alpha+\beta}{n}(1-T'(y(w_k)))g(y(w_k)) > 0, \tag{A9}
$$

since the marginal tax rate is lower than one by Lemma 1.

Regarding (*iii*), for a continuous ability distribution where the density goes to zero at the minimum and maximum ability level, respectively, it follows from equation (A9) that

$$
T''(y(w_{\min})) = T''(y(w_{\max})) = 0 .
$$

Consider next (*iv*)–(*vi*). From equation (13), we obtain

$$
\frac{T'(y(w_{\max}))}{1-T'(y(w_{\max}))} = \frac{\alpha}{n}\int_{w_{\min}}^{w_{\max}}\frac{f(w)}{1-T'(y(w))}dw \tag{A10}
$$

and

$$
\frac{T'(y(w_{\min}))}{1-T'(y(w_{\min}))} = -\frac{\beta}{n}\int_{w_{\min}}^{w_{\max}}\frac{f(w)}{1-T'(y(w))}dw. \tag{A11}
$$

Equations (A10) and (A11) together imply

$$
\frac{\dfrac{T'(y(w_{\max}))}{1-T'(y(w_{\max}))}}{\dfrac{T'(y(w_{\min}))}{1-T'(y(w_{\min}))}} = -\frac{\alpha}{\beta}. \tag{A12}
$$

Now,

$$
\operatorname{sign}\left(T'(y(w_{\max}))\right) = \operatorname{sign}\left(\frac{T'(y(w_{\max}))}{1-T'(y(w_{\max}))}\right)
$$

$$\operatorname{sign}\left(T'(y(w_{\min}))\right) = \operatorname{sign}\left(\frac{T'(y(w_{\min}))}{1-T'(y(w_{\min}))}\right)$$

together with (A12) imply $\operatorname{sign}\left(T'(y(w_{\min}))\right) = -\operatorname{sign}\left(T'(y(w_{\max}))\right)$, which together with (*ii*) implies that $T'(y(w_{\min})) < 0$ and $T'(y(w_{\max})) > 0$. (*vi*) then follows directly from equation (A12).

Consider next the second order derivative of the marginal tax function and properties (*vii*) and (*viii*). By differentiating both sides of equation (A9) with respect to $y(w_k)$, we obtain

$$T'''(y(w_k)) = \frac{\alpha+\beta}{n}\left(-T''(y(w_k)g(y(w_k) + (1-T'(y(w_k))g'(y(w_k))\right). \qquad \text{(A13)}$$

Thus, for a continuous ability distribution where both the density and its derivative go to zero at the minimum and maximum ability level, so that

$$g(y(w_{\min}) = g'(y(w_{\min}) = g(y(w_{\max}) = g'(y(w_{\max}),$$

we obtain (*xiii*) from equation (A13), i.e.

$$T'''(y(w_{\min})) = T'''(y(w_{\max})) = 0.$$

We also have from equation (A13) that $T'''(y(w_k)) = 0$ implies

$$T''(y(w_k)g(y(w_k) = (1-T'(y(w_k))g'(y(w_k), \qquad \text{(A14)}$$

so

$$\frac{T''(y(w_k)}{1-T'(y(w_k)} = \frac{g'(y(w_k)}{g(y(w_k)}. \qquad \text{(A15)}$$

Since the left-hand side of equation (A15) is positive, we must have $g'(y(w_k) > 0$ for the right-hand-side to be positive as well, implying (*xii*). Finally, (*ix*) is obtained directly from equations (A10) and (A11).

QED

Proof of Corollary1

Substituting equations (A6) and (A7) into equation (10) yields

$$\begin{aligned} T'(y(w_k)) &= \frac{\alpha}{n}\int_{w_{\min}}^{w_k} f(c(w))dw - \frac{\beta}{n}\int_{w_k}^{w_{\max}} f(c(w))dw = \alpha R(c(w_k)) - \beta(1-R(c(w_k))) \\ &= -\beta + (\alpha+\beta)R(c(w_k)) \end{aligned}. \qquad \text{(A16)}$$

QED

Proof of Proposition 2

When the marginal tax rates are low, it follows from equation (10) that

$$T'(y(w_k)) = \frac{1}{n\overline{c}^2} \int_{w_{\min}}^{w_{\max}} c(w) \frac{\dfrac{\partial u(c(w), z(w), C)}{\partial C}}{\dfrac{\partial u(c(w), z(w), C)}{\partial c(w)}} f(w) dw .$$

(A17)

If based on the more general equation (9) we instead obtain

$$\frac{T'(y(w_k))}{1 - T'(y(w_k))} = \frac{1}{n\overline{c}^2} \int_{w_{\min}}^{w_{\max}} c(w) \frac{\dfrac{\partial u(c(w), z(w), C)}{\partial C}}{\dfrac{\partial u(c(w), z(w), C)}{\partial c(w)}} \frac{f(w)}{1 - T'(y(w))} dw , \qquad \text{(A18)}$$

which is also the same for all, implying that equation (A18) reduces to equation (A17) and hence to equation (16). QED

Derivation of equation (18)

From equation (17) follows that

$$\frac{\dfrac{\partial u(c(w), z(w), C)}{\partial C}}{\dfrac{\partial u(c(w), z(w), C)}{\partial c(w)}} = 2\phi \left(\frac{\overline{c}}{c(w)} \right)^2 \left(\frac{\overline{c}}{c(w)} - 1 \right). \qquad \text{(A19)}$$

Substituting equation (A19) into equation (16) gives equation (18). QED

Derivation of equation (16')

Using the sub-utility functions, we obtain

$$T'(y(w_k)) = T' = \frac{\phi^{Below}}{n\overline{c}^2} \int_{w_{\min}}^{w_{\overline{c}}} c(w)\, f(w) dw - \frac{\phi^{Above}}{n\overline{c}^2} \int_{w_{\overline{c}}}^{w_{\max}} c(w)\, f(w) dw$$

$$= \frac{\phi^{Below}}{n\overline{c}^2} \int_{w_{\min}}^{w_{\overline{c}}} f(w) dw \frac{\int_{w_{\min}}^{w_{\overline{c}}} c(w)\, f(w) dw}{\int_{w_{\min}}^{w_{\overline{c}}} f(w) dw} - \frac{\phi^{Above}}{n\overline{c}^2} \int_{w_{\overline{c}}}^{w_{\max}} f(w) dw \frac{\int_{w_{\overline{c}}}^{w_{\max}} c(w)\, f(w) dw}{\int_{w_{\overline{c}}}^{w_{\max}} f(w) dw} .$$

$$= \frac{\phi^{Below}}{n\overline{c}} R(\overline{c}) \frac{\overline{c}_{below\ mean}}{\overline{c}} - \frac{\phi^{Above}}{n\overline{c}} \left(1 - R(\overline{c})\right) \frac{\overline{c}_{above\ mean}}{\overline{c}}$$

Proof of Proposition 3

Equation (19) with $\eta = 1$ implies

$$\frac{\frac{\partial u(c(w),z(w),C)}{\partial C}}{\frac{\partial u(c(w),z(w),C)}{\partial c(w)}} = -\gamma \frac{c(w)}{1-G}. \tag{A20}$$

Now,

$$D = \frac{1}{n^2}\int_0^\infty f(\hat{w})\int_0^{\hat{w}} f(\breve{w})(c(\hat{w})-c(\breve{w}))d\breve{w}d\hat{w} + \frac{1}{n^2}\int_0^\infty f(\hat{w})\int_{\hat{w}}^\infty f(\breve{w})(c(\breve{w})-c(\hat{w}))d\breve{w}d\hat{w}, \tag{A21}$$

such that[9]

$$\begin{aligned}\frac{\partial D}{\partial c(w_k)} &= \frac{1}{n^2}\left(\frac{\partial\left(\int_0^\infty f(\hat{w})\int_0^{\hat{w}} f(\breve{w})(c(\hat{w})-c(\breve{w}))d\breve{w}d\hat{w} + \int_0^\infty f(\hat{w})\int_{\hat{w}}^\infty f(\breve{w})(c(\breve{w})-c(\hat{w}))d\breve{w}d\hat{w}\right)}{\partial c_k}\right)_{norm} \\ &= \frac{2}{n^2}\left(\int_0^{w_k} f(\hat{w})d\hat{w} - \int_{w_k}^\infty f(\hat{w})d\hat{w}\right) = \frac{2}{n}\left(2R(c(w_k))-1\right)\end{aligned} \tag{A22}$$

By using $G = 0.5D/\bar{c}$, we can correspondingly derive

$$\begin{aligned}\frac{\partial C}{\partial c(w_k)} = \frac{\partial G}{\partial c(w_k)} &= 0.5\frac{1}{\bar{c}}\frac{\partial D}{\partial c(w_k)} - 0.5\frac{D}{\bar{c}^2}\frac{\partial \bar{c}}{\partial c(w_k)} \\ &= \frac{1}{n\bar{c}}\left(2R(c(w_k))-1-0.5\frac{D}{\bar{c}}\right) = \frac{1}{n\bar{c}}\left(2R(c(w_k))-1-G\right)\end{aligned} \tag{A23}$$

and

$$\begin{aligned}MWTP(c(w),c(w_k)) &= -\frac{\frac{\partial u(c(w),z(w),C)}{\partial C}}{\frac{\partial u(c(w),z(w),C)}{\partial c(w)}}\frac{1}{n\bar{c}}\left(2R(c(w_k))-1-G\right) \\ &= \frac{\gamma}{1-G}\frac{c(w)}{n\bar{c}}\left(2R(c(w_k))-1-G\right)\end{aligned}. \tag{A24}$$

Substituting equation (A24) into equation (9) yields equation (20).

The monotonicity property can be shown by rewriting equation (20) to read

[9] Where this partial derivative should thus be interpreted as the effect on D due to a consumption change *for an individual of type k* (and hence not due to a consumption change *for all individuals of type k* in the continuous distribution, which is zero).

$$\begin{aligned}\frac{T'(y(w_k))}{1-T'(y(w_k))} &= \gamma\frac{2R(c(w_k))-1-G}{n(1-G)\overline{c}}\int_{w_{\min}}^{w_{\max}}\frac{c(w)f(w)}{1-T'(y(w))}dw \\ &= \frac{2R(c(w_k))-1-G}{(1-G)\overline{c}}K\end{aligned}, \tag{A25}$$

where K is a constant. By differentiating both sides of equation (A25) with respect to $c(w_k)$, we obtain

$$\frac{T''(y(w_k))}{\left(1-T'(y(w_k))\right)^2}\frac{dy(w_k)}{dc(w_k)}\frac{1}{K}=\frac{2f(w_k)}{(1-G)\overline{c}}>0. \tag{A26}$$

If the density function approaches zero at the end points (also if the upper end is infinite), such that $f(w_{\min})=f(w_{\max})=0$, then we obtain (*ii*) since we can write

$$T''(y(w_k))=\frac{K}{\frac{dy(w_k)}{dc(w_k)}}\frac{2f(w_k)}{(1-G)\overline{c}}\left(1-T'(y(w_k))\right)^2>0. \tag{A27}$$

It follows from equation (A27) that $T''(y(w_{\min}))=T''(y(w_{\max}))=0$, and hence (*iii*).

(*iv*) follows directly from equation (20) while noting that $R(c(w_{median}))=0.5$ so that

$$T'(y(w_{median}))=-\gamma\frac{G}{n(1-G)\overline{c}}\int_{w_{\min}}^{w_{\max}}c(w)\frac{1-T'(y(w_k))}{1-T'(y(w))}f(w)dw<0. \tag{A28}$$

Furthermore, $R(c(w_{\max}))=1$, so it also follows from equation (20) that

$$T'(y(w_{\max}))=\frac{\gamma}{n\overline{c}}\int_{w_{\min}}^{w_{\max}}\frac{1-T'(y(w_{\max}))}{1-T'(y(w))}c(w)f(w)dw. \tag{A29}$$

Since $\frac{1-T'(y(w_{\max}))}{1-T'(y(w))}<1$, equation (A29) implies $T'(y(w_{\max}))<\gamma$.

Since $R(c(w_{\min}))=0$, it follows that

$$T'(y(w_{\min}))=-\gamma\frac{1+G}{n(1-G)\overline{c}}\int_{w_{\min}}^{w_{\max}}c(w)\frac{1-T'(y(w_k))}{1-T'(y(w))}f(w)dw<1 \qquad \forall k. \tag{A30}$$

Therefore,

$$-\frac{\frac{T'(y(w_{\min}))}{1-T'(y(w_{\min}))}}{\frac{T'(y(w_{\max}))}{1-T'(y(w_{\max}))}}=\frac{1+G}{1-G}=1+\frac{2G}{1-G}>1 \tag{A31}$$

and

$$-\frac{T'(y(w_{\min}))}{T'(y(w_{\max}))} = \frac{1-T'(y(w_{\min}))}{1-T'(y(w_{\max}))}\frac{1+G}{1-G} > 1 + \frac{2G}{1-G} > 1 \tag{A32}$$

since $T'(y(w_{\min})) < 0, \quad T'(y(w_{\max})) > 0$.

QED

Proof of Corollary 2

Follows directly from Proposition 3 when $\left(1-T'(y(w_k))\right)/\left(1-T'(y(w_k))\right) \approx 1$.

QED

Proof of Proposition 4

By using $C = V = \sigma / \overline{c}$, we obtain

$$\begin{aligned}\left(\frac{\partial C}{\partial c(w_k)}\right)_{norm} &= \frac{1}{n\overline{c}}\left(\frac{\partial \sigma}{\partial c(w_k)}\right)_{norm} - \frac{\sigma}{n\overline{c}^2}\left(\frac{\partial \overline{c}}{\partial c(w_k)}\right)_{norm} \\ &= \frac{1}{n\overline{c}}\left(\frac{\partial\left(\left(\int_{w_{\min}}^{w_{\max}} (c(\widehat{w})-\overline{c})^2 f(\widehat{w})d\widehat{w}\right)^{0.5}\right)}{\partial c(w_k)}\right)_{norm} - \frac{\sigma}{n\overline{c}^2} f(w_k). \\ &= \frac{f(w_k)}{n\overline{c}}\left(\frac{c(w_k)-\overline{c}}{\sigma} - \frac{\sigma}{\overline{c}}\right)\end{aligned} \tag{A33}$$

Using equation (A33) in equation (8) then implies

$$MWTP(c(w), c(w_k)) = -\frac{\dfrac{\partial u(c(w), z(w), C)}{\partial C}}{\dfrac{\partial u(c(w), z(w), C)}{\partial c(w)}}\frac{1}{n\overline{c}}\left(\frac{c(w_k)-\overline{c}}{\sigma} - \frac{\sigma}{\overline{c}}\right). \tag{A34}$$

By using equation (19), we obtain the following marginal rate of substitution:

$$\frac{\dfrac{\partial u(c(w), z(w), C)}{\partial C}}{\dfrac{\partial u(c(w), z(w), C)}{\partial c(w)}} = -\frac{1}{n}\frac{\gamma c(w)}{\eta - V} = -\frac{1}{n}\gamma c(w)\frac{\overline{c}}{\eta\overline{c} - \sigma}. \tag{A35}$$

Substituting equation (A35) into equation (A34) gives

$$
\begin{aligned}
MWTP(c(w),c(w_k)) &= \frac{1}{n}\frac{\gamma c(w)}{\eta\bar{c}-\sigma}\left(\frac{c(w_k)-\bar{c}}{\sigma}-V\right) \\
&= \gamma\frac{c(w)}{n\bar{c}}\frac{V}{\eta-V}\left(\frac{c(w_k)/\bar{c}-1}{V^2}-1\right)
\end{aligned}. \qquad \text{(A36)}
$$

Substituting finally equation (A36) into equation (9) yields equation (22).

The monotonicity property (*ii*) can be shown by rewriting equation (22) to read

$$
\begin{aligned}
\frac{T'(y(w_k))}{1-T'(y(w_k))} &= \frac{1}{n}\frac{\gamma}{\bar{c}}\frac{V}{\eta-V}\left(\frac{c(w_k)-\bar{c}}{\bar{c}}\frac{1}{V^2}-1\right)\int_{c_{\min}}^{c_{\max}}\frac{c(w)}{1-T'(y(w))}f(w)dw \\
&= \frac{1}{\bar{c}}\frac{V}{\eta-V}\left(\frac{c(w_k)-\bar{c}}{\bar{c}}\frac{1}{V^2}-1\right)K
\end{aligned}. \qquad \text{(A37)}
$$

By differentiating both sides of equation (A37) with respect to $c(w_k)$ and simplifying, we obtain

$$
\frac{T''(y(w_k))}{\left(1-T'(y(w_k))\right)^2}\frac{dy(w_k)}{dc(w_k)}\frac{1}{K} = \frac{1}{\bar{c}^2}\frac{V}{\eta-V},
$$

so

$$
T''(y(w_k)) = \left(1-T'(y(w_k))\right)^2\frac{K}{\bar{c}^2}\frac{V}{\eta-V}\bigg/\frac{dy(w_k)}{dc(w_k)} > 0. \qquad \text{(A38)}
$$

Consider next the third-order derivative, (*iii*). Differentiating both sides of

$$
T''(y(w_k))\frac{dy(w_k)}{dc(w_k)} = \frac{1}{\bar{c}^2}\frac{V}{\eta-V}K\left(1-T'(y(w_k))\right)^2
$$

with respect to $c(w_k)$ gives

$$
\begin{aligned}
&T'''(y(w_k))\frac{dy(w_k)}{dc(w_k)} + T''(y(w_k))\frac{d^2y(w_k)}{dc^2(w_k)} \\
&= -\frac{2}{\bar{c}^2}\frac{V}{\eta-V}K\left(1-T'(y(w_k))\right)T''(y(w_k))\frac{dy(w_k)}{dc(w_k)}
\end{aligned},
$$

implying

$$
T'''(y(w_k)) = -T''(y(w_k))\left(\frac{2K}{\bar{c}^2}\frac{V}{\eta-V}\left(1-T'(y(w_k))\right) + \frac{\dfrac{d^2y(w_k)}{dc^2(w_k)}}{\dfrac{dy(w_k)}{dc(w_k)}}\right). \qquad \text{(A39)}
$$

A sufficient, but not necessary, condition for this is thus that $\frac{d^2 y(w_k)}{dc^2(w_k)} \Big/ \frac{dy(w_k)}{dc(w_k)} > 0$. This is, in turn, equivalent to $\frac{d^2 c(w_k)}{dy^2(w_k)} \Big/ \frac{dc(w_k)}{dy(w_k)} < 0$, and since the monotonicity part is assumed, i.e., $\frac{dc(w_k)}{dy(w_k)} > 0$, it follows that $\frac{d^2 c(w_k)}{dy^2(w_k)} < 0$.

As for the crossing point on the horizontal axis, (*iv*), where the marginal tax rate is zero, we immediately get $T'(y(w_k)) = 0$ for $c(w_k) = \left(V^2 + 1\right)\bar{c}$ from equation (22). Monotonicity ensures the other parts of (*iv*).

Part (*v*), i.e., $T'((y(\bar{c})) < 0$, follows by substituting $c(w_k) = \bar{c}$ into equation (22), implying

$$T'(y(\bar{c})) = -\frac{\gamma}{n\bar{c}} \frac{V}{\eta - V} \int_{w_{\min}}^{w_{\max}} c(w) \frac{1 - T'(y(w_k))}{1 - T'(y(w))} f(w) dw < 0, \qquad \text{(A40)}$$

where $y(\bar{c})$ denotes the income level corresponding to the mean consumption level.

Consider next property (*vi*), which means that the marginal income tax rate may approach infinity when consumption approaches a finite level from the right. Note that if a finite optimal marginal tax rate exists when consumption approaches zero, we have

$$\frac{T'(y(w_{\min}))}{1 - T'(y(w_{\min}))} = -\frac{\gamma}{n\bar{c}} \frac{V}{\eta - V} \left(\frac{1}{V^2} + 1\right) \int_{w_{\min}}^{w_{\max}} \frac{c(w)}{1 - T'(y(w))} f(w) dw < 0. \qquad \text{(A41)}$$

Note that $T'(y(w_{\min})) / \left(1 - T'(y(w_{\min}))\right)$ approaches minus one when $T'(y(w_{\min}))$ approaches minus infinity. Hence, it cannot be smaller than minus one. Yet for a sufficiently large γ, the right-hand side of equation (A41) satisfies

$$-\frac{\gamma}{n\bar{c}} \frac{V}{\eta - V} \left(\frac{1}{V^2} + 1\right) \int_{w_{\min}}^{w_{\max}} \frac{c(w)}{1 - T'(y(w))} f(w) dw < -1 \qquad \text{(A42)}$$

regardless of the stream of finite marginal tax rates. By continuity, this can also happen for a slightly higher consumption level. Thus, for a sufficiently large γ, there exists a finite consumption level $c^* > 0$ such that

$$\lim_{c \to c^*} T'(y(\tilde{w})) = -\infty .$$

Finally, consider the maximum marginal tax rate, (*vii*). We have

$$\frac{T'(y(w_{\max}))}{1-T'(y(w_{\max}))} = \frac{\gamma}{n\overline{c}} \frac{V}{\eta - V} \left(\frac{c(w_{\max}) - \overline{c}}{\overline{c}} \frac{1}{V^2} - 1 \right) \int_{w_{\min}}^{w_{\max}} \frac{c(w)}{1-T'(y(w))} f(w) dw . \text{ (A43)}$$

For $c(w_{\max})$ approaching infinity, it follows that the right-hand side of equation (A43) approaches infinity, too. For the left-hand side to approach infinity, this means that $T'(y(w_{\max}))$ must approach one.

QED

Proof of Corollary 3

Follows directly from Proposition 4 when $\left(1-T'(y(w_k))\right) / \left(1-T'(y(w_k))\right) \approx 1$.

QED

References

Abel, A.B. (2005) Optimal Taxation When Consumers Have Endogenous Benchmark Levels of Consumption. Review of Economic Studies 72, 1–19.

Alvarez-Cuadrado, F., N.V. Long,. (2011). The Relative Income Hypothesis. Journal of Economic Dynamics and Control 35, 1489–1501.

Alvarez-Cuadrado, F., N.V. Long (2012) Envy and Inequality. Scandinavian Journal of Economics 114, 949–973.

Andreoni, J., B. Erand, J. Feinstein (1998). Tax Compliance. Journal of Economic Literature 36, 818–860.

Andreoni, J., A.A. Payne (2013). Charitable Giving. In Handbook of Public Economics, vol. 5, edited by A. Auerbach, R. Chetty, M. Feldstein, and E. Saez, Elsevier B.V. pp. 1–50.

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

Aronsson, T., O. Johansson-Stenman (2010) Positional Concerns in an OLG Model: Optimal Labor and Capital Income Taxation. International Economic Review 51, 1071–1095.

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

Aronsson, T., O. Johansson-Stenman (2018) Paternalism Against Veblen: Optimal Taxation and Non-Respected Preferences for Social Comparisons. American Economic Journal: Economic Policy 10, 39-76.

Aronsson, T., O. Johansson-Stenman (2020) Optimal Second-Best Taxation When Individuals Have Social Preferences. Working Paper: Umeå Economic Studies no 973.

Bolton, G.E., A. Ockenfels (2000) ERC: A Theory of Equity, Reciprocity and Competition. American Economic Review 90, 166–193.

Boskin, M.J., E. Sheshinski (1978) Individual Welfare Depends upon Relative Income. Quarterly Journal of Economics 92, 589–601.

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

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

Corneo, G., O. Jeanne (1997) Conspicuous Consumption, Snobbism and Conformism. Journal of Public Economics 66, 55–71.

Dupor, B., W.F. Liu (2003) Jealousy and Overconsumption. American Economic Review 93, 423–428.

Eckerstorfer, P., R. Wendner (2013) Asymmetric and Non-atmospheric Consumption Externalities, and Efficient Consumption Taxation. Journal of Public Economics 106, 42–56.

Fehr, E., K. Schmid (1999) A Theory of Fairness, Competition, and Cooperation. Quarterly Journal of Economics 114, 817–868.

Fong, C. (2001) Social Preferences, Self-Interest, and the Demand for Redistribution. Journal of Public Economics 82: 225–246.

Frank, R.H. (1985a) The Demand for Unobservable and Other Nonpositional Goods. American Economic Review 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. American Economic Review 95, 137–141.

Frank, R.H. (2008) Should Public Policy Respond to Positional Externalities? Journal of Public Economics 92, 1777–1786.

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

Kanbur, R., M. Tuomala (2014) Relativity, Inequality, and Optimal Nonlinear Income Taxation. International Economic Review 54, 1199–1217.

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

Ljungqvist, L., H. Uhlig (2000) Tax Policy and Aggregate Demand Management Under Catching Up with the Joneses. American Economic Review 90, 356–366.

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

Nyborg Støstad, M. (2019) Inequality as an Externality: Existence and Consequences. Master's thesis at Paris School of Economics.

Nyborg Støstad, M., F. Cowell (2020) Inequality as an Externality: Consequences for Tax design. Mimeo.

Oswald, A. (1983) Altruism, Jealousy and the Theory of Optimal Non-Linear Taxation. Journal of Public Economics 20, 77–87.

Persson, M. (1995) Why are Taxes so High in Egalitarian Societies? Scandinavian Journal of Economics 97, 569–580.

Pommerehne, W.W., H. Weck-Hannemann (1996) Tax Rates, Tax Administration and Income Tax Evasion in Switzerland. Public Choice 88, 161–170.

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. Journal of Public Economics 94, 1093–1107.

Wendner, R. (2014) Ramsey, Pigou, Heterogeneous Agents, and Non-atmospheric Consumption Externalities. Journal of Public Economic Theory 16, 491–521.