

# Optimal Tax Theory with Endogenous Social Marginal Welfare Weights\*

Emmanuel Saez, UC Berkeley and NBER

Stefanie Stantcheva, MIT

November 2012

## Abstract

This paper proposes a generalized theory of optimal taxation using the tax reform approach and endogenous social marginal welfare weights. A tax system is optimal if no budget neutral marginal reform can increase a weighted sum of (money metric) gains and losses across individuals. However, the social marginal welfare weights used for aggregating gains and losses are not derived from a standard social welfare function based on individual utilities but instead directly specified to reflect society's views for justice. Hence, our optimal tax formulas are no longer obtained by maximizing an objective function. Nevertheless, optimum tax formulas take the same form as standard utilitarian tax formulas by simply substituting standard marginal social welfare weights by those *endogenous marginal social welfare weights*. Hence our theory nests standard theory and is equally tractable. Through a series of examples, we show how the use of suitable endogenous social welfare weights can help resolve most of the puzzles of the traditional welfarist approach and account for existing tax policy debates and structures while retaining Pareto constrained efficiency. In particular, endogenous welfare weights can provide a rich theory of optimal taxation even absent any behavioral responses. Endogenous welfare weights can be derived from social justice principles, leading to a normative theory of taxation or can be derived from estimating actual social preferences of the public, leading to a positive theory of taxation. Our theory brings back social preferences as a critical element for optimal tax theory analysis.

---

\*Stefanie Stantcheva, MIT, stefanie@mit.edu; Emmanuel Saez, University of California, Department of Economics, 530 Evans Hall #3880, Berkeley, CA 94720, saez@econ.berkeley.edu. We thank Thomas Piketty, Louis Kaplow and Matthew Weinzierl for useful discussions and comments. We acknowledge financial support from the Center for Equitable Growth at UC Berkeley.

# 1 Introduction

The dominant approach in optimal tax theory is to use the standard *welfarist* framework in which the government sets taxes and transfers to maximize a social welfare function which is an explicit function of individual utilities. Social welfare is maximized subject to a government budget constraint and taking into account how individuals respond to taxes and transfers. The theory derives optimal tax formulas expressed as a function of the size of the behavioral responses to taxation and the value of redistribution. Behavioral responses to taxation are typically measured using elasticities such as the elasticity of pre-tax income with respect to tax rates and there is a wide empirical literature estimating them. The benefit of redistribution is measured by the variation of marginal social welfare weights across individuals. The marginal social welfare weight on a given individual captures the value that society puts on providing an additional dollar of consumption to this individual. With standard concave individual utilities, those marginal welfare weights decrease in after-tax income under the *utilitarian* criterion<sup>1</sup> so that redistributing from high to low incomes is socially valued.

Such optimal tax formulas do indeed capture the key equity-efficiency trade-off that is at the core of the public debate on the fair distribution of the tax burden and the proper level of means-tested transfers. Right-wing oriented citizens sometimes oppose redistributive taxation either because they view as unfair that high earners necessarily carry lower social marginal welfare weights than low earners, or because they believe that redistribution discourages economic activity. More left-wing oriented people typically support progressive taxation both because they value redistribution and do not consider the efficiency costs to be too high. However, the standard utilitarian optimal tax approach also generates predictions for optimal taxes and transfers that are clearly at odds with actual tax systems.

First, if individuals do not respond to taxes, i.e., if pre-tax incomes are fixed, and individual utilities are concave, then utilitarianism recommends a 100% tax and full redistribution, a point originally made by Edgeworth (1897). In reality, even absent behavioral responses, many and perhaps even most people would still object to such levels of taxes on the grounds that it is unfair to fully confiscate individual incomes. Therefore, even absent behavioral responses, optimal taxation is probably not as trivial as standard utilitarian theory would suggest.

---

<sup>1</sup>The utilitarian criterion is the sum of individual utilities. It is the most commonly used one in the welfarist approach.

Second, views on taxes and redistribution seem largely shaped by views on whether the income generating process is fair and whether individual incomes are deserved or not. The public tends to dislike the redistribution of fairly earned income but is in favor of redistributing income earned unfairly or due to pure luck. Such distinctions are irrelevant for utilitarianism.<sup>2</sup>

Third and related, society assesses the value of transfers or the costs of taxes not only based on the actual economic situation of a given individual but also based on what this individual would have done absent that transfer or tax. For example, most people would value transfers to those unable to work (the so called “deserving poor”) while fewer people value transfers to those who would stop working because of transfers themselves (the “undeserving free loaders”) (Will 1993, Larsen 2008, Jeene, van Oorschot, and Uunk, 2011).

Fourth, under utilitarianism, optimal taxes should depend not only on income but also on all other observable characteristics which are correlated with intrinsic earning ability. Many such characteristics exist and are readily observable. Examples are height, gender, or race. Yet society seems highly reluctant to make taxes depend on such “tags”.<sup>3</sup> Intuitively, the public would find it unfair to tax differently people with the same ability to pay. This is known as the horizontal equity concern and it cannot be easily reconciled with the welfarist approach.

Moreover, society seems to inherently value taxpayers relative to those receiving transfers, even beyond what would be captured by a person’s net income. In addition, when assessing reforms, losers tend to carry more weight than winners, and it seems harder in practice to raise taxes than to lower them. Finally, the policy debate often revolves around specific but highly visible statistics such as the poverty rate so that there is a particular focus on poverty alleviation programs.

In this paper, we propose an alternative approach that nests the standard approach. Our approach does not maximize a specific social objective, yet allows to retain standard individual utilities, and is flexible enough to resolve most discrepancies between optimal tax theory and practice. It is motivated by the fact that governments certainly do not explicitly posit and maximize a social welfare function based on individual utilities. Actual tax policy debates

---

<sup>2</sup>Alternative theories of social justice have put emphasis on notions of compensation (individuals should be compensated for outcomes they are not responsible for) and responsibility (individuals should not be compensated for outcomes due to their choices). See in particular Roemer (1998), Fleurbaey and Maniquet (2006, 2007) and the comprehensive treatment of responsibility in Fleurbaey (2008).

<sup>3</sup>Akerlof (1978) pointed out that the use of such tags can help the government improve the tax system under a standard utilitarian social welfare criterion. More recently, Mankiw and Weinzierl (2010), using the case of height, point out that this reveals a deep weakness of standard welfarist optimal tax theory.

tend to focus instead on specific tax reforms, starting from a given situation, considering who the winners and losers are, and the broader consequences of the reform on economic activity and tax revenue. As is well known, the standard optimal tax approach can be recast in such tax reform terms, since the optimal tax system is precisely the one around which no further reform is desirable. Indeed, the first order conditions from maximizing a standard social welfare function always imply that a small budget neutral tax reform around the optimum has no overall impact on social welfare. In this tax reform approach, the gains and losses for each individual are aggregated using the social marginal welfare weights. The standard approach imposes a specific structure on those weights based on the individual utility functions and the social welfare function. Our approach replaces those standard weights by alternative *endogenous social marginal welfare weights* that instead directly reflect society’s views for justice. They are not necessarily derived from the underlying individual utility functions or the social welfare function. Our optimum tax formulas hence take the same form as standard welfarist tax formulas, but with the new *endogenous social welfare weights* replacing standard social welfare weights. Therefore, our theory nests the standard theory and remains equally tractable. The optimal taxes we obtain can be seen as an equilibrium around which no marginal reform is desirable given the endogenous social marginal welfare weights. Because the latter are always non-negative, our optimal taxes retain the key Pareto efficiency property.<sup>4</sup> At the optimum, there is no balanced budget reform that the government could undertake and that could increase everybody’s welfare, given its informational constraints.

Through a series of examples, we show how the use of suitable endogenous social welfare weights can resolve most of the puzzles of the traditional utilitarian approach and account for existing tax policy debates and structures. First, we show that making social endogenous weights depend negatively on net taxes paid—in addition to net disposable income—eliminates the 100% tax result of Edgeworth and generates a non-trivial optimal tax theory even absent behavioral responses. In particular, this can be used to determine optimal family taxation (i.e., the treatment of couples and children) based on family equivalence scales. Second, we show that our theory allows to introduce an asymmetry in how losses vs. gains from tax reforms are valued by society. If gains from a tax reform are socially less valued than losses, perhaps because losers

---

<sup>4</sup>Note however that due to the local nature of those weights, we can only guarantee local Pareto efficiency, that is, Pareto efficiency relative to small reforms around the status quo. The evaluation of large reforms requires non-marginal welfare weights and is left for a follow-up project.

tend to complain more loudly than winners rejoice, there can be a wide range of tax equilibria rather than a single tax equilibrium. Third, our theory allows to make social weights depend on what individuals would have done absent taxes and transfers. Hence, we can capture the idea that society dislikes transfers toward free loaders who would work absent means-tested transfers. Fourth, endogenous social welfare weights can also capture the fact that society prefers to tax income due to luck rather than income earned through hard work, and hence can capture the principles of compensation and responsibility that have been developed by Roemer (1998) and Fleurbaey (2008) in alternative normative theories. Fifth, we show that endogenous social welfare weights can capture horizontal equity concerns as well. A reasonable criterion is that introducing horizontal inequities is acceptable only if it benefits the group discriminated against. This dramatically limits the scope for using non-income based tags. Finally, we show how endogenous social welfare weights can be specified to capture poverty alleviation objectives.

In the examples we present, the endogenous welfare weights might appear ad-hoc and specified exactly so as to explain the puzzle at hand. However, we see this flexibility of our approach as a virtue that opens two new avenues of investigation. First, endogenous welfare weights can be derived from social justice principles, leading to a normative theory of taxation. The most famous example is the Rawlsian theory which can also be seen as a particular case of our theory where the endogenous social marginal welfare weights are concentrated solely on the most disadvantaged members of society. We show that a “locally Rawlsian” theory that endogenously and systematically divides the population into those deserving of support vs. not based on their situation and the tax/transfer system can generate a rich and realistic set of normative criteria. Second, endogenous welfare weights could also be derived empirically, by estimating actual social preferences of the public, leading to a positive theory of taxation. There is indeed a small body of work trying to uncover perceptions of the public about various tax policies. Those approaches either start from the existing tax and transfers system and reverse-engineer it to obtain the underlying social preferences (Christiansen and Jansen 1978, Bourguignon and Spadaro 2012) or directly elicit preferences on various social issues in surveys.<sup>5</sup>

Naturally, many previous studies have proposed alternatives to the standard welfarist ap-

---

<sup>5</sup>See Frohlich and Oppenheimer (1992), Cowell and Shoekkart (2001), Fong (2001), Devooght and Shoekkart (2003), Engelmann and Strobel (2004), Ackert, Martinez-Vazquez, and Rider (2007), Kuziemko, Norton, Saez, and Stantcheva, (2012). Note also that the focus on tax reform and on (local) marginal welfare weights might make it much easier to elicit social preferences than if trying to calibrate a global objective function.

proach (see Kaplow (2008), Chapter 15, Fleurbaey and Maniquet (2011), and Piketty and Saez (2013), Section 6 for surveys). The next section reviews those previous papers in detail, highlighting the differences with our. Here, we just briefly emphasize that the most common alternative approach is to define some modified social objective but to keep the maximization of an objective principle intact. There have indeed been several attempts along these lines in the literature (Auerbach and Hassett (2002) for horizontal equity, Besley and Coate (1992) for poverty rate focus, Fleurbaey and Maniquet (2006, 2007) for responsibility, and Weinzierl (2012) for libertarianism). While such studies do succeed to some extent in extending social preferences over and above pure utilitarianism, this approach faces limitations. In fact, it greatly restricts the number of phenomena that can be captured, because it is necessary to retain individualistic social welfare functions, i.e., a welfarist objective, (as defined by Kaplow and Shavell, (2001)) if one does not want to conflict with the Pareto principle.<sup>6</sup> Indeed, Kaplow and Shavell (2001) show that any non individualistic social function, that includes non-welfarist elements, necessarily leads to a violation of the Pareto principle in some circumstances.

Our paper is organized as follows. Section 2 describes in more detail the related literature and our contribution. Section 3 presents the new framework in a simplified setting with no behavioral responses to taxation. Section 3 considers the general case with behavioral responses and builds examples dealing with several specific puzzles observed in real tax policy. Section 4 considers various extensions. Section 5 concludes.

## 2 Related Literature

There have been many papers trying to move beyond utilitarianism and provide a more satisfactory treatment of social preferences for redistribution. Mirrlees himself (Mirrlees (1974)) highlighted a problem with the utilitarian approach. In the first best, high-skilled agents “envy” low-skilled agents because they are forced to work more for the same disposable income, a result extended to the case with heterogeneous preferences by Choné and Laroque (2005). Indeed, apart from inequality aversion, captured by the concavity of the social welfare function, there are no other fairness requirements or concepts of justice embodied in the utilitarian welfare

---

<sup>6</sup>An individualistic social welfare function is a function of individual utilities only, and does not depend on other than utilities. However, many notions of fairness cannot be accounted for through an individualistic welfare function. For example the aforementioned ‘free loaders’ concept, which fundamentally depends on what people *would have done* in a different tax system, or a reasonable notion of horizontal equity.

function. The papers presented here have taken one of four possible avenues, other than ours, to remedy the shortcomings of the utilitarian framework.

First, one could focus on Pareto efficient taxation in order to find properties of optimal tax systems that are true for any social welfare function. This approach, adopted by Werning (2007) and Stiglitz (1987), remains very agnostic and unfortunately does not yield many practical policy recommendations. Our endogenous weights are always non-negative, hence guaranteeing Pareto efficient taxation, but we impose more structure on them depending on social preferences we try to capture.

Second, one can try to directly augment the social welfare function to include other concerns society might have. This is the approach adopted in a recent paper by Weinzierl (2012) who argues that people, and hence society, exhibit ‘normative diversity’, that is, simultaneously use several normative concepts to judge policies. In addition to the utilitarian concept, he focuses on the libertarian “Equal Sacrifice” principle and considers as a social objective the minimization of a weighted sum of the utilitarian and libertarian functions. This allows him to explain why tagging is limited in the real world and why tags more strongly correlated with earning potential are used more. Our approach nests his specific normative reasoning with the weights set to equal to the derivatives of his social welfare function. Auerbach and Hassett (2002) consider a welfare function ‘à la Atkinson’ (1970) in which the arguments are the net consumptions of agents of different types, and there is aversion to after-tax income inequality, both across agents of the same ‘type’ (horizontal equity) and across agents of different income levels (vertical equity). In general, however, augmenting the social criterion with other considerations of fairness (or with any consideration other than individual welfare) can lead to Pareto inefficient outcomes in some circumstances, which greatly restricts the generalizability of this approach. Indeed, as already mentioned in the introduction, and as shown by Kaplow and Shavell (2001) and Kaplow (2001), any non-welfarist social welfare function necessarily leads to a violation of the Pareto principle in some cases.<sup>7</sup> In that sense, we are nesting this approach, which consists in directly augmenting the social welfare function by extra-individualistic elements, only for the cases in which it generates Pareto efficient outcomes.<sup>8</sup>

---

<sup>7</sup>For example, but not only, when one introduces uncertainty (Kaplow (2011)). As explained in Kaplow (2001), in Auerbach and Hassett (2002), the welfare function depends on income differences between individuals derived using income levels in a reference distribution, and not just on individual welfare.

<sup>8</sup>Conversely and as we explain in the next section, while it is possible to obtain our outcomes within the traditional approach *ex-post* by setting the traditional Pareto weights equal to our endogenous welfare weights,

Third, one can abandon the maximization of a social objective directly based on welfare and instead specify another objective, derived from specific normative concepts and core principles.<sup>9</sup> Fleurbaey and Maniquet use this axiomatic approach in a series of studies (Fleurbaey and Maniquet 2006, 2007, 2011). In Fleurbaey and Maniquet (2006), they consider the tradeoff between fairness and responsibility when agents differ both in their intrinsic earnings potential and their preferences for work. The two fundamental principles posited are the Pigou–Dalton principle (that transfers reducing income inequalities are acceptable, if performed between agents with the same preferences and labor supply), and the principle that “Laissez-faire” should be the social optimum if all agents had the same skill. Hence, unlike the utilitarian criterion, inequality in outcomes is acceptable if driven by different preferences for work. This leads to a measure of individual well-being in terms of an ‘equivalent wage’, the hypothetical wage rate that would allow an individual to reach her indifference curve at that current allocation if she could choose labor supply freely. The appropriate social criterion is then to minimize the maximal average tax rate paid by those with the lowest equivalent wage. The optimum involves the ‘hardworking poor’ receiving the greatest subsidy. In the last section in particular, we show how our endogenous welfare weights can capture such a type of maximin social preferences. The three main advantages of their approach are that it does not require cardinal, inter-personal utility comparisons, that it preserves Pareto efficiency, and that the axioms postulated for social preferences are the foundation upon which the objective to be maximized is built. The difficulty is that many axioms might not always give rise to a tractable objective and that the objective needs to be re-derived from scratch for any new set of axioms considered. In our view, our approach seems simpler and more transparent. As we discuss, we think that a combination of Fleurbaey and Maniquet’s axiomatic approach and our endogenous welfare weights could lead to very interesting avenues for future research, if the axioms on social preferences and the principles of fairness were imposed directly on the social marginal welfare weights.

Besley and Coate (1992), as well as Kanbur, Keen, and Tuomala (1994), start from the fairness principle that everyone should be entitled to a minimal level of consumption and hence adopt as a criterion the poverty rate, that is the number of people living below the poverty line. An issue is that their objective does not guarantee Pareto efficiency in all cases, as people could

---

it is not possible to choose the Pareto weights *ex-ante*, without knowing the tax system that arises in equilibrium.

<sup>9</sup>The two examples chosen here illustrate that Pareto efficiency may or may not be one of those driving principles.

be forced to work despite prohibitively high disutility of labor, to push them above the poverty line. In our framework, we can also address poverty alleviation programs (see Section 5).

The fourth method, which always guarantees that the Pareto principle holds, is to include the alternative, non-standard considerations directly into the *individual* utility functions and keep a utilitarian social welfare function. For example, Alesina and Angeletos (2005) introduce a disutility term stemming from the amount of ‘unfair’ income in the economy. This however has two drawbacks. Firstly, it is not clear that social preferences are equivalent to (nor even respectful of) individual preferences. Secondly, it could lead to non-standard individual behaviors, if individual choice and social preference are non-separable in the utility function. In our framework, we remain completely agnostic about individual utilities, which can either be standard or incorporate any behavioral considerations one wishes.

### 3 Optimal Tax Theory with Fixed Incomes

We start with the simple case in which pre-tax incomes are completely inelastic to taxes and transfers. This puts the focus solely on the redistributive issues, and is useful as an introduction to our approach, especially as contrasted with the standard welfarist approach. We start by briefly reviewing the standard utilitarian setting.

#### 3.1 Standard Utilitarian Approach

Consider an economy with a population normalized to one, an exogenous pre-tax earnings distribution  $H(z)$ , and a homogenous utility function  $u(c)$  increasing and concave in disposable income  $c$ . Disposable income is equal to pre-tax earnings minus taxes so that  $c = z - T(z)$ . Assume that the government chooses the tax function  $T(z)$  to maximize the utilitarian social welfare function:

$$SWF = \int_0^{\infty} u(z - T(z))dH(z) \quad \text{subject to} \quad \int T(z)dH(z) \geq 0 \quad (p),$$

where  $p$  is the Lagrange multiplier of the government budget constraint. As incomes  $z$  are fixed, a point-wise maximization with respect to  $T(z)$  yields:

$$u'(z - T(z)) = p \quad \Rightarrow \quad c = z - T(z) = \text{constant across } z.$$

Hence, utilitarianism with inelastic earnings and concave individual utility functions leads to complete redistribution of incomes. The government confiscates 100% of earnings and redistributes income equally across individuals (Edgeworth, 1897).

Let us denote by  $g_i = u'(c_i)/p$  the social marginal welfare weight on individual  $i$  with consumption  $c_i$ .  $g_i$  measures the monetary value that society puts on an extra dollar of consumption for individual  $i$ . The optimum is such that all marginal welfare weights are set equal to one.

This simple case highlights three of the drawbacks of utilitarianism. First, complete redistribution seems too strong a result. In reality, even absent behavioral responses, many and perhaps even most people would still object to 100% taxation on the grounds that it is unfair to fully confiscate individual incomes. Second, the outcome is extremely sensitive to the specification of individual utilities, as linear utility calls for no taxes at all, while introducing just a bit of concavity leads to complete redistribution. Third, and as is well known, the utilitarian approach cannot handle well heterogeneity in individual utility functions. With heterogeneous utilities  $u_i(c)$ , the optimum is such that  $u'_i(c_i)/p = 1$  for all  $i$ . Hence, consumption is no longer necessary equal across individuals and is higher for individuals more able to enjoy consumption. This issue is known as the problem of inter-personal utility comparisons. In reality, society would be reluctant to redistribute based on preferences.<sup>10</sup>

### 3.2 Endogenous Welfare Weights

A simple way to generalize the utilitarian approach is as follows. Instead of assuming that  $g_i = u'(c_i)/p$ , we can write directly  $g_i = g(c_i, T_i)$  as a function of disposable income  $c_i$  as well as net taxes  $T_i = T(z_i)$  paid by individual  $i$ .

It is natural to assume that  $g(c, T)$  decreases in  $c$  to reflect the fact that society values additional consumption less (and hence accepts additional taxes more readily) for those with more disposable income, as under utilitarianism with a concave utility of consumption. This captures the old notion of “ability to pay”. However, we can also assume that  $g(c, T)$  increases with  $T$  as taxpayers contribute more to society’s well being and are hence more deserving of additional consumption. Another interpretation is that individuals are in principle entitled to their income and hence become more deserving as the government taxes away their income.

---

<sup>10</sup>Redistribution based on marginal utility is socially acceptable if there are objective reasons a person might need more disposable income, such as having a medical condition requiring high expenses, or a large family with many dependents.

Conversely, those receiving a net subsidy from the government are perceived to be less deserving as they are debtors to society. The utilitarian case is a polar case in which  $g$  depends solely on  $c$ . The alternative polar case in which  $g$  depends solely on  $T$  would reflect the libertarian view according to which the level of one's disposable income is irrelevant and only the tax contribution matters for how socially deserving an individual is.<sup>11</sup>

Note that  $g(c, T)$  is only defined up to a multiplicative constant. To simplify the presentation in some cases, we will assume that the latter is chosen so that the average  $g(c, T)$  across the population is equal to one. Note also that we could have equivalently specified  $g$  as a function of  $c$  and  $z$  (instead of  $c$  and  $T$  as  $c = z - T$ ).

The optimal tax system is such that no reform can increase social welfare at the margin, when the value of transfers is measured using the  $g$  weights. With no behavioral responses, the optimal rule is very simple: the social welfare weights  $g(z - T(z), T(z))$  should be constant across all income levels  $z$ . To see this, suppose by contradiction that  $g(z_1 - T(z_1), T(z_1)) > g(z_2 - T(z_2), T(z_2))$ . Then transferring a dollar from those earning  $z_2$  toward those earning  $z_1$  (by adjusting  $T(z_1)$  and  $T(z_2)$  correspondingly and in a budget balanced manner) would be desirable. Hence, the optimal tax schedule is characterized by:

$$g_c \cdot (1 - T'(z)) + g_T \cdot T'(z) = 0 \quad \text{so that} \quad T'(z) = \frac{-g_c}{g_T - g_c} = \frac{1}{1 - g_T/g_c} \quad (1)$$

where  $g_c$  and  $g_T$  denote the partial derivatives of  $g$  with respect to its first and second argument respectively. Note that  $0 \leq T'(z) \leq 1$  as  $g_c \leq 0$  and  $g_T \geq 0$ .

The standard utilitarian case, with  $g(c, T) = g(c)$  implies that  $T'(z) \equiv 1$  and the libertarian case with  $g(c, T) = g(T)$  implies that  $T'(z) \equiv 0$ .

The specification  $g(c, T) = g(c - \alpha T) = g(z - (1 + \alpha)T(z))$  where  $\alpha$  is a constant parameter delivers an optimal linear tax rate with  $T'(z) = 1/(1 + \alpha)$ . In that case, paying one extra \$1 in taxes and consuming  $\alpha$  more dollars leaves the person equally deserving. This means that, when earning an extra dollar, each person should be entitled to keep  $\alpha/(1 + \alpha)$  for consumption and pay an extra  $1/(1 + \alpha)$  in taxes. This case intuitively captures the preferences of a society which finds everybody equally deserving at the margin when they are contributing a fraction  $1/(1 + \alpha)$  of their incomes to taxes used to fund a uniform demogrant.

The optimal tax has an increasing marginal tax rate if  $-g_T/g_c$  decreases with income, i.e.,

---

<sup>11</sup>We assume away government funded public goods in our set-up for simplicity.

if society feels that a higher income person should be entitled to keep a smaller fraction of her income than a lower income person.

**Actual empirical test.** Such preferences for redistribution could be elicited through a survey by asking simple questions such as: Who is most deserving of a tax break? Person  $i$  earning  $z_i$  and paying  $T_i$  in taxes and having  $c_i = z_i - T_i$  as disposable income vs. person  $j$  earning  $z_j$  and paying  $T_j$  in taxes and having  $c_j = z_j - T_j$  as disposable income. If the public has utilitarian preferences, then desert should depend solely on the relative ranks of  $c_i$  and  $c_j$ . Beyond testing for purely utilitarian preferences, well-designed questions could in principle allow us to recover  $\alpha$ . In the conclusion, we present a more detailed description of an empirical test and application of our theory.

**Link with the standard Pareto frontier approach.** Instead of maximizing a given utilitarian social welfare function, another common approach is to derive the full Pareto frontier, by considering as an objective a weighted sum of utilities, where the weights are exogenous, nonnegative, and allowed to vary in an arbitrary fashion across individuals. Denoting by  $\omega_z \geq 0$  the weight on an individual with income  $z$ , the government chooses  $T(z)$  to maximize:

$$SWF((\omega_z)_z) = \int_0^\infty \omega_z u(z - T(z)) dH(z) \quad \text{subject to} \quad \int_0^\infty T(z) dH(z) \geq 0 \quad (p),$$

This program leads to the first order condition  $\omega_z u'(z - T(z))/p \equiv 1$ . By varying the exogenous weights  $(\omega_z)$ , one can recover all possible Pareto optima. However, it is not possible to specify the  $(\omega_z)$  that deliver the same outcome as the endogenous weights  $g(c, T)$  without first solving the optimal tax using the  $g(c, T)$  endogenous weights. Furthermore, as we illustrate in examples below, the Pareto weights  $(\omega_z)$  cannot depend directly or indirectly on tax policy, but instead only on parameters or characteristics exogenous to tax policy. Hence, specifying Pareto weights cannot be a substitute to our approach with endogenous welfare weights.

Finally, it is not possible to obtain our optimum by considering a generalized social welfare function of the form

$$SWF = \int_0^\infty G(u(z - T(z)), T(z)) dH(z),$$

that directly incorporates  $T(z)$  in the social objective, over and beyond its effect on individual utility  $z - T(z)$ . Indeed, such an approach can lead to outcomes that can be Pareto dominated, as a simple thought-experiment can highlight. Suppose for example that  $G(u, T)$  increases with

taxes  $T$  (see the next subsection for an example of when this can realistically happen). In that case, increasing  $T$  has value per-se, independent of its effect on the individual disposable income. Hence, it could be desirable to raise taxes to burn money, i.e., have a government budget constraint that is slack and hence a situation that is Pareto dominated. Our approach using endogenous non-negative marginal social welfare weights ensures that the optimum is always Pareto efficient.

### 3.3 Extensions

**Luck income vs. deserved income: Endogenous desert criterion.** An important belief society seems to hold is that it is fairer to tax income due to ‘luck’ than income earned through hard effort, and conversely, that it is fairer to insure against losses in income beyond an individual’s control (see e.g., Fong, 2001, as well as Devooght and Shoekkart 2003 for how the notion of control over one’s income is crucial to identify what is deserved income and Cowell and Shoekkart 2001 for how perceptions of risk and luck inform redistributive preferences). Our framework can allow in a tractable way for such social preferences, which differentiate income streams according to their source. These preferences can also provide a micro-foundation for endogenous social welfare weights  $g(c, T)$  which are increasing in taxes  $T$ , as presented above.

Let us consider two sources of income  $z = w + y$  where  $w$  is deserved income (due to one’s own effort) and  $y$  is luck income (due to one’s luck). Suppose first that the government can observe separately  $y$  and  $w$  and impose separate taxes  $T_y(y)$  and  $T_w(w)$ . Then naturally,  $g(c, T_w)$  is independent of  $T_y$  because it is perceived as fair to tax lucky income. The optimum is to naturally confiscate 100% of luck income  $y$  and to tax  $w$  according to the standard rule.

Suppose next that the government cannot observe  $y$  and  $w$  separately and can only observe total income  $z$  so that taxes have to be based on total income  $T(z)$ . Consider a society with sharp preferences for redistribution which considers that, ideally, all luck income should be fully redistributed, but that, by contrast, individuals are fully entitled to their deserved income. Let us denote by  $Ey$  average luck income in the economy. These social preferences can be captured by the following binary set of weights. A person is seen as deserving and has a weight of one if  $c = z - T(z) \leq w + Ey$ , i.e., disposable income is less than deserved income plus average luck income. Conversely, a person is seen as non-deserving and has a social weight of zero if  $z - T(z) > w + Ey$ . The average weight, i.e., the fraction of deserving individuals, at income

level  $z$  for a given tax system (such that  $z - T(z) = c$ ) is then given by:

$$\pi(c, z) = \int_{w+y=z} 1(c \leq w + Ey) f(w, y | w + y = z) dw dy.$$

This function is naturally decreasing in disposable income  $c$  (keeping total income  $z$  constant). Equivalently as  $c = z - T$ ,  $\pi$  is increasing in  $T$  (keeping  $z$  constant). Increasing the tax on those with income  $z$  makes them more deserving on average. Hence, this model can provide a micro-foundation for the endogenous weights  $g(c, T) = \pi(c, c + T)$ . Note importantly that, despite the absence of behavioral effects here, the social weights are endogenous to the tax system.

As above, the optimal tax system  $T(z)$  is such that the average weight should be the same, across all income levels  $z$ . Hence, the presence of both (indistinguishable) deserved income and luck income is enough to generate a non-trivial theory of optimal taxation, even in the absence of behavioral responses. Beliefs about what constitutes luck income vs. deserved income will naturally play a large role in the level of optimal redistribution with two polar cases. If all income is deserved, as Libertarians believe in a well-functioning free market economy, the optimal tax is zero. Conversely, if all income were due to luck, the optimal tax is 100% redistribution. If social beliefs are such that high incomes are primarily due to luck while lower incomes are deserved, then the optimal tax system will be progressive.

This simple model already captures both the concept of compensation (i.e., individuals should be compensated for outcomes such as luck income for which they are not responsible for) and the principle of responsibility (i.e., individuals should not be compensated for outcomes due to their merit or effort such as deserved income).

**Taxpayers vs. transfer receivers.** In practice, taxpayers—those for whom  $T(z_i) \geq 0$ —are perceived as more deserving than benefits recipients for whom  $T(z_i) < 0$ .<sup>12</sup> This can be captured by having  $g(c, T)$  change discontinuously in  $T$  at  $T = 0$ . The optimum will then have a range of income levels for which  $T(z) = 0$ , that is a transfer program at the very bottom up to  $z_1$ , and an income tax above an exemption  $z_2$  with  $z_2 > z_1$  and  $T(z) = 0$  in the range  $(z_1, z_2)$ . This fits with current practice where income tax exemptions for bottom earners are common and more readily accepted than direct transfers.

**Family taxation.** Our model can be used to discuss optimal family taxation, i.e., the treatment

---

<sup>12</sup>Besley and Coate (1992) study the stigmas of being a net welfare recipient and taxpayer 'resentment' towards those who do not pay taxes. Reutter et. al. (2009) document this empirically.

of couples and children, in a simple yet realistic way. Suppose  $g(c, T)$  is the profile of weights for singles with no children. At the optimum, as we have seen,  $T(z)$  is such that  $g(z - T(z), T(z)) \equiv g$  is constant with  $z$  across all single individuals.

Consider now couples. If the couple has earnings  $z_1, z_2$ , fully shares consumption, with no economies of scale in consumption for couples relative to singles, and pays total tax  $T$  then  $c_1 = c_2 = (z_1 + z_2 - T)/2$ . If social welfare weights only depend on consumption and the tax system (that is, society does not put intrinsic value on living as a couple as opposed to living single), the social marginal welfare weight for each member of the couple would naturally be  $g((z_1 + z_2 - T)/2, T/2)$ . At the optimum, those welfare weights should be equal to the (common) welfare weight  $g$  on singles and in particular on singles earning exactly the average income of the couple,  $(z_1 + z_2)/2$ , implying that  $g((z_1 + z_2 - T)/2, T/2) = g((z_1 + z_2)/2 - T((z_1 + z_2)/2), T((z_1 + z_2)/2))$ . This immediately implies that  $T(z_1 + z_2)/2 = T((z_1 + z_2)/2)$ . i.e., there should be perfect splitting of earnings, and then split earnings should be taxed according to the standard single schedule.<sup>13</sup>

Suppose now that couples do not split their incomes evenly. If couples do not share their incomes at all (and there are no economies of scale), then each member of the couple should be treated as a single individual.<sup>14</sup> Many countries do use such individualized tax systems. Hence, even if we abstract from considerations related to the intrinsic value of living as a couple versus living as a single, our analysis is consistent with the general observation that those who believe that families fully share economic resources tend to support family-based taxation while those who believe that family members act more independently tend to support individual taxation.

Suppose now that there are economies of scale in consumption so that if the couple has a total disposable income of  $c$ , the per-person consumption equivalent is  $c_1 = c_2 = (1 + \delta)c/2$  with  $0 \leq \delta \leq 1$ , measuring the strength of economies of scales. In that case, the optimal tax  $T$  on the couple, which satisfies neutrality with respect to family choice, should be such that  $g((1 + \delta)(z_1 + z_2 - T)/2, T/2) = g$ . This implies that there is a single income equivalent  $z$  such that  $T(z) = T/2$  and  $z - T(z) = (1 + \delta)(z_1 + z_2 - T)/2$ , i.e., the single equivalent income  $z$  is such that  $(z_1 + z_2)/2 = (z + \delta T(z))/(1 + \delta)$ . Once  $z$  is found, each member of the couple

<sup>13</sup>This is the basic model of the French income tax system.

<sup>14</sup>If there is imperfect sharing, e.g., couples only share a fraction  $\lambda$  of their income, then the optimal system is to assume that member  $i$  earns  $z'_i = \lambda(z_1 + z_2)/2 + (1 - \lambda)z_i$  and tax each member of the couple as if he/she were earning  $z'_i$ .

should pay  $T(z)$  in taxes. In the case of a linear income tax at rate  $\tau$  with demogrant  $R$ , the optimal tax on couples should also be linear at rate  $\tau \cdot \frac{1+\delta}{1+\delta\tau} > \tau$  with demogrant (per person)  $R/(1 + \delta\tau) < R$ . Effectively, couples are taxed more and receive a smaller demogrant because they benefit from economies of scales. Naturally, economies of scale are likely more important at the low end than at the high end of the income distribution. If  $\delta$  is small for high incomes, then the tax rate on high-earning couples should asymptotically be the same as for high-earning singles.

The analysis of economies of scale can easily be extended to account for children. For simplicity, let us now ignore the distinction between couples and singles and instead focus on whether individuals have children or not. Suppose also that society does not put intrinsic value on children, so that the social welfare weights do not depend directly on whether someone has children or not.<sup>15</sup> Let us denote by  $T(z, n)$  the tax on an individual with income  $z$  and  $n$  children. If having  $n$  children absorbs a fraction  $\delta(n)$  of one's resources ( $\delta(n)$  can be made dependent on the level of resources  $z$  as well), then the optimal neutral tax/transfer on a parent with  $n$  children should be such that  $g((z - T(z, n)) \cdot (1 - \delta(n)), T(z, n)) = g$ . For any pair  $(z, n)$  there is a "no child" equivalent income  $z^n$  such that  $z^n - T(z^n, 0) = (z - T(z, n)) \cdot (1 - \delta(n))$  and  $T(z, n) = T(z^n, 0)$ . Hence, given the optimal tax schedule for childless individuals, we can derive the optimal tax schedule for those with children.

This discussion shows that our conceptual framework can easily be mapped into the equivalence scales used in applied welfare economics to obtain a simple and realistic theory of the taxation of families, even abstracting from behavioral responses. The standard utilitarian model cannot handle the issue of family taxation in a satisfactory way for two reasons (see also Kaplow, 2008, chapter 12 for a detailed review of the utilitarian approach to family taxation). First, without behavioral responses, the optimum always implies full redistribution so that the problem becomes degenerate, unlike in our setup. Second, with behavioral responses, the problem quickly becomes intractable and calls for redistribution across family types due to arguments related to tagging, which violate horizontal equity concerns (see below).

**Tax increases vs. tax decreases.** Often, during actual tax reform discussions, potential losers tend to complain more loudly than potential winners rejoice. To capture such effects,

---

<sup>15</sup>Of course, since having children affects both the resources and consumption available to a family, it enters the social marginal welfare weights indirectly.

we need to specify two distinct sets of social marginal weights,  $g^+(c, T) > g^-(c, T) \geq 0$ , where  $g^+$  is the social marginal weight for tax increases and  $g^-$  is the social marginal weight for tax decreases.<sup>16</sup> There will now be an interval set of equilibria instead of a single equilibrium. Any tax system such that  $g^+(z - T(z), T(z)) \geq g^-(s - T(s), T(s))$  for all  $z, s$  is an equilibrium.

Note that such asymmetric effects can also be used to capture lack of trust in the government whereby the public does not trust that the proposed reform can actually be implemented under as favorable terms as advertised. For example, if the public believes that part of taxes collected and then redistributed are captured or wasted by the government (through inefficient bureaucracies, rewards to special interests or lobby groups, etc.), this would lead the public to discount projected gains for winners and inflate projected losses for losers.

This example also highlights one potential danger, which is the non-existence of an equilibrium for some sets of differential welfare weights, that is welfare weights depending on the nature of the reform itself. Here for example there would be no equilibrium if winners were weighted more heavily than losers (i.e., if  $g^+(c, T) < g^-(c, T)$  for all  $c$  and  $T$ ), since it would always be desirable to have additional transfers.

Conversely, it is also possible for *any* tax system to be an equilibrium if society is sufficiently averse to tax increases. An example is the case in which the weights only depend on  $c$  (that is:  $g^+(c, T) = g^+(c)$  and  $g^-(c, T) = g^-(c)$ ) and for every  $c_1$  and  $c_2$  such that  $c_1 > c_2$ , we have:  $g^+(c_2) > g^+(c_1) > g^-(c_2) > g^-(c_1)$ . Put differently, no matter how little taxes someone is paying, raising taxes is undesirable, even if it is done in order to give it to someone poorer.<sup>17</sup>

**Making weights depend on non-income characteristics: Exogenous desert criterion with an application to equality of opportunity.** One of the advantages of our approach is that we can allow the social marginal welfare weights to depend on any characteristics which are deemed relevant by society—other than disposable income—and to nevertheless preserve Pareto efficiency. This model can be usefully applied whenever individuals differ along several dimensions and society perceives some of these differences as unfair but others as fair. Across characteristics which are deemed to be unfairly distributed (for example, family background), society’s

---

<sup>16</sup>Note that in this situation, the two sets of weights cannot possibly both average to 1. If we normalize the set  $g^-$  such that  $\int g_i^-(c, T) d\nu(i) = 1$ , then the weights  $g^+$  will average to more than 1. This makes sense if we consider that the weights  $g^-$  measure the value of \$1 uniformly distributed, while  $g^+$  measure the value of \$1 uniformly taken. In a society averse to tax increases *per se*, the latter is costlier.

<sup>17</sup>This type of situation can lead to a dead-lock in policy making, where any status quo, even if very suboptimal, can be sustained.

preferences are redistributive (and *locally Rawlsian* in the case of a binary characteristic such as disadvantaged versus advantaged background). On the other hand, across characteristics which are deemed to be fairly distributed (for example, income conditional on family background), society's preferences are libertarian.

Suppose that there is a number  $M$  of distinct groups of people which society can rank in terms of desert or merit, with group 1 being the most valued group and group  $M$  the least valued group. These social preferences are taken as given and may or may not be related to earnings potential, although most reasonable social preferences will likely focus on characteristics related to earnings. As an example, groups could be delimited by family background, with children from advantaged backgrounds being privileged all throughout life (see Roemer et. al. 2003 and the example in the next section). Let  $P_m$  be the proportion of people from group  $m$  in the population. The density of individuals in group  $m$  who earn income  $z$  is given by  $f_m(z)$ , which is independent of the tax system due to the absence of behavioral effects. Within each group  $m$ , we can define the marginal social welfare weights as above, as a function of disposable income and taxes  $g_m(z - T, T)$ . Society's preferences imply that, at any given disposable income and tax level, a transfer to a group deemed more deserving is more valuable than a transfer to a less deserving group, i.e.,  $g_m(z - T, T) \geq g_n(z - T, T)$  for all  $z, T$  and all  $m < n$ . Let the fraction of members from group  $m$  at income level  $z$ , be denoted by  $f(m|z) = f_m(z) P_m / \sum_{j=1}^M f_j(z) P_j$ . The average welfare weight at income level  $z$  is then:

$$g(z - T(z), T(z)) = \sum_{m=1}^M f(m|z) g_m(z - T(z), T(z)) \quad (2)$$

Consider two cases. First, suppose that all inequality within a group is viewed as fairly based on merit and no redistribution is desired within groups. Inequality across groups by contrast is viewed as unfair. This implies that society places equal weights on all people (at all income levels) within a group but places higher weights on people with the same income level but in more deserving groups. Put differently, society's preferences are redistributive across groups,<sup>18</sup> and fully libertarian within a given group. Given that all individuals within a group are equally valued, we have  $g_m(z - T(z), T(z)) = g_m(T(z))$ , decreasing in  $T$ . If there can be group specific tax schedules  $T^m$ , (i.e., if there was the possibility of 'perfect tagging'), then the

---

<sup>18</sup>With only two groups or with several groups where only one group is valued, we have a *locally Rawlsian* case, in which we only care about the least advantaged group. With more than two groups, there are still positive weights on groups less valued than the least advantaged one, which we call redistributive preferences.

optimal tax system involves zero marginal taxes (i.e., constant taxes or transfers  $T^m(z) = T^m$  for all  $z$ ) within groups. Furthermore, the tax and transfers will be set such that there is perfect equalization of the (decreasing)  $g_m(T^m)$  weights across groups. At any income level, this will involve taxing less individuals which are considered more deserving. On the other hand, suppose taxes can only depend on income and not on group belonging. At the optimum, given the absence of behavioral responses, the weights  $g(z - T(z), T(z)) = \sum_{m=1}^M f(m|z) g_m(T(z))$  need to be equalized across income levels which involves higher taxes for income levels which are likely to have been earned by people in less deserving groups. This is a “reversed tagging” situation, in which, instead of using the group as a tag for income, we use income itself as a tag for group belonging (being able to do so is of course due to no behavioral responses).<sup>19</sup>

Secondly, and more realistically, suppose that society also cares about within-group inequality so that welfare weights within the groups,  $g_m(z - T(z), T(z))$ , are strictly decreasing in disposable income. If taxes can be made dependent on the group, the optimal tax schedule for group  $m$ ,  $T^{*m}(z)$  would be set according to (1), that is:  $T^{*m}(z) = -g_{mc}/(g_{mT} - g_{mc})$  so that weights are fully equalized within each group.<sup>20</sup> The level of the group-specific demogrant, or equivalently, of inter-group transfers are set to equalize all group-specific weights  $g_m(T^{*m})$ . On the other hand, if taxes can only depend on total income, they will be set according to (1), using as weights  $g$  as defined in (2).

As an application, consider family background, ranked from poorest ( $m = 1$ ) to richest ( $m = M$ ).<sup>21</sup> While low earners are valued more, those coming from poorer family backgrounds are viewed as more deserving at any income level. Since a better background provides a boost to earnings all throughout life,  $f(m|z)/f(n|z)$  for  $m < n$  will be decreasing in  $z$ . Hence, higher income levels act as an imperfect tag for a better family background and will be taxed more, even if within a given family background, society has perfectly libertarian preferences.<sup>22</sup>

---

<sup>19</sup>As an application of this ‘reverse tagging’, consider groups based on how fairly people earn their income. In the extreme, group 1 could be pure rent-seekers who exert no productive effort but instead extract resources from others. Group 0 could be hardworking people. If rent-seekers exhibit higher incomes, society might use high incomes as a reverse tag for rent-seeking and want high taxes on top incomes. In a less stark way, this sort of argument is sometimes heard, especially in disfunctional economies, when the elite is considered to be extractive. Piketty, Saez and Stantcheva (2011) also show that within the standard framework, the presence of rent-seeking among top earners tends to push up the top tax rate.

<sup>20</sup>This determines the tax schedule within group  $m$  only up to a constant, which is the group-specific demogrant,  $T^m(0)$ . Given the tax system  $T^{*m}$ , this leads to a common, equalized group-specific welfare weight  $g_m(T^{*m})$ .

<sup>21</sup>Income of individual  $i$  in group  $m$  might for example take the form  $z_i = m + l_i$  where  $l_i$  is an inelastic and heterogeneous labor supply (due to individual preferences, assumed orthogonal to family background).

<sup>22</sup>The next section will consider the exogenous desert model in a more realistic fashion, including behavioral

## 4 Optimal Tax Theory with Behavioral Responses

In this section, we introduce behavioral responses. To keep the presentation as simple as possible, we consider the case with no income effects, so that the utility function of individual  $i$  takes the form  $u(c - h_i(z))$  where  $h_i(z)$  is the disutility of earning  $z$ . We assume that  $h'_i(z) > 0$ ,  $h''_i(z) > 0$  and  $u'(\cdot) > 0$ ,  $u''(\cdot) < 0$ . The absence of income effects simplifies the presentation because the average social marginal welfare weight is equal to one. We focus on linear taxes, at a rate  $\tau$  and assume that all tax revenues are rebated as a uniform demogrant  $R$ . Hence, tax policy can be summarized by the one dimensional variable  $\tau$ .

Faced with a linear tax rate  $\tau$ , individual  $i$  chooses his income  $z_i$  so as to maximize his utility  $u(z_i(1 - \tau) + R - h_i(z_i))$ . Thus,  $h'_i(z_i) = 1 - \tau$ , and we can rewrite taxable income as a function of the net retention rate  $(1 - \tau)$  only, i.e.  $z_i = z_i(1 - \tau)$ .<sup>23</sup> Aggregating over the population, total earnings, denoted by  $Z(1 - \tau)$ , are also a function of the net retention rate  $(1 - \tau)$  with elasticity  $e = [(1 - \tau)/Z]dZ/d(1 - \tau)$ . Using the government's budget constraint, the lumpsum demogrant is then  $R = \tau Z(1 - \tau)$ .

### 4.1 Standard Utilitarian Approach

In the standard utilitarian approach, the government chooses  $\tau$  and  $R$  to maximize

$$SWF = \int_i u((1 - \tau)z_i(1 - \tau) + R - h_i(z_i(1 - \tau)))d\nu(i) \quad \text{s.t.} \quad R \leq \tau Z(1 - \tau) \quad (p)$$

with  $p$  the multiplier on the budget constraint, and  $\nu(i)$  the CDF of individual types  $i$ . Using the envelope theorem applied to each individual  $i$ 's utility maximization problem, the first order conditions with respect to  $R$  and  $\tau$  for the government are simply:

$$\int u'_i d\nu(i) = p$$

$$\int_i u'_i z_i d\nu(i) = p \left[ Z - \frac{dZ}{d(1 - \tau)} \right]$$

Denoting by  $g_i = (u'_i)/p$  the normalized social marginal welfare weight on person  $i$ , the first equation states that the  $g_i$ 's average to one. This is a consequence of the absence of income effects: the government is indifferent between \$1 of public funds and \$1 uniformly distributed

---

responses.

<sup>23</sup>This is due to the absence of income effects.

to all. The two equations can be combined to obtain the standard optimal tax formula

$$\tau = \frac{1 - \bar{g}}{1 - \bar{g} + e} \quad \text{with} \quad \bar{g} = \frac{\int g_i z_i d\nu(i)}{Z} \quad (3)$$

where  $\bar{g}$  is the average social marginal welfare weight weighted by pre-tax income  $z$ .  $\bar{g}$  can also be seen as the average normalized income  $z_i/Z$  weighted by the social welfare weights  $g_i$ .

The optimal tax rate  $\tau$  balances the equity-efficiency tradeoff. It decreases with the elasticity of income  $e$ , which measures the efficiency costs of taxation and increases with  $(1 - \bar{g})$  which measures the value of redistribution. When utility is linear,  $g_i = 1$ , and hence  $\bar{g} = 1$ . Since there is no value for redistribution,  $\tau = 0$ . When utility is *any* concave function,  $g_i$  decreases with  $z_i$  and hence  $\bar{g} < 1$  and  $\tau > 0$ . Note also that  $\bar{g}$  increases with  $\tau$  as increasing taxation combined with an increased lump-sum demogrant reduces the difference in utility across individuals with different earnings abilities. Therefore, in general, equation (3) defines a unique solution.

## 4.2 Endogenous Social Welfare Weights

Instead of specifying the social welfare weights as derived from the social welfare function, i.e.,  $g_i = u'_i((1 - \tau)z_i + R - h_i(z_i))/p$ , in our new approach, we directly write them as a function of disposable income and taxes,  $g_i = g(c_i, T(z_i))$  with  $T(z_i) = \tau z_i - R$  as in the prior section. Formula (3) continues to apply with such weights. To see this, let us derive it again, using the tax reform (or "perturbation") approach to optimal taxation. Suppose the government considers changing  $\tau$  by some small amount  $d\tau$ . The impact of the reform on the utility of individual  $i$ , measured in monetary terms is just the direct effect on consumption  $-z_i d\tau + dR$ , since, by the envelope theorem, the change in behavior  $dz_i$  has no first order impact on utility. Budget balance  $R = \tau Z(1 - \tau)$  hence requires that  $dR = [Z - \tau dZ/d(1 - \tau)]d\tau = Zd\tau[1 - e\tau/(1 - \tau)]$ . Therefore, using weight  $g_i$  to measure the net social benefit of the reform for individual  $i$ , at the optimum we must have that the reform has a null value:

$$\int_i g_i \cdot \left[ -z_i + Z \cdot \left( 1 - e \frac{\tau}{1 - \tau} \right) \right] d\nu(i) = 0$$

This can be immediately rewritten as equation (3).

The standard approach does hence successfully capture the key equity-efficiency trade-off. If this were the only concern for optimal tax policy, our approach would not add much to the standard one. However, as we shall see in the next subsections, our approach can capture a

number of additional effects. Before moving on to those, let us first show how our approach can nest the most widely used alternatives to utilitarianism in a more natural way than social welfare maximization.

**Libertarian case.** Under libertarianism, any individual is fully entitled to his pre-tax income and society should not be responsible for those with lower earnings.<sup>24</sup> This can be captured again in our framework by assuming that  $g_i = g(T(z_i))$  is increasing in  $T(z_i)$ . This will immediately deliver  $T(z_i) = 0$  (hence,  $\tau = 0$ ) as the optimal policy. In the standard framework, the way to obtain a zero tax at the optimum is to either assume that utility is linear or to specify a *convex* transformation of  $u(\cdot)$  in the social welfare function which undoes the concavity of  $u(\cdot)$ . This seems much more artificial than directly stating that society considers redistribution as unjust confiscation and hence that how deserving a person is depends on her net tax contribution.

**Rawlsian case.** The Rawlsian case is the polar opposite of the Libertarian case. Society cares most about the least fortunate and hence sets the tax rate to maximize her welfare. In terms of a social welfare function, this can be captured by a maxi-min criterion. In our framework, it can be done instead by assuming that all the social welfare weight is concentrated on the least advantaged. If the latter have zero earnings, independently of taxes, then a reform is desirable if and only if it increases the demogrant  $R$ , leading to the revenue maximizing tax rate  $\tau = 1/(1 + e)$  at the optimum.<sup>25</sup>

**Political economy.** Our framework can also naturally incorporate political economy considerations. The most widely used model for policy decisions among economists is the median-voter model. In our model, each individual has single peaked preferences about the tax rate  $\tau$ . This is because indirect utility  $\psi_i(\tau) = u((1 - \tau)z_i(1 - \tau) + \tau Z(1 - \tau) - h_i(z_i(1 - \tau)))$  is single peaked at  $\tau_i^*$  which is the solution to  $-z_i + Z - \tau dZ/d(1 - \tau)$ , i.e.,  $\tau_i^* = (1 - z_i/Z)/(1 - z_i/Z + e)$ . Hence, the median voter is the voter with median income  $z_m$  and the political equilibrium has:

$$\tau_m = \frac{1 - z_m/Z}{1 - z_m/Z + e}.$$

---

<sup>24</sup>This view could for example be justified in a world where individuals differ solely in their preferences for work but not in their ability to earn. In that case, there is no good normative reason to redistribute from goods lovers to leisure lovers (exactly as there would be no reason to redistribute from apple lovers to orange lovers in an exchange economy where everybody starts with the same endowment).

<sup>25</sup>This can also be seen as follows. Under the Rawlsian criterion and assuming the most disadvantaged have no earnings,  $\bar{g} = 0$  because  $g_i = 0$  whenever  $z_i > 0$ , hence formula (3) also leads immediately to  $\tau = 1/(1 + e)$ .

Note that  $\tau_m > 0$  when  $z_m < Z$  which is the standard case with empirical income distributions.

This case can be seen as a particular case of endogenous weights where all the weight is concentrated at the median voter. As with the Rawlsian case,  $g$  is not strictly speaking a function but a distribution with mass one at the median.<sup>26</sup>

**Role of behavioral responses.** Suppose there are no behavioral responses so that  $e = 0$ . In that case, the optimum  $\tau$  is such that  $\bar{g} = 1$ , i.e., social welfare weights  $g_i$  should not be correlated with pre-tax income  $z_i$ . Under a standard welfarist model with  $g_i$  decreasing with  $c_i$ ,  $\bar{g} = 1$  only when  $\tau = 1$ . In contrast, with endogenous social welfare weights  $\bar{g} = 1$  is possible even with  $\tau < 1$ . For example, with social weights  $g_i = g(c_i, T_i)$  as in Section 2,  $\bar{g} = 1$  when  $\tau$  is set high enough so that the correlation between deservedness and pre-tax income disappears.<sup>27</sup> With behavioral responses  $e > 0$ , then  $\bar{g} = 1 - e \frac{\tau}{1-\tau} < 1$ , i.e., the optimum  $\tau$  is set so that the correlation between deservedness and pre-tax income remains negative. Hence, as in the standard model, the presence of behavioral responses lowers the optimum tax rate  $\tau$  everything else being the same.

**Equality of opportunity.** Consider again the model of exogenous desert from the previous section, as applied to family background, now adding behavioral responses. Individuals can either come from an advantaged (in the notation of the previous section,  $m = 1$ ) or disadvantaged background ( $m = 0$ ). A good background gives an unfair advantage to earnings, such as through better access to schools from early on, but conditional on background, the level of earnings is based purely on merit (e.g., taste for work). The natural way to set the marginal welfare weights in this situation is to assume uniform and positive weights (say, equal to one) among those from a disadvantaged background ( $g_0(z - T, T) = 1$  for all  $z$ ) and weights equal to zero among those from an advantaged background ( $g_1(z - T, T) = 0$  for all  $z$ ). Those weights capture the two normative principles already mentioned. First, that there is value in redistributing from those coming from an advantaged background to those coming from a disadvantaged background conditional on having the same rank in the distribution of earnings

---

<sup>26</sup>One may object that this result can also be recovered using a standard Pareto weighted sum of utilities with all the weight placed on the median earner. However, the difficulty of this approach is that the median could be endogenous to tax policy  $\tau$  and hence impossible to specify ex-ante, before solving the problem using our endogenous social marginal welfare weights (since, recall, that the standard exogenous Pareto weights should not depend directly on the tax policy).

<sup>27</sup>Another way to see this is as follows. With  $\tau = 1$ , deservedness and pre-tax incomes would be positively correlated (when  $g$  increases with  $T$ ) so that  $\bar{g} > 1$ , calling for a lower tax rate.

in each background group. Second, that there is no value in redistributing across individuals with different earnings conditional on background because those different earnings are purely due to merit. If tax policy cannot be conditioned directly on background, then the social marginal welfare weight at a given earnings level is given by the formula in (2) which simplifies to  $g(z - T(z), T(z)) = f(0|z) = f_0(z) P_0 / \sum_{j=0}^1 f_j(z) P_j$ , the fraction of individuals with a disadvantaged family background at the given earnings level. Since people with a disadvantage background typically have lower earnings, their frequency is decreasing at higher pre-tax earnings and so are the weights. Those are the implicit social welfare weights used in the analysis of Roemer et al. (2003) (see Piketty and Saez, 2013 for an application to the case of the optimum top income tax rate in such a model).

**Skill or preferences for work.** The exogenous desert model can also capture different notions of fairness analyzed in the literature when people differ according to both ability and taste for work (see for example Fleurbaey and Maniquet 2006, 2007, 2011). Suppose that the desert groups are split according to intrinsic ability or skill, which leads to a higher wage for any hour of work. In addition, within each skill group, people have heterogeneous disutilities from work, with ‘lazier’ people disliking work more. For simplicity, consider two skill levels (high and low skill) and two preference types (lazy and hard-working). If society is libertarian along the preference dimension (that is, conditional on a skill level it views differences in earnings due to different preferences for work as fair), but Rawlsian along the skill dimension (that is, at any given level of income, it values most a transfer towards the low skill agent), then we can again capture this in a natural way with binary weights which are constant and equal to 1 for all low-skilled individuals and equal to 0 for all high-skilled individuals. The marginal social welfare weight at income level  $z$  will then be equal to the fraction of low skill individuals at income level  $z$ , so that the marginal social welfare weight will again be decreasing in income, even though conditional on a skill level (that is, given equality of opportunity), society is not bothered by inequality in outcomes. As shown by Fleurbaey (1994), the principle of compensation and the principle of responsibility are generally mutually incompatible. This example shows that endogenous social marginal welfare weights can be used very simply to model the trade-off between those two principles.

### 4.3 Transfers and Free Loaders

In practice, behavioral responses are inherently tied to social welfare weights since one of the biggest complaint against redistribution is that it benefits “free loaders”, that is those who stop working precisely because of the generosity of the redistributive system. Yet, standard welfarism cannot capture such effects, since the social welfare weight on a given individual depends solely on her current situation, and not on whether her current situation arises from responses to taxes and transfers.

The simplest way to illustrate this is to consider a simple binary model in which individuals can either work and earn a uniform wage  $w$ , or choose to not work. Utility is  $u(c_l - \theta l)$  where  $l \in \{0, 1\}$  takes the value 1 if an individual works and 0 otherwise and consumption is equal to  $c_0$  if an individual is out of work and to  $c_1 = w \cdot (1 - \tau) + c_0$  if she works, where  $\tau$  is the linear earnings tax rate. Taxes fund the transfer  $c_0$  to non-workers. The cost of work  $\theta$  is distributed according to a cdf  $P(\theta)$ . Individual  $\theta$  works if and only if  $\theta \leq c_1 - c_0 = (1 - \tau) \cdot w$ . Hence, the fraction of people working is  $P(w(1 - \tau))$ . As before, let  $e$  be the elasticity of aggregate earnings  $Z(1 - \tau) = wP(w(1 - \tau))$  with respect to the retention rate  $(1 - \tau)$ .

Under the utilitarian objective, the government maximizes:

$$SWF = \int_{\theta > (1-\tau)w} u(c_0) dP(\theta) + \int_{\theta \leq (1-\tau)w} u(c_0 + (1-\tau)w - \theta) dP(\theta) \quad \text{s.t.} \quad c_0 = \tau w P((1-\tau)w) \quad (p).$$

Routine computations show that the optimal tax formula takes the form:

$$\frac{\tau}{1 - \tau} = \frac{(1 - P)(\bar{g}_0 - \bar{g}_1)}{e},$$

where  $\bar{g}_0 = u'(c_0)/p$  is the average social welfare weight on non-workers and  $\bar{g}_1 = \int_{\theta \leq (1-\tau)w} u(c_0 + (1 - \tau)w - \theta) dP(\theta) / (p \cdot P)$  the average social welfare weight on workers. As before, because of no income effects, the average social welfare weight across the population is one, so that  $(1 - P)\bar{g}_0 + P\bar{g}_1 = 1$ .

In the utilitarian framework, the social welfare weight placed on the unemployed depends only on  $c_0$  and is completely independent of whether they would have worked absent taxes and transfers. By contrast, the public policy debate focuses on whether the unemployed are deserving of support or not. Transfer beneficiaries are only deemed deserving if they are truly unable to work, that is, if absent any transfers, they would still not work and live in great poverty without resources. Conversely, they are considered non-deserving, or “free loaders”

if they could work and would do so absent more generous transfers. The presence of such “free loaders”, perceived to take undue advantage of a generous transfer system, is precisely why many oppose welfare (see e.g., Ellwood (1988), Ellwood and Bane (1996)). It is also the reason why many welfare programs try to target populations which are deemed more vulnerable and less prone to taking advantage of the system. Historically, disabled people, widows, and later on single parents have been most likely to receive support from the government. The origins of the US welfare system since 1935, starting with the Aid to Families with Dependent Children (AFDC) and continuing with the Temporary Assistance to Needy Families (TANF) federal assistance programs highlights exactly that logic. The goal was to help children of single parents or whose families had low or no income, rather than a more general population, among which there could be many free-loaders.

Naturally, the fraction of “free loaders” among the unemployed increases with the generosity of transfers and with the behavioral elasticity  $e$ . Under standard utilitarianism, free loaders and deserving poor are treated equally in the social welfare function. With endogenous welfare weights, it is possible to treat those two groups differentially.

Formally, let us define the deserving poor as those with  $\theta > w$ , (those who would not work, even absent any transfer), and the free loaders as those with  $w \geq \theta > w \cdot (1 - \tau)$  (those who do not work because of the welfare program only). Denoting by  $P_0 = P(w)$  the fraction working when  $\tau = 0$ , there are  $P(w(1 - \tau))$  workers,  $1 - P_0$  deserving poor, and  $P_0 - P(w(1 - \tau))$  free loaders.

Let us assume that society sets social welfare weights for the deserving poor,  $\bar{g}_0$  and those who work,  $\bar{g}_1$  as under utilitarianism, but sets weights to zero for free loaders. Weights still average to one so that  $(1 - P_0)\bar{g}_0 + P\bar{g}_1 = 1$ . The optimum tax rate becomes

$$\frac{\tau}{1 - \tau} = \frac{(1 - P) \left[ \frac{1 - P_0}{1 - P} \cdot \bar{g}_0 - \bar{g}_1 \right]}{e}.$$

Two points are worth noting about this formula. First, since  $(1 - P_0)/(1 - P) < 1$  taxes will naturally be lower relative to the utilitarian case. In the extreme case in which all unemployed are free-loaders, the optimal transfer (and hence the taxes financing it) is zero. This corresponds to the (admittedly extreme) view that all unemployment is created by an over-generous welfare system. As long as there are some deserving poor though, taxes and transfers will be positive. Second, when  $e$  is larger,  $(1 - P_0)/(1 - P)$  is smaller and hence a higher elasticity reduces the

optimal tax rate not only through the standard efficiency effect but also through the social welfare weight channel as it negatively affects society’s view on how deserving the poor are.

#### 4.4 Luck versus Deserved Income

Let us consider again the case with luck vs. deserved income. Suppose that total income of individual  $i$ ,  $z_i$ , has two components: standard income  $w_i$ , earned at a disutility cost  $h_i(w_i)$  and a stochastic ‘luck’ income,  $y_i$  received at no cost. Faced with a linear tax  $\tau$ , agent  $i$  again generates earnings  $w_i(1 - \tau)$  and receives a random luck shock,  $y_i$ , independent of taxes, leading to total observed earnings  $z_i(1 - \tau) = w_i(1 - \tau) + y_i$ , which are decreasing in  $\tau$ .

Let us assume as in Section 2.3 that society believes that individuals are entitled to deserved income but not to luck income with binary social welfare weights based on  $c = z - T(z) \leq w + Ey$  so that the average social marginal welfare weight  $g(z)$  at pre-tax income level  $z$  is given by the fraction of individuals with income  $z$  such that  $z - T(z) < w + Ey$ . Given such welfare weights, the derivation of the optimal tax formula is the same as above, leading to:

$$\tau = \frac{1 - \bar{g}}{1 - \bar{g} + e} \quad \text{with} \quad \bar{g} = \frac{\int g(z_i)z_i}{Z \int g(z_i)} \quad (4)$$

Under a standard utilitarian criterion,  $\bar{g}$  increases with  $\tau$  so that the solution of (4) is unique.<sup>28</sup>

In contrast, in this luck vs. deserved income model,  $\bar{g}$  is not necessarily increasing with  $\tau$  so that multiple equilibria (i.e., both low tax rate and high tax rate equilibria) are possible as in Alesina and Angeletos (2005).

To see this, consider the case where deserved income is highly elastic to  $\tau$ . Assume that, when taxes are low (e.g.,  $\tau \simeq 0$ ), luck income is small relative to deserved income on average, i.e.,  $Ey \ll Ew$ . In that case, with low  $\tau$ ,  $e$  is large (as elastic deserved income is the main component of total income). Furthermore, a low tax rate is desirable in that case, as income is mostly deserved.<sup>29</sup> Hence, a close to zero tax rate is a stable optimum.

Conversely, suppose the tax rate  $\tau$  is very high (and close to one). Because  $w$  is highly elastic, deserved income is now small. Let us assume it is quantitatively small relative to luck income. In that case, the tax base  $z$  has low elasticity and hence  $e \simeq 0$ . This implies that the optimal tax rate should indeed be high. The welfare effect further reinforces this as most income is undeserved so that a close to one tax rate is also a stable optimum.

<sup>28</sup>This also requires that the elasticity  $e$  does not depend on  $\tau$  (or at least does not vary too quickly with  $\tau$ ).

<sup>29</sup>More precisely, imposing a positive tax rate would make higher incomes (above average) more deserving than lower incomes (below average), leading to a  $\bar{g}$  above 1.

Thus, economies with social preferences favoring hard-earned income over luck income can end up in two possible situations. In the low tax equilibrium, people work hard, luck income makes up a small portion of total income and hence, in a self-fulfilling manner, social preferences tend to favor low taxes. In the alternative equilibrium, high taxes lead people to work less, which implies that luck income represents a larger fraction of total income. This in turn pushes social preferences to favor higher taxes, to take away that unfair luck income (itself favored by the high taxes in the first place). This shows that our framework can encompass the important multiple equilibria outcomes of Alesina and Angeletos (2005) without departing as drastically from optimal income tax techniques as in the model of Alesina and Angeletos (2005).

## 5 Extensions

In this section, we illustrate how our framework can capture additional considerations which matter in the public debate on taxes and yet cannot be satisfactorily accounted for by the standard welfarist approach.

### 5.1 Horizontal Equity Concerns

The standard utilitarian framework leads to the conclusion that if agents can be separated into different groups, based on attributes, so-called ‘tags’, which are correlated with income and exogenous to taxes, then an optimal tax system should have differentiated taxes for those groups. Some attributes can be perfect tags in the sense of being impossible to influence by the agent. An example would be height, which has been shown to be positively correlated with earnings (see Mankiw and Weinzierl (2010)), or gender. Others are only mildly elastic to taxes (such as, arguably, the number of children or marital status). Mankiw and Weinzierl (2010) explore a tax schedule differentiated by height and use this stark example as a critique of the standard utilitarian framework. In practice, society seems to oppose taxation based on such characteristics, probably because it is deemed unfair to tax differently people with the same ability to pay. These ‘horizontal equity’ concerns, or the wish to treat ‘equals as equals’ seem important in practice and a framework for optimal tax policy which wishes to respect society’s preferences needs to be able to include them.<sup>30</sup>

---

<sup>30</sup>A few comments, essentially based on Kaplow (2001), seem important. Kaplow (2001) forcefully argues that Horizontal Equity (HE) per se does not have an independent normative appeal, but that it often only proxies for losses in welfare due to unequal treatment. We are refraining from any judgment here on how important

To see how our approach allows us to think about horizontal equity, consider two groups which differ according to some observable and perfectly inelastic attribute  $m \in \{1, 2\}$  and according to their taxable income elasticities (respectively denoted  $e_1$  and  $e_2$ ).<sup>31</sup> Let  $\lambda_1$  and  $\lambda_2$  be the fraction each group represents in the total population.

As a benchmark, the standard utilitarian approach would lead to two different tax rates,  $\tau_1$  and  $\tau_2$  for the two groups, such that:

$$\tau_m = \frac{1 - \bar{g}_m}{1 - \bar{g}_m + e_m} \quad \text{with} \quad \bar{g}_m = \frac{\lambda_m \int_{i \in m} g_i z_i d\nu_m(i)}{Z_m}$$

where  $\bar{g}_m$  is the income weighted average social marginal welfare weight for group  $m$  and  $\nu_m$  is the CDF of types, conditional on being in group  $m$ . A group would tend to be taxed more if it is less elastic, in the spirit of the traditional Ramsey inverse elasticity rule.

In our framework, we can however incorporate society's belief that different taxes on people with the same earnings are not fair. To do so, the social marginal welfare weight on each group can be specified as a function of both tax rates  $\tau_1$  and  $\tau_2$ . Suppose we start from a situation with equal tax rates  $\tau_1 = \tau_2 = \tau$ , and consider a reform introducing differential taxes (and hence, horizontal inequity). Using standard weights  $g_1(c_1, \tau_1) = g_2(c_2, \tau_2)$  at  $c_1 = c_2$  and  $\tau_1 = \tau_2$ , if  $e_1 < e_2$ , it would be desirable in the standard utilitarian framework to perform a small reform  $(d\tau_1, d\tau_2)$ , increasing taxes on group 1 and reducing them on group 2 (i.e.,  $d\tau_1 > 0 > d\tau_2$ ). But with endogenous welfare weights and a preference for horizontal equity,  $\tau_1 = \tau_2$  can be an equilibrium, despite different elasticities. Consider for example "differential weights" for reforms

---

HE truly is as a normative criterion, or on where the social concern for it stems from. We simply take as given, driven by casual empirical observations, that society values it, whether it is reasonable or not, and we show how our framework can capture it. Secondly, Kaplow (2001), and Kaplow and Shavell (2001) highlight that Horizontal Equity considerations, in particular as modeled in Auerbach and Hassett (2000) will conflict with the Pareto principle in some cases. Our non-negative social welfare weights guarantee that this can never occur in our setup, so that the pursuit of horizontal equity will not come at the expense of welfare. Of course, the concept of horizontal equity per se remains subject to the valid criticisms raised in Kaplow (2001). To name a few, the benchmark against which society judges horizontal inequity (most often, pre-tax income) may itself be endogenous to tax policy or unfairly achieved (through other luck shocks), as well as arbitrary (which order one assigns to luck shocks matters for what is considered to be the fair 'status quo' distribution. Ideally, one would like to have some fundamental, morally justifiable definition of what constitutes "equals", rather than just based on pre-tax income. Our approach is somewhat more flexible than previous ones in that our groups could be based on any characteristic. A somewhat appealing interpretation for horizontal equity in our view is the fear of *unfair* discrimination by the government of otherwise "equals" such as based on religion, gender, interest groups lobbying, etc..

<sup>31</sup>Note that there are two possible ways to think about tags. The first and more standard one, in line with the aforementioned papers, considers two groups which differ in terms of their average earning abilities, in a world in which the government is unable to observe individual abilities. With nonlinear taxation, self-selection constraints are then relaxed thanks to tagging. The second approach, pursued here and more relevant for linear taxation, is to consider groups which differ in terms of their *elasticities* to taxes.

introducing horizontal inequity, with which consumption at the margin for a group gaining from the horizontal inequity would be less valued than for a group losing from it. A simple way is to assume that  $g(c_i, \tau, d\tau_2 > d\tau_1) = 0$  for all  $i \in 1$  and  $g(c_i, \tau, d\tau_2 > d\tau_1) = g(c_i, \tau) / \int_{j \in 2} g(c_j, \tau)$  for all  $i \in 2$  (and vice versa for a reform such that  $d\tau_2 < d\tau_1$ ).

To see what taxes  $\tau$  can be equilibria, consider that the value of any reform  $(d\tau_1, d\tau_2)$  around an equilibrium with equal treatment  $\tau_1 = \tau_2 = \tau$  must be non-positive, that is:

$$Z_1(-\bar{g}_1 + [1 - e_1\tau/(1 - \tau)])d\tau_1 + Z_2(-\bar{g}_2 + [1 - e_2\tau/(1 - \tau)])d\tau_2 \leq 0$$

where:

$$\bar{g}_m = \frac{\lambda_m \int_{i \in m} g(c_i, \tau_i, d\tau_2, d\tau_1) z_i d\nu_m(i)}{Z_m}$$

Since this must hold for any reform, it is sufficient to check that it holds for four 'basic' reforms, namely for  $d\tau_1 > 0 = d\tau_2$ , for  $d\tau_2 > 0 = d\tau_1$ , for  $d\tau_2 < 0 = d\tau_1$  and for  $d\tau_1 < 0 = d\tau_2$ . Checking those four cases, leads to the following range of possible equilibrium taxes:

$$\min \left\{ \frac{1}{1 + e_1}, \frac{1}{1 + e_2} \right\} \geq \tau \geq \max \left\{ \frac{1 - \bar{g}_1(c, \tau, d\tau_1 > d\tau_2)}{1 - \bar{g}_1(c, \tau, d\tau_1 > d\tau_2) + e_1}, \frac{1 - \bar{g}_2(c, \tau, d\tau_2 > d\tau_1)}{1 - \bar{g}_2(c, \tau, d\tau_2 > d\tau_1) + e_2} \right\}$$

Hence, if the weight on the group threatened to suffer from a potential horizontal inequity is sufficiently large, there is a large possible interval of equilibrium taxes, despite the differential elasticities of both groups. This is because any deviation from equal taxes is penalized sufficiently heavily. In the limit, suppose that  $\bar{g}_1(c, \tau, d\tau_1 > d\tau_2)$  tends to 1. Then any non-negative tax below the smallest revenue maximizing level across the two groups can be sustained as an equilibrium. Note however, that due to the fact that our social welfare weights are always non-negative, we cannot have a Pareto-dominated situation (as illustrated by the upper bounds equal to the revenue maximizing rates for each group).

We can also address the more general question of what *different* tax rates on the two groups can be sustained in equilibrium. To do so, we need to define the weights more generally as functions of both the tax levels and the tax changes. Let the weight for person  $i$  be  $g_i = g(c_i, \tau_m, \tau_n, d\tau_m, d\tau_n)$ . Consider the following example. Start from a set of standard welfare weights  $\{g(c_i, \tau_m)\}_i$  for each agent which sum to 1 across the population:  $\int_{i \in m} g(c_i, \tau) d\nu(i) + \int_{i \in n} g(c_i, \tau) d\nu(i) = 1$ . Now define the weights capturing horizontal equity concerns as follows:

i) If in the status quo  $\tau_m > \tau_n$ , a reform is introduced with  $d\tau_m < d\tau_n$  (so that the group which is already taxed more is helped by a reduction in its taxes), then let  $g(c_i, \tau_m > \tau_n, d\tau_m < d\tau_n) =$

$g(c_i, \tau_m) / \int_{i \in m} g(c_i, \tau_m)$  for  $i \in m$  and  $g(c_i, \tau_m > \tau_n, d\tau_m > d\tau_n) = 0$  for  $i \in n$ .

ii) If in the status quo  $\tau_m > \tau_n$ , a reform is introduced with  $d\tau_m > d\tau_n$  (so that the group which is already taxed more is hurt further by a tax increase), then let:  $g_i = g(c, \tau_m > \tau_n, d\tau_m > d\tau_n) = g_i^+ = g^+(c, \tau_m > \tau_n, d\tau_m < d\tau_n)$  for all  $i \in m$ , where  $g^+$  is a mean-preserving spread of  $g$  (informally put, the set of weights  $g^+$  place even less value on high consumption people, and even more weight on low consumption people). This means that the average weight on group  $m$ , weighted by income will be higher for the proposed reform ii) than for reform i). Again, let  $g_i = 0$  for all  $i \in n$ .

With this set of weights, a situation with  $\tau_1 > \tau_2$  can be an equilibrium, if and only if:

$$Z_1(-\bar{g}_1 + [1 - e_1\tau_1/(1 - \tau_1)])d\tau_1 + Z_2(-\bar{g}_2 + [1 - e_2\tau_2/(1 - \tau_2)])d\tau_2 \leq 0$$

It is again sufficient to check that this equality holds for the following ‘basic’ tax reforms (since all possible other reforms can be expressed as combinations of these basic reforms) as explained in the Appendix. The equilibrium range of taxes obtained for a case with  $\tau_1 < \tau_2$  is then:

$$\frac{1 - \bar{g}_1^+}{1 - \bar{g}_1^+ + e_1} \leq \tau_1 \leq \frac{1 - \bar{g}_1}{1 - \bar{g}_1 + e_1} \quad \text{and} \quad \tau_2 = \frac{1}{1 + e_2}$$

The last equality implies that if the government wants to set  $\tau_2$  at a lower level than  $\tau_1$ , then it must necessarily be set at the revenue maximizing rate.

For a tax  $\tau_1$  to exist in this range, we need to have a valid range, so that we require:

$$\frac{1}{1 + e_2} < \frac{1 - \bar{g}_1}{1 - \bar{g}_1 + e_1}$$

which reduces to  $e_2(1 - \bar{g}_1) > e_1$ . Despite this condition being itself endogenous to  $\tau_1$  (since  $\bar{g}_1$  is), it means broadly speaking that  $e_2$  will have to be sufficiently larger than  $e_1$  to justify  $\tau_2$  being set lower (and the more so, the more people are averse to horizontal inequity as captured in a larger  $\bar{g}_1$ ).

Intuitively, our endogenous social marginal welfare weights punish the group which benefits from tagging (i.e., which is taxed less as a result of tagging), so that the latter will only gain if the its elasticity is really sufficiently large relative to the other group’s. This is reminiscent of a Rawlsian setup, in which society only cares about the least well-off. Here, the set of people whom society cares about is endogenous to the tax system. Namely, they are the ones discriminated against because of tagging. In that case, the revenue-maximizing rate is imposed

on the group ‘favored’ by the tax system, in the same way that the revenue-maximizing rate is imposed on everyone except the poorest with Rawlsian social preferences. In this framework, a tradeoff appears between efficiency (setting taxes based on the differential elasticities) and social preferences for horizontal equity (pushing for taxes to be equalized despite different elasticities). In other words, we can rephrase the Rawlsian famous criterion as follows:

“It is permissible to discriminate against a group using taxes and transfers only in the case where such discrimination allows to improve the welfare of the group discriminated against.”

One possible application of this analysis would be reforms focusing on gender-differentiated taxation. Indeed, there is ample empirical evidence that single mothers or secondary earners are more elastic in their labor supply than prime age men (see e.g., Blundell and MaCurdy, 1999). Yet, almost no country has adopted gender-differentiated taxes, despite the standard Ramsey consideration. Our welfare weights can make sense of the absence of such taxes.

## 5.2 Poverty Alleviation

The poverty rate, defined as the fraction of households below a given disposable income threshold (the poverty threshold) gets substantial attention in the public debate. Hence, it is conceivable that governments aim to either reduce the poverty gap (defined as the amount of money needed to lift all households out of poverty) or reduce the poverty rate (the number of households below the poverty threshold). A few studies have considered used government objectives incorporating such government objectives. Besley and Coate (1992) and Kanbur, Keen, and Tuomala (1994) show how adopting poverty minimization indexes affects optimal tax analysis. Importantly, they show that the outcomes can be Pareto dominated. In this section, we show how the use of endogenous welfare weights allows to incorporate in a simple way in the traditional optimal tax analysis poverty alleviation considerations while maintaining the Pareto principle.

We consider the optimal nonlinear income tax problem as we are particularly interested in the profile of taxes and transfers.

We assume that individuals differ solely through their ability parameter  $n$ , distributed with cdf  $F(n)$  and density  $f(n)$ . Individual  $n$  has utility

$$u^n(c, z) = c - nh(z/n)$$

with  $h(\cdot)$  denoting the increasing and convex cost of work function. We normalize  $h$  so that  $h'(1) = 1$  and  $h(0) = 0$ . Each individual chooses  $z$  to maximize  $z - T(z) - nh(z/n)$  leading

to first order condition  $1 - T'(z) = h'(z/n)$ . Hence, we have  $z = n\phi(1 - T')$  where  $\phi(\cdot)$  is the inverse of  $h'(\cdot)$ . When  $T' = 0$ ,  $z = n$  so that we can interpret  $n$  as “potential income.” Positive marginal tax rates depress real income relative to potential income. We denote by  $e$  the elasticity of earnings  $z$  with respect to the net-of-tax rate  $1 - T'$ .

The government maximizes

$$SWF = \int \omega_n G(z - T(z) - nh(z/n)) dF(n)$$

Denoting by  $g_n = \omega_n G'(u_n)/p$  the social marginal welfare weight on individual  $n$ , no income effects implies that  $\int g_n dF(n) = 1$ . The standard optimal tax formula takes the following form (see the derivation in appendix):

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{1}{e} \cdot \frac{\int_n^\infty (1 - g_m) dF(m)}{nf(n)}. \quad (5)$$

The demogrant is then defined by the government budget constraint.

Let us now consider criteria of poverty alleviation. Let us denote the poverty threshold by  $\bar{c}$ . Anybody with disposable income  $c < \bar{c}$  is poor. If the demogrant can be made bigger than  $\bar{c}$ , then the optimum way to fight poverty is to raise enough taxes to set the demogrant equal to  $\bar{c}$ . Once the poverty threshold has been attained, there is no reason to have differences in social welfare weights and hence the weights would all be equal to a fixed  $g$ . Hence, the marginal tax rates would take the form

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{1 - g}{e} \cdot \frac{1 - F(n)}{nf(n)} \quad \text{or} \quad T'(z_n) = \frac{1 - g}{1 - g + a_n \cdot e},$$

where  $a_n = nf(n)/[1 - F(n)]$  is the local Pareto parameter.  $g$  is set so that total taxes collected raise enough revenue to fund the demogrant  $\bar{c}$ . The less trivial case is when even with  $g = 0$  (which corresponds to the Rawlsian case), tax revenue cannot fund a demogrant as large as  $\bar{c}$ . Let us assume that  $n^*$  is the ability level at the poverty threshold so that  $F(n^*)$  individuals are poor at the optimum. There are two ways to consider poverty alleviation, one is to minimize the poverty gap, the other to minimize the poverty rate.

**Poverty gap alleviation.** Suppose the government cares about the consumption of people living in poverty. A natural way to capture this is to assume that social welfare weights are concentrated among those living in poverty, i.e., with disposable income  $c$  below the poverty threshold  $\bar{c}$ . We can therefore specify the welfare weights as follows:  $g(c) = \bar{g} > 0$  if  $c < \bar{c}$  and

$g(c) = 0$  if  $c \geq \bar{c}$ .<sup>32</sup> The normalization of social welfare weights implies that  $\bar{g}F(n^*) = 1$ . The optimum marginal tax rates take the following form:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{1}{e} \cdot \frac{1 - F(n)}{nf(n)} \quad \text{if } n > n^*$$

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{\bar{g} - 1}{e} \cdot \frac{F(n)}{nf(n)} \quad \text{if } n < n^*$$

Because  $\bar{g}F(n^*) = 1$ , the marginal tax rate is continuous at the poverty level. The marginal tax rate is Rawlsian above  $n^*$  and positive (and typically large) below  $n^*$ . The shape of optimal tax rates is quite similar to the standard utilitarian case.

**Poverty rate minimization.** Suppose the government cares only about the number of people living in poverty, that is the poverty rate. In that case, the government puts more value in lifting people above the poverty line than helping those substantially below the poverty line. Hence, the social marginal welfare weights are concentrated solely at the poverty threshold  $\bar{c}$ . Hence  $g(c) = 0$  below  $\bar{c}$  and above  $\bar{c}$ , and  $g(c) = \bar{g}$  at  $\bar{c}$  ( $\bar{g}$  is finite if a positive fraction bunch at the poverty threshold as we shall see, otherwise  $g(c)$  would be a Dirac distribution). The optimum marginal tax rates take the following form:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{1}{e} \cdot \frac{1 - F(n)}{nf(n)} \quad \text{if } n > n^{**}$$

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{1}{e} \cdot \frac{-F(n)}{nf(n)} \quad \text{if } n < n^*$$

Hence, there is a kink in the optimal tax schedule with bunching at the poverty threshold  $\bar{c}$ . All individuals with  $n$  such that  $n^* \leq n \leq n^{**}$  bunch at the poverty level so that  $z_n$  is constant and  $c_n = \bar{c}$  in that range. The marginal tax rate is Rawlsian above the poverty threshold and is negative below the poverty threshold so as to push as many people just above poverty. Hence, the optimum would take the form of an EITC designed so that at the EITC maximum, earnings plus EITC equal the poverty rate.

### 5.3 Empirical Analysis

The next step in this research agenda is to provide empirical foundation for our theory, in addition to the existing papers cited in Section 2. Despite this part still being work in progress,

---

<sup>32</sup>A less extreme version of this assumption would set  $g(c) = \underline{g}$  above  $\bar{c}$  with  $\underline{g} < \bar{g}$ . It is easy to adapt our results to that case.

we briefly describe our approach here. The basic tool we use is a series of online surveys destined to elicit people’s preferences for redistribution and their concepts of fairness. The questions are clustered in two main groups. The first one serves to find out what notions of fairness people use to judge tax and transfer systems. We focus on the themes addressed in this paper, namely, whether luck income is less deserved than earned income, whether horizontal equity is important, or whether the notion of free loaders has empirical grounding, among others. Through a series of questions, as well as randomized display of various types of information, we try to get a better grasp of what drives society’s preferences, i.e., which elements should enter the marginal social welfare weights. The second group of tests has a more quantitative ambition. As described in Section 3, a simple specification of the sort  $g(c, T) = g(c - \alpha T) = g(z - (1 + \alpha)T(z))$  where  $\alpha$  is a constant parameter delivers an optimal linear tax rate with  $T'(z) = 1/(1 + \alpha)$ . The parameter  $\alpha$  can be calibrated by asking people to choose whom to give a \$1 transfer/tax break to, among a number of different groups earning different different income levels  $z_i$  and being taxed different amounts  $T_i$ . First, this allows to test whether the public has utilitarian preferences, in which case the choice of the recipient of the transfer should depend solely on the relative levels of disposable income  $c_i = z_i - T_i$ . We strongly conjecture that people’s choices will also depend on the amount of taxes the various groups pay, justifying our move beyond utilitarianism. Second, over and above simply testing for utilitarian preferences, well-designed questions can in principle allow us to recover  $\alpha$  or a schedule for  $\alpha$  as a function of other characteristics.

## 6 Conclusion

This paper has proposed a generalized theory of optimal taxation using the tax reform approach and endogenous social marginal welfare weights. A tax system is optimal if no budget neutral marginal reform can increase the sum of (money metric) gains and losses across individuals weighted using the endogenous social marginal welfare weights. Optimum tax formulas take the same form as standard utilitarian tax formulas by simply substituting standard marginal social welfare weights by those *endogenous marginal social welfare weights*. Hence our theory nests standard theory and is equally tractable.

In addition, as we have shown through a series of examples, the use of suitable endogenous social welfare weights can help resolve most of the puzzles of the traditional welfarist approach and account for existing tax policy debates and structures while retaining Pareto constrained ef-

iciency. In particular, endogenous welfare weights can provide a rich theory of optimal taxation even absent any behavioral responses. Our theory brings back social preferences as a critical element for optimal tax theory analysis. Naturally, this flexibility of endogenous social weights begs the question of what social welfare weights ought to be and how they are formed.

Endogenous welfare weights can be derived from social justice principles, leading to a normative theory of taxation. The most famous example is the Rawlsian theory where the endogenous social marginal welfare weights are concentrated solely on the most disadvantaged members of society. As we have discussed, binary weights (equal to one for those deserving more support and zero otherwise) have normative appeal and can be used in a broad range of cases. The Rawlsian case can also be extended to a discrete number of groups, ranked according to desert, such that society has redistributive preferences across groups but libertarian preferences within groups. Naturally, who is deserving might itself be endogenous to the tax system. Such weights can also prioritize justice principles in a lexicographic form.

First, injustices created by tax policy (such as violations of horizontal equity) may have the highest priority. In that case, those deserving of support are those discriminated against whenever horizontal inequities arise. This implies that horizontal inequities can only arise if they help the group discriminated against, dramatically lowering the scope for such policies (such as tagging) that arise with the standard welfarist approach and that are not observed as frequently in the real world.

Second, deserving individuals will be those who face difficult economic situations through no fault of their own. This captures the principle of compensation. Health shocks come to mind, explaining why virtually all advanced countries adopt generous public health insurance that effectively compensate individuals for the bad luck of having poor health and facing high health expenses. Once disparities in health care costs have been compensated by public health insurance provision, this element naturally drops out of social welfare weights. Family background is obviously another element that affects outcomes and that individuals do not choose. This explains why equality of opportunity has wide normative appeal both among liberals and conservatives. Policies aiming directly to curb such inequities such as public education or inheritance taxation (Piketty and Saez, 2012)<sup>33</sup> can therefore be justified on such grounds. Naturally, public education or inheritance taxation cannot fully erase inequalities due to background. This

---

<sup>33</sup>Stantcheva (2012) considers optimal education and human capital policies, distinguishing between policies which improve equality of opportunity versus those who exacerbate already existing skill differences.

leaves a role for taxes and transfers based on income as in the theory of Roemer et al. (1993), which can also be nested in our choice of welfare weights.

Third, even conditional on background, there remains substantial inequality in incomes. Part of this inequality is due to choices (preferences for leisure vs. preferences for goods) but part is due to luck (ability and temperament are often not based on choice). Naturally, there is a debate on the relative importance of choices vs. luck, which impacts the resulting social welfare weights. The endogenous social welfare weights have the advantage of highlighting which differences society considers unfair (for example, due to intrinsic skill differences) and which it considers fair (for example, due to different preferences for work).

Finally, there might be scope for redistribution based on more standard utilitarian principles, i.e., the fact that an additional dollar of consumption matters more for lower incomes than for higher incomes. This principle might be particularly strong at the low income end to justify the use of anti-poverty programs.

Social preferences of the public are indeed shaped by beliefs about what drives disparities in individual economic outcomes (effort, luck, background, etc.) as in the model of Piketty (1995). Endogenous welfare weights could also be derived empirically, by estimating actual social preferences of the public, leading to a positive theory of taxation. There is indeed a small body of work trying to uncover perceptions of the public about various tax policies using surveys (see e.g., Fong, 2001 and Frohlich and Oppenheimer, 1992). More ambitiously, economists can also cast light on those mechanisms and hence enlighten public perceptions so as to move the debate up to higher level normative principles.

# A Appendix

## A.1 Horizontal Equity Proofs

The value of any differential reform on the two groups,  $d\tau_1$  and  $d\tau_2$  can be derived as usual: There is a direct consumption effect on group  $i$  equal to:  $-z_i d\tau + dR$ . By budget balance, we have that:  $R = \tau_1 Z_1(1 - \tau_1) + \tau_2 Z_2(1 - \tau_2)$  ( $Z_i$  denotes the total income earned by group  $i$ ), so that  $dR = [Z_1 - \tau dZ_1/d(1 - \tau_1)]d\tau_1 + [Z_2 - \tau dZ_2/d(1 - \tau_2)]d\tau_2 = Z_1 d\tau_1 [1 - e_1 \tau_1 / (1 - \tau_1)] + Z_2 d\tau_2 [1 - e_2 \tau_2 / (1 - \tau_2)]$

Weights sum to 1 over the whole population,  $\int_i g_i d(v_i) = 1$ . and hence  $\int_{i \in 1} g_i d\nu(i) = 1 - \int_{i \in 2} g_i d\nu(i)$ . Also, letting  $v_m(i)$  denoted the CDF of types conditional on being in group  $m$ , and by  $\lambda_m$  the fraction of types  $m$  in the population, we have:  $\lambda_1 \int_{i \in 1} g_i d\nu(i) + \lambda_2 \int_{i \in 2} g_i d\nu(i) = 1$ . A situation with  $\tau_1 = \tau_2 = \tau$  is an equilibrium iff:

$$\begin{aligned} & -\lambda_1 \int_{i \in 1} g_i z_i d\nu_1 d\tau_1 + \lambda_1 \int_{i \in 1} g_i Z_1 \left[ 1 - \frac{e_1 \tau}{1 - \tau} \right] d\nu_1 d\tau_1 + \lambda_1 \int_{i \in 1} g_i Z_2 \left[ 1 - \frac{e_2 \tau}{1 - \tau} \right] d\nu_1 d\tau_2 + \\ & -\lambda_2 \int_{i \in 2} g_i z_i d\nu_2 d\tau_2 + \lambda_2 \int_{i \in 2} g_i Z_1 \left[ 1 - \frac{e_1 \tau}{1 - \tau} \right] d\nu_2 d\tau_1 + \lambda_2 \int_{i \in 2} g_i Z_2 \left[ 1 - \frac{e_2 \tau}{1 - \tau} \right] d\nu_2 d\tau_2 \leq 0 \end{aligned}$$

or, simplifying:

$$Z_1 (-\bar{g}_1 + [1 - e_1 \tau / (1 - \tau)]) d\tau_1 + Z_2 (-\bar{g}_2 + [1 - e_2 \tau / (1 - \tau)]) d\tau_2 \leq 0$$

for any  $d\tau_1$  and  $d\tau_2$ , where  $\bar{g}_m$  is as defined in the main text. It is sufficient to check that this holds for four 'basic' reforms, namely i)  $d\tau_1 > 0 = d\tau_2$ , ii)  $d\tau_2 > 0 = d\tau_1$ , iii)  $d\tau_1 < 0 = d\tau_2$ , iv)  $d\tau_2 < 0 = d\tau_1$ . All other reforms can be written as combinations of these four basic reforms.<sup>34</sup>Hence, for  $d\tau_1 > 0 = d\tau_2$ ,  $\bar{g}_1 = \bar{g}_1(c, \tau, d\tau_1 > d\tau_2)$  and  $g_2 = 0$ , the following condition must hold:

$$\tau \geq \frac{1 - \bar{g}_1}{(1 - \bar{g}_1 + e_1)}$$

And symmetrically, one should not be able to profitably increase  $\tau_2$ , that is for  $d\tau_2 > 0 = d\tau_1$  the following condition must hold:

$$\tau \geq \frac{1 - \bar{g}_2}{(1 - \bar{g}_2 + e_2)}$$

In addition, it should not be possible to reduce either  $\tau_1$  or  $\tau_2$  leading to:

$$\min \left\{ \frac{1}{1 + e_2}, \frac{1}{1 + e_1} \right\} \geq \tau$$

---

<sup>34</sup>For example consider a more general reform  $d\tau_2 > d\tau_1 > 0$ . If the effect of only  $d\tau_2$  is positive, then we are ruling that out by the basic reform  $d\tau_2 > 0$ . If the effect of  $d\tau_2$  is itself negative, we can not 'compensate' it by increasing  $d\tau_1$  so much that the net effect is positive. Because that would require the effect of  $d\tau_1$  alone to be positive, which we rule out as well in one of the four basic reforms.

Existence here is not a problem, since this is a well-defined interval for  $\tau$  as long as both average welfare weights are less than 1, weakly.

**General Case.** With the set of weights as defined in the main text, a situation with  $\tau_1 > \tau_2$  can be an equilibrium, if and only if:

$$Z_1(-\bar{g}_1 + [1 - e_1\tau_1/(1 - \tau_1)])d\tau_1 + Z_2(-\bar{g}_2 + [1 - e_2\tau_2/(1 - \tau_2)])d\tau_2 \leq 0$$

It is again sufficient to check that this equality holds for the following 'basic' tax reforms (since all possible other reforms can be expressed as combinations of these basic reforms).

1. For  $d\tau_1 > d\tau_2 = 0$  :

$$\frac{1 - \bar{g}_1^+}{1 - \bar{g}_1^+ + e_1} \leq \tau_1$$

2. For  $d\tau_1 < d\tau_2 = 0$  :

$$\tau_1 \leq \frac{1 - \bar{g}_1}{1 - \bar{g}_1 + e_1}$$

3. And finally, both  $d\tau_2 > d\tau_1 = 0$  and  $d\tau_2 < d\tau_1 = 0$  taken together lead to:

$$\tau_2 = \frac{1}{1 + e_2}$$

## A.2 Poverty Alleviation

The individual first order condition  $h'(z/n) = 1 - T'$  implies that the elasticity of earnings with respect to  $1 - T'$  is  $e = \frac{1-T'}{z} \frac{dz}{d(1-T')} = h'(z/n)/[(z/n)h''(z/n)]$ . We denote by  $c_n$ ,  $z_n$ , and  $u_n$  the consumption, earnings, and utility of individual  $n$ . Using the envelope theorem, we have  $du_n/dn = -h(z/n) + (z/n)h'(z/n)$ .

The government maximizes a social welfare function,

$$W = \int \omega_n G(u_n) f(n) dn \quad \text{s.t.} \quad \int c_n f(n) dn \leq \int z_n f(n) dn \quad (p),$$

where  $\omega_n$  is the Pareto weight on individual  $n$ . Following Mirrlees (1971), in the maximization program of the government,  $u_n$  is regarded as the state variable,  $z_n$  as the control variable, while  $c_n = u_n + nh(z_n/n)$ . Therefore, the Hamiltonian is

$$H = [\omega_n G(u_n) + p \cdot (z_n - u_n - nh(z_n/n))] f(n) + \psi(n) \cdot [-h(z_n/n) + (z_n/n)h'(z_n/n)],$$

where  $\psi(n)$  is the multiplier of the state variable. The first order condition with respect to  $z$  is

$$p(1 - h'(z_n/n)) f(n) + \frac{\psi(n)}{n} \cdot ((z_n/n)h''(z_n/n)) = 0.$$

The first order condition with respect to  $u$  is

$$-\frac{d\psi(n)}{dn} = [\omega_n G'(u_n) - p] f(n).$$

Denoting by  $g_n = \omega_n G'(u_n)/p$  the social marginal welfare weight on individual  $n$  and using the transversality condition  $\psi(\infty) = 0$ , we can integrate this equation to

$$\psi(n) = p \int_n^\infty [g_m - 1]f(m)dm.$$

Plugging this expression into the first order condition with respect to  $z$ , noting that  $h'(z_n/n) = 1 - T'(z_n)$ , and  $(z_n/n)h''(z_n/n) = h'(z_n/n)/e$ , we have:

$$pT'(z_n)f(n) = \frac{1}{n \cdot e}(1 - T'(z_n)) \int_n^\infty [1 - g_m]f(m)dm,$$

which can be rearranged as equation (5) in the text. Note that the transversality condition  $\psi(0) = 0$  implies that  $\int g_n f(n)dn = 1$ , i.e., social welfare weights average to one.

In the poverty alleviation case,  $g_n = 0$  when  $c_n > \bar{c}$ , which leads to the formula in the text.

In the poverty minimization case, all the weights are concentrated at  $\bar{c}$ . If individuals with  $n \in (n^*, n^{**})$  bunch at  $\bar{c}$  then  $g_n = 0$  below  $n^*$  and above  $n^{**}$  and  $g_n = 1/[F(n^{**}) - F(n^*)]$  for  $n \in (n^*, n^{**})$  which leads to the formulas in the text.

## References

- Ackert, Lucy, Jorge Martinez-Vazquez, and Mark Rider.** 2007. "Social Preferences and Tax Policy Design: Some Experimental Evidence," *Economic Inquiry* 45(3), 487-501.
- Akerlof, George.** 1978. "The Economics of Tagging as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," *American Economic Review* 68(1), 8-19.
- Alesina, Alberto and George-Marios Angeletos.** 2005. "Fairness and Redistribution," *American Economic Review* 95(3), 960-980.
- Atkinson, A.B.** 1970. "On the Measurement of Inequality," *Journal of Economic Theory* 2(3), 244-63
- Almas, Ingvild, Cappelen, Alexander, Sorensen, Erik and Bertil Tungodden.** 2010. "Fairness and the Development of Inequality Acceptance," *Science* 328(5982), 1176-1178
- Auerbach, Alan, and Kevin Hassett.** 2002. "A New Measure of Horizontal Equity," *American Economic Review*, 92(4), 1116-1125.
- Besley, Timothy, and Steven Coate.** 1992. "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs," *American Economic Review*, 82(1), 249-261.
- Besley, Timothy, and Stephen Coate.** 1992 "Understanding welfare stigma: taxpayer resentment and statistical discrimination," *Journal of Public Economics* 48(2), 165-183.
- Blundell, Richard, MaCurdy, Thomas** 1999. "Labour Supply: A Review of Alternative Approaches," In: Ashenfelter, O., Card, D. (Eds.), *Handbook of labour Economics*, Volume 3A, North-Holland, Amsterdam.
- Bourguignon, François and Amedeo Spadaro.** 2012. "Tax-benefit Revealed Social Preferences," *Journal of Economic Inequality* 10(1), 75-108.
- Bosmans, K., and Schokkaert, E.,** 2004. "Social welfare, the veil of ignorance and purely individual risk: an empirical examination," *Research on Economic Inequality* 11, 85-114
- Choné, Philippe, and Guy Laroque** 2005 "Optimal incentives for labor force participation," *Journal of Public Economics* 89.(2), 395-425
- Cowell, F., and Schokkaert, E.,** 2001. "Risk perceptions and distributional judgments," *European Economic Review* 45 (4-6), 941-952
- Christiansen, Vidar, and E. Jansen,** 1978. "Implicit social preferences in the Norwegian system of indirect taxation," *Journal of Public Economics* 10(2), 217-245.
- Chone, P., and G. Laroque,** 2005. "Optimal Incentives for Labor Force Participation," *Journal of Public Economics* 89, 395-425.
- Devooght, K., and Schokkaert, E.,** 2003. "Responsibility-sensitive fair compensation in different cultures," *Social Choice and Welfare* 21, 207-242
- Edgeworth, F. Y.** 1897. "The Pure Theory of Taxation," *Economic Journal* 7, 46-70, 226-238, and 550-571.
- Ellwood, David.** 1988. *Poor Support: Poverty and the American Family*, New York: Basic

Books.

- Ellwood, David and Bane, Mary J.**, 1996. *Welfare Realities: from Rhetoric to Reform*, Cambridge: Harvard University Press.
- Engelmann, Dirk and Martin Strobel.** 2004. "Inequality Aversion, Efficiency and Maximin Preferences in Simple Distribution Experiments," *American Economic Review* 94(4), 857–69.
- Fleurbaey, Marc.** 1994. "On Fair Compensation", *Theory and Decision* 36, 277–307.
- Fleurbaey, Marc.** 2008. *Fairness, Responsibility and Welfare*, Oxford: Oxford University Press.
- Fleurbaey, Marc, and François Maniquet.** 2006. "Fair Income Tax," *Review of Economic Studies* 73, 55-83.
- Fleurbaey, Marc, and François Maniquet.** 2007. "Help the Low Skilled or let the Hard-working Thrive? A Study of Fairness in Optimal Income Taxation," *Journal of Public Economic Theory* 9(3), 467-500.
- Fleurbaey, Marc, and François Maniquet.** 2011. *A Theory of Fairness and Social Welfare*, Cambridge: Cambridge University Press.
- Fong, Christina.** 2001. "Social Preferences, Self-interest, and the Demand for Redistribution," *Journal of Public Economics* 82(2), 225–246.
- Frohlich, Norman and Joe A. Oppenheimer.** 1992. *Choosing Justice: An Experimental Approach to Ethical Theory*, Berkeley University of California Press.
- Jeene, Marjolein, Wim van Oorschot, and Wilfred Uunk.** 2011. "Popular Criteria for the Welfare Deservingness of Disability Pensioners: The Influence of Structural and Cultural Factors," *Journal of Social Indicators Research*, 1-15.
- Kanbur, Ravi, Michael Keen, and Matti Tuomala.** 1994. "Optimal Nonlinear Income Taxation for the Alleviation of Income-Poverty," *European Economic Review* 38(8), 1613-1632.
- Kaplow, Louis** 2008. *The Theory of Taxation and Public Economics*, Princeton University Press: Princeton.
- Kaplow, Louis, and Steven Shavell.** 2001. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle," *Journal of Political Economy* 109(2), 281–86.
- Kaplow, Louis.** 2001. "Horizontal Equity: New Measures, Unclear Principles (Commentary)," *Inequality and Tax Policy*, Hassett and Hubbard, eds., American Enterprise Institute, 75–97.
- Kuziemko, I., Norton, M., Saez., E., and S. Stantcheva** 2012. "The effects of Information about Inequality and Taxes on Preferences for Redistribution: Evidence from a Randomized Survey Experiment", draft in preparation
- Larsen, Christian Albrekt.** 2008. "The Institutional Logic of Welfare Attitudes: How Welfare Regimes Influence Public Support," *Comparative Political Studies* 41(2), 145-168.
- Mankiw, N. Gregory, and Matthew Weinzierl.** 2010. "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, 2(1), 155-76.

- Mirrlees, James A.** 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies*, 38: 175-208.
- Mirrlees, James A.** 1974. "Notes on Welfare Economics, Information and Uncertainty" in M. Balch, D. McFadden and S. Wu (eds.) *Essays in Equilibrium Behavior under Uncertainty* (Amsterdam: North-Holland), 243-258
- Piketty, Thomas.** 1995 "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics*, 110(3), 551-584.
- Piketty, Thomas and Emmanuel Saez.** 2012. "A Theory of Optimal Capital Taxation," NBER Working Paper No. 17989.
- Piketty, Thomas and Emmanuel Saez.** 2013. "Optimal Labor Income Taxation," *Handbook of Public Economics*, Volume 5, (Amsterdam: North Holland).
- Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva.** 2011. "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," NBER Working Paper No. 17616
- Reutter, L. I., Stewart, M. J., Veenstra, G., Love, R., Raphael, D., and Makwarimba, E.** (2009). "Who Do They Think We Are, Anyway?": Perceptions of and Responses to Poverty Stigma". *Qualitative Health Research*, 19(3), 297-311.
- Roemer, John** 1998. *Equality of Opportunity*, Cambridge: Harvard University Press.
- Roemer, John et al.,** 2003. "To What Extent Do Fiscal Systems Equalize Opportunities for Income Acquisition Among Citizens?", *Journal of Public Economics*, 87, 539-565.
- Stantcheva, Stefanie** 2012 "Optimal Human Capital Policies and Redistribution over the Lifecycle", MIT Working Paper
- Stiglitz, Joseph.** 1987. "Pareto Efficient and Optimal Taxation and the New New Welfare Economics," *Handbook of Public Economics*, Volume 2, (Amsterdam: North Holland).
- Weinzierl, Matthew C.** 2012. "Why Do We Redistribute So Much But Tax So Little? The Principle of Equal Sacrifice and Optimal Taxation," Harvard Business School Working Paper, No. 12-64.
- Werning, Ivan.** 2007. "Pareto Efficient Taxation," MIT working paper
- Will, Jeffrey.** 1993. "The Dimensions of Poverty: Public Perceptions of the Deserving Poor," *Social Science Research* 22, 312-332.