Fairness and Paretian Social Welfare Functions

Kristoffer Berg* and Paolo G. Piacquadio†

August 19, 2022

Abstract

Treating individuals equitably is often more appropriate than treating them equally. Fairness is a key concern for individuals and policymakers, but practically absent from the analysis of second-best policies. We enrich the welfare-analysis toolbox by accommodating fairness concerns in Paretian social welfare functions, while including standard welfare criteria as special cases. To illustrate our axiomatic characterization, we investigate various fairness views in the context of labor-income taxation. Utilitarianism implicitly assumes individuals do not deserve their income opportunities; in contrast, our criteria allow any degree of deservingness. Our simulation analysis shows the US tax system is rationalized by a large degree of deservingness and little concern for progressivity.

JEL codes: D63, H21, I31.

Keywords: fairness, welfare criteria, generalized utilitarianism, optimal income taxation.

1 Introduction

Unequal allocations are not necessarily unfair. It is generally considered fair that: larger effort be rewarded with larger benefits; heavier polluters bear a larger share of the environmental costs; people with special needs or who suffered unfair treatment be compensated; etc. In fact, fairness is a key concern of individuals and policymakers.[1]

*Centre for Business Taxation, Saïd Business School, University of Oxford.

†School of Economics and Political Sciences, University of St. Gallen, and Department of Economics, University of Oslo (paolo.piacquadio@unisg.ch).

[1] A wide range of fairness views have emerged from recent experimental and survey-based research, including Schokkaert and Devooght (2003); Cappelen et al. (2007); Cavaillé and Trump (2015); Weinzierl (2014); Saez and Stantcheva (2016); Schokkaert and Tarroux (2021); Hvidberg et al. (2020); Stantcheva (2021).

Yet, fairness concerns are practically absent from the economic criteria adopted for the evaluation and design of second-best policies.[2]

In this paper, we generalize existing social welfare functions and enrich the toolbox of welfare analysis to account for fairness concerns. More precisely, we propose and axiomatically characterize a large family of welfare criteria that accommodate wide-spread fairness views, while including existing criteria—such as utilitarianism, generalized utilitarianism, and maximin—as special or limit cases.

Intuitively, fairness is achieved when individuals are treated the same, unless there are morally relevant reasons for a differential treatment. Needs, deserts, efforts, status quo, and special ties are often recognized as morally relevant. Yet, fairness is necessarily context specific, making it difficult to provide a general definition. In fact, we might consider it to be fair that those who contributed most to climate change exert more effort in combatting it; that individuals jointly share the burden of an economic crisis; that trade agreements deliver benefits to all signatories; that inheritances depend on the degree of kinship. In all these cases, there is some heterogeneity—past emission, pre-crisis situation, no-trade situation, degree of kinship—which might be considered morally relevant. Our approach allows for these differences to matter.

For the sake of simplicity and without loss of generality, we derive our main results in the context of income taxation. What is fairness in taxation? In "La decima scalata in Firenze nel 1497," Francesco Guicciardini defends the introduction of a progressive tax system. He argues a proportional tax is unfair: the poor will have to cut on necessities, while the rich will only reduce superfluous goods. As he writes: "*...the equality of the tax burden does not require that everyone pays pro rata as much as everybody else, but that the payment is such that it causes as much inconvenience to one as it does to all others.*"[3] These views define tax fairness based on both pre-tax and after-tax incomes, while standard welfare criteria consider pre-tax income morally irrelevant.[4]

[2]The earlier literature mainly focused on the selection of fair alternatives in specific domains (see Moulin, 2004; Thomson, 2011); yet, such approaches are not suited for the analysis of second-best policies. The closest approach to ours is developed by Fleurbaey and Maniquet (2006). However, their welfare criteria justify differential treatment of individuals only when individuals have different preferences (see also Fleurbaey and Maniquet 2011, 2018).

[3]Our translation from the Italian: "Però la egualità di una gravezza non consiste in questo; che ciascuno paghi per rata tanto l'uno quanto l'altro; ma, che el pagamento sia di sorte, che tanto s'incomodi l'uno, quanto l'altro" (Guicciardini, 1867).

[4]The importance of pre-tax income for tax fairness is shared by a large number of survey respondents (Schokkaert and Devooght, 2003; Weinzierl, 2014; Saez and Stantcheva, 2016; Schokkaert

Guicciardini's view on tax fairness was later supported by Mill (1848) and gained prominence as the equal-sacrifice principle (Cohen Stuart, 1889; Edgeworth, 1897; Pigou, 1928; Frisch, 1932; Vickrey, 1947; Musgrave, 1959). Yet, like other fair allocation rules, the equal-sacrifice principle is not readily applicable to modern optimal income taxation. Labor-supply responses and asymmetric information impose a constraint on feasible policies and restrict tax system to second-best optima. As a result, the equalization of individuals' sacrifices leaves efficiency gains unrealized (Berliant and Gouveia, 1993; da Costa and Pereira, 2014).

In this paper we address this limitation of fair allocation rules. We characterize the family of *fair social welfare functions*, which prioritize fair allocations while respecting the Pareto principle. When efficiency gains are not possible, the fair allocation is optimal. When efficiency gains are possible, our criterion trades off deviations from the fair allocation with efficiency. Thus, our results bridge the gap between fair allocation rules and social welfare functions. Our criteria can then be applied to optimal policies in models with behavioral responses and asymmetric information, such as the income taxation model of Mirrlees (1971).

Our social welfare functions are the sum of the integrals of individuals' losses from the fair allocation. These minimize individuals' losses—thus respecting efficiency—while prioritizing individuals with larger losses (or smaller gains)—thus respecting equity.

To illustrate, assume the social welfare function W takes the standard form of the sum of transformed individuals utilities

$$W = \sum_i g_i \left(u_i \left(c_i, -\ell_i \right) \right),$$

where c_i and ℓ_i denote consumption and labor supply, u_i is i's utility function, and g_i is a strictly increasing function such that $g_i \left(u_i \right)$ is strictly concave in its arguments.[5] These social welfare functions include most criteria adopted in the literature.

and Tarroux, 2021). In particular, Saez and Stantcheva (2016) ask who most deserves a $\$1,000$ tax break among the following families. Family A earns $\$50,000$ per year, pays $\$15,000$ in taxes, and hence nets out $\$35,000$; family B earns $\$40,000$ per year, pays $\$5,000$ in taxes, and hence nets out $\$35,000$. As they write "overall, subjects overwhelmingly find family A more deserving than family B" (p.43) and conclude that "this contradicts the basic utilitarian model".

[5]This functional form emerges from imposing that social preferences are Paretian, individualistic, and strictly convex. Strict convexity is more standard for individual preferences. It is similar in spirit, but significantly weaker, than the Pigou-Dalton transfer principle—holding that a transfer from a

Utilitarianism emerges when the g_i functions are the identity function. Generalized utilitarianism requires that the g_is are equal across individuals. Weighted utilitarianism is the case when g_is are constant individual-specific weights.[6] Our family of fair social welfare functions lead to new restrictions on the g_is, and emerge from three intuitive principles of distributive justice: optimality, weak progressivity, and horizontal equity.

First, *optimality* tells that, ruling out efficiency considerations, it is optimal to assign to each individual i her bundle at the fair allocation. Put differently, reallocating consumption across individuals from the fair allocation cannot improve social welfare. Importantly, the choice of the fair allocation remains an open ethical choice: this flexibility accommodates a broad applicability to various economic issues and the rich spectrum of rule identified by the fair allocation literature. Second, *weak progressivity* tells that it is socially more desirable to assign smaller tax burdens to individuals with smaller fair consumption. The intuition is that taking a dollar from an individual who is assigned her fair consumption c_i^* determines a smaller welfare cost than taking a dollar from another individual who is assigned her fair consumption c_j^* when $c_j^* < c_i^*$. Finally, *horizontal equity* tells that, for two individuals with $c_i^* > c_j^*$, it is socially undesirable if i ends up with a smaller consumption than j (similar to Feldstein, 1976; Rosen, 1978).

Our axioms jointly imply that there exists a comparable measure of individuals' losses—a *loss function* L with specific properties—which depends on each individual's assigned consumption c_i and fair consumption c_i^*. In fact, at fair labor supply ℓ_i^*, the value for society of giving one additional dollar to i—called i's social marginal welfare

poorer to a richer individual reduces welfare. The main difference is that the Pigou-Dalton transfer principle additionally imposes an anonymity condition, which rules out most fairness considerations and reasons for treating individuals differently.

[6]Kaplow and Shavell (2001) impose the restriction that social welfare only depends on utility levels and, thus, rule out that the g_i functions might depend on individuals situation and/or on fairness considerations. In contrast, the family of social welfare functions W also include the criteria and fairness concerns discussed by Fleurbaey and Maniquet (2018) and Heathcote and Tsujiyama (2021). An example of excluded criteria are the rank-dependent social welfare functions, as discussed by **?**. A broader approach is proposed by Saez and Stantcheva (2016): rather than defining a social welfare function, they suggest directly setting the social values of marginal changes in individuals' consumptions. Sher (2021) analyses the conditions for this "social marginal welfare weights" approach to violate the Pareto principle.

weight (Saez and Stantcheva, 2016)—is given by

$$\frac{\partial W}{\partial c_i} = \frac{\partial g_i\left(u_i\left(c_i, -\ell_i^*\right)\right)}{\partial c_i} = 1 + L\left(c_i, c_i^*\right)$$

and identifies the g_i functions. When $c_i = c_i^*$, the loss of individual i is 0 and a marginal change in her consumption has social value 1. When $c_i < c_i^*$, the loss of i is positive and a marginal increase in her consumption is socially more valuable. Thus, society places a larger priority on individuals with larger losses. In the absence of efficiency costs, the government reaches the first-best optimum by equalizing the levels of losses across individuals and chooses the fair allocation. With efficiency costs, society accepts some inequalities in losses to pursue efficiency gains.

Our welfare criteria are constructed to deal with second-best settings, such as the seminal model of optimal non-linear income taxation (Mirrlees, 1971). To illustrate our results and discuss policy implications, we conduct a standard optimal tax simulation for the US economy, following the exercise by Mankiw et al. (2009). In line with the literature, the utilitarian criterion recommends extensive redistribution with marginal tax rates above 60% (and up to 80%): it aims at equalizing marginal utilities of consumption, but is constrained by individuals' behavioral responses. When behavioral responses are small, efficiency losses are small and second-best redistribution is large. In contrast, the "opportunity-equalization social welfare functions" accommodate the view that individuals deserve a part—if not all—of their income opportunities. This reduces the scope for redistribution: the criterion aims at equalizing individuals' losses relative to their fair share. Our simulations show that second–best taxes are very responsive to different fairness views. With a large degree of deservingness and a small concern for progressivity, our criterion recommends marginal tax rates that are about 20 percentage points lower than the utilitarian optimum and roughly in line with the US tax system. The more opportunity-equalization is introduced and the larger the concern for progressivity, the smaller the gap with the utilitarian recommendation.

The rest of the paper is organized as follows. Section 2 illustrates our principles of distributive justice and main intuitions. Section 3 presents the framework. Section 4 formalizes the axioms and the fair social welfare function, provides the main characterization result, and presents a simple parametric family. Section 5 explores the implications of our criteria for optimal income taxation, including a simulation of

the optimal tax system for the US economy. Section 6 concludes. All the proofs are in the appendix.

2 Principles of distributive justice

2.1 Three principles

Our starting point is that there exists a fair allocation, which we denote a^*. The choice of fair allocations depends on the context, on the preferences of individuals, on the feasibility constraint and, not least, on ethical views (Moulin, 2004; Thomson, 2011). In this section, we take a^* as given and discuss how to rank allocations in light of the *fairness* of a^*. We discuss later the choice of a^* in the context of income taxation (Section 5) and in other domains (Section 6).

We propose three principles of distributive justice. The first principle tells that differences at the fair allocation a^* are morally justified. The fair allocation a^* ought to be top-ranked among all those allocations that can be obtained from a^* by a pure redistribution, that is, a reassignment of the available total income of the individuals.

Principle of optimality: *Pure redistribution from the fair allocation a^* cannot be welfare improving.*

The main novelty is to recognize that individuals may be treated differently at the fair allocation a^*. For example, at a^* there might be higher-income and lower-income individuals. If this is the case, these income differences are fair and, by the principle of optimality, a transfer from an income rich to an income poor is not called for.

We next consider situations where all individuals obtain a smaller income than at the fair allocation a^*.[7] Let the tax burden of an individual be the difference between her income at the fair allocation and her assigned income. Our second principle tells that it is unfair if higher-income individuals have a smaller tax burden than lower-income individuals.

Principle of weak progressivity: *Higher-income individuals deserve a larger tax burden than lower-income individuals. If this is not the case, decreasing the*

[7]One can define a similar property for situations where all individuals obtain a larger income than at the fair allocation. Focusing on losses is natural in the context of income taxation and, more generally, in settings where the fair allocation cannot be implemented.

tax burden of a low-income individual and correspondingly increasing that of a high-income individual improves social welfare.

The third principle tells that rerankings are unfair. If an individual deserves a higher income at a^*, she should not face such a large tax burden as to end up with a smaller after-tax income.

Principle of horizontal equity: *Lower-income individuals do not deserve a larger after-tax income than higher-income individuals. If this is the case, decreasing the tax burden of a high-income individual and correspondingly increasing that of a low-income individual improves social welfare.*

2.2 Income distributions, equity, and equal losses

We illustrate these principles for income distributions. Assume individuals i and j are assigned incomes c_i^* and c_j^* at the fair income distribution. Without loss of generality, let $c_i^* \geq c_j^*$. The principle of optimality tells that $\left(c_i^*, c_j^*\right)$ is socially at least as desirable as any other income distribution (c_i, c_j) with $c_i + c_j = c_i^* + c_j^*$, as represented in Figure 1.

The principle of weak progressivity deals with the unfair situations (c_i, c_j) where the tax burden of i is smaller than the tax burden of j, i.e., $c_i^* - c_i < c_j^* - c_j$ (the shaded triangle in the bottom-right of Figure 1). In such situations, a transfer of consumption from i to j cannot decrease social welfare, as illustrated by the arrow.

Finally, the principle of horizontal equity deals with the unfair situations (c_i, c_j) where the tax burden of i is so large that her consumption is smaller than that of j, i.e., $c_i < c_j$ (the shaded triangle in the top-left of Figure 1). In such situations, a transfer of consumption from j to i cannot decrease social welfare, again as illustrated by the arrow.

In contrast to fair allocation theory, our principles of justice do not identify when individuals i and j are treated equitably. Yet, if social welfare is averse to inequity—as we shall assume—we can narrow down such situations to the **area of justifiable inequalities**, as represented in Figure 1.[8] The intuition is as follows. When individuals are treated equitably, redistributing their incomes increases unfair inequality. By

[8]The same area emerges in the literature on the allocation of conflicting claims, see Aumann and Maschler (1985).

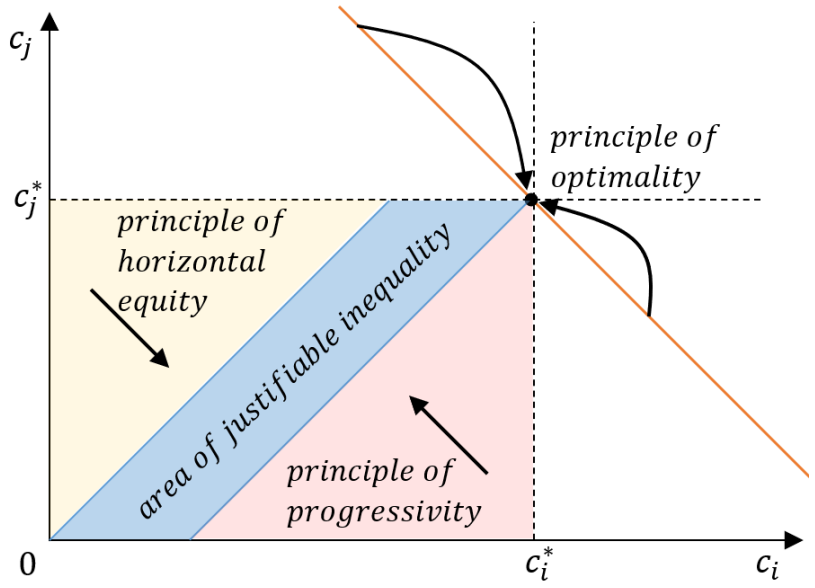

Figure 1: Fairness principles and the area of justifiable inequality.

inequity aversion, this reduces social welfare. Thus, by the principle of optimality, the fair income distribution $\left(c_i^*, c_j^*\right)$ is equitable. By the principle of weak progressivity, an equitable allocation cannot assign to i a smaller tax burden than that of j. By the principle of horizontal equity, an equitable allocation cannot assign to i a smaller consumption than that of j.

When social preferences are continuous, the set of equitable allocations constitutes a path which starts at the origin, stays in the area of justifiable inequalities, and reaches the fair income distribution $\left(c_i^*, c_j^*\right)$. We refer to the set of equitable allocations as allocations of **equal losses**.[9] The simplest example is the "proportional rule." It demands that each individual be assigned benefits and losses in proportion to the fair allocation. Then, the set of equitable allocations takes the form of a straight line from the origin through $\left(c_i^*, c_j^*\right)$. Other examples include the constrained equal-award rule—the most favorable path for j—and the constrained equal-loss rule—the most favorable path for i (Thomson, 2019). We follow Young (1988) and classify equal losses based on progressivity. Equal losses are **(relatively) progressive** if $c_i^* \geq c_j^*$ implies that the tax burden of i is proportionally larger than that of j, that is, $\left(c_i^* - c_i\right)/c_i^* \geq \left(c_j^* - c_j\right)/c_j^*$; these are **(relatively) regressive** when the tax burden

[9]In the context of the allocation of conflicting claims, the focus is more on the benefits—rather than on the losses—and such allocations constitute the "path of awards" of a rule (Thomson, 2019).

of i is proportionally smaller that that of j, that is, $(c_i^* - c_i)/c_i^* \leq \left(c_j^* - c_j\right)/c_j^*$. The **proportional** equal losses constitute the knife-edge case for which $(c_i^* - c_i)/c_i^* = \left(c_j^* - c_j\right)/c_j^*$.

Our first main contribution is to move from the choice of equitable allocations—those of equal losses in our terminology—to complete rankings. Our second main contribution is to generalize the above analysis from the one-dimensional setting—income distribution—to multidimensional settings with behavioral responses and heterogeneous preferences. In the next subsections, we illustrate these results.

2.3 The measurement of losses and social welfare

Our axioms imply there exists a **loss functions** L with specific properties. This function measures the loss of each individual based on the assigned and fair consumptions. Individuals have equal losses when $L(c_i, c_i^*) = L\left(c_j, c_j^*\right)$; the larger the consumption of an individual, the smaller her loss. Our loss function generalizes the "sacrifice rule" characterized by Young (1988), requiring that $L(c_i, c_i^*) = v(c_i^*) - v(c_i)$ for some "utility" function v, and includes as special cases most rules discussed in the context of conflicting claims (Thomson, 2019).

Our principles of justice force social preferences to prioritize individuals with larger losses: social welfare increases more when assigning a dollar of consumption to an individual with larger loss. Assume the social welfare function W is increasing in consumptions, continuous, and additively separable. Then, the social value created by giving a marginal increase in consumption to an individual—the social marginal welfare weight (Saez and Stantcheva, 2016)—is an increasing function of her loss. Thus, social preferences can be represented by the **fair social welfare function**

$$W \equiv \sum_i \int_0^{c_i} \left(1 + L(x, c_i^*)\right) dx.$$

Earlier attempts to extend the allocation of equal losses to complete rankings are problematic. The most prominent is to minimize the sum of individuals' losses (see, among others, Musgrave, 1959). Yet, this leads to prioritizing individuals with the largest *marginal loss* and not those with the largest *loss*. For example, when the loss of individual i is measured by Young (1988)'s proposal $L_i(c_i) = v(c_i^*) - v(c_i)$, minimizing the sum of losses turns out to be equivalent to maximizing the utilitarian

criterion $W^u = \sum_i v(c_i)$.[10]

2.4 Generalized utilitarianism

An important special case emerges when the fair allocation assigns the same consumption to all individuals. When $c_i^* = c_j^*$, the area of justifiable inequality shrinks to allocations where $c_i = c_j$. Since allocations of equal losses assign equal consumption, the loss functions are equal across individuals. As discussed, the fair social welfare function prioritizes individuals with larger losses. Here, this is equivalent to prioritizing individuals with smaller consumption. Thus, when the fair allocation a^* gives the same consumption to all individuals, the fair social welfare function is equivalent to the generalized utilitarian criterion. That is, there exists an increasing and concave transformation U such that

$$W = \sum_i U(c_i).$$

Our results thus clarifies that generalized utilitarianism is appropriate when there are no moral reasons—merits, deserts, efforts, etc.—for assigning different amounts to individuals.

2.5 Multi-commodity settings and heterogeneous preferences

In multi-commodity settings, two key challenges emerge. First, the welfare criterion should account for all the dimensions that matter for individuals, as not doing so leads to a violation of the Pareto principle. Second, the welfare criterion needs to establish how to compare individuals with heterogeneous preferences. Yet, existing proposals—such as Samuelson and Swamy (1974), Fleurbaey and Maniquet (2006), and Piacquadio (2017)—treat equal-preference individuals equally: these disregard their differences at the fair allocation a^* and thus violate our principles of justice.

Our solution is to introduce a ceteris-paribus condition. The principles of optimality, weak progressivity, and horizontal equity hold only when the non-income

[10]The more general proposal of minimizing the sum of convexly-transformed losses introduces some prioritiy to those individuals with largest losses. However, unless one takes the limit case of infinite convexity, these criteria violate the above-introduced fairness principles: these recommend moving away from the area of justifiable inequalities for no efficiency gains.

dimensions of individuals' bundles are fixed to the level of the fair allocation a^*. As we show, the Pareto principle and this ceteris-paribus condition address the challenges of multi-commodity settings.

To illustrate, consider allocations of consumption and labor supply. Let the fair allocations of i and j be $a_i^* = (c_i^*, -\ell_i^*)$ and $a_j^* = \left(c_j^*, -\ell_j^*\right)$. When labor supply is fixed at the level of the fair allocation, our above fairness principles hold and lead to the same loss function as before: social preferences prioritize individuals with larger loss.

What if labor supply is different? By the Pareto principle, we can compute the level of consumption that—at the fair-allocation level of labor supply—makes the individual indifferent. Then, we use such counterfactual consumption to find and compare the levels of losses across individuals. Said differently, the consumption reduction from the fair allocation, that is $c_i^* - c_i$, needs to be adjusted by i's willingness to pay for the the labor supply increase $\ell_i^* - \ell_i$. Let the equivalent consumption of i be measured by the function $e_i(c_i, \ell_i)$, defined by setting $e_i(c_i, \ell_i) = k$ if $u_i(c_i, -\ell_i) = u_i(k, -\ell_i^*)$. Then, the relevant consumption reduction for i is $c_i^* - e_i(c_i, \ell_i)$. The fair social welfare function identified by our principles is then

$$W = \sum_i \int_0^{e_i(c_i,\ell_i)} (1 + L(x, c_i^*))\, dx.$$

3 The framework

3.1 Model

The set of individuals is $I \subset \mathbb{N}$; it is finite and satisfies $|I| \geq 3$. Each individual $i \in I$ supplies labor $\ell_i \geq 0$ and consumes $c_i \geq 0$. Her preferences are represented by an additively separable utility function $u(c_i, -\ell_i)$, which is continuous, strictly increasing, and strictly concave in its arguments. We assume consumption is an essential good, that is, $\lim_{c\to 0} u_c = \infty$. Equal preferences is without loss of generality: the proof of the main result does not use this restriction.

An **allocation** $a \equiv \left(\{a_i\}_{i\in I}\right)$ specifies a bundle $a_i \equiv (c_i, -\ell_i)$ for each individual $i \in I$. Let A be the set of all allocations and $A^+ \subset A$ be the set of allocations with strictly positive consumptions.

3.2 Paretian and additively separable social welfare functions

We first restrict the attention to a broad class of welfare criteria; in the next section we formalize the concern for fairness. Social preferences $\succsim$ are a complete, transitive, and continuous binary relation on the set of allocations A. For each pair $a, a' \in A$, $a \succsim a'$ means that a is socially at least as desirable as a'. The asymmetric and symmetric counterparts of $\succsim$ are denoted $\succ$ and $\sim$. Social preferences can be represented by a continuous social welfare function $W : A \to \mathbb{R}$. Thus, for each pair $a, a' \in A$, $a \succsim a'$ if and only if $W(a) \geq W(a')$.

As standard, we require social preferences to satisfy the Pareto principle. In other words, if individuals are made better off, social welfare cannot decrease.

Efficiency: *For each pair $a, a' \in A$, if $u(a_i) \geq u(a'_i)$ for each $i \in I$ and $u(a_i) > u(a'_i)$ for some $i \in I$, then $a \succsim a'$.*

Next, we impose inequity aversion on social preferences by requiring social preferences to be strictly convex.[11]

Inequity aversion: *For each pair $a, a' \in A$ and each $\beta \in (0,1)$, $a \sim a'$ implies $\beta a + (1-\beta) a' \succ a$.*

Finally, we impose that social welfare comparisons do not depend on the bundle assigned to an unconcerned individual. Denote by (a_i, a_{-i}) the allocation $a \in A$ that assigns a_i to individual i and $a_{-i} \equiv (a_j)_{j \in I \setminus \{i\}}$ to the other individuals.

Separability: *For each $a, a' \in A$, each $i \in I$, and each $\bar{a}_i = \left(\bar{c}_i, -\bar{\ell}_i\right)$, $(a_i, a_{-i}) \succsim \left(a_i, a'_{-i}\right)$ if and only if $(\bar{a}_i, a_{-i}) \succsim \left(\bar{a}_i, a'_{-i}\right)$.*

By *efficiency*, society evaluates individuals through their own preferences: W can be written as a function of the utilities achieved by each individual. By *inequity aversion*, social preferences are strictly convex with respect to the allocation and, thus, W is strictly concave in its arguments. By *separability*, the assignment of individual i does

[11] Convexity is significantly weaker than what is generally assumed in the literature, where this condition is supplemented with some form of anonymity or symmetry. For example, the stronger axiom of "equal-preference transfer" in Piacquadio (2017) forces social preferences to treat all equal-preferences individuals equally; then, fairness considerations only matter for comparisons across individuals with different preferences. Strict convexity—rather than convexity—avoids a technical issue: when social preferences are only convex, inequalities may be socially irrelevant and fairness considerations may be disregarded.

not matter for how society trades off the utility of individuals j and k; thus, W is additively separable.

Lemma 1. *Social preferences $\succsim$ satisfy efficiency, inequity aversion, and separability if and only if $\succsim$ can be represented by a sum-of-utilities social welfare function $W : A \to \mathbb{R}$, defined by setting for each $a \in A$*

$$W(a) = \sum_{i \in I} g_i(u(a_i)), \tag{1}$$

for some real-valued functions $(g_i)_{i \in I}$ such that $g_i(u)$ is continuous, strictly increasing, and strictly concave for each $i \in I$.

Let the functions $(g_i)_{i \in I}$ be called **Pareto functions** to emphasize these generalize the standard Pareto weights. Here, the degree of freedom in the choice of the Pareto functions is essential to accommodating fairness views. As discussed in Section 2, our focus is to characterize rankings—and the corresponding social welfare functions—to evaluate and assess all feasible and non-feasible allocations. The main novelty is to study how the ranking can reflect the view that one specific allocation is fair—namely the **fair allocation** $a^* \in A^+$. Our results are valid for any choice of the fair allocation. In Section 5, we discuss a parametric family of fair allocations that ultimately allow us to capture the most influential ethical views expressed in the context of income taxation.

4 The fair social welfare function

4.1 Social aversion to unequal losses

In this subsection, we introduce axioms that: *(i)* identify how to compare individuals' bundles in terms of their fairness; *(ii)* discipline how social preferences ought to react to unfair allocations; and, jointly with the previous axioms, *(iii)* characterize the Pareto functions $(g_i)_{i \in I}$ consistent with social aversion to unfair inequalities.[12]

[12] A different approach is to start with a cardinally measurable and interpersonally comparable index of losses of individuals, such as the utility loss from the fair allocation. With such rich information, one could just require that social welfare decreases when losses are transfered from a low-loss individual to a high-loss individual. However, without a theory of how to measure losses at each allocation, the corresponding welfare criterion would not be applicable. Our approach is more

The first axiom captures the principle of optimality and establishes that—absent efficiency gains—the **fair allocation** a^* is optimal. We impose that, at unchanged labor-supply, the fair allocation is socially at least as desirable as those allocations obtained by reallocating consumption across individuals.

Optimality: *For each* $a \in A$ *such that* $\ell_i = \ell_i^*$ *for each* $i \in I$ *and* $\sum_i c_i = \sum_i c_i^*$, $a^* \succsim a$.

At the fair allocation, each individual i works ℓ_i^* and consumes c_i^*. At the bundle $(c_i, -\ell_i^*)$ with unchanged labor supply, i has a consumption c_i. Let the difference in consumption $b_i \equiv c_i^* - c_i$ be the **tax burden** of individual i.

Without loss of generality, assume individual i's consumption at the fair allocation a^* is larger than j's, that is, $c_i^* \geq c_j^*$. Since a^* is fair, this means that i has a rightful claim to a larger consumption than j. Assume—at unchanged labor supply—the tax burden of i is smaller than that of j, that is, $0 \leq b_i < b_j$. Then, while both individuals consume less than at the fair allocation, the different tax burdens exacerbate the differences in consumption between individuals. Consider now increasing further the tax burden of j, while further reducing that of i. Our next axiom—embodying the principle of weak progressivity—tells that such transfer of consumption from j to i is unfair and cannot improve social welfare.

Weak progressivity: *For each pair* $a, a' \in A$, *each pair* $i, j \in I$ *with* $c_i^* \geq c_j^*$, *and each* $\varepsilon > 0$, *such that:*

- $0 \leq b_i' + \varepsilon = b_i < b_j = b_j' - \varepsilon$;
- $\ell_i = \ell_i' = \bar{\ell}_i$ *and* $\ell_j = \ell_j' = \bar{\ell}_j$; *and*
- $a_k = a_k' = a_k^*$ *for each* $k \in I / \{i, j\}$;

then, $a \succsim a'$.

Weak progressivity deals with situations whereby the tax burden of individual i is too small relative to that of some other individual. Our next axiom deals with situations whereby the tax burden of individual i is too large and determines a change in individuals' relative standings at the fair allocation. This idea builds on the principle of

ambitious. Here, the measure of losses emerges endogenously from the axioms as a way to represent the social ranking of allocations. As we clarify in the discussion of the characterization result, our approach includes as a special case the utility-based definition of losses.

horizontal equity (Feldstein, 1976; Rosen, 1978; Auerbach and Hassett, 2002; Berg, 2021).

As before, individual i's consumption at the fair allocation is larger than j's, that is, $c_i^* \geq c_j^*$. At allocation $a \in A$, individual i's consumption is smaller than j's, that is, $c_i < c_j$; labor supply is that of the laissez-faire allocation. The tax burden of i is so large that her consumption turns out to be smaller than that of j. The next axiom imposes that further reducing i's consumption for the benefit of j cannot improve social welfare.

Horizontal equity: *For each pair $a, a' \in A$, each pair $i, j \in I$ with $c_i^* \geq c_j^*$, and each $\varepsilon > 0$, such that:*

- $c_i' + \varepsilon = c_i < c_j = c_j' - \varepsilon$;
- $\ell_i = \ell_i' = \bar{\ell}_i$ *and* $\ell_j = \ell_j' = \bar{\ell}_j$; *and*
- $a_k = a_k' = a_k^*$ *for each* $k \in I / \{i, j\}$;

then, $a \succsim a'$.

4.2 Loss function and equivalent consumption

Let individual $i \in I$ be assigned consumption c_i and the fair labor supply ℓ_i^*. Then, the **loss function** measures the loss made by i as a function of the consumption c_i and the laissez-faire consumption c_i^*. Formally, the loss function is a continuous function $L : \mathbb{R}_+ \times \mathbb{R}_{++} \to (-1, +\infty)$.[13] At the fair consumption, i's loss is normalized to zero: $L(c_i, c_i^*) = 0$ for $c_i = c_i^*$. The smaller the consumption, the larger the loss: $L(c_i, c_i^*) > L(c_i', c_i^*)$ if $c_i < c_i'$. The loss function is increasing in the fair consumption: of two individuals i and j with $c_i^* > c_j^*$, $L(c_i, c_i^*) > L\left(c_j, c_j^*\right)$ if $c_i = c_j$. Finally, equality of losses requires that individuals with higher fair consumption be assigned a larger tax burden: of two individuals i and j with $c_i^* > c_j^*$, $L(c_i, c_i^*) = L\left(c_j, c_j^*\right) > 0$ implies $c_i^* - c_i \geq c_j^* - c_j$. Let $\mathcal{L}$ be the domain of the loss functions.

The loss function measures and compares the losses of individuals when labor supply is at the levels of the fair allocation. To extend the comparison to all allocations, we can move along individuals' indifference curves up to the point where this level of

[13] A bounded negative loss function is necessary for the welfare criterion to satisfy *efficiency*, as explained later.

labor supply is obtained. Formally, let the **equivalent consumption of i at** a_i be $e_i(c_i, \ell_i)$, defined by setting

$$e_i(c_i, \ell_i) = k \iff u(c_i, -\ell_i) = u(k, -\ell_i^*).$$

4.3 The welfare criterion

The **fair social welfare function** $W^L : A \to \mathbb{R}$ is defined by setting for each allocation $a \in A$,

$$W^L(a) \equiv \sum_{i \in I} \int_0^{e_i(c_i, \ell_i)} (1 + L(x, c_i^*))\, dx. \tag{2}$$

where the $L \in \mathcal{L}$ is a loss function. Said differently, the fair social welfare function is the sum of the integral of individuals' (unit-translated) losses $1 + L(\cdot, c_i^*)$, evaluated at their equivalent consumption $e_i(c_i, \ell_i)$.

To illustrate the logic behind (2), consider the sum-of-utilities criteria of Lemma 1:

$$W(a) = \sum_{i \in I} g_i(u(c_i, -\ell_i)).$$

The fair social welfare function is uniquely identified by imposing that the Pareto functions $(g_i)_{i \in I}$ satisfy

$$\frac{\partial g_i(u(c_i, -\ell_i^*))}{\partial c_i} = 1 + L(c_i, c_i^*) \tag{3}$$

for each individual $i \in I$. The left hand side of (3) is i's social marginal welfare weight—i.e., the value for society of assigning a marginal increase of consumption to individual i—at the bundle $(c_i, -\ell_i^*)$.[14] Since L takes values in the interval $(-1, \infty)$, society gives positive value to increasing the consumption of all individuals. Moreover, society is indifferent between giving a dollar to any two individuals with the same loss. Finally, society prioritizes individuals making larger losses.

Integrating over consumption gives $g_i(u(c_i, -\ell_i^*)) = \int_0^{c_i} (1 + L(x, c_i^*))\, dx$ (up to an additive constant, which is welfare irrelevant). We can extend the definition to all allocations by moving along the indifference curve of each individual. By def-

[14] Note that extending condition (3) to all levels of labor supply leads to a violation of the Pareto principle, unless individuals' preferences are quasi-linear in consumption.

inition of equivalent consumption, $u(c_i, -\ell_i) = u(e_i(c_i, \ell_i), -\ell_i^*)$. It follows that $g_i(u(c_i, -\ell_i)) = \int_0^{e_i(c_i,\ell_i)} (1 + L(x, c_i^*))\, dx$, as required in (2).

Our first result shows our axioms uniquely characterize the family of fair social welfare functions.

Theorem 1. *Social welfare $\succsim$ satisfies efficiency, inequity aversion, separability, optimality, weak progressivity, and horizontal equity if and only if it can be represented by a fair social welfare function W^L.*

Remark 1. The result directly extends when preferences differ across individuals. In fact, the proof does not use the Mirrleesian assumption that utilities are equal across individuals. Intuitively, interpersonal comparisons of losses are established in terms of the consumption reduction (or increase) when labor supply is at the level of the fair allocation. Clearly, at fixed labor supply, how individuals trade off consumption and labor is irrelevant. To satisfy *efficiency,* all other bundles are assessed by moving along the indifference curves of individuals. This is the role of the equivalent consumption e_i. When utilities are individual specific, the definition of the equivalent consumption function simply changes to

$$e_i(c_i, \ell_i) = k \iff u_i(c_i, -\ell_i) = u_i(k, -\ell_i^*),$$

where u_i is the utility of individual i.

Remark 2. When individuals' preferences are not additively separable, further restrictions need to be introduced to ensure that the social welfare function W^L satisfies *inequity aversion.* In our setting, additive separability guarantees that

$$\int_0^{e_i(c_i,\ell_i)} (1 + L(x, c_i^*))\, dx$$

is a strictly concave representation of individuals' preferences. Without additive separability, one has to restrict the set of admissible loss functions in $\mathcal{L}$ to ensure strict concavity holds for each individual.[15]

[15] An alternative (but equivalent) approach is to introduce a concave transformation of these integrals. Then, the characterization result imposes that: (i) such transformation function g be equal across individuals; and (ii) the concavity of g is bounded below by ensuring strict concavity for all individuals. The existence of such a function is proven in Piacquadio (2017) for a more general setting.

Remark 3. Theorem 1 generalizes Young (1988)'s characterization of equal-sacrifice allocations to a complete ranking of all conceivable allocations, where equality of losses (sacrifice in Young's paper) is traded off with efficiency. Our results share with Young (1988) the adoption of a fairness justification of losses and its measurement. An alternative approach is to rely on individuals' utility function and measure losses by the absolute utility loss

$$L^a\left(c_i, c_i^*\right) = u\left(c_i^*, -\ell_i^*\right) - u\left(c_i, -\ell_i^*\right),$$

or the relative utility loss

$$L^r\left(c_i, c_i^*\right) = \frac{u\left(c_i^*, -\ell_i^*\right) - u\left(c_i, -\ell_i^*\right)}{u\left(c_i^*, -\ell_i^*\right)}.$$

Such an approach accommodates the views of early proponents of equality of sacrifice, but relies on cardinal utility information, which some may object to. Note that, to respect *efficiency*, the social marginal welfare weights need to be positive and impose restrictions on the admissible utility functions.[16]

4.4 A parametric family

Each social welfare function in our family is uniquely identified by the loss function $L \in \mathcal{L}$. Next, we propose a parametric specification of the loss function that is tractable and, at the same time, rich enough to capture two key ethical choices. First, the **degree of (relative) progressivity** is captured by the parameter p and identifies the allocations of equal losses. Second, the **degree of inequity aversion** is captured by the parameter γ and defines the trade-off between equality of losses and efficiency.

The parametric sacrifice function $L^{p,\gamma}$ is defined by

$$L^{p,\gamma}\left(c_i, c_i^*\right) \equiv \left(\frac{(1+c_i)^{1+p} - 1}{(1+c_i^*)^{1+p} - 1}\right)^{-\gamma} - 1. \tag{4}$$

[16]For absolute utility-based sacrifice, the utility function needs to be bounded above. With unbounded utility, individuals' sacrifice is unbounded below and social marginal welfare weights might be negative. In this case, well-being gains for sufficiently well-off individuals decrease social welfare. For relative utility-based sacrifice, we additionally need that utility at the fair allocation is positive.

Two individuals i and j make equal losses when $L^{p,\gamma}(c_i, c_i^*) = L^{p,\gamma}\left(c_j, c_j^*\right)$. Let v be the isoelastic function defined by setting $v(x) = (1+p)^{-1}\left((1+x)^{1+p} - 1\right)$ for each $x \in \mathbb{R}$. Then, equal losses requires that

$$\frac{v(c_i^*) - v(c_i)}{v(c_i^*)} = \frac{v\left(c_j^*\right) - v(c_j)}{v\left(c_j^*\right)}.$$

When $p = 0$, the function v is linear and equal losses are proportional to the fair allocation a^*. In the context of the claims problem, these allocations correspond to the proportional rule. When $p > 0$, the function v is convex and equal losses are progressive. At the limit for $p \to \infty$, the allocations of equal losses correspond to the constrained-equal award rule. When $p < 0$, the function v is concave and losses are regressive. At the limit for $p \to -\infty$, the allocations of equal losses correspond to the constrained-equal loss rule.

The parameter γ governs the social aversion to inequality of losses or, simply, inequity aversion. At the limit when $\gamma = 0$, society is indifferent to inequity and maximizes the sum of equivalent consumptions. As γ increases, society increases the weight placed on individuals with larger losses. At the limit for $\gamma = \infty$, society places infinite priority on the individual making the largest losses.

The **parametric fair social welfare function** is obtained by substituting $L^{p,\gamma}$ in (2). The special case of proportional losses ($p = 0$) takes the simple form[17]

$$W^{0,\gamma}(a) = \sum_{i \in I} (c_i^*)^{\gamma} \frac{e_i(c_i, \ell_i)^{1-\gamma}}{1-\gamma}. \tag{5}$$

5 Tax fairness and optimal income taxes

We next discuss optimal non-linear income taxation in the Mirrleesian model. Individuals differ by their labor skills, reflected in their wage rates: for each $i \in I$, let $w_i > 0$ denote the wage rate of individual i; her income is given by $y_i \equiv w_i \ell_i$.

Standard welfare criteria—such as utilitarianism and generalized utilitarianism—assume differences in wages do not justify treating individuals differently. In fact, if all individuals supply the same labor time, it is optimal to share the available income equally across individuals. The opposite viewpoint is taken by equal-sacrifice criteria,

[17] Berg and Piacquadio (2022) axiomatically characterize this criterion when the fair allocation is the laissez-faire.

which broadly argue that differences in incomes due to wages are instead justified and do not call for redistribution (Musgrave, 1959). According to equal sacrifice, if the government has no budget requirement, no taxes should be imposed.

In this section, we propose a parametric family of welfare criteria that bridge the gap between these two extremes. These criteria are the fair social welfare functions characterized in Theorem 1, for different choices of the fair allocation a^*.

5.1 Fair allocation

We suggest individuals only partially deserve the income opportunities provided by their wages. Said differently, there is a component of wages that individuals deserve—say due to past effort—and a component of wages that individuals do not deserve—say due to circumstances.[18]

Let the degree of opportunity equalization be $\theta \in [0,1]$. Then, each individual i with wage w_i deserves to freely choose their preferred bundle from the income opportunity $B_i^\theta \equiv \{(c_i, -\ell_i) \ \ s.t. \ c_i \leq (\theta w^m + (1-\theta)\, w_i)\, \ell_i\}$, where w^m is the median income. Said differently, individuals deserve only a fraction $1-\theta$ of the advantages/disadvantages provided by their income opportunities.[19] Let $\left(c_i^\theta, -\ell_i^\theta\right)$ be the bundle that maximizes each individual i's preferences over B_i^θ and let the corresponding allocation be $a^\theta \in A$.

When $\theta = 0$, the income opportunity B_i^0 is identified by the budget constraint $c_i \leq w_i \ell_i$. Then, the fair allocation $a^0 \in A$ is the laissez-faire allocation. When $\theta = 1$, the income opportunity B_i^1 is identified by $c_i \leq w^m \ell_i$ and is equal across individuals. In the current framework with homogeneous preferences, the preferred bundle $(c_i^1, -\ell_i^1)$ is also equal across individuals. Then, the fair allocation $a^1 \in \mathcal{A}$ is an allocation of perfect equality.[20]

[18]The importance of distinguishing between efforts and circumstances is recognized in the literature on equality of opportunity, pioneered by Roemer (1998). The main difference is that our approach respects individuals preferences (by *efficiency*), while equality of opportunity violates the Pareto principle.

[19]Our results generalize to group- or individual-specific degrees of wage equalization $\left(\{\theta_i\}_{i\in I}\right) \in [0,1]^{|I|}$. For example, if a category of individuals is discriminated agains or suffers from exploitation, or if some individuals' wages are not sufficient to participate to society, it could be appropriate to correct upwards their deserved income opportunities.

[20]A different but closely related fair allocation has been proposed by Kolm (2005) and later studied by Maniquet (2011) and, within a maximin social welfare function, by Fleurbaey and Maniquet (2018). According to Kolm's "equal-labor income equalization rule", individuals deserve their income opportunities plus a bonus or a penalty, depending on the difference between their wage and the

5.2 Opportunity-equalization social welfare function

We next present the welfare criteria identified by our axioms for the above parametrical choice of fair allocations.

The losses of individuals are now measured with respect to the fair allocation $a^\theta \in A$ for some $\theta \in [0,1]$. Similarly, the equivalent consumption of individual i is now computed at the fair labor supply ℓ_i^θ, i.e., $e_i^\theta(c_i, \ell_i) \equiv k$ iff $u(c_i, -\ell_i) = u\left(k, \ell_i^\theta\right)$.

Then, the **opportunity-equalization social welfare function** $W^\theta : A \to \mathbb{R}$ is defined by setting for each allocation $a \in A$,

$$W^\theta(a) \equiv \sum_{i \in I} \int_0^{e_i^\theta(c_i,\ell_i)} \left(1 + L\left(x, c_i^\theta\right)\right) dx, \tag{6}$$

for a loss function $L \in \mathcal{L}$. The opportunity-equalization social welfare functions make three ethical choices:

1. the degree of opportunity equalization θ, determining the income opportunities individuals deserve and, thus, the fair allocation a^*;

2. the progressivity of the loss function, identifying the allocations of equal losses; and

3. the social aversion to inequity, disciplining the trade-off between unequal losses and efficiency.

Few special cases are particularly interesting. When $\theta = 1$, the fair allocation requires income opportunities to be perfectly equalized. Then, the opportunity-equalization social welfare function W^1 is **generalized utilitarian** and can be expressed as

$$W^1 = \sum_{i \in I} g\left(u\left(c_i, -\ell_i\right)\right),$$

with $g(u(\cdot,\cdot))$ strictly concave in consumption and labor. **Utilitarianism** is the special case when g is linear.

---

average wage in the economy. Kolm's approach introduces a direct concern for poverty, but implies that different-wage individuals are always differently deserving, ruling out standard criteria such as utilitarianism. We leave to future work the extension of our criteria to account for poverty concerns.

When $\theta = 0$, the fair allocation a^0 is the laissez-faire allocation, whereby income opportunities are the budget set of individuals. The corresponding criterion is the **equal-sacrifice social welfare function** (see Berg and Piacquadio, 2022).

5.3 Optimal tax formulas

For our next results, we assume individuals' wage rates are continuously distributed according to the density function $f(w)$ on the interval $\left[w^b, w^t\right]$.[21] We assume the standard isoelastic utility function $u(c_i, -\ell_i) = \left(c_i^{1-\rho} - 1\right)/(1-\rho) - \alpha(\ell_i)^\sigma/\sigma$, with $\alpha, \rho, \sigma \geq 0$. The equivalent consumption of i is

$$e_i(c_i, \ell_i) = \left(c_i^{1-\rho} + \frac{\alpha(1-\rho)}{\sigma}\left((\ell_i^*)^\sigma - (\ell_i)^\sigma\right)\right)^{\frac{1}{1-\rho}}.$$

Let the tax function be denoted by $T : \mathbb{R} \to \mathbb{R}$. For each $i \in I$, after-tax income is $y_i - T(y_i)$. Let ε_w^c and ε_w^u denote the compensated and uncompensated labor supply elasticities of an individual with wage rate w at her optimal income y_w. Let $e(w)$ and $\bar{L}(w)$ be the equivalent income and loss of the individual with wage rate w. Let $u_c(w)$ and $e_c(w)$ be the marginal effect of a change in consumption on, respectively, the level of utility and on equivalent income. Finally, we denote by λ the Lagrangean multiplier of the revenue requirement in the government optimization problem. Following Saez (2001), we can now characterize the optimal non-linear income taxes.

Proposition 1. *The first-order condition for the optimal tax rate at income level y_w can be written as follows,*

$$\frac{T'(y_w)}{1 - T'(y_w)} = A(w)\, B(w),$$

where

$$A(w) \equiv \frac{1 + \varepsilon_w^u}{\varepsilon_w^c} \frac{u_c(w)}{w f(w)},\ B(w) \equiv \int_w^{w^t} \left(1 - \frac{\left(1 + \bar{L}(\tau)\right) e_c(\tau)}{\lambda}\right) \frac{f(\tau)}{u_c(\tau)} d\tau.$$

The only difference with Saez (2001)'s formula is that the social marginal welfare weights of each individual are here given by $\left(1 + \bar{L}(w)\right) e_c(w)$, as defined by our

[21] The continuity of the welfare criterion with respect to the types of individuals—here identified by their wage rate—ensures that the continuous-type versions of our criteria are well-defined.

criteria (and thus omit the proof). Consequently, the standard results of the literature apply, including no negative tax rates and the zero marginal tax rate at the upper limit. Moreover, the necessary condition is also sufficient when the single-crossing condition is satisfied, that is pre-tax income is non-decreasing in wage rates.

Diamond and Saez (2011) highlight that, for utilitarianism, marginal welfare weights converge to 0 for very high income levels (since their consumption levels are also very high due to binding self-selection constraints). This is not the case for the opportunity-equalization social welfare functions with $\theta < 1$. In fact, a non-negligible weight is assigned even to very high income levels. Hence, marginal tax rates for very high earners are lower for our criteria than for the generalized utilitarian ones. For lower income individuals, the differences between utilitarianism and our social welfare functions cannot be characterized algebraically. We thus turn to a numerical simulation.

5.4 Simulation exercise for the US economy

For the sake of comparability, we set the same utility parameters as in Mankiw et al. (2009): $\rho = 1.5$, $\alpha = 2.55$, and $\sigma = 3$ (further specifications are discussed in Appendix C).[22] We also adopt the same income distribution parameters (for the US in 2007). We deviate from their study by including an exogenous revenue requirement, R, set to 30% of total laissez-faire income.

We simulate the optimal non-linear income tax for several specifications of our opportunity-equalization social welfare function. Our baseline is the case of complete opportunity equalization ($\theta = 1$), corresponding to the standard generalized utilitarian criterion. We also consider the opposite extreme of no opportunity equalization ($\theta = 0$), when the criterion is the equal-sacrifice social welfare function, and an intermediate case of partial opportunity equalization ($\theta = 1/2$). For the form of the loss function, we adopt the parametric specification of Eq. (4). In our baseline specification, we assume proportional losses ($p = 0$) and logarithmic inequity aversion ($\gamma = 1$).

The results are summarized by the graphs in Figure 2, representing the marginal tax rate and the average tax rate for the above criteria. The optimal tax system

[22]We thank Gregory Mankiw, Matthew Weinzierl, and Danny Yagan for making their data and code available.

derived with the proportional-sacrifice social welfare function ($\theta = 0$, $p = 0$, $\gamma = 1$) is less redistributive than the one derived with the 50 percent opportunity-equalization social welfare function ($\theta = 1/2$, $p = 0$, $\gamma = 1$) and even less redistributive than the generalized utilitarian social welfare function ($\theta = 1$, $p = 0$, $\gamma = 1$). In particular, the utilitarian second-best policy supports marginal tax rates above 60% for all individuals (and up to 80% for high-income individuals). In contrast, the equal-sacrifice criterion supports marginal tax rates that are about 20 percentage points lower.[23]

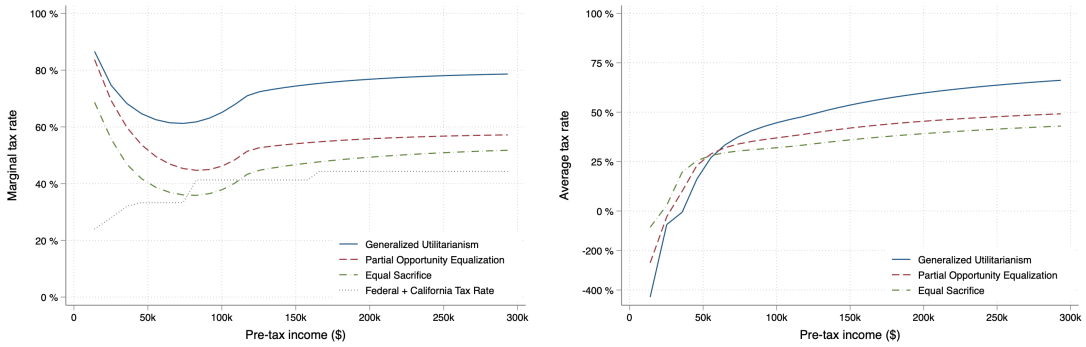

Figure 2: Marginal tax rates and average tax rates for the generalized utilitarian, partial opportunity-equalization, and equal-sacrifice criteria.

We also plot the 2007 marginal tax rates for the combined federal and California state taxes on the income of singles (since California sets the highest state income taxes, it provides the strongest case for the tax schedule implied by utilitarianism). The marginal tax rates derived from the proportional-sacrifice criterion are slightly more progressive than what is observed in the Californian tax system. Since all marginal top tax rates across US states rates are between the Federal and Californian systems, a regressive-loss function combined with no opportunity equalization (that is, equal sacrifice) seem necessary to explain the US income tax system in our setting. This finding is in line with Heathcote and Tsujiyama (2021): to rationalize the US income tax, a less redistributive criterion than utilitarianism is needed (even if no inequity aversion is assumed). Note that, for low incomes, the discrepancy of marginal tax rates with the US tax system may be partly explained by the absence of an

[23]The result of our simulation is a standard u-shaped marginal tax schedule for all criteria. The u-shape has recently been challenged by Heathcote and Tsujiyama (2021), who argue that an increasing marginal tax schedule is more in line with the empirical evidence on the size of the revenue requirement and the shape of the productivity distribution. In our setting, the u-shape is due to using the data from Mankiw et al. (2009), who also find a u-shaped schedule.

extensive labor supply margin (see the discussion in Diamond and Saez, 2011).[24]

The average tax rates are also informative. Utilitarianism suggests subsidies (negative average taxes) ought to be distributed to the bottom 35% of the population, while the proportional-sacrifice criterion does so for only the bottom 15% of the population. Contrary to what one might expect, the tax systems implied by our fair criteria redistribute on net to the lowest income levels. This is due to the presence of income effects, such that two contrasting forces define their second-best after-tax income. On the one hand, the negative loss (due to lump-sum transfer) suggests society ought to increase their taxes for the benefit of higher income individuals. On the other hand, the income effects amplify the welfare effect of changes in their after-tax income. Thus, while the equity motive suggests an additional dollar be given to high-skill individuals, the efficiency motive dominates for the lowest income earners. This result disappears when society is infinitely averse to inequity in losses ($\gamma \to \infty$), at which point the criterion requires minimizing the largest losses.

The degree of opportunity equalization determines what each individual deserves at the fair allocation. We next illustrate the role of the degree of progressivity, which determines how the tax burden—measured as the income reduction from the fair allocation at unchanged labor supply—should be shared across individuals. Above, we assumed that losses should be proportional, that is, absent efficiency costs, the tax burden should be proportionally equal across individuals ($p = 0$). We next compare it with the cases of progressive losses ($p = 1/3$) and regressive losses ($p = -1/3$).

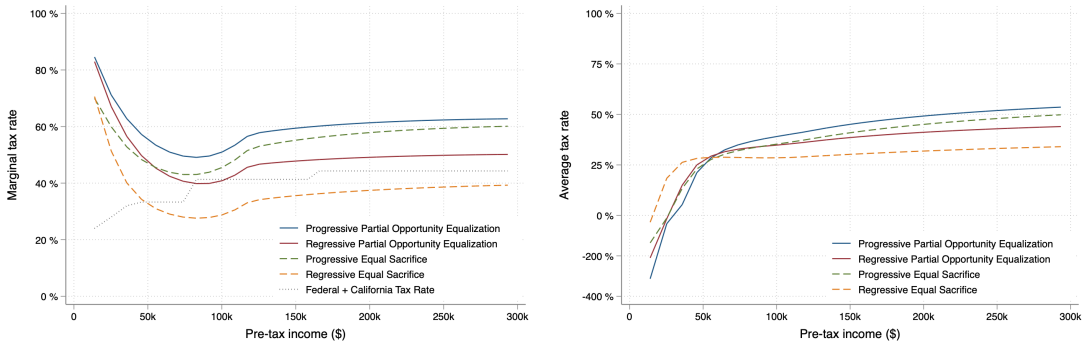

Figure 3: Progressive and regressive equal-sacrifice and partial opportunity-equalization criteria: marginal tax rates and average tax rates.

[24]When the extensive margin is accounted for, optimal marginal tax rates at the lowest income levels are lower (Saez, 2002; Jacquet, Lehman, and van der Linden, 2013).

As expected, demanding that losses be more progressive than proportional leads to more redistribution, while imposing that these be more regressive leads to less redistribution. This holds for both the partial opportunity-equalization and for the equal-sacrifice criteria. This figure also highlights that the effect of the degree of opportunity equalization and that of progressivity are different: opportunity equalization assigns a larger priority to the lower income individuals. Finally, a regressive definition of equal-sacrifice losses is now less redistributive than the Californian tax schedule.

The effect of inequity aversion and different specifications of the utility parameters are discussed in Appendix C.

6 Discussion

Treating individuals *equitably* is often more appropriate than treating them *equally*. Fairness considerations might justify that some individual is assigned more than another, even if they share the same preferences or utility function. Despite the important role of fairness in the public and political arenas, fairness views remain mostly absent from the economists toolbox of welfare analysis. In fact, virtually all existing social welfare functions—including utilitarianism, generalized utilitarianism, rank-dependent utilitarianism, maximin, etc.—trade off *efficiency* with *equality*, rather than with *equity*. In contrast, the theory of justice we develop here allows us to discuss the tradeoff between efficiency and equity.

Situations justifying different treatment are ubiquitous, as those discussed in the literature on fair allocation theory (Moulin, 2004; Thomson, 2011). These include the division of an estate between heirs, the dissolution of a partnership, the allocation of jobs to workers, etc. In comparison to fair allocation rules, our approach is more general and particularly suited to study economies with market imperfections, such as asymmetric information and externalities, where fair and efficient allocations are either not feasible or do not exist.

Consider the issue of allocating the cost of climate mitigation across countries. Standard equality-oriented criteria—such as utilitarianism—prescribe massive transfers of capital and consumption from rich countries to poor countries, mainly addressing the global inequality problem rather than the climate change externalities. To address this difficulty, Nordhaus and Yang (1996) propose to attach "Negishi weights"

(the inverse of marginal utilities) to the utilities of regions, thereby assigning larger weights to richer countries and avoiding the inequality-reducing transfers. Our family of fair social welfare functions include Negishi weighting as a special case and clarify that such criterion takes the *status quo* distribution across countries as fair.

The evaluation and design of trade agreements face similar ethical challenges. Standard criteria cannot accommodate the view that countries ought to share the benefits from trade. The mainstream approach is thus to limit the welfare analysis to (Pareto) efficiency or to adopt the perspective of one country. Similarly, when dealing with an economic crisis, the analysis of optimal policy intervention needs to confront the ethical question of the entitlement of individuals to the pre-crisis situation. Standard criteria are indifferent to people switching their situation, whereby some gain and other lose. Yet, a different view considers it morally wrong if someone benefits from the crisis, while others lose. Our criteria can accommodate such an aversion to unfair gains, while respecting efficiency.

While our results apply very generally, we develop our analysis in the context of labor-income taxation. The basic objection to equality-promoting social welfare functions was already laid out by Edgeworth (1897). When two equal-preference workers supply the same amount of labor, utilitarianism demands these earn the same after-tax income, independently of their wage rates (Mill, 1848; Musgrave, 1959; Saez and Stantcheva, 2016; Fleurbaey and Maniquet, 2018). As Feldstein (1976) clarifies, utilitarianism implicitly assumes that all differences in wage rates across individuals are undeserved: society jointly owns everyone's potential earnings.[25]

What if some wage differences are deserved? Our theory of justice captures these views. The main result of the paper is the axiomatic characterization of the family of fair social welfare functions. These criteria prioritize individuals making larger losses, measuring the relative stand of each individual's bundle compared to what said individual deserves. By allowing individuals to partially deserve the consumption-labor opportunities provided by their wage rates, our criteria can redeem utilitarianism's counterintuitive instances and thus can have large impacts on optimal tax policy.

[25] A related objection is the "slavery of the talented:" individuals should not be penalized for their talent (Musgrave, 1959). The utilitarian first best, however, requires the high-income individual to work more for a lower level of utility. For example, with additively separable utilities, after-tax income is equalized, while labor supply differs and penalizes the high-wage workers. This objection extends to prioritiarianism and is avoided only at the limit case with infinite priority to the worse-off, that is, with maxmin.

To speak to those impacts, we numerically simulate the optimal tax schedule in a standard Mirrlees model. Our criteria can justify a large range of tax schedules and help us understand the ethical views supporting observed tax schedules. In particular, we can adjust the extent to which individuals deserve their income opportunities, the social attitudes to progressivity, and the trade-off between equity and efficiency. The US income tax system belongs to the least redistributive schedules we can rationalize. In fact, even when individuals are regarded as entirely deserving their income opportunities, proportional losses, and intermediate inequity aversion, our simulated optimal tax system is more redistributive than the Californian one.

Acknowledgments: The authors wish to thank Geir Asheim, Ashley Craig, Marc Fleurbaey, Bård Harstad, Etienne Lehmann, Amy McCall, Magne Mogstad, Morten Håvarstein, Itai Sher, Yves Sprumont, Kjetil Storesletten, and Matt Weinzierl for comments, in addition to the seminar audiences at Welfare Economics Seminars, Deakin University, University of St. Andrews, University of Milano - Bicocca, Michigan University, CREST, BI. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme ERC Starting Grant VALURED (Grant agreement No. 804104).

References

Auerbach, A. J. and K. A. Hassett (2002). A new measure of horizontal equity. *The American Economic Review 92*(4), 1116–1125.

Aumann, R. J. and M. Maschler (1985). Game theoretic analysis of a bankruptcy problem from the talmud. *The Journal of Economic Theory 36*(2), 195–213.

Berg, K. (2021). Revealing inequality aversion from tax policy. *CBT Working Papers 2021-18*.

Berg, K. and P. G. Piacquadio (2022). Equal sacrifice and second-best taxation. *mimeo*.

Berliant, M. and M. Gouveia (1993). Equal sacrifice and incentive compatible income taxation. *The Journal of Public Economics 51*(2), 219–240.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2007). The pluralism of fairness ideals: An experimental approach. *The American Economic Review 97*(3), 818–827.

Cavaillé, C. and K.-S. Trump (2015). The two facets of social policy preferences. *The Journal of Politics 77*(1), 146–160.

Cohen Stuart, A. J. (1889). *Bijdrage tot de Theorie der Progressieve Inkomstenbelastung*. Amsterdam: Martinus Nijhoff.

da Costa, C. E. and T. Pereira (2014). On the efficiency of equal sacrifice income tax schedules. *The European Economic Review 70*, 399–418.

Diamond, P. and E. Saez (2011). The case for a progressive tax: from basic research to policy recommendations. *The Journal of Economic Perspectives 25*(4), 165–90.

Edgeworth, F. Y. (1897). The pure theory of taxation. *The Economic Journal 7*(25), 46–70.

Feldstein, M. (1976). On the theory of tax reform. *The Journal of Public Economics 6*(1-2), 77–104.

Fleurbaey, M. and F. Maniquet (2006). Fair income tax. *The Review of Economic Studies 73*(1), 55–83.

Fleurbaey, M. and F. Maniquet (2011). *A Theory of Fairness and Social Welfare*, Volume 48. Cambridge University Press.

Fleurbaey, M. and F. Maniquet (2018). Optimal income taxation theory and principles of fairness. *The Journal of Economic Literature 56*(3), 1029–79.

Frisch, R. (1932). Méthodes nouvelles pour mesurer l'utilité marginale. *Revue d'Économie Politique 46*(1), 1–28.

Golosov, M., M. Graber, M. Mogstad, and D. Novgorodsky (2021). How americans respond to idiosyncratic and exogenous changes in household wealth and unearned income. *NBER: WP 29000*.

Gorman, W. M. (1968). The structure of utility functions. *The Review of Economic Studies 35*(4), 367–390.

Guicciardini, F. (1867). La decima scalata in firenze. *Opere Inedite 10*, 382–388.

Heathcote, J. and H. Tsujiyama (2021). Optimal income taxation: Mirrlees meets ramsey. *The Journal of Political Economy 129*(11).

Hvidberg, K. B., C. Kreiner, and S. Stantcheva (2020). Social position and fairness views. *NBER: WP 28099*.

Jacquet, L., E. Lehmann, and B. Van der Linden (2013). Optimal redistributive taxation with both extensive and intensive responses. *The Journal of Economic Theory 148*(5), 1770–1805.

Kaplow, L. and S. Shavell (2001). Any non-welfarist method of policy assessment violates the pareto principle. *The Journal of Political Economy 109*(2), 281–286.

Kolm, S.-C. (2005). *Macrojustice: The Political Economy of Fairness*. Cambridge University Press.

Maniquet, F. (2011). An axiomatic study of the elie allocation rule. In *On Kolm's Theory of Macrojustice*, pp. 189–206. Springer.

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009). Optimal taxation in theory and practice. *The Journal of Economic Perspectives 23*(4), 147–74.

Mill, J. S. (1848). *Principles of Political Economy*. London: John W. Parker.

Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*(2), 175–208.

Moulin, H. (2004). *Fair Division and Collective Welfare*. MIT press.

Musgrave, R. (1959). *Theory of Public Finance: A Study in Public Economy*. New York: McGraw-Hill.

Nordhaus, W. D. and Z. Yang (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *The American Economic Review*, 741–765.

Piacquadio, P. G. (2017). A fairness justification of utilitarianism. *Econometrica 85*(4), 1261–1276.

Pigou, A. C. (1928). *A Study in Public Finance*. Read Books Ltd.

Roemer, J. E. (1998). *Theories of Distributive Justice*. Harvard University Press.

Rosen, H. S. (1978). An approach to the study of income, utility, and horizontal equity. *The Quarterly Journal of Economics*, 307–322.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies 68*(1), 205–229.

Saez, E. (2002). Optimal income transfer programs: intensive versus extensive labor supply responses. *The Quarterly Journal of Economics 117*(3), 1039–1073.

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *The American Economic Review 106*(1), 24–45.

Samuelson, P. A. and S. Swamy (1974). Invariant economic index numbers and canonical duality: survey and synthesis. *The American Economic Review 64*(4), 566–593.

Schokkaert, E. and K. Devooght (2003). Responsibility-sensitive fair compensation in different cultures. *Social Choice and Welfare 21*(2), 207–242.

Schokkaert, E. and B. Tarroux (2021). Empirical research on ethical preferences: how popular is prioritarianism? In M. D. Adler and O. F. Norheim (Eds.), *Prioritarianism in Practice*. Cambridge University Press.

Sher, I. (2021). Generalized social marginal welfare weights imply inconsistent comparisons of tax policies. *arXiv preprint arXiv:2102.07702*.

Stantcheva, S. (2021). Understanding tax policy: How do people reason? *The Quarterly Journal of Economics 136*(4), 2309–2369.

Thomson, W. (2011). Fair allocation rules. In *Handbook of Social Choice and Welfare*, Volume 2, pp. 393–506. Elsevier.

Thomson, W. (2019). *How to Divide When There Isn't Enough: From Aristotle, the Talmud, and Maimonides to the Axiomatics of Resource Allocation*, Volume 62. Cambridge University Press.

Vickrey, W. (1947). *Agenda for Progressive Taxation*. New York: Ronald Press.

Weinzierl, M. (2014). The promise of positive optimal taxation: Normative diversity and a role for equal sacrifice. *The Journal of Public Economics 118*, 128–142.

Young, H. (1988). Distributive justice in taxation. *The Journal of Economic Theory 44*(2), 321–335.

A Proof of Lemma 1

Inequity aversion implies that any subset of individuals is "strictly essential:" for each $I' \subseteq I$ and each $\{a_i^*\}_{i \in I \backslash I'}$, allocations $a, a' \in A$ with $a_i = a_i' = a_i^*$ for each $i \in I \backslash I'$ are not all indifferent. By continuity of the social preferences, *separability*, and *strict essentiality*, Theorem 1 and Theorem 2 in Gorman (1968) apply and prove the existence of a representation $W(a) = \sum_{i \in I} H_i(c_i, -\ell_i)$, where H_i is continuous for each $i \in I$. By *efficiency*, $H_i(c_i, -\ell_i)$ is an order preserving transformation of $u(c_i, -\ell_i)$. By *inequity aversion*, $H_i(c_i, -\ell_i)$ is strictly concave in its arguments. Thus, there exist a continuous function g_i such that $H_i(c_i, -\ell_i) = g_i(u(c_i, -\ell_i))$. Substituting gives the result.

B Proof of Theorem 1

Part 1. We first show the *fair social welfare function* satisfies the axioms. Let the fair allocation be $a^* \in A_+$and let

$$W^L(a) \equiv \sum_{i \in I} \int_0^{e_i(c_i, \ell_i)} (1 + L(x, c_i^*))\, dx,$$

for some loss function $L \in \mathcal{L}$.

Efficiency. By definition of the loss function, $1 + L > 0$ at any allocation $a \in A$. By definition of equivalent consumption, $e_i(c_i, \ell_i)$ is a representation of the utility function of i with codomain $[0, \infty)$. Thus, for each $i \in I$ and each pair a_i, a_i', $u(a_i) \geq u(a_i')$ if and only if

$$\int_0^{e_i(c_i, \ell_i)} (1 + L(x, c_i^*))\, dx \geq \int_0^{e_i(c_i', \ell_i')} (1 + L(x, c_i^*))\, dx.$$

Consider a pair of allocations $a, a' \in A$ such that $u(a_i) \geq u(a_i')$ for each $i \in I$ and $u(a_i) > u(a_i')$ for some $i \in I$. It follows that $W^L(a) > W^L(a')$ and $a \succ a'$, proving that the criterion satisfies *efficiency*.

Inequity aversion. Let $i \in I$. Since $L(x, c_i^*)$ is strictly decreasing in x,

$$\int_0^{e_i(c_i, \ell_i)} (1 + L(x, c_i^*))\, dx$$

is strictly concave with respect to $e_i(c_i, \ell_i)$. Moreover, by concavity and additive separability of the utility function, $e_i(c_i, \ell_i)$ is concave with respect to c_i and convex with respect to ℓ_i. It follows that

$$\int_0^{e_i(c_i,\ell_i)} (1 + L(x, c_i^*))\, dx$$

is strictly concave with respect to $(c_i, -\ell_i)$ and, thus, also W^L. This proves *inequity aversion* holds.

Separability. *Separability* follows from the additivity of the function W^L: the bundle of an unconcerned individual is irrelevant for the ranking of two allocations.

Optimality. Let $a \in A$ be such that $\ell_i = \ell_i^*$ for each $i \in I$ and $\sum_i c_i = \sum_i c_i^*$. By definition of the criterion,

$$W^L(a^*) - W^L(a) \;=\; \textstyle\sum_{i\in I} \left[\int_{c_i}^{c_i^*} (1 + L(x, c_i^*))\, dx\right].$$

Since $L(x, c_i^*)$ is strictly decreasing in x, $\int_0^{c_i} (1 + L(x, c_i^*))\, dx$ is strictly concave in c_i. Thus,

$$\sum_{i\in I} \left[\int_{c_i}^{c_i^*} (1 + L(x, c_i^*))\, dx\right] > \sum_{i\in I} \left[(1 + L(c_i^*, c_i^*))(c_i^* - c_i)\right].$$

Since $L(c_i^*, c_i^*) = 0$ for each $i \in I$ and $\sum_i (c_i^* - c_i) = 0$, $W^L(a^*) - W^L(a) > 0$ and proves *optimality* holds.

Weak progressivity. Consider a pair of allocations $a, a' \in A$ satisfying the requirements in the definition of *weak progressivity*: for some pair of individuals $i, j \in I$ with $c_i^* \geq c_j^*$ and some $\varepsilon > 0$: $0 \leq b_i' + \varepsilon = b_i < b_j = b_j' - \varepsilon$; $\ell_i = \ell_i' = \ell_i^*$ and $\ell_j = \ell_j' = \ell_j^*$; and $a_k = a_k' = a_k^*$ for each $k \in I/\{i, j\}$. Substituting for the definition of tax burden, we obtain that $c_i' = c_i + \varepsilon$ and $c_j' = c_j - \varepsilon$. Since individuals $k \in I/\{i, j\}$ are unaffected, the difference in social welfare is

$$\begin{aligned} W^L(a) - W^L(a') \;=\; & \textstyle\int_0^{c_i} (1 + L(x, c_i^*))\, dx - \int_0^{c_i+\varepsilon} (1 + L(x, c_i^*))\, dx + \\ & + \textstyle\int_0^{c_j} \left(1 + L\left(x, c_j^*\right)\right) dx - \int_0^{c_j-\varepsilon} \left(1 + L\left(x, c_j^*\right)\right) dx \end{aligned}$$

Now, by first-degree Taylor expansion and concavity of social welfare,

$$\int_0^{c_i+\varepsilon} \left(1+L\left(x,c_i^*\right)\right) dx \leq \int_0^{c_i} \left(1+L\left(x,c_i^*\right)\right) dx + \varepsilon\left(1+L\left(c_i,c_i^*\right)\right)$$

and

$$\int_0^{c_j-\varepsilon} \left(1+L\left(x,c_j^*\right)\right) dx \leq \int_0^{c_j} \left(1+L\left(x,c_j^*\right)\right) dx - \varepsilon\left(1+L\left(c_j,c_j^*\right)\right).$$

Thus,

$$W^L(a) - W^L(a') \geq \varepsilon\left[L\left(c_j,c_j^*\right) - L\left(c_i,c_i^*\right)\right].$$

Finally, since $c_i \leq c_i^*$, $L\left(c_i,c_i^*\right) \geq 0$. Furthermore, $c_i - c_j \geq c_i^* - c_j^*$. Thus, by the definition of L, $L\left(c_i,c_i^*\right) \leq L\left(c_j,c_j^*\right)$. Thus, $W^L(a) \geq W^L(a')$ and $a \succsim a'$. This proves *weak progressivity* holds.

Horizontal equity. Consider a pair of allocations $a, a' \in A$ satisfying the requirements in the definition of *horizontal equity*: for some pair of individuals $i, j \in I$ with $c_i^* \geq c_j^*$ and some $\varepsilon > 0$: $c_i' + \varepsilon = c_i < c_j = c_j' - \varepsilon$; $\ell_i = \ell_i' = \ell_i^*$ and $\ell_j = \ell_j' = \ell_j^*$; and $a_k = a_k' = a_k^*$ for each $k \in I/\{i,j\}$. Since individuals $k \in I/\{i,j\}$ are unaffected, the difference in social welfare is

$$\begin{aligned} W^L(a) - W^L(a') \;=\; & \int_0^{c_i} \left(1+L\left(x,c_i^*\right)\right) dx - \int_0^{c_i-\varepsilon} \left(1+L\left(x,c_i^*\right)\right) dx + \\ & + \int_0^{c_j} \left(1+L\left(x,c_j^*\right)\right) dx - \int_0^{c_j+\varepsilon} \left(1+L\left(x,c_j^*\right)\right) dx \end{aligned}$$

As before, by first-degree Taylor expansion and concavity of social welfare,

$$\int_0^{c_i-\varepsilon} \left(1+L\left(x,c_i^*\right)\right) dx \leq \int_0^{c_i} \left(1+L\left(x,c_i^*\right)\right) dx - \varepsilon\left(1+L\left(c_i,c_i^*\right)\right)$$

and

$$\int_0^{c_j+\varepsilon} \left(1+L\left(x,c_j^*\right)\right) dx \leq \int_0^{c_j} \left(1+L\left(x,c_j^*\right)\right) dx + \varepsilon\left(1+L\left(c_j,c_j^*\right)\right).$$

Thus,

$$W^L(a) - W^L(a') \geq \varepsilon\left[L\left(c_i,c_i^*\right) - L\left(c_j,c_j^*\right)\right].$$

Finally, L is decreasing in the first argument and increasing in the second: $c_i < c_j$ and $c_i^* \geq c_j^*$ imply that $L\left(c_i,c_i^*\right) > L\left(c_j,c_j^*\right)$. Thus, $W^L(a) \geq W^L(a')$ and $a \succsim a'$. This proves *horizontal equity* holds.

Part 2. We now show social preferences satisfying the axioms admit a representation by means of a *fair social welfare function.* The proof is organized in steps.

Step 1. *Assume social preferences $\succsim$ satisfy the axioms. Then, there exists real-valued increasing and strictly concave functions $(h_i)_{i\in I}$ such that social welfare W representing $\succsim$ is defined by setting for each $a \in A$,*

$$W(a) \equiv \sum_{i\in I} h_i\left(e_i\left(c_i, \ell_i\right)\right).$$

Proof. By Lemma 1, there exist real-valued increasing functions $(g_i)_{i\in I}$ such that $g_i\left(u\left(c_i, -\ell_i\right)\right)$ is strictly concave in its arguments for each $i \in I$ and such that, for each pair $a, a' \in A$, $a \succsim a'$ if and only if

$$W(a) = \sum_{i\in I} g_i\left(u\left(c_i, -\ell_i\right)\right) \geq \sum_{i\in I} g_i\left(u\left(c_i', -\ell_i'\right)\right) = W\left(a'\right).$$

Next, for each $i \in I$, $e_i\left(c_i, \ell_i\right)$ is the consumption-equivalent representation of preferences of i. Since e_i is continuous, there exists a real-valued increasing and continuous function h_i such that $h_i\left(e_i\left(c_i, \ell_i\right)\right) = g_i\left(u\left(c_i, -\ell_i\right)\right)$ for each $\left(c_i, -\ell_i\right)$. Since $e_i\left(c_i, \ell_i^*\right) = c_i$ for each $c_i \in \mathbb{R}_+$, h_i is strictly concave. This shows social preferences can be represented by the social welfare function W, as defined above. □

Let $\bar{A}$ be the set of allocations $a \in A$ such that $\ell_i = \ell_i^*$ for each $i \in I$. Then, by the definition of consumption equivalent functions, for each $a \in \bar{A}$, $W(a) = \sum_{i\in I} h_i\left(c_i\right)$. Let the choice correspondence C be defined as follows: for each $k \geq 0$, $C(k)$ is the set of consumption vectors $(c_i)_{i\in I}$ with $\sum_{i\in I} c_i \leq k$ that maximize W. Let $k^* \equiv \sum_{i\in I} c_i^*$. The following steps characterize the properties of C (with a slight abuse of notation, we shall use C also to denote the choice function, after showing the correspondence C is single-valued).

Step 2. *The choice correspondence C satisfies the following properties:*

1. *it is non-empty, single-valued, and continuous with respect to k;*
2. *it is strictly monotonic, $k > k'$ implies $C(k) \gg C\left(k'\right)$;*
3. *$C\left(k^*\right) = \left(c_i^*\right)_{i\in I}$;*
4. *$(c_i)_{i\in I} = C(k)$ implies $c_i > c_j \iff c_i^* > c_j^*$ for each $i, j \in I$;*

5. *for* $k < k^*$, $(c_i)_{i\in I} = C(k)$ *implies* $c_i - c_j \leq c_i^* - c_j^*$ *for each* $i, j \in I$.

Proof. 1. W is increasing, continuous, and strictly concave, and so is $\sum_{i\in I} h_i(c_i)$. Thus, the choice correspondence C is non-empty, single-valued, and continuous with respect to k.

2. For each $i \in I$ and each $c_i \in \mathbb{R}_+$, denote $h_i'(c_i^-)$ and $h_i'(c_i^+)$ the left and right first-order derivatives, respectively, of h_i at c_i. Let $(c_i)_{i\in I} = C(k)$ and $(c_i')_{i\in I} = C(k')$. By contradiction of strict monotonicity, assume $k > k'$ and $C(k) \not\gg C(k')$. Then, there exists a pair of individuals $i, j \in I$ such that $c_i' \leq c_i$ and $c_j' > c_j$. At the optima, $h_i'(c_i^-) \geq h_j'(c_j^+)$ and $h_i'(c_i^+) \leq h_j'(c_j^-)$ and, similarly, $h_i'(c_i'^-) \geq h_j'(c_j'^+)$ and $h_i'(c_i'^+) \leq h_j'(c_j'^-)$. By strict concavity, $h_i'(c_i'^-) \geq h_i'(c_i'^+) \geq h_i'(c_i^-) \geq h_i'(c_i^+)$ and $h_j'(c_j^-) \geq h_j'(c_j^+) > h_j'(c_j'^-) \geq h_j'(c_j'^+)$. Combining these conditions leads to the following contradiction:

$$h_i'(c_i^-) \geq h_j'(c_j^+) > h_j'(c_j'^-) \geq h_i'(c_i'^+) \geq h_i'(c_i^-).$$

3. Given single-valuedness of C, $C(k^*) = (c_i^*)_{i\in I}$ directly follows from *optimality*.

4. By contradiction, assume$(c_i)_{i\in I} = C(k)$ with $c_i < c_j$ and $c_i^* > c_j^*$ for some k and some $i, j \in I$. Let $a' \in \bar{A}$ be such that: $c_i + \varepsilon = c_i' < c_j' = c_j - \varepsilon$ for some $\varepsilon > 0$; and $c_k' = c_k$ for each $k \neq i, j$. Let $a'', a''' \in A$ be obtained from a and a' respectively by assigning the fair bundle a_k^* to each $k \neq i, j$. By *horizontal equity*, $a''' \succsim a''$. By *separability*, $a' \succsim a$. By construction, $\sum_i c_i = \sum_i c_i^*$. Thus, by single-valuedness of C, $(c_i)_{i\in I} \neq C(k)$: this is a contradiction.

5. By contradiction, assume$(c_i)_{i\in I} = C(k)$ with $c_i - c_j > c_i^* - c_j^*$ for some $k < k^*$ and some $i, j \in I$ and, without loss of generality, assume $c_i^* \geq c_j^*$. In terms of tax burdens, $\bar{c}_i - c_i = b_i < b_j = \bar{c}_j - c_j$. Let $a' \in \bar{A}$ be such that $b_i + \varepsilon = b_i' < b_j' = b_j - \varepsilon$ for some $\varepsilon > 0$; and $c_k' = c_k$ for each $k \neq i, j$. Let $a'', a''' \in A$ be obtained from a and a' respectively by assigning the fair bundle a_k^* to each $k \neq i, j$. By *weak progressivity*, $a''' \succsim a''$. By *separability*, $a' \succsim a$. By construction, $\sum_i c_i = \sum_i c_i^*$. Thus, by single-valuedness of C, $(c_i)_{i\in I} \neq C(k)$: this is a contradiction. □

Before constructing the loss function L and the welfare criterion (Step 4), we introduce a "pseudo-loss function" $\bar{L}$ and show its properties (Step 3).

Let $\bar{L} : \mathbb{R}_+ \times \mathbb{R}_{++} \to \mathbb{R}$ satisfy the following properties. First, for each $i \in I$ and each $k \geq 0$, let $\bar{L}(c_i, c_i^*) = k^* - k$ whenever c_i is i's consumption at the allocation

$(c_j)_{j\in I} = C(k)$. Second, we complete the $\bar{L}$ function *"linearly"* for non-observed levels of fair consumption as specified hereafter.

Reorder individuals in increasing order of fair consumption: individual i is ranked before j, denoted $(i) \leq (j)$, if $c_i^* \leq c_j^*$. Then, $c_{(i)}$ and $c_{(i)}^*$ denote the consumption and fair consumption of the ith-ranked individual. Let $y \in \mathbb{R}_{++}$. Three cases are possible: either (1) there exists $i \in I$ with $c_{(i-1)}^* \leq y \leq c_{(i)}^*$ or (2) $y < c_{(1)}^*$ or (3) $y > c_{(|I|)}^*$.

Case 1. Let $i \in I$ be such that $c_{(i-1)}^* \leq y \leq c_{(i)}^*$ and let $\alpha \in [0,1]$ be such that $y = \alpha c_{(i-1)}^* + (1-\alpha) c_{(i)}^*$. Then, for each $x \in \mathbb{R}_+$, $\bar{L}(x,y) = \bar{L}\left(c_{(i)-1}, c_{(i)-1}^*\right) = \bar{L}\left(c_{(i)}, c_{(i)}^*\right)$ if and only if $x = \alpha c_{(i)-1} + (1-\alpha) c_{(i)}$.

Case 2. Let $\alpha \in (0,1)$ be such that $y = \alpha c_{(1)}^*$. Then, for each $x \in \mathbb{R}_+$, $\bar{L}(x,y) = \bar{L}\left(c_{(1)}, c_{(1)}^*\right)$ if and only if $x = \alpha c_{(1)}$.

Case 3. Let $i \in I \bigcup \{0\}$ be the individual with second-largest fair consumption $c_{(i)}^* \neq c_{(|I|)}^*$, if it exists; otherwise set $c_{(0)}^* = 0$. Let $\alpha > 1$ be such that $y = \alpha c_{(|I|)}^* + (1-\alpha) c_{(i)}^*$. If $(i) > 0$, for each $x \in \mathbb{R}_+$, $\bar{L}(x,y) = \bar{L}\left(c_{(i)}, c_{(i)}^*\right) = \bar{L}\left(c_{(|I|)}, \bar{c}_{(|I|)}\right)$ if and only if $x = \alpha c_{(|I|)} + (1-\alpha) c_{(i)}$. If $(i) = 0$, for each $x \in \mathbb{R}_+$, $\bar{L}(x,y) = \bar{L}\left(c_{(|I|)}, \bar{c}_{(|I|)}\right)$ if and only if $x = \alpha c_{(|I|)}$.

Step 3 proves $\bar{L}$ satisfies all properties of the loss function, except the range restriction to $(-1, \infty)$.

Step 3. *The function $\bar{L}$ satisfies the following conditions:*

1. *a) decreasing in the first argument, b) increasing in the second argument, and c) continuous;*
2. *$x = y$ implies $\bar{L}(x,y) = 0$; and*
3. *$\bar{L}(x,y) = \bar{L}(x',y') > 0$ implies $|x - x'| \leq |y - y'|$.*

Proof. 1a) For each i, the function $\bar{L}(c_i, c_i^*)$ is decreasing in c_i by strict monotonicity of $C(k)$: more precisely, let $k < k'$; then, $(c_i)_{i\in I} = C(k) \ll C(k') = (c_i')_{i\in I}$, $c_i < c_i'$, and $\bar{L}(c_i, c_i^*) = k^* - k > \bar{L}(c_i', c_i^*) = k^* - k'$. By the construction of $\bar{L}$, the function $\bar{L}(x,y)$ is decreasing in x for each $y \in \mathbb{R}_+$.

1b) Property 4 of Step 2 states that $(c_i)_{i\in I} = C(k)$ implies $c_i > c_j \iff c_i^* > c_j^*$ for each $i, j \in I$. By construction of $\bar{L}$, this implies that $\bar{L}(x,y) = \bar{L}(x',y')$ with $y < y'$ if and only if $x < x'$. Since $\bar{L}$ is decreasing in the first argument, $\bar{L}(x,y) < \bar{L}(x,y')$ whenever $y < y'$.

1c) Since $C(k)$ is continuous in k, for each i, the function $\bar{L}(c_i, c_i^*)$ is continuous in c_i. Continuity of $\bar{L}$ then follows by construction.

2) By definition of $C(k^*)$ and $\bar{L}$, $\bar{L}(c_i^*, c_i^*) = k^* - k^* = 0$ for each $i \in I$. By construction of $\bar{L}$, the result extends to $\bar{L}(y, y)$ for each $y \in \mathbb{R}_{++}$.

3) By contradiction, let $\bar{L}(x, y) = \bar{L}(x', y') > 0$ and $|x - x'| > |y - y'|$. Without loss of generality, let $x > x'$ and $y > y'$ and let $k \equiv k^* - \bar{L}(x, y)$. By construction, the implicit function defined by setting $\bar{L}(x'', y'') = k^* - k$ is piecewise linear: it may change slope only in correspondence to $y'' = c_i^*$ for some $i \in I$. By the mean value theorem, $x - x' > y - y'$ implies there exists a pair $i, j \in I$ such that $c_i - c_j > c_i^* - c_j^*$ with $\bar{L}(c_i, c_i^*) = \bar{L}(c_j, c_j^*) = k^* - k$. Clearly, c_i and c_j belong to $(c_m)_{m \in I} = C(k)$. Thus, $c_i - c_j > c_i^* - c_j^*$ is a violation of *horizontal equity* (as shown by Property 5 of Step 2). □

The proof is completed by the following step, which constructs the loss function L and derives the functional form of the fair social welfare function.

Step 4. *The social welfare function $W(a)$ is ordinally equivalent to a fair social welfare function*

$$W^L(a) \equiv \sum_{i \in I} \int_0^{e_i(c_i, \ell_i)} (1 + L(x, c_i^*))\, dx,$$

where the loss function $L : \mathbb{R}_+ \times \mathbb{R}_{++} \to (-1, \infty)$ is an increasing transformation of $\bar{L}$ such that $L(x, y) = \bar{L}(x, y)$ whenever $x = y$.

Proof. Let $a \in \bar{A}$ and $k \geq 0$. Then, $(c_m)_{m \in I} = C(k)$ requires that $\partial h_i(c_i)/c_i = \partial h_j(c_j)/c_j$ for each $i, j \in I$. Thus, there exists a positive-valued and strictly increasing transformation f such that $\partial h_i(c_i)/c_i = f\left(\bar{L}(c_i, c_i^*)\right)$ for each $i \in I$. Define the loss function by setting $L \equiv f\left(\bar{L}\right)/f(0) - 1$. Thus, the loss function takes values in $(-1, \infty)$ and $L(x, y) = \bar{L}(x, y)$ whenever $x = y$. Since $e_i(c_i, \ell_i^*) = c_i$, for each $a \in \bar{A}$,

$$W(a) = \sum_i h_i(c_i) = \sum_i t_i + \sum_i \int_0^{c_i} (1 + L(x, c_i^*))\, dx,$$

for some vector of welfare-irrelevant constants $(t_i)_{i \in I}$. Removing the additive constants, let W^L be defined by setting for each $a \in A$,

$$W^L(a) \equiv W(a) - \sum_i t_i = \sum_i \int_0^{e_i(c_i, \ell_i)} (1 + L(x, c_i^*))\, dx.$$

□

C Further Simulation Results

In this section, we address the sensitivity of the simulation results to different specifications of the utility parameters and the inequity aversion of social preferences.

First, we set a utility function that is consistent with a higher income effect and lower labor supply elasticity: $\rho = 2$ and $\sigma = 5$.[26] The lower distortions from behavioral responses lead to higher tax rates for high incomes, as confirmend by the simulation results in Figure 4.

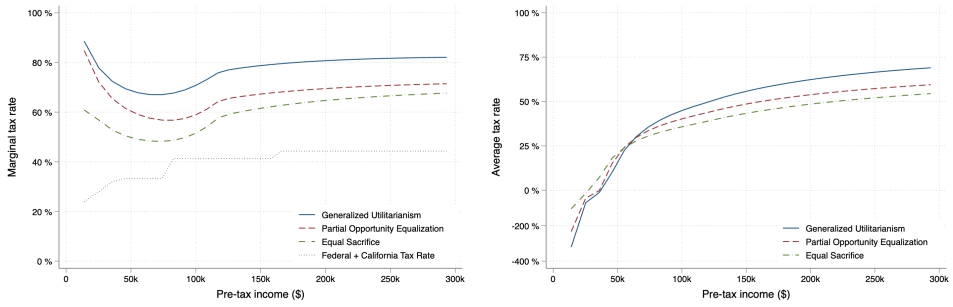

Figure 4: Marginal tax rates and average tax rates for the generalized utilitarian, partial opportunity-equalization, and equal-sacrifice criteria with different utility function.

Next, we compare higher and lower levels of inequity aversion: $\gamma = 1.2$ and $\gamma = 0.5$. All other parameters are the same as in Section 5.3.

[26] Using lottery data for the US, Golosov et al. (2021) find that income effects might be significantly larger than earlier estimates suggest.

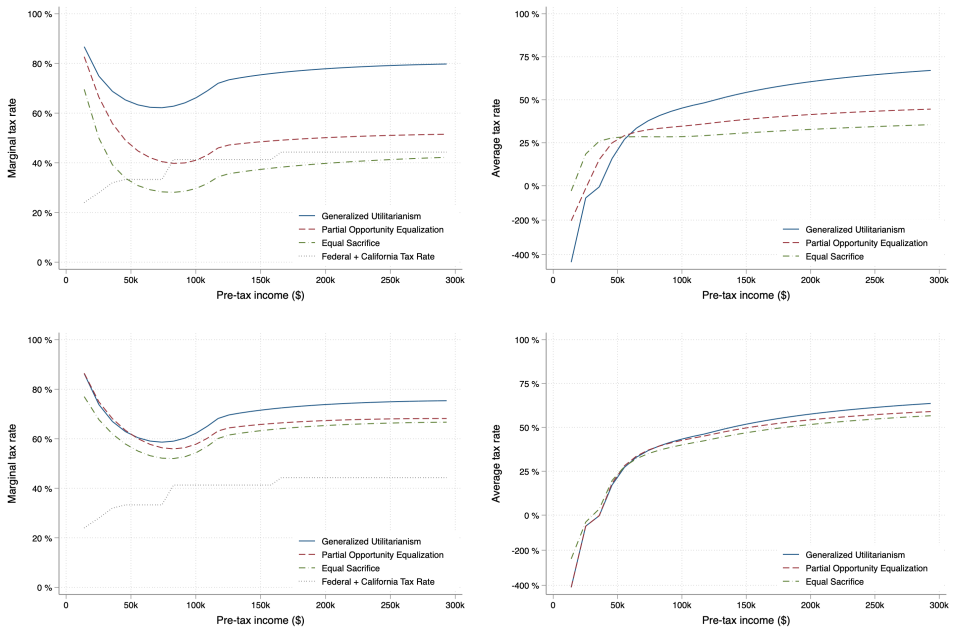

Figure 5: Marginal tax rates and average tax rates for the generalized utilitarian, partial opportunity-equalization, and equal-sacrifice criteria. High (low) inequity aversion for the top (bottom) tax schedules.

For the generalized utilitarian criterion, a higher γ leads to more redistribution, albeit this criterion is notoriously not very sensitive to this ethical parameter in the context of labor income taxation. In contrast, for the partial opportunity-equalization and for the equal-sacrifice criteria, a higher γ leads to a (sizable) reduction in redistribution. This is due to the different social marginal welfare weights attached to individuals by the different criteria. For generalized utilitarianism, γ is a measure of the priority to the lowest income individuals. For out fairness-based criteria, higher income individuals may face larger losses and, as in the simulation results, may be prioritized over lower income individuals. In this case, a larger aversion to unequal losses reduces the progressivity of the tax system.