KU LEUVEN

DEPARTMENT OF ECONOMICS

Fairness gaps for earnings tax design

Erwin Ooghe, Erik Schokkaert, Hannes Serruys

FACULTY OF ECONOMICS AND BUSINESS

DISCUSSION PAPER SERIES DPS 23.15

September 2023

Fairness gaps for earnings tax design*

Erwin Ooghe – Erik Schokkaert – Hannes Serruys

September 5, 2023

Abstract

In a setting with skill and preference heterogeneity, we characterize a family of social welfare measures that aggregate fairness gaps, defined as the difference between the money-metric utilities that individuals have and the money-metric utilities they should have in a fair society. Each welfare measure can be decomposed into government revenues (size), excess burden (inefficiency), and unfair inequality (inequity). As a proof of concept, we evaluate four hypothetical earnings tax reforms based on two normative parameters: the degree of unfairness aversion and the degree of compensation for productive skills.

JEL-codes: D3, D6, D7, H2, I3, J2

Keywords: social welfare, fairness gaps, money-metric utility, excess burden, unfair inequality, unfairness aversion, degree of compensation, libertarianism, resource-egalitarianism

1 Introduction

The traditional earnings tax design literature is welfarist, i.e., the evaluation of earnings tax schemes is based on an increasing and concave social welfare function defined over individual utilities.[1] The only normative parameter is inequality aversion, ranging from no aversion to inequality to absolute priority for the worst off. Its weaknesses are twofold. First, the (preference-based) welfarist approach does not provide a clear guideline how to choose a specific cardinal and interpersonally comparable utility function for each individual.[2] Such a guideline becomes especially

*We would like to thank Felix Bierbrauer, Kristof Bosmans, Bart Capéau, Koen Decancq, Liebrecht De Sadeleer, Maya Eden, Antoine Germain, Esteban Munoz, Andreas Peichl, Paolo Piacquadio, Itai Sher as well as seminar participants at KU Leuven, CESifo (Munich), and ECINEQ (Aix-en-Provence) for helpful comments and suggestions. An earlier version of this paper circulated under the title 'Fair earnings tax reforms'. E-mail addresses and affiliations: erwin.ooghe@kuleuven.be (Department of Economics, KU Leuven), erik.schokkaert@kuleuven.be (Department of Economics, KU Leuven), hannes.serruys@ec.europa.eu (JRC Sevilla, European Commission).

[1] For overviews of the traditional earnings tax design literature, see, e.g., Salanié (2003), Kaplow (2008), Mankiw, Weinzierl, and Yagan (2009), Diamond and Saez (2011), Boadway (2012), Tuomala (2016), and Tuomala and Weinzierl (2022).

[2] The happiness literature does offer a specific proposal in this regard (Layard and De Neve, 2023), but has until now not been applied in an explicit optimal tax model.

pressing when individuals differ in preferences. Second, welfarism, being based on utility information only, cannot treat different sources of utility in a different way: whether individual utility differences arise from differences in endowments or differences in ambitions does not matter for welfarists. The welfarist view is therefore not only subject to philosophical critique, but also in sharp contrast with the values that people use to assess redistribution.[3]

Fleurbaey and Maniquet's (2011, 2018a) fairness approach to social welfare handles both shortcomings. First, the choice of a specific cardinal and interpersonally comparable utility function follows from explicit normative fairness principles. Second, the fairness approach allows to treat different sources of utility in different ways. The so-called resource-egalitarian view on fairness, for example, advocates to compensate for differences in endowments (the compensation principle), while keeping people responsible for differences in ambitions (the responsibility principle). But also other fairness views, such as the libertarian view, which rejects the compensation principle, can be axiomatized. From a practical point of view, it may be seen as a disadvantage that each alternative theory of justice must be tailor-made by adjusting the axioms appropriately. More importantly, the fairness principles—and especially the compensation principle—combined with a focus on ordinal preferences usually imply that the worst-off individual must receive absolute priority, which is seen as too extreme by many.

Saez and Stantcheva (2016) deal with the shortcomings of the welfarist approach in a different way. They propose to evaluate earnings tax reforms by "aggregating money-metric losses and gains of different individuals using generalized social marginal welfare weights." First, they pragmatically choose money-metric utilities to cardinalize preferences. Second, their generalized social marginal welfare weights allow to easily incorporate alternative fairness views. Unfortunately, as mentioned by Fleurbaey and Maniquet (2018a) and worked out by Sher (2021), this marginal approach is local and may lead to intransitivities when extended to the global level. In other words, beyond small earnings tax reforms, the approach may run into trouble.

Our approach—the gaps approach—is based on money metric utilities, but, contrary to Saez and Stantcheva (2016)'s generalized social marginal welfare weights, it is a globally transitive approach. Moreover, our gaps approach is derived axiomatically, but, contrary to Fleurbaey and Maniquet (2011, 2018a)'s fairness approach, we adjust the compensation principle to allow for different degrees of inequality aversion and for a more practical way to introduce alternative fairness views. We discuss these two adjustments in more detail.

Compensation principles are usually formalized as transfer principles that require to approve of mean-preserving and progressive transfers of resources between individuals. In a unidimensional setting with, say, only income, mean-preservingness guarantees that the income allocation before and after the transfer is efficient; and progressivity guarantees that the income transfer goes from

[3] For philosophical critiques on welfarism, see, e.g., Rawls (1971), Dworkin (1981), and Sen (1985). For overviews of the empirical research on people's opinions on distributive justice, see, e.g., Konow (2001), Gaertner and Schokkaert (2012), Sheffrin (2013), and Schokkaert and Tarroux (2021).

a richer to a poorer individual so that it reduces inequality. In such a setting, a mean-preserving and progressive transfer preserves efficiency and improves equity and should therefore be approved of according to the transfer principle. In production economies, however, individual bundles are (at least) twodimensional as they also contain labour (or leisure) besides income. This gives rise to two problems.

First, progressivity is no longer unambiguously defined. For a transfer to be progressive, resources must be transferred from a better off to a worse off individual, requiring a metric to define who is better off and who is worse off.[4] Compensation principles therefore often restrict transfers to take place between individuals with the same preferences.[5] Indeed, given the same preferences, a progressive transfer can be unambiguously defined as a transfer from an individual on a higher indifference curve to an individual on a lower indifference curve. Moreover, the individual on a lower indifference curve is not only unambiguously worse off, but also unfairly treated according to resource-egalitarians: given the same preferences, being on a lower indifference curve can only be caused by a difference in endowments for which the individual is not held responsible. Defined this way, compensation principles are limited in scope as they can only implement the resource-egalitarian ideal that aims to compensate for differences in endowments. To see this, consider the libertarian view, which states that the laisser-faire allocation in a production economy is fair. Individuals with the same preferences, but different productive skills end up at different indifference curves in the laisser-faire. Yet, libertarians would not want to transfer resources in that situation; and they would probably like to transfer resources in other situations, e.g., if the transfer moves the current allocation more closely to the laisser-faire allocation. To allow for alternative fairness views, our version of progressivity adapts the proposal of Bosmans, Lauwers, and Ooghe (2009, 2018) to a production economy. As will become clear later on, it does not exogenously impose who is better off and who is worse off and remains therefore open to different fairness views, including resource-egalitarian and libertarian views.

Second, mean-preservingness no longer guarantees that allocations before and after a transfer are equally efficient in a multidimensional setting. To see this, imagine a simple exchange economy with two goods and two individuals. Switching between any two allocations in the Edgeworth box corresponds with a mean-preserving transfer of resources, but the two allocations may differ substantially in terms of allocative efficiency. Most compensation principles in the fairness literature neglect allocative efficiency considerations. This neglect implies that equity receives priority over efficiency, which is only possible by giving absolute priority to the worse off. To avoid such an extreme egalitarian position, some compensation principles impose additional restrictions (on top of the commonly made assumption that the transfers must take place between individuals with the same preferences). Fleurbaey and Tadenuma (2014) restrict the transfers (i) to be uniform (i.e.,

[4] A straightforward proposal is to use bundle dominance: if an individual posesses more of each good, then that individual is better off. This principle is, however, not compatible with the Pareto principle (Fleurbaey and Trannoy, 2003).

[5] For an overview of transfer principles, see Fleurbaey and Maniquet (2011, chapter 3).

the bundles after the transfer are a convex combination of the bundles before the transfer) and (ii) to take place between individuals with homothetic preferences. By construction, the allocation after the transfer cannot be less efficient because the average utility (as measured by a linearly homogenous utility representation) cannot be lower after the transfer. Piacquadio (2017) restricts preferences to be representable by a concave utility function such that transfers cannot decrease average utility (as measured by a concave utility representation). Bosmans, Decancq, and Ooghe (2018) restrict the transfers to be efficiency-preserving, i.e., not only the societal bundle (i.e., the sum of resources), but also the Scitovsky set (i.e., the Minkowski sum of all better-than sets) must remain the same before and after the transfer.[6] As in Bosmans, Decancq, and Ooghe (2018), our transfer principle will preserve efficiency, but in a simpler way: we require the allocations before and after the transfer to be Pareto efficient.

We show that our modified transfer principle—together with standard efficiency, impartiality, and separability requirements—is satisfied if and only if social welfare is equal to the average transformed fairness gap, i.e., $\frac{1}{n}\sum_{i=1}^{n}\phi(m_i - m_i^*)$, with (i) ϕ a differentiable, strictly increasing, and strictly concave transformation function and (ii) $m_i - m_i^*$ the fairness gap, i.e., the difference between the money-metric utility in the current bundle and in the fair bundle of individual $i = 1, 2, \ldots, n$. The fair allocation—the allocation that collects the fair bundles of the different individuals—is implicitly defined as a Pareto efficient and anonymous allocation that maximizes social welfare over the set of feasible allocations.

We highlight the role of the two modifications of the transfer principle. First, restricting transfers to be efficiency-preserving implies that the aggregation of the fairness gaps can flexibly range from averaging gaps to (lexicographically) focusing on the smallest fairness gap. Second, not exogenously imposing who is better off and worse off, our social welfare measure must be based on fairness gaps, which are compatible with different fairness views. The fair allocation can indeed be the laisser-faire allocation (as in the libertarian view), the equal-wage-equivalent allocation (as in the resource-egalitarian view), or any other allocation that is Pareto efficient and anonymous.[7] Our main result can therefore be interpreted as a structural result that bridges the gap between the social welfare literature and the fair allocation literature.[8]

Our paper relates closely to two strands of the literature. First, there exists a literature that measures unfair inequality as the divergence between the current and a fair distribution of outcomes; see, e.g., Devooght (2008), Fleurbaey and Schokkaert (2009), Almas et al. (2011), Magdalou and Nock (2011), and Hufe, Kanbur, and Peichl (2022).[9] This literature does not consider indi-

[6]Fleurbaey and Maniquet (2018b) propose a similar principle to characterize an individual well-being measure that can be used in a social welfare function with any degree of inequality aversion.

[7]The equal-wage-equivalent allocation (Fleurbaey and Maniquet, 1999) is the Pareto efficient allocation in which everyone is indifferent between her bundle and the bundle she would have chosen if everyone had the same wage in the laisser-faire. We will present this allocation in more detail in section 2.

[8]For an overview of the fair allocation literature, see Moulin (2004), Fleurbaey (2008), and Thomson (2011).

[9]Our approach is also similar to the notion of distributional change, as proposed by Cowell (1985) in a setting of economic mobility.

vidual preferences, let alone preference heterogeneity. Moreover, it focuses on unidimensional inequality, rather than multidimensional social welfare, but these two can be related to each other: we show how our social welfare measure can be additively decomposed into government revenues (size), excess burden (inefficiency), and unfair inequality (inequity).[10] This decomposition also allows to obtain an expression for the welfare gain per invested euro of government revenues, a concept that is closely related to the marginal value of public funds proposed by Hendren and Sprung-Keyser (2020) and Finkelstein and Hendren (2020). Second, there exists a literature that aims to justify the use of fairness gaps in (preference-based) social welfare measures in the specific context of earnings tax design. Weinzierl (2014, 2018) starts from a common cardinal utility function and proposes to minimize a weighted sum of losses from utilitarianism and losses from equal sacrifice, where these losses are based on the utility gaps between the current bundle and the bundle that would be first-best optimal (according to either utilitarianism or equal sacrifice theory). There is no preference heterogeneity and no axiomatic justification for the use of utility gaps. Berg and Piacquadio (2023) start from ordinal and possibly different utility functions, as we do, but adopt a claims approach.[11] They introduce an exogenous fair allocation, containing the legitimate claims of the different individuals, and use it to define their main axioms.[12] These axioms characterize a social welfare measure that is equal to the sum of the integrals of (unit-translated) individual losses relative to the fair consumption level.

To apply our social welfare measure to the design of tax-benefit schemes, we must choose a transformation function and a fair allocation. First, we propose to use the exponential (Kolm-Pollak) transformation function. This choice is natural as fairness gaps can be negative. The curvature of the exponential transformation function—the so-called degree of unfairness aversion—will be our first normative parameter. Second, following Fleurbaey and Maniquet (1999, 2018a), we propose to use a fair allocation with a flexible degree of compensation for productive skills that ranges from libertarianism (no compensation for productive skills) to resource-egalitarianism (full compensation for productive skills). The degree of compensation is our second normative parameter.

As a proof of concept, we use EU-SILC (European Union Statistics on Income and Living Conditions) data to estimate the productive skills (hourly gross wage rates) and preferences (tastes for working) of Belgian singles without children. For the estimation of preferences, we consider a labour market without rationing (all jobs are available to everyone) and one with rationing (jobs are available with probabilities that may differ among individuals). We use gross earnings to define four groups: a group with zero gross earnings (the unemployed) and three equally sized groups among the working (the working poor, the working middle, and the working rich). We simulate four hypothetical tax reforms that give a small amount of extra net income to each individual whose

[10] As will become clear, this decomposition requires the transformation function to be invariant to additions.
[11] For an overview of the claims approach, see, e.g., Thomson (2019).
[12] As will become clear, an endogenous fair allocation follows from the axioms in our approach.

gross income falls within one of the four groups. Our evaluation of these four reforms highlights three results. First, changes in welfare are mainly driven by changes in unfair inequality. Second, the degree of compensation for productive skills plays an important role in the evaluation, probably more important than the degree of unfairness aversion. Third, the evaluation changes drastically if one allows for rationing, including the possibility of involuntary unemployment.

Section 2 characterizes the average transformed fairness gap as a measure of social welfare. With an eye to application, section 3 introduces additional specifications (the fair allocation and the transformation function), discusses the resulting marginal social welfare weights, and provides a decomposition of social welfare into government revenues (size), excess burden (inefficiency), and unfair inequality (equity). Section 4 illustrates the use of our social welfare measure—including the marginal social welfare weights and the decomposition—for the design of earnings taxes. Section 5 concludes.

2 Social welfare as average transformed fairness gap

In this section, we characterize average transformed fairness gaps as a measure of social welfare. To do so, we first introduce notation and axioms.

2.1 Notation

Let $I = \{1, 2, \dots, n\}$ be a set of $n \geq 3$ individuals. An allocation $x = (x_1, x_2, \dots, x_n)$ contains bundles $x_i = (c_i, \ell_i) \in X = \mathbb{R} \times \mathbb{R}_+$ of net income c_i and labour ℓ_i for each individual i in I. A type profile $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ contains a type θ_i for each individual i in I. Each type $\theta_i = (s_i, u_i)$ consists of a productive skill s_i and a utility function u_i. Skill levels $s_i \geq 0$ map labour ℓ_i into gross incomes y_i in a linear way, i.e., $y_i = s_i \ell_i$ for each individual. Utility functions $u_i : X \to \mathbb{R}$ map bundles into ordinal and non-comparable utility levels. The utility functions are continuously differentiable, strictly increasing in consumption, strictly decreasing in labour, and strictly quasi-concave.

A society $\mathscr{S} = (\theta, R_0)$ is fully described by a type profile $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ and an exogenous per capita revenue requirement, denoted R_0. The set of feasible allocations is defined as

$$F(\mathscr{S}) = \{x \in X^n \,|\, \frac{1}{n} \sum_{i \in I} c_i + R_0 \leq \frac{1}{n} \sum_{i \in I} s_i \ell_i\}, \tag{1}$$

and the set of Pareto efficient allocations is

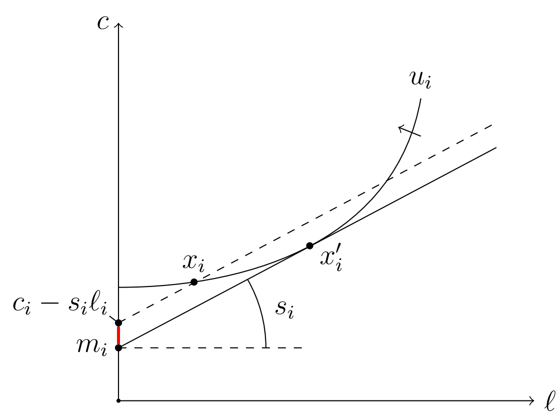
$$\begin{aligned} P(\mathscr{S}) \;=\; & \{x \in F(\mathscr{S}) \,|\, \nexists x' \in F(\mathscr{S}) \text{ s.t.} \\ & \forall i \in I, u_i(x'_i) \geq u_i(x_i) \text{ and } \exists i \in I, u_i(x'_i) > u_i(x_i)\}\,. \end{aligned} \tag{2}$$

The money-metric utility of an individual is defined as the minimal unearned income that is needed to guarantee that this individual, choosing from a budget set based on this unearned income and her own skill level, would not be worse off (according to her own utility function) compared to obtaining her actual bundle (Samuelson and Swamy, 1974; Deaton and Muellbauer, 1980; King, 1983). Formally, money-metric utility—given a bundle $x_i = (c_i, \ell_i)$ and the type θ_i of individual i—is defined as

$$m(x_i, \theta_i) = \min_{(c,\ell)\in X} (c - s_i\ell) \text{ subject to } u_i(c, \ell) \geq u_i(c_i, \ell_i). \tag{3}$$

Figure 1 illustrates the construction of money-metric utility. Given the bundle $x_i = (c_i, \ell_i)$, individual i earns $s_i\ell_i$ and the unearned income is therefore $c_i - s_i\ell_i$. Yet, individual i could reach the same utility level in bundles that exhibit a lower unearned income. Among these bundles, the bundle x_i' on the indifference curve through x_i provides the same utility level and has the lowest unearned income. This is the money-metric utility of individual i in bundle x_i, denoted as $m_i = m(x_i, \theta_i)$.

Figure 1: Money-metric utility and excess burden



In equation (3), the expression $\min(c - s_i\ell)$ can be replaced by $-\max(s_i\ell - c)$. So, (minus) money-metric utility can be interpreted as the maximal tax amount that one can hypothetically extract from an individual—or, as in the figure, the minimal subsidy amount one has to give to an individual—without him or her losing utility.[13] The difference between the maximal and actual tax revenue is called (individual) excess burden and is defined as

$$EB(x_i, \theta_i) = -m(x_i, \theta_i) - (s_i\ell_i - c_i) \geq 0. \tag{4}$$

Figure 1 shows the individual excess burden on the vertical axis as the thick (red) line. By construction it is always nonnegative, and zero if the marginal rate of substitution is equal to the

[13] We deliberately write "hypothetically," as this maximal extraction is possible only with first-best lump-sum taxes. Moreover, to simplify the theoretical reasoning, we assume that the (non-tax) unearned income of the individual is zero. In the empirical exercise, we will take into account the unearned incomes of the individuals in our sample.

productive skill.

2.2 Axioms

A social welfare measure $W(x;\mathscr{S})$ is used to judge the goodness of an allocation x in a given society $\mathscr{S}$. We focus on a class of smooth and additively separable welfare measures, formalized by the next representation axiom.

Representation. For a given society $\mathscr{S}$, for any allocation x in X, social welfare can be represented as

$$W(x;\mathscr{S}) = \frac{1}{n}\sum_{i\in I} v_i(x_i;\mathscr{S}),$$

with $v_1, v_2, \ldots, v_n$ evaluation functions that are continuously differentiable in net earnings and labour.

The interpretation of the evaluation functions will become clear later on. The separability assumption underlying the representation axiom is very common in welfare analysis. Its main limitation is that it excludes rank-dependent social welfare measures.[14]

The Pareto principle imposes that higher utility for all is better; and strictly higher utility for some (in addition to higher utility for all) is strictly better.

Pareto. For a given society $\mathscr{S}$, for any two allocations x and x' in X, if $u_i(x_i) \geq u_i(x'_i)$ for all i in I, then $W(x;\mathscr{S}) \geq W(x';\mathscr{S})$; if, in addition, $u_i(x_i) > u_i(x'_i)$ holds for some individual i in I, then $W(x;\mathscr{S}) > W(x';\mathscr{S})$.

The anonymity principle requires that permuting bundles of individuals with the same type (skills and preferences) does not matter for social welfare. Let $\theta_i = \theta_j$ mean that individuals i and j have the same skills (i.e., $s_i = s_j$) and the same preferences (i.e., $u_i = \varphi(u_j)$ for some strictly increasing transformation function φ).

Anonymity. For a given society $\mathscr{S}$, for any allocation x in X, for any two individuals i, j in I, if $\theta_i = \theta_j$ holds, then $W(\ldots, x_i, \ldots, x_j, \ldots;\mathscr{S}) = W(\ldots, x_j, \ldots, x_i, \ldots;\mathscr{S})$.

The transfer principle requires that a *progressive* and *Pareto-efficiency-preserving* transfer of resources between two individuals improves social welfare. A transfer is called *progressive* if it is directed from a "better-off" to a "worse-off" individual without changing their relative position. The measures that determine who is better off and worse off are the functions $v_1, v_2, \ldots, v_n$ defined in the representation axiom.[15] A transfer is called *Pareto-efficiency-preserving* if the allocations before and after the transfer are Pareto efficient. This additional condition is inspired by Bosmans,

[14] See, e.g., the discussion in Adler (2022).

[15] This "implicit" choice guarantees that the transfer principle is "consistent" as defined in Bosmans, Lauwers, and Ooghe (2009, 2018).

Decancq, and Ooghe (2018). As we will see, it allows us to avoid maximin-type results, while still leading to an inequality averse welfare specification.

Transfer. For a given society $\mathscr{S}$, for any two Pareto efficient allocations $x, x' \in P(\mathscr{S})$, for any two individuals i, j in I, if the transition from allocation x' to x is based on a progressive transfer of resources from i to j, i.e.,

$$v_i(x'_i; \mathscr{S}) > v_i(x_i; \mathscr{S}) \geq v_j(x_j; \mathscr{S}) > v_j(x'_j; \mathscr{S}),$$

without affecting other individuals, i.e.,

$$x_k = x'_k, \text{ for all } k \neq i, j,$$

then $W(x; \mathscr{S}) > W(x'; \mathscr{S})$.

2.3 Result

Theorem 1 is our main theoretical result: a social welfare measure satisfies all axioms if and only if it can be represented as the average transformed fairness gap (with the fairness gap defined as the difference in money-metric utility between the actual and the fair bundle of an individual). A proof can be found in Appendix A.

Theorem 1. A social welfare measure $W(x; \mathscr{S})$ satisfies Representation, Pareto, Anonymity, and Transfer if and only if it can be represented as

$$W(x; \mathscr{S}) = \frac{1}{n} \sum_{i \in I} \phi(m_i - m_i^*; \mathscr{S}),$$

with

1. $\phi(\cdot; \mathscr{S})$ satisfying $\phi'(\cdot; \mathscr{S}) > 0$ and $\phi''(\cdot; \mathscr{S}) < 0$,
2. $m_i = m(x_i, \theta_i)$ and $m_i^* = m(x_i^*(\mathscr{S}), \theta_i)$ for each individual i in I, and
3. $x^*(\mathscr{S}) = (x_1^*(\mathscr{S}), x_2^*(\mathscr{S}), \ldots, x_n^*(\mathscr{S}))$ a Pareto efficient and anonymous allocation (i.e., $x^*(\mathscr{S}) \in P(\mathscr{S})$ and $x_i^*(\mathscr{S}) = x_j^*(\mathscr{S})$ for all i, j in I with $\theta_i = \theta_j$).

Several remarks apply.

We shall refer to $x^*(\mathscr{S})$ as the 'fair' allocation from now on. Our axioms only impose that $x^*(\mathscr{S})$ is a Pareto efficient and anonymous allocation. This means that Theorem 1 remains open to very different fairness views. In the next section we will illustrate this flexibility and show

how it can encompass the libertarian view (that does not want to compensate for differences in productive skills), the resource-egalitarian view (that wants to fully compensate for differences in productive skills), as well as intermediate views between both extremes. The desired degree of compensation is a first normative parameter in our approach.

The fairness gap $m_i - m_i^*$ measures whether individual i is treated better than fairly (if the gap is positive), exactly fairly (if zero), or worse than fairly (if negative). We like to stress however that these gaps do not measure individual well-being. Two individuals with the same preferences can be on the same indifference curve (and can therefore be said to have the same individual well-being), but one of them can still be treated less fairly than the other. This would be the case, e.g., if one adopts a libertarian view in which the laisser-faire market allocation is considered fair. According to this view, a high-skilled is treated less fairly than a low-skilled if both have the same preferences and reach the same indifference curve.

The transformation function $\phi(\cdot;\mathscr{S})$ is strictly increasing and strictly concave. Increasingness ensures that Pareto holds: a more preferred bundle leads to a higher actual money-metric utility (without changing the fair money-metric utility), which, in turn, implies a higher social welfare, ceteris paribus. Concavity implies that (money-metric utility) transfers that reduce unfairness, i.e., transfers that are directed from more fairly to less fairly treated individuals, are approved of. The curvature of the transformation function is called unfairness aversion, and is a second normative parameter.

3 Towards application

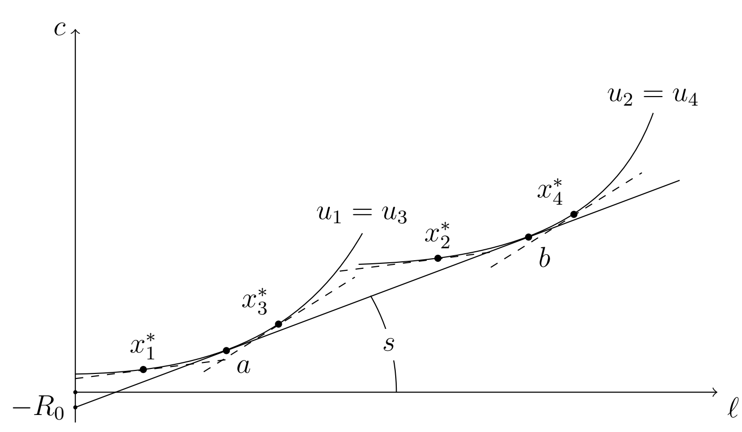
In this section we introduce additional specifications with an eye to application. We first propose a fair allocation that is flexibly parametrized by the degree of compensation for productivity differences. Afterwards, we introduce a transformation function that is flexibly parametrized by the degree of unfairness aversion. We also discuss the resulting marginal social welfare weights and provide a simple decomposition of social welfare.

3.1 A flexible fair allocation

The choice of the fair allocation $x^*(\mathscr{S})$, allows for many different fairness views. In a setting with differences in both skills and preferences, the most important issue is the degree to which one considers these differences to lead to unfair outcomes. Many economists and most lay people (see, e.g., Konow, 2003, and Gaertner and Schokkaert, 2012) accept that earnings differences following from differences in preferences, i.e., in the taste for working, are ethically legitimate. Opinions differ, however, about the acceptability of earnings differences reflecting differences in skills. Libertarians and resource-egalitarians are at the extreme sides of that debate. Libertarians state that, in the absence of market failures, the market allocation is fair. This can be interpreted as an assumption that individuals are fully responsible for both skills and preferences. Resource-egalitarians hold individuals fully responsible for differences in outcomes caused by differences in preferences (ambitions), but not responsible at all for differences in outcomes caused by differences in skills (endowments). The latter outcome differences must therefore be fully compensated.

The optimal libertarian solution is the laisser-faire allocation, defined as no intervention, except possibly for a head tax to finance the exogenous per capita revenue requirement R_0. A prominent resource-egalitarian solution is the so-called equal-skill-equivalent allocation rule, proposed by Fleurbaey and Maniquet (1999).[16] This rule selects the Pareto efficient allocation in which everyone is indifferent between her bundle and the bundle she would have chosen if everyone had the same skill in the laisser-faire.[17] Figure 2 illustrates for the case of four individuals differentiated by productive skills and preferences. Skill heterogeneity is such that individuals 1 and 2 have the same low skills, while individuals 3 and 4 have the same high skills. Preference heterogeneity is such that individuals 1 and 3 have the same low tastes for working, while 2 and 4 have the same high tastes for working. The fair allocation x^* is indeed Pareto efficient (the dashed lines indicate that the marginal rate of substitution is equal to the skill for each individual). Moreover, each individual is indifferent between her bundle and the bundle she would have chosen in the laisser-faire with a common skill level, denoted s in Figure 2, and a head tax equal to R_0. This equal-skill-equivalent rule offers full compensation for skills: individuals with the same preferences, but different skills end up at the same indifference curve. In terms of responsibility, if an individual has the same skills as another individual, but a stronger taste for working, he/she will end up with a higher consumption level. Indeed, they will end up on the indifference curves through the bundle where consumption is proportional to labour, being the bundles a and b in Figure 2.

Figure 2: The equal-skill-equivalent allocation

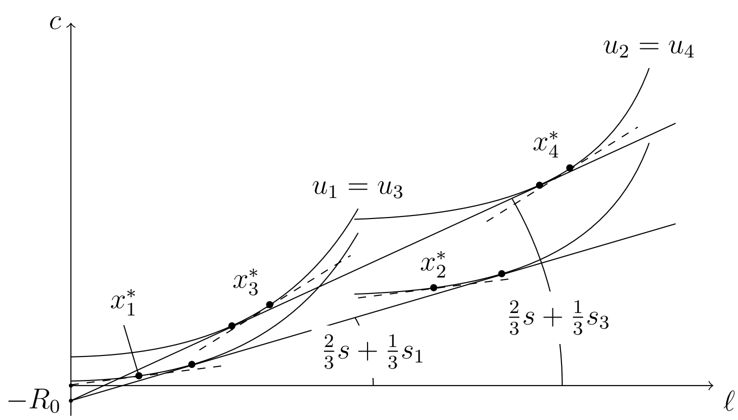


[16]A closely related rule is implemented in the fair income tax model of Fleurbaey and Maniquet (2006).

[17]As the chosen allocation is Pareto efficient, the common skill is defined by the feasibility constraint. In other words, the sum of money-metric utilities must be equal to $-R_0$ at the Pareto-optimal allocation.

Opinions in society differ about the degree of responsibility for skills. We therefore propose to adjust the equal-skill-equivalent allocation rule as proposed by Fleurbaey and Maniquet (2018a, p. 1060) to allow for a partial degree of compensation in between the extremes of no compensation (libertarianism) and full compensation (resource-egalitarianism). We introduce a (normative) parameter γ that captures the degree of compensation for skills. The resulting γ-skill-equivalent allocation rule selects the Pareto efficient allocation in which everyone is indifferent between her bundle and the bundle she would have chosen in the laisser-faire if everyone had a skill equal to $\gamma s + (1-\gamma)s_i$, with s a common skill level. The parameter γ lies between no compensation ($\gamma = 0$, consistent with the libertarian view) and full compensation ($\gamma = 1$, the typical resource-egalitarian view), and the γ-skill-equivalent allocation contains the (libertarian) market allocation ($\gamma = 0$) and the (unadjusted) equal-skill-equivalent allocation ($\gamma = 1$) as extreme cases. Figure 3 presents the γ-skill-equivalent allocation with $\gamma = \frac{2}{3}$ for the case of the same four individuals. Here there is some compensation for skill differences, but it is incomplete in that individuals with the same preferences will no longer end up on the same indifference curve. Varying the parameter γ makes it possible to cover different ethical perspectives. In applied policy analysis, it is possible to perform a sensitivity analysis with respect to the parameter γ.

Figure 3: The γ-skill-equivalent allocation, with $\gamma = \frac{2}{3}$



3.2 A flexible transformation function

Without loss of generality, we rewrite social welfare in its equally-distributed-equivalent form, i.e., $W(x; \mathscr{S}) = \phi^{-1}[\frac{1}{n}\sum_{i \in I} \phi(m_i - m_i^*)]$, with $\phi(\cdot; \mathscr{S})$ abbreviated as $\phi(\cdot)$ from now on. As the fairness gaps can be negative, it is natural to choose a transformation function that can handle negative values. In our empirical illustration, we will use the Kolm-Pollak specification defined as $\phi(m) = \exp(-rm)$, where $r > 0$ measures unfairness aversion. The Kolm-Pollak specification is invariant to additions, i.e., for arbitrary scalars m, m', c, if $\phi(m) \geq \phi(m')$, then $\phi(m+c) \geq \phi(m'+c)$. In the remainder of this section, we discuss the marginal social welfare weights that result from this specification. We also show that addition-invariant transformation functions (such as the Kolm-Pollak specification) allow for a decomposition of social welfare into government revenues (size), excess burden (inefficiency), and unfair inequality (inequity).

3.2.1 The marginal social welfare weights

The marginal social welfare weight of an individual is usually defined as the social welfare impact of a small increase in the consumption (or net income) of that individual. These weights are key for redistribution because a small income transfer from an individual with a lower weight to an individual with a higher weight will increase social welfare. To focus here on the normative aspects, we neglect the impact of a small income increase on the money-metric utility of individual i, and consider only the impact of a small increase of that money-metric utility on social welfare. With this definition, the marginal social welfare weight of individual i becomes equal to

$$msww_i = \frac{\frac{1}{n}\phi'(m_i - m_i^*)}{\phi'\{\phi^{-1}[\frac{1}{n}\sum_{i\in I}\phi(m_i - m_i^*)]\}},$$

for the equally-distributed-equivalent welfare formulation. For the Kolm-Pollak specification, we get

$$msww_i = \frac{\exp(-r(m_i - m_i^*))}{\sum_{i\in I}\exp(-r(m_i - m_i^*))}. \tag{5}$$

Consider two individuals i and j who are both treated unfairly, but individual i is treated more unfairly than j, i.e., $m_i - m_i^* < m_j - m_j^* < 0$. Equation (5) tells us that individual i has a higher marginal social welfare weight than j (given $r > 0$). Moreover, the higher the degree of unfairness aversion r, the higher the marginal social welfare weight of i relative to j (i.e., the higher $msww_i/mssw_j$).

3.2.2 A decomposition of social welfare

Let $R(x;\mathscr{S}) = \frac{1}{n}\sum_{i\in I}(s_i\ell_i - c_i)$ be the average tax revenues. Let ϕ be a transformation function that is invariant to additions. In appendix B we show that social welfare can be decomposed as follows:

$$W(x;\mathscr{S}) = RS(x;\mathscr{S}) - EB(x;\mathscr{S}) - UI(x;\mathscr{S}), \tag{6}$$

with

$$\begin{aligned} RS(x;\mathscr{S}) &= R_0 - R(x;\mathscr{S}), \\ EB(x;\mathscr{S}) &= -\frac{1}{n}\sum_{i\in I} m_i - R(x;\mathscr{S}), \\ UI(x;\mathscr{S}) &= -\phi^{-1}\{\frac{1}{n}\sum_{i\in I}\phi(m_i - m_i^* - \frac{1}{n}\sum_{i\in I}(m_i - m_i^*))\}. \end{aligned}$$

The first component $RS(x;\mathscr{S})$ measures the average revenue shortage as the difference between the amount of taxes that should be raised and the amount that is currently raised on average. The second component $EB(x;\mathscr{S})$ measures inefficiency as average excess burden, i.e., the difference between the revenue that can be maximally raised (without losses in utility) and the revenue that is currently raised on average. It is the average of the individual excess burdens defined in equation (4). The third component $UI(x;\mathscr{S})$ measures unfair inequality, defined as the inequality in fair treatment between individuals. If all individuals are treated equally fairly or equally unfairly (i.e., the fairness gap $m_i - m_i^*$ is the same for all individuals), then unfair inequality is zero. We stress that zero unfair inequality does not correspond to zero unfairness: all individuals can be treated unfairly, but in an equal way.

To assess tax reforms, we focus on the difference in welfare after and before the reform. Let x_0 be the allocation before the reform and x_1 the allocation after the reform. Minus the change in welfare is equal to

$$\begin{aligned} -\triangle W &= W(x_0;\mathscr{S}) - W(x_1;\mathscr{S}), \\ &= (R(x_1;\mathscr{S}) - R(x_0;\mathscr{S})) + (EB(x_1;\mathscr{S}) - EB(x_0;\mathscr{S})) + (UI(x_1;\mathscr{S}) - UI(x_0;\mathscr{S})), \\ &= \triangle R + \triangle EB + \triangle UI. \end{aligned} \tag{7}$$

Changes in welfare depend negatively on changes in revenues, changes in excess burden, and changes in unfair inequality. If the tax reform is budget-neutral, then the first term at the right-hand side is equal to zero and the change in welfare depends only on changes in excess burden and changes in unfair inequality:

$$-\triangle W = \triangle EB + \triangle UI. \tag{8}$$

If the tax reform is not budget-neutral, the decomposition can be rewritten as

$$-\frac{\triangle W}{\triangle R} = 1 + \frac{\triangle EB}{\triangle R} + \frac{\triangle UI}{\triangle R}. \tag{9}$$

In our empirical illustration, we will focus on hypothetical tax reforms that reduce taxes locally, ceteris paribus. So, we can expect welfare to increase ($\triangle W > 0$) and government revenues to decrease ($\triangle R < 0$). The left-hand side of equation (9) will therefore be positive and is called the welfare gain per euro of public funds. The first term on the right-hand side shows the direct welfare effect of giving one euro, which is by definition equal to one euro. But the overall welfare gain of giving one euro will be different from one. Given $\triangle R < 0$, efficiency costs ($\triangle EB > 0$) or equity costs ($\triangle UI > 0$) will push the overall welfare effect down, and vice-versa in case of efficiency gains or equity gains.[18]

[18] For small reforms, the welfare gain per euro in equation (9) becomes the marginal welfare gain per euro,

4 An empirical illustration

We first introduce the data, the estimation of the productive skills (gross hourly wages) and the tastes for working (preferences), and the simulation methodology. Afterwards we propose four hypothetical earnings tax reforms. Finally, we evaluate and compare the four tax reforms on the basis of excess burden, unfair inequality, and social welfare. To better understand the evaluation, we also compute the marginal social welfare weights.

4.1 Data, skills, preferences, and simulations

4.1.1 Data

We use the cross-sectional EU-SILC data of Belgium for 2016 (wave 2017). We select all singles without children between 18 and 65 years old, who are active on the labour market, i.e., either working (but not self-employed) or unemployed (but searching for work) in 2017. Further details on the data selection can be found in appendix C. Table 1 provides descriptive statistics for our sample of 861 individuals (about 6% of the total EU-SILC sample of 13974 individuals for Belgium).

Table 1: Descriptive statistics

gender (%)	male	43.5
	female	56.5
age (years)	mean	44.0
nationality (%)	Belgian	85.8
	EU (& not Belgian)	9.4
	not EU	4.8
area (%)	urban	43.1
	middle	44.6
	rural	12.3
highest degree (%)	< secondary	22.1
	= secondary	34.7
	tertiary	43.2
experience (years)	mean	19.7
labour market status (%)	unemployed	13.3
	working	86.7

Notes: "= secondary" also includes individuals with a highest degree in non-tertiary higher education. Area is based on Eurostat's degree of urbanization (DEGURBA) classification.

which is closely related to the marginal value of public funds introduced in Hendren and Sprung-Keyser (2020) and Finkelstein and Hendren (2020).

4.1.2 Skills

Gross earnings are in theory equal to the product of skill and labour. Defining labour as labour hours, skill corresponds to the gross hourly wage rate. For singles with non-zero gross earnings in 2016, the wage rate is computed as the ratio of yearly gross earnings and (an estimate of) yearly labour hours in 2016. For singles with zero gross earnings, we predict hourly wages based on a simple OLS regression model.[19] Further details on the computation and prediction of hourly gross wages can be found in appendix D.

Labour hours are discretized in four groups. The resulting set of possible labour hours is $L = \{0, 24, 38, 51\}$ and the set of bundles reduces to $X = \mathbb{R} \times L$.[20] Individuals are assigned to 0 hours if their reported labour hours are equal to 0 (13.3% of the sample), 24 hours if reported hours lie in $]0, 30[$ (16.8%), 38 hours if reported hours lie in $[30, 45[$ (63.3%), and 51 hours if reported hours lie in $[45, 61[$ (6.6%). We compute gross earnings based on (computed or predicted) wages and (re-assigned) labour hours and use EUROMOD (version 3.3.8) to simulate the corresponding net disposable incomes.[21] Table 2 provides descriptive statistics for the hourly wage rate, labour hours, gross earnings, and net incomes.

Table 2: Gross wage rates, labour hours, gross earnings, and net incomes

		p25	p50	p75
gross hourly wages (euro)	all	14.48	17.82	23.26
	working	14.38	17.89	23.75
	unemployed	14.72	17.66	20.75
actual weekly labour hours	all	29	38	40
	working	34	38	40
assigned weekly labour hours	all	24	38	38
	working	38	38	38
monthly gross earnings (euro)	all	1624.20	2692.38	3599.75
	working	2160.58	2893.21	3756.00
monthly net income (euro)	all	1498.10	1837.81	2213.45
	working	1659.71	1923.31	2290.10
	unemployed	1433.83	1466.18	1469.12

Notes: gross hourly wages of the unemployed are predicted using a simple OLS regression. Assigned weekly labour hours are recomputed after assigning individuals to one of the four possible labour hours choices that we use in the discrete choice model. Gross earnings are computed as the product of the (computed or predicted) hourly gross wage rate and the assigned weekly labour hours, multiplied with 52/12. Monthly net incomes are simulated using EUROMOD.

[19] We also experimented with Heckman selection models. However, the inverse Mills ratio was never statistically significant. Output is available upon request.

[20] The classification is based on a visual inspection of the (local peaks of the) labour hours distribution.

[21] In these simulations, unearned incomes do not change with labour hours.

4.1.3 Preferences

We estimate preferences using a discrete choice model. Individuals maximize utility specified as

$$u(c_i(\ell), \ell; z_i) + \epsilon_i(\ell), \tag{10}$$

with (i) $u(c_i(\ell), \ell; z_i)$ the deterministic utility as a function of net income $c_i(\ell)$, labour hours ℓ, and a vector of individual characteristics z_i, and (ii) $\epsilon_i(\ell)$ a random utility term (independent and identically distributed over individuals and choices according to an extreme value type I distribution). Preference heterogeneity enters in two ways: the deterministic utility part captures observed preference heterogeneity via observed characteristics in z_i and the random utility terms reflect unobserved preference heterogeneity.[22]

Preferences are estimated without and with rationing in the labour market. Without rationing means that everyone faces the same discrete opportunity set $L = \{0, 24, 38, 51\}$. As a consequence, labour choice, including unemployment, is entirely voluntary. In case of no rationing, we assume that all observed characteristics (gender, age, nationality, area, and highest educational degree) may influence someone's taste for working.

To model rationing, we assume that the opportunity sets of individuals, denoted $O_i \subseteq L$, are probabilistic and may depend on observable characteristics.[23] This raises a difficult identification problem. Observed choices will reflect both available opportunities and preferences. In our illustration, we assume that the tastes for working can only vary with gender and age and that the opportunities vary with region, population density, gender, education, and nationality. This is only meant as an illustration. Further technical details and estimation results can be found in appendix E.

4.1.4 Simulation methodology

Simulations are based on an artificial sample in which we replace each individual by 809 artificial individuals with the same observable characteristics, gross hourly wages, and deterministic utility functions, but with a randomly drawn vector of unobserved utility terms to create unobserved preference heterogeneity.[24] In case we allow for rationing, we also simulate the opportunity sets using the estimated probabilities for the opportunity sets.

Table 3 presents the actual and simulated probabilities under the assumption that there is no rationing in the labour market.[25] The model predicts almost exactly the overall probabilities. Yet,

[22]This interpretation of the random utility terms as unobserved preference heterogeneity is only one possibility. An alternative interpretation would be to see them as optimization errors. This would of course change the welfare evaluation. See, e.g., Creedy, Hérault, and Kalb (2011) for a discussion.

[23]The technical details can be found in appendix E.

[24]This replication number is equal to the ratio of the Belgian population (11303528) and the EU-SILC sample size for Belgium (13974) in 2016.

[25]The pseudo-R^2 is equal to 0.31, which is reasonable: compare, e.g., with Bargain, Orsini, and Peichl (2014),

it clearly overpredicts working half-time for men and underpredicts it for women. The opposite is true for working full-time.

Table 3: Actual and predicted probabilities (no rationing)

all	$\ell = 0$	$\ell = 24$	$\ell = 38$	$\ell = 51$
actual (%)	13.3	16.8	63.3	6.6
predicted (%)	13.2	16.8	63.3	6.6
predicted - actual	-0.1	0.0	0.1	0.0
men	$\ell = 0$	$\ell = 24$	$\ell = 38$	$\ell = 51$
actual (%)	14.5	10.1	66.8	8.7
predicted (%)	14.4	16.4	62.5	6.7
predicted - actual	-0.1	6.3	-4.3	-2.0
women	$\ell = 0$	$\ell = 24$	$\ell = 38$	$\ell = 51$
actual (%)	11.8	25.7	58.6	3.9
predicted (%)	11.7	17.4	64.3	6.5
predicted - actual	-0.1	-8.2	5.7	2.6

Table 4 presents the actual and predicted probabilities in case we allow for rationing in the labour market. As expected, allowing for rationing gives a much better fit. The over- and underprediction of working half-time or full-time for men and women have almost completely disappeared.

Table 4: Actual and predicted probabilities (rationing)

all	$\ell = 0$	$\ell = 24$	$\ell = 38$	$\ell = 51$
actual (%)	13.3	16.8	63.3	6.6
predicted (%)	13.3	16.8	63.3	6.6
predicted - actual	0.0	0.0	0.0	0.0
men	$\ell = 0$	$\ell = 24$	$\ell = 38$	$\ell = 51$
actual (%)	14.5	10.0	66.8	8.7
predicted (%)	14.7	10.3	66.2	8.8
predicted - actual	0.2	0.3	-0.6	0.1
women	$\ell = 0$	$\ell = 24$	$\ell = 38$	$\ell = 51$
actual (%)	11.8	25.7	58.6	3.9
predicted (%)	11.5	25.3	59.5	3.7
predicted - actual	-0.2	-0.4	0.8	-0.2

4.2 Four hypothetical earnings tax reforms

We distinguish four groups of individuals on the basis of gross earnings: the unemployed with zero gross earnings and three equally sized groups among the working, called the working poor, the working middle, and the working rich. Each reform assigns 75 euro extra to the monthly net

who obtain 0.28 for singles (on average across a selection of countries).

Table 5: Behavioral reactions (no rationing)

	unemployed	working poor	working middle	working rich
monthly gross income (euro)	$y = 0$	$0 < y \leq 2409$	$2409 < y \leq 3493$	$3493 < y$
avg. monthly net income (euro)	1415	1513	1920	2925
fraction before reform (%)	13.5	28.8	28.8	28.8
reform 1 (euro)	+75			
fraction after the reform (%)	14.6	28.4	28.4	28.6
percentage point change	+1.2	-0.5	-0.4	-0.2
elasticity (p.p. change / 5.3)	+**0.22**	-0.09	-0.08	-0.04
reform 2 (euro)		+75		
fraction after the reform (%)	13.0	30.1	28.2	28.7
percentage point change	-0.5	+1.2	-0.6	-0.1
elasticity (p.p. change / 5.0)	-0.09	+**0.25**	-0.13	-0.03
reform 3 (euro)			+75	
fraction after the reform (%)	13.1	28.3	30.2	28.5
percentage point change	-0.4	-0.6	+1.3	-0.4
elasticity (p.p. change / 3.9)	-0.10	-0.15	+**0.34**	-0.09
reform 4 (euro)				+75
fraction after the reform (%)	13.3	28.7	28.5	29.5
percentage point change	-0.2	-0.1	-0.3	+0.6
elasticity (p.p. change / 2.6)	-0.07	-0.05	-0.12	+**0.25**

Notes: the denominators of the elasticities can be obtained by dividing 75 euro by the average monthly net income of each group (reported in the second row of the table); the monthly gross incomes of the different groups in Tables 5 and 6 are slightly different because the replicated sample of individuals may have different preferences, random terms, and opportunity sets.

disposable income of individuals whose gross income belongs to one of these four income groups. Of course, individuals may change income group after the reform depending on their behavioral reactions.

Table 5 reports the behavioral reactions if we do not allow for rationing. The own-elasticities (on the diagonal in bold) indicate that the behavioral reactions in the different income groups are fairly similar, except for the working middle. If the net income of a certain income group increases with 1%, then the fraction of individuals in that group increases with 0.22 (the unemployed), 0.25 (the working rich), 0.25 (the working poor), and 0.34 (the working middle) percentage points. The cross-elasticities (off the diagonal) indicate where these increases come from. If the net income of a certain income group increases with 1%, then the fraction of individuals in the other groups decreases with between 0.03 and 0.15 percentage points. As expected, the cross-elasticities are usually stronger if the income group where the change occurs is 'closer'. For example, assigning extra net income to the unemployed, leads to the strongest behavioral reactions among the working poor, then among the working middle, and finally among the working rich.

Table 6 reports the behavioral reactions if we allow for rationing. As individuals will typically choose among a lower number of labour options, the behavioral responses to the tax reforms are in general weaker. For example, the elasticity with respect to an increase in the unemployment benefit is 0.07 in the model with rationing, compared to 0.22 in the model without rationing. This is also true for the other elasticities on the diagonal, but to a somewhat lesser extent. As people get richer, they are typically less constrained in their labour choices.

Table 6: Behavioral reactions (rationing)

	unemployed	working poor	working middle	working rich
monthly gross income (euro)	$y = 0$	$y \leq 2404$	$2404 < y \leq 3463$	$3463 < y$
avg. monthly net income (euro)	1415	1509	1907	2913
fraction before reform (%)	13.4	28.9	28.9	28.8
reform 1 (euro)	+75			
fraction after the reform (%)	13.8	28.7	28.8	28.7
percentage point change	+0.4	-0.1	-0.1	-0.1
elasticity (p.p. change / 5.3)	+**0.07**	-0.03	-0.02	-0.02
reform 2 (euro)		+75		
fraction after the reform (%)	13.2	29.4	28.7	28.7
percentage point change	-0.1	+0.5	-0.3	-0.1
elasticity (p.p. change / 5.0)	-0.03	+**0.09**	-0.05	-0.01
reform 3 (euro)			+75	
fraction after the reform (%)	13.3	28.6	29.5	28.6
percentage point change	-0.1	-0.3	+0.5	-0.2
elasticity (p.p. change / 3.9)	-0.03	-0.06	+**0.14**	-0.04
reform 4 (euro)				+75
fraction after the reform (%)	13.3	28.8	28.8	29.1
percentage point change	-0.1	-0.1	-0.2	+0.3
elasticity (p.p. change / 2.6)	-0.04	-0.03	-0.07	+**0.13**

Notes: the denominators of the elasticities can be obtained by dividing 75 euro by the average monthly net income of each group (reported in the second row of the table); the monthly gross incomes of the different groups in Tables 5 and 6 are slightly different because the replicated sample of individuals may have different preferences, random terms, and opportunity sets.

4.3 An evaluation of the earnings tax reforms

As the proposed hypothetical tax reforms are not budget-neutral, our evaluation will be based on the welfare gain expressed per invested euro of public funds, being the left-hand side of equation (9). To better understand the overal welfare changes of the different reforms, and its efficiency and equity components, we first take a look at the marginal social welfare weights. Afterwards, we look at the efficiency, equity and welfare changes of the different reforms.

4.3.1 Marginal social welfare weights

Equation (5) shows that the marginal social welfare weights depend on the fairness gaps and society's aversion to unfairness. Given $r > 0$, the social planner is inequality averse, and she will approve of a shift of resources of someone with a high fairness gap to someone with a low fairness gap, even if a part of that resource "leaks away" during the transfer. For ease of interpretation, we re-express the unfairness aversion r in terms of the maximal leak ρ (a fraction of the transferred amount) that society is willing to accept in case of progressive transfers. This maximal leak is determined by the value of r.[26] If r approaches zero, there is minimal aversion to unfairness and ρ becomes close to zero. If r approaches infinity, there is maximal aversion to unfairness and ρ becomes close to 1.

Figure 4 plots the average marginal social welfare weights of the four income groups before reform as a function of the degree of compensation (γ) in case there is no rationing. Each figure corresponds to a different value of the maximal leak (ρ). Increasing ρ magnifies the differences between the different groups, but does not change the social priority ranking. For low to middle degrees of compensation (roughly between 0 and 0.65), the working rich have the highest social priority, followed by the working middle, the working poor, and the unemployed. Over a small range (roughly between 0.65 and 0.9), the working middle get the highest social priority. For high degrees of compensation (between 0.9 and 1), the working poor have the highest social priority, followed by the working middle. The unemployed always have the lowest social priority.

Figure 5 shows that the average marginal social welfare weights change drastically, especially for the unemployed, if we allow for rationing. While they always had the lowest social priority if we do not allow for rationing, they now have the highest priority, except when the degree of compensation is close to libertarian (say, γ smaller than 0.2-0.3, depending on the maximal leak ρ). The reason is that the fraction of rationed workers is especially high among the unemployed, which implies that their fairness gap is much smaller (as their actual money metric utility will be lower and their fair money metric utility will typically be higher).[27]

[26] The maximal leak ρ also depends on the difference in the fairness gaps of the donor and the receiver of the transfer. We choose 500 units (roughly equal to the average difference in fairness gap in our sample for $\gamma = 0.5$) such that the maximal leak is defined as $\rho = 1 - \exp(-500r)$.

[27] We deliberately write 'typically' as it requires that their wage rate is lower than the average wage rate in society, a condition that typically holds among the unemployed.

Figure 4: The average marginal social welfare weights before reform (no rationing)

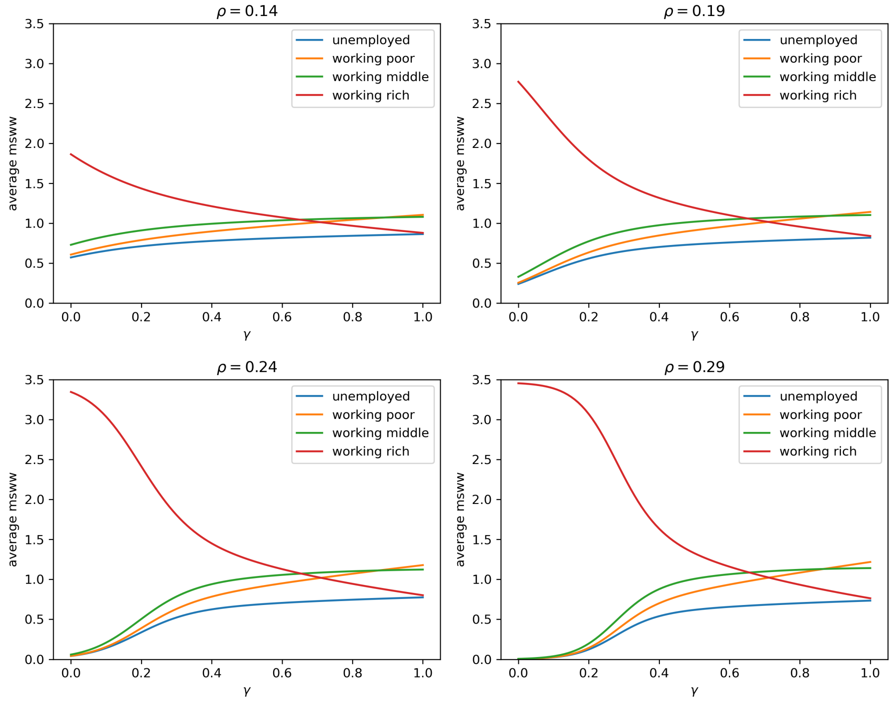


Figure 5: The average marginal social welfare weights before reform (rationing)

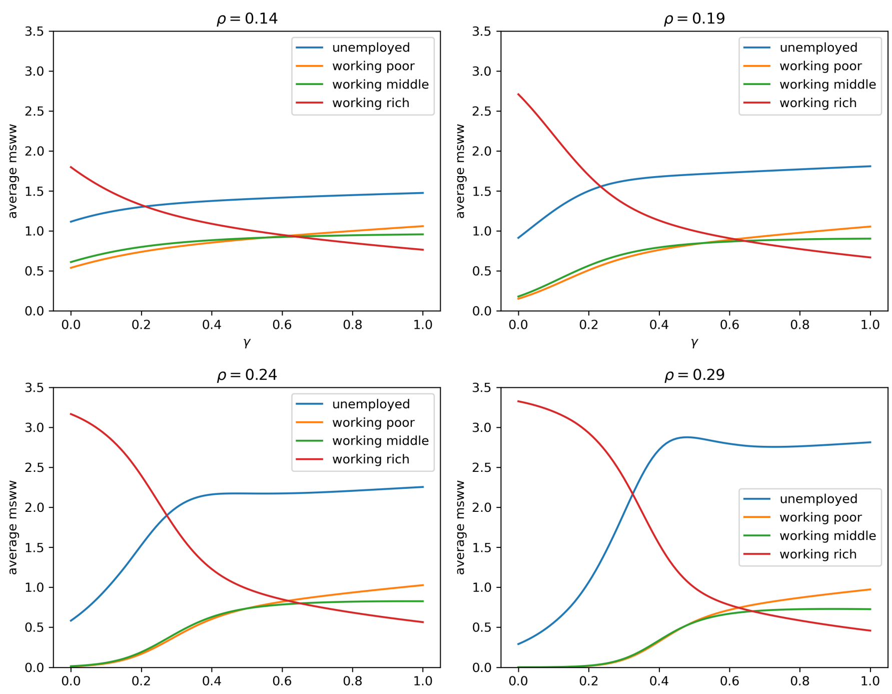


Table 7: Changes in per capita revenues, excess burden, and the efficiency gain per euro

four reforms		no rationing				rationing			
+75 euro for the	ΔR_M	ΔR_B	ΔR	ΔEB	$\frac{\Delta EB}{\Delta R}$	ΔR_B	ΔR	ΔEB	$\frac{\Delta EB}{\Delta R}$
1 - unemployed	-10.13	-28.13	-38.26	25.75	-0.67	-10.15	-20.28	10.72	-0.53
2 - working poor	-21.60	0.77	-20.83	-1.98	0.09	-0.52	-22.12	0.78	-0.04
3 - working middle	-21.60	9.76	-11.84	-10.27	0.87	-2.55	-19.05	-2.61	0.14
4 - working rich	-21.60	9.36	-12.24	-10.96	0.90	5.23	-16.37	-5.62	0.34

Notes: Given $\Delta R < 0$, $\frac{\Delta EB}{\Delta R}$ measures the efficiency gain per euro; efficiency losses appear as negative numbers.

4.3.2 Efficiency

Table 7 shows the changes in (per capita) government revenues, the changes in (average) excess burden, and the efficiency gain per euro (the ratio of the previous two numbers) for each reform and each labour market assumption. The changes in government revenues are split into the mechanical and behavioral revenue changes (denoted ΔR_M and ΔR_B respectively).

First, we discuss the revenue effects. The mechanical revenue effects do not depend on labour market status. They are negative (by definition), smaller for the first reform (as the unemployed form a smaller group), and equal for the other three reforms (given equally sized groups among the working). The behavioral revenue effect is large and negative if one increases the subsidies of the unemployed (the first reform). This is especially true if there is no rationing.[28] So, for the first reform, both the mechanical and the behavioral revenue effect is negative and large, implying that the overall revenue effect is also negative and large. In contrast, if one reduces the taxes of the working, the behavioral revenue effects are sometimes positive, indicating that the (positive) revenue effects of those who increase labour hours are larger than the (negative) revenue effects of those who reduce labour hours. This is, by definition, true if one reduces the taxes of the working rich (because no one reduces labour hours in this reform). It is also true for the other two reforms that "make work pay," but only if there is no rationing.

Second, if we transfer to the unemployed, the change in (average) excess burden is positive and high, indicating a strong efficiency loss, under both labour market regimes. In contrast, the change in excess burden is always negative (an efficiency gain) if we transfer to the working middle and the working rich. For the working poor it depends on the labour market regime, but the change in excess burden is, in both cases, relatively small.

Third, given $\Delta R < 0$ everywhere, the ratio of the change in excess burden and the change in government revenues has to be interpreted as an efficiency gain (or efficiency loss, if negative) expressed per invested euro of government revenues. Extra transfers to the unemployed generate large efficiency losses per euro, while extra transfers to the working middle and rich generate

[28]Indeed, as was seen in Tables 5 and 6 already, the percentage point change in unemployment for the first reform is three times as large if there is no rationing (compared to rationing).

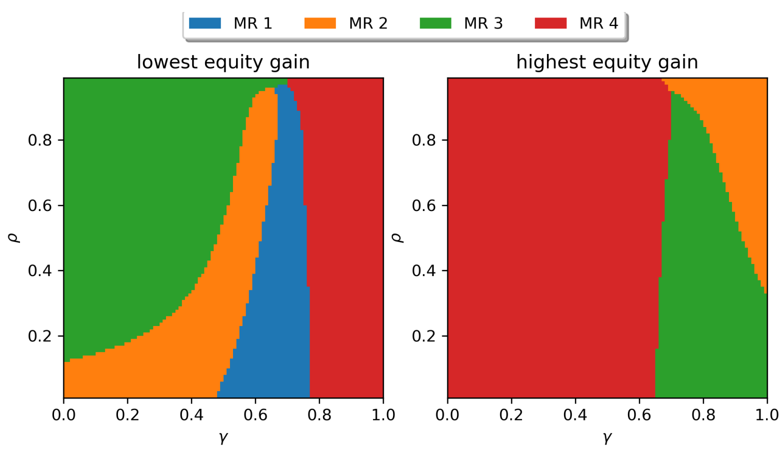
efficiency gains, especially if there is no rationing. For the working poor it depends again on the labour market status. The efficiency gains tell us that there exist small regressive transfers (from those who earn less to those earn more) that can be made budget neutral and that will increase efficiency. This will be especially the case if one transfers from the unemployed to the working rich. This result stands to reason: regressive tax reforms improve efficiency as they bring us closer to the efficient laisser-faire situation.

4.3.3 Equity

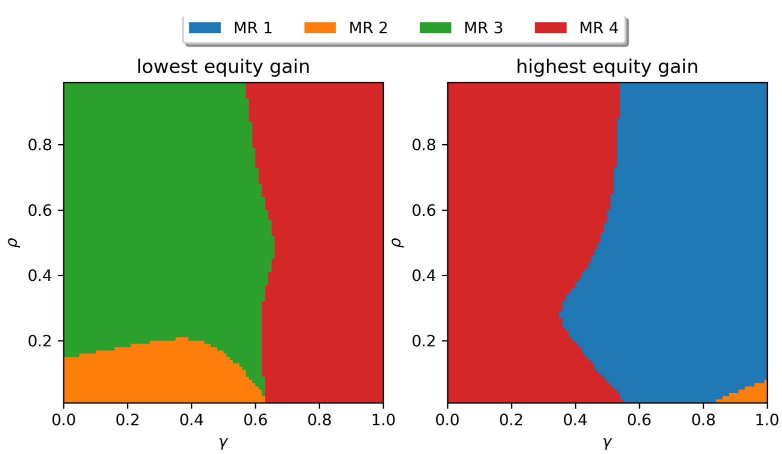
Figure 6 shows the reform with the lowest and the highest equity gain per euro, without rationing (upper panel) and with rationing (lower panel). At a first glance, these coloured zones are more vertically oriented, suggesting that the degree of compensation is a more important normative parameter compared to the degree of unfairness aversion.

Figure 6: Lowest and highest equity gain

(a) no rationing



(b) rationing



Notes: Given $\Delta R < 0$, $\frac{\Delta UI}{\Delta R}$ measures the equity gain per euro; equity losses appear as negative numbers.

Without rationing, the highest equity gain per euro always occurs if one reduce the taxes for the working: for the working rich if the degree of compensation is low (below 0.6 approximately) or for the working poor or middle (depending on the level of unfairness aversion) if the degree of compensation is high (above 0.6). Increasing the subsidies of the unemployed is never the most equitable reform.

With rationing, the picture changes drastically: the most valuable reform is either to reduce the taxes of the working rich if the degree of compensation is low (below 0.5 approximately) or to increase the subsidies of the unemployed if the degree of compensation is high (above 0.5).[29] Moreover, the lowest equity gains always realize if one reduces the taxes of one of the working groups.

The lowest and highest equity gains tell us that, without rationing, small transfers to the working rich (if the degree of compensation is lower than 0.6) or from the working rich to the working poor or working middle (if the degree of compensation is higher than 0.8) can be made budget neutral and will improve equity most. However, with rationing, small transfers from the working poor or working middle to the working rich (if the degree of compensation is lower than 0.5) or from the working rich to the unemployed (if the degree of compensation is higher than 0.6) can be made budget neutral and will improve equity most.

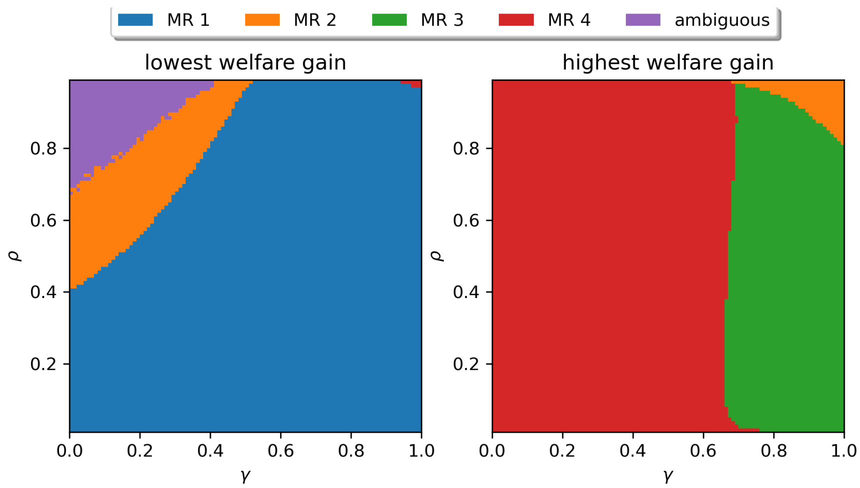
4.3.4 Welfare

Figure 7 shows the reform with the lowest and the highest welfare gain per euro without rationing (upper panel) and with rationing (lower panel). We start with the upper right panel. If the degree of compensation is not too high (below 0.7), the social planner prefers the reform in favor of the working rich, whatever the degree of unfairness aversion. For higher degrees of compensation (above 0.7), either the working poor or the working middle yield the highest welfare gain per euro. As seen already before, the social planner will favor the working poor only if she adopts a very high degree of both unfairness aversion and compensation. Reforms in favor of the unemployed never yield the highest welfare gain. In fact, the upper left panel indicates that helping the unemployed yields the lowest welfare gain for a wide range of normative positions, because of its inefficient nature.

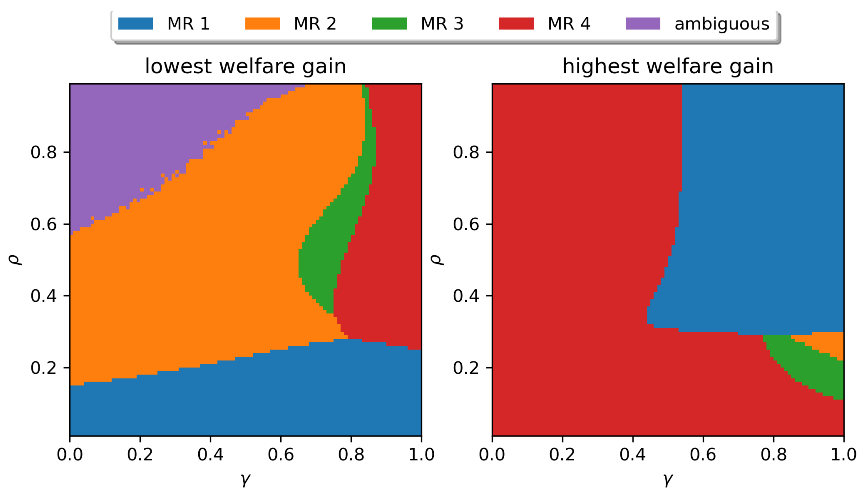
[29] We neglect the small zone characterized by a very low inequality aversion and a very high degree of compensation where the highest equity gain per euro occurs for the working poor.

Figure 7: Lowest and highest welfare gain

(a) no rationing



(b) rationing



Notes: Given $\Delta R < 0$, $-\frac{\Delta W}{\Delta R}$ measures the welfare gain per euro (and whelfare costs appear as negative numbers). The term 'ambiguous' refers to the fact that due to machine precision we are not able to rank reforms 1, 2 and 3.

The lower right panel (with rationing) shows that the social planner will favor marginal reforms in favor of the working rich as long as unfairness aversion is low (below 0.3) and the degree of compensation is not too high (below 0.8). With higher unfairness aversion, the preferred policy will depend on the degree of compensation: for ethical positions close to libertarianism, increasing the income of the working rich is the most desirable policy. For larger values of the degree of compensation (say, above 0.5), social welfare is most increased by increasing the transfers to the unemployed. Increasing the income of the working poor is almost never the best policy. On the contrary, as shown in the lower left panel, it has the lowest welfare gain per euro for a large range of parameter values.

Figure 7 illustrates the advantages of our flexible ethical framework. If we were to accept maximin combined with resource-egalitarianism and no rationing, then the upper right panel tells us that we must increase the transfers to the working poor. This confirms the result of Fleurbaey and Maniquet (2006). Yet, changing the normative parameters slightly, changes the results drastically. A lower degree of compensation (below 0.7) or a lower degree of unfairness aversion (below 0.8) would advocate transfers to the working rich or the working middle. Moreover, also the empirical setting matters. Once we allow for rationing, the same ethical position (maximin combined with resource egalitarianism) would advocate to increase transfers to the unemployed.

5 Conclusion

The traditional welfarist approach based on revealed preferences runs into difficulties with interpersonal utility comparisons under the (obviously realistic) assumption that preferences differ between individuals. One then needs a notion of well-being that can reflect such preference differences. Moreover, differences in outcomes will not only reflect differences in innate productivities (as in traditional optimal tax theory), but also differences in preferences, i.e., in the motivation to work. As there are different views in society about the ethical status of differences in innate productivities and in preferences, this raises the challenge of developing a social welfare framework that is sufficiently flexible to accommodate these different views.

We characterize a social welfare measure that aggregates fairness gaps, defined as the difference between the actual money-metric utilities of the individuals and the money-metric utilities they should have in a fair society. The fair allocation is only restricted to be Pareto efficient and anonymous. We propose to use a flexible fair allocation that is parametrized by the degree of compensation for productive skills, ranging from the libertarian laisser-faire allocation (with no compensation) to the resource-egalitarian equal-wage-equivalent allocation (with full compensation). Moreover, the aggregation of the fairness gaps is based on a strictly increasing and concave transformation function. This allows for different degrees of unfairness aversion, ranging from no aversion to unfairness to absolute priority to the worst off.

The flexibility of our social welfare function offers more possibilities than the ones we have worked out in this paper, as the fair money-metric utilities can be defined in many ways. Of course, it will never be as flexible as the direct specification of the welfare weights proposed by Saez and Stantcheva (2016), but, since it is the representation of a transitive social welfare ordering, it will never lead to inconsistencies when going beyond local approximations. Moreover, maximizing a well-defined social welfare function is perfectly in line with the public economics tradition on optimal taxation.

The applicability of the approach is illustrated with the empirical analysis of four hypothetical tax reforms. The degree of compensation turns out to be an important normative choice. We find that for libertarians the present tax burden of the rich is too large and should be lowered. For resource-egalitarians the evaluation depends on whether we allow for rationing on the labour market or not. If we assume that there is no rationing and if unfairness aversion is sufficiently high, then transfers should be directed to the working poor. This is in line with earlier findings in the fairness literature (Fleurbaey and Maniquet, 2006). However, if we allow for rationing and if unfairness aversion remains high, then transfers should be directed to the unemployed.

Our empirical analysis is only an illustration. It obviously does not make much sense to consider tax reforms for singles in isolation. The econometric analysis can also be refined. As it stands, however, our empirical analysis shows the relevance of a social welfare framework that goes beyond welfarism and can integrate ethical perspectives (such as libertarianism and resource-egalitarianism) that are often taken in society, but almost never found in traditional optimal tax analysis.

References

Adler, M. (2022). Theory of prioritarianism. Chapter 2 in: Adler, M., Norheim, O., eds., *Prioritarianism in Practice*. Cambridge: Cambridge University Press.

Almås, I., Cappelen, A., Lind, J.T., Sørensen, E., Tungodden, B. (2011). Measuring unfair (in)equality. *Journal of Public Economics* 95, 488–499.

Bargain, O., Orsini, K., Peichl, A. (2014). Comparing labor supply elasticities in Europe and the United States: new results. *Journal of Human Resources* 49(3), 723–838.

Berg and Piacquadio (2023). *Fairness and Paretian Social Welfare Functions*. Mimeo.

Boadway (2012). *From Optimal Tax Theory to Tax Policy: Retrospective and Prospective Views*, The Munich Lectures. Cambridge, MA: MIT Press.

Bosmans, K., Decancq, K., Ooghe, E. (2018). Who ís afraid of aggregating money metrics? *Theoretical Economics* 13, 467–484.

Bosmans, K., Lauwers, L., Ooghe, E. (2009). A consistent multidimensional Pigou-Dalton transfer principle. *Journal of Economic Theory* 144, 1358–1371.

Bosmans, K., Lauwers, L., Ooghe, E. (2018). Prioritarian poverty comparisons with cardinal and ordinal attributes. *Scandinavian Journal of Economics* 120, 925–942.

Cowell, F. (1985). Measures of distributional change: an axiomatic approach. *Review of Economic Studies* 52, 135–151.

Creedy, J., Hérault, N., Kalb, G. (2011). Measuring welfare changes in behavioural microsimulation modelling: accounting for the random utility component. *Journal of Applied Economics* 14, 5–34.

Deaton, A., Muellbauer, J. (1980). *Economics and consumer behaviour*. Cambridge University Press.

Devooght, K. (2008). To each the same and to each his own. A proposal to measure responsibility-sensitive income inequality. *Economica* 75, 280–295.

Diamond, P., Saez, E. (2011). The case for a progressive tax: from basic research to policy recommendations. *Journal of Economic Perspectives* 25(4), 165-190.

Dworkin, R. (1981). What is equality? Part 1: equality of welfare. *Philosophy and Public Affairs* 10(3), 185–246.

Finkelstein, A., Hendren, N. (2020). Welfare analysis meets causal inference. *Journal of Economic Perspectives* 34, 246–167.

Fleurbaey, M. (2008). *Fairness, Responsibility, and Welfare*. Oxford: Oxford University Press.

Fleurbaey, M., Maniquet, F. (1999). Fair allocation with unequal production skills: the solidarity approach to compensation. *Social Choice and Welfare* 16, 569–583.

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

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

Fleurbaey, M., Maniquet, F. (2018a). Optimal income taxation theory and principles of fairness. *Journal of Economic Literature* 56(3), 1029–1079.

Fleurbaey, M., Maniquet, F. (2018b). Inequality-averse well-being measurement. *International Journal of Economic Theory 14,* 35–50.

Fleurbaey, M., Schokkaert, E. (2009). Unfair inequalities in health and health care. *Journal of Health Economics* 28, 73–90.

Fleurbaey, M., Tadenuma, K. (2014). Universal social orderings: an integrated theory of policy evaluation, inter-society comparisons, and interpersonal comparisons. *Review of Economic Studies* 81(3), 1071–1101.

Fleurbaey, M., Trannoy, A. (2003). The impossibility of a Paretian egalitarian, *Social Choice and Welfare* 21, 243–264.

Gaertner, W., Schokkaert, E. (2012). *Empirical Social Choice*. Cambridge: Cambridge University Press.

Hendren, N., Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *Quarterly Journal of Economics* 135(3), 1209–1318.

Hufe, P., Kanbur, R., Peichl, A. (2022). Measuring unfair inequality: reconciling equality of opportunity and freedom from poverty. *Review of Economic Studies*, https://doi.org/10.1093/restud/rdab101

Layard, R., De Neve, E. (2023) *Wellbeing: Science and Policy.* Cambridge: Cambridge University Press.

Kaplow, L. (2008). *The Theory of Taxation and Public Economics.* Princeton University Press.

King, M. (1983). An index of inequality: with application to horizontal equity and social mobility. *Econometrica* 51, 99–115.

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature* 41, 1188–1239.

Magdalou, B., Nock, R. (2011). Income distributions and decomposable divergence measures. *Journal of Economic Theory* 146 (6), 2440–2454.

Mankiw, N., Weinzierl, M., Yagan, D. (2009) Optimal taxation in theory and practice. *The Journal of Economic Perspectives* 23(4), 147–174.

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

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

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.

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

Salanié, B. (2003). *The Economics of Taxation*. MIT Press.

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

Schokkaert, E., Tarroux, B. (2021). Empirical research on ethical preferences: how popular is prioritarianism? Chapter 10 in: Adler, M., Norheim, O., eds., *Prioritarianism in Practice*. Cambridge: Cambridge University Press.

Sen, A. (1985). *Commodities and Capabilities*. North-Holland: Elsevier.

Sheffrin, S. (2013). *Tax Fairness and Folk Justice*. Cambridge: Cambridge University Press.

Sher, I. (2021). *Generalized social marginal welfare weights imply inconsistent comparisons of tax policies*. Mimeo.

van Soest, A. (1995). Structural models of family labor supply: A discrete choice approach. *Journal of Human Resources* 30(1), 63–88.

Thomson, W. (2011). Fair allocation rules. Chapter 21 in: Arrow, K., Sen, A., Suzumura, K., eds., *Handbook of Social Choice and Welfare*, Volume 2, North-Holland: Elsevier.

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

Tuomala, M. (2016). *Optimal Redistributive Taxation*. Oxford: Oxford University Press.

Tuomala, M., Weinzierl, M. (2022). Prioritarianism and optimal taxation. Chapter 4 in: Adler, M., Norheim, O., eds., *Prioritarianism in Practice*. Oxford: Oxford University Press.

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

Weinzierl, M. (2018). Revisiting the classical view of benefit-based taxation. *Economic Journal* 128, 37–64.

A Proof of theorem 1

First, Representation allows us to represent social welfare as

$$W(x;\mathscr{S}) = \frac{1}{n}\sum_{i\in I} v_i(x_i;\mathscr{S}),$$

with $v_1, v_2, \ldots, v_n$ evaluation functions that are continuously differentiable in resources (earnings and labour).

Second, Pareto requires that the evaluation function v_i is a strictly increasing transformation of money-metric utility, i.e., for each individual $i = 1, 2, \ldots, n$, we have $v_i(x_i;\mathscr{S}) = \phi_i(m(x_i,\theta_i);\mathscr{S})$, with $\phi_i'(\cdot;\mathscr{S}) > 0$. Social welfare can thus be represented as

$$W(x;\mathscr{S}) = \frac{1}{n}\sum_{i\in I} \phi_i(m(x_i,\theta_i);\mathscr{S}), \tag{11}$$

where the transformation functions satisfy $\phi_i'(\cdot;\mathscr{S}) > 0$ for all $i = 1, 2, \ldots, n$.[30]

Third, the (Pareto-efficiency-preserving) transfer principle can be stated in terms of regressive transfers as follows. A regressive transfers of resources between two individuals i and j, i.e.,

$$v_i(x_i';\mathscr{S}) > v_i(x_i;\mathscr{S}) \geq v_j(x_j;\mathscr{S}) > v_j(x_j';\mathscr{S}),$$

that does not affect other individuals, i.e.,

$$x_k = x_k', k \neq i, j,$$

and satisfying $x, x' \in P(\mathscr{S})$ should be disapproved of, i.e., $W(x;\mathscr{S}) > W(x';\mathscr{S})$. We proceed in several steps.

1. Because x, x' are Pareto efficient allocations for the same society $\mathscr{S}$, we must have

$$\frac{1}{n}\sum_{i\in I} m(x_i,\theta_i) = -R_0 = \frac{1}{n}\sum_{i\in I} m(x_i',\theta_i),$$

and thus, given $x_k = x_k'$, for all $k \neq i, j$, we get $m(x_i,\theta_i) + m(x_j,\theta_j) = m(x_i',\theta_i) + m(x_j',\theta_j)$. In other words, any transfer of resources between two individuals that preserves Pareto efficiency corresponds with a mean-preserving transfer of money-metric utilities between

[30] Note that it is not restrictive to use money-metric utility computed at own skills and tastes because the transformation functions $\phi_i(\cdot;\mathscr{S})$ can depend on society. In other words, the current money-metric utility function can be transformed into any other possible function that cardinalizes preferences from knowledge of the money-metric utility, the skill, and the taste of an individual.

these two individuals. For later use, define this transfer in money-metric utility as

$$T = m(x_i', \theta_i) - m(x_i, \theta_i) = m(x_j, \theta_j) - m(x_j', \theta_j) > 0. \tag{12}$$

2. Using Representation and given $x_k = x_k', k \neq i, j$, the requirement $W(x; \mathscr{S}) > W(x'; \mathscr{S})$ can be rewritten as
$$v_i(x_i; \mathscr{S}) + v_j(x_j; \mathscr{S}) > v_i(x_i'; \mathscr{S}) + v_j(x_j'; \mathscr{S}). \tag{13}$$

3. Using $v_i(x_i; \mathscr{S}) = \phi_i(m(x_i, \theta_i); \mathscr{S})$ for each individual i, as derived before, and using equations (12) and (13), the (regressive) transfer principle can now be rewritten as follows. Suppose $x \in P(\mathscr{S})$ holds. A regressive transfer $T > 0$ of money-metric utility from individual j and i (ceteris paribus), i.e.,
$$\phi_i(m(x_i, \theta_i) + T; \mathscr{S}) > \phi_i(m(x_i, \theta_i); \mathscr{S}) \geq \phi_j(m(x_j, \theta_j); \mathscr{S}) > \phi_j(m(x_j, \theta_j) - T; \mathscr{S}),$$
should be disapproved of, i.e.,
$$\phi_i(m(x_i, \theta_i); \mathscr{S}) + \phi_j(m(x_j, \theta_j); \mathscr{S}) > \phi_i(m(x_i, \theta_i) + T; \mathscr{S}) + \phi_j(m(x_j, \theta_j) - T; \mathscr{S}).$$

4. In the limit $T \to 0$, Transfer requires that $\phi_i(m(x_i, \theta_i); \mathscr{S}) \geq \phi_j(m(x_j, \theta_j); \mathscr{S})$ implies $\phi_i'(m(x_i, \theta_i); \mathscr{S}) \leq \phi_j'(m(x_j, \theta_j); \mathscr{S})$ for any two individuals i, j and for any Pareto efficient allocation x. As the opposite is true as well, we get that $\phi_i(m(x_i, \theta_i); \mathscr{S}) = \phi_j(m(x_j, \theta_j); \mathscr{S})$ implies $\phi_i'(m(x_i, \theta_i); \mathscr{S}) = \phi_j'(m(x_j, \theta_j); \mathscr{S})$ for any two individuals i, j and for any Pareto efficient allocation x. As it must hold for any pair of individuals and for any Pareto efficient allocation x, we can reformulate the condition as follows: for each individual i, the derivative $\phi_i'(\cdot; \mathscr{S})$ must be a common, continuous, and strictly positive function, say $f(\cdot; \mathscr{S}) > 0$, of the level function $\phi_i(\cdot; \mathscr{S})$, i.e., $\phi_i'(m; \mathscr{S}) = f(\phi_i(m; \mathscr{S}); \mathscr{S})$ must hold for each individual i and each money-metric utility level m.

5. This differential equation can be solved as follows.[31] Define $\varphi'(\cdot; \mathscr{S}) = 1/f(\cdot; \mathscr{S}) > 0$. We can now rewrite the differential equation as
$$\frac{\phi_i'(m; \mathscr{S})}{f(\phi_i(m; \mathscr{S}); \mathscr{S})} = \varphi'(\phi_i(m); \mathscr{S})\phi_i'(m; \mathscr{S}) = 1.$$
Integrating both sides with respect to m, we get $\varphi(\phi_i(m; \mathscr{S}); \mathscr{S}) = m + k_i(\mathscr{S})$ for some constant $k_i(; \mathscr{S})$, leading to $\phi_i(m; \mathscr{S}) = \varphi^{-1}(m + k_i(\mathscr{S}); \mathscr{S})$ for each i and m. Define a

[31] This step of the proof is based on Bosmans, Lauwers, and Ooghe (2009).

common function $\phi(\cdot;\mathscr{S}) = \varphi^{-1}(\cdot;\mathscr{S})$ to obtain that social welfare can be represented as

$$W(x;\mathscr{S}) = \frac{1}{n}\sum_{i\in I}\phi(m(x_i,\theta_i)+k_i(\mathscr{S});\mathscr{S}), \tag{14}$$

with $\phi'(\cdot;\mathscr{S}) > 0$ and $k_1(\mathscr{S}), k_2(\mathscr{S}), \ldots, k_n(\mathscr{S})$ arbitrary constants.[32]

6. Finally, note that $\phi''(\cdot;\mathscr{S}) < 0$ must hold as well to ensure that the transfer principle also works beyond the limiting case $T \to 0$ that we considered before. To see this, we can use $\phi_i(m(x_i,\theta_i);\mathscr{S}) = \phi(m(x_i,\theta_i)+k_i(\mathscr{S});\mathscr{S})$ to rewrite the transfer principle as follows: a regressive transfer $T > 0$ of money-metric utility from individual j and i (ceteris paribus), i.e.,

$$\phi(m(x_i,\theta_i)+T+k_i(\mathscr{S});\mathscr{S}) > \phi(m(x_i,\theta_i)+k_i(\mathscr{S});\mathscr{S}) \geq$$
$$\phi(m(x_j,\theta_j)+k_j(\mathscr{S});\mathscr{S}) > \phi(m(x_j,\theta_j)-T+k_j(\mathscr{S});\mathscr{S}),$$

should be disapproved of, i.e.,

$$\phi(m(x_i,\theta_i)+k_i(\mathscr{S});\mathscr{S}) + \phi(m(x_j,\theta_j)+k_j(\mathscr{S});\mathscr{S}) >$$
$$\phi(m(x_i,\theta_i)+T+k_i(\mathscr{S});\mathscr{S}) + \phi(m(x_j,\theta_j)-T+k_j(\mathscr{S});\mathscr{S}).$$

As we are allowed to choose $m(x_i,\theta_i)+k_i(\mathscr{S}) = m(x_j,\theta_j)+k_j(\mathscr{S}) \equiv \mu$, the transfer principle requires

$$\phi(\mu;\mathscr{S}) + \phi(\mu;\mathscr{S}) > \phi(\mu+T;\mathscr{S}) + \phi(\mu-T;\mathscr{S}),$$

for any μ, which requires $\phi(\cdot;\mathscr{S})$ to be strictly concave and thus, given differentiability, leads to $\phi''(\cdot;\mathscr{S}) < 0$.

Fourth, Anonymity requires $W(\ldots,x_i,\ldots,x_j,\ldots;\mathscr{S}) = W(\ldots,x_j,\ldots,x_i,\ldots;\mathscr{S})$ if $\theta_i = \theta_j$. Using equation (14), this implication can be spelled out as

$$\phi(m(x_i,\theta_i)+k_i(\mathscr{S});\mathscr{S}) + \phi(m(x_j,\theta_j)+k_j(\mathscr{S});\mathscr{S}) =$$
$$\phi(m(x_j,\theta_i)+k_i(\mathscr{S});\mathscr{S}) + \phi(m(x_i,\theta_j)+k_j(\mathscr{S});\mathscr{S}).$$

As $\theta_i = \theta_j$ implies $m(x_j,\theta_i) = m(x_j,\theta_j) \equiv \mu_j$ and $m(x_i,\theta_j) = m(x_i,\theta_i) \equiv \mu_i$, the implication is

$$\phi(\mu_i+k_i(\mathscr{S});\mathscr{S}) + \phi(\mu_j+k_j(\mathscr{S});\mathscr{S}) = \phi(\mu_j+k_i(\mathscr{S});\mathscr{S}) + \phi(\mu_i+k_j(\mathscr{S});\mathscr{S}),$$

which is, given strict concavity (and thus non-linearity) of ϕ, only possible if $k_i(\mathscr{S}) = k_j(\mathscr{S})$. To sum up, if $\theta_i = \theta_j$, then $k_i(\mathscr{S}) = k_j(\mathscr{S})$ must hold.

[32]Note indeed that $\phi'(m;\mathscr{S}) = 1/\varphi'(\varphi^{-1}(m;\mathscr{S});\mathscr{S}) = f(\varphi^{-1}(m;\mathscr{S});\mathscr{S}) > 0$ for all m.

Fifth, to sum up, a social welfare measure $W(x;\mathscr{S})$ satisfies Representation, Pareto, Anonymity, and Transfer if and only if it can be written as

$$W(x;\mathscr{S}) = \frac{1}{n}\sum_{i\in I}\phi(m(x_i,\theta_i)+k_i(\mathscr{S});\mathscr{S}),$$

where the transformation function $\phi(\cdot;\mathscr{S})$ satisfies $\phi'(\cdot;\mathscr{S}) > 0$ and $\phi''(\cdot;\mathscr{S}) < 0$ and $k_i(\mathscr{S})$ are individual-specific constants satisfying $k_i(\mathscr{S}) = k_j(\mathscr{S})$ if $\theta_i = \theta_j$. Let $x^*(\mathscr{S})$ be an allocation that maximizes welfare $W(x;\mathscr{S})$ subject to the feasibility constraint $x \in F(\mathscr{S})$. Given the properties of the welfare function (implied by Pareto and anonymity), the allocation $x^*(\mathscr{S})$ must be a Pareto efficient and anonymous allocation (i.e., $x^*(\mathscr{S}) \in P(\mathscr{S})$ and $x_i^*(\mathscr{S}) = x_j^*(\mathscr{S})$ for all i, j in I with $\theta_i = \theta_j$). Using $\lambda > 0$ as the Lagrange multiplier, the first-order conditions of the social planner, evaluated at allocation $x^*(\mathscr{S})$, are

$$\phi'(m(x_i^*(\mathscr{S}),\theta_i)+k_i(\mathscr{S});\mathscr{S})\frac{\partial m(x_i^*(\mathscr{S}),\theta_i)}{\partial c} = \lambda,$$

$$\phi'(m(x_i^*(\mathscr{S}),\theta_i)+k_i(\mathscr{S});\mathscr{S})\frac{\partial m(x_i^*(\mathscr{S}),\theta_i)}{\partial \ell} = -\lambda s_i,$$

for each individual. Because the optimal allocation is Pareto efficient, we have $\frac{\partial m(x_i^*(\mathscr{S}),\theta_i)}{\partial c} = 1$ and $\frac{\partial m(x_i^*(\mathscr{S}),\theta_i)}{\partial \ell} = -s_i$ for each individual. The system of first-order conditions reduces therefore to

$$\phi'(m(x_i^*(\mathscr{S}),\theta_i)+k_i(\mathscr{S});\mathscr{S}) = \lambda,$$

for each individual. These conditions can be satisfied only if there is a constant $k(\mathscr{S})$ such that $k_i(\mathscr{S}) = -m(x_i^*(\mathscr{S}),\theta_i) + k(\mathscr{S})$ for each individual i. If we take up the common constant $k(\mathscr{S})$ in the transformation function $\phi(\cdot;\mathscr{S})$, welfare becomes

$$W(x;\mathscr{S}) = \frac{1}{n}\sum_{i\in I}\phi(m(x_i,\theta_i)-m(x_i^*(\mathscr{S}),\theta_i);\mathscr{S}),$$

as required.

B A decomposition

Let ϕ be a transformation function that is invariant to additions, i.e., either ϕ is linear or $\phi(m+c) = \phi(m)\phi(c)$ holds for arbitrary scalars m, c. Two remarks. First, if $\phi(m+c) = \phi(m)\phi(c)$ holds for any two scalars m, c, then $\phi^{-1}(m'c') = \phi^{-1}(m') + \phi^{-1}(c')$ holds for any two scalars m', c'. To see this, $\phi(m+c) = \phi(m)\phi(c)$ implies $m + c = \phi^{-1}(\phi(m)\phi(c))$. Defining $m' = \phi(m)$ and $c' = \phi(c)$ leads to the desired result. Second, note that the fair allocation x^* is Pareto efficient and thus exactly feasible, i.e., $\frac{1}{n}\sum_i(c_i^* - s_i\ell_i^*) = -R_0$. For Pareto efficient bundles, we have $m_i^* = c_i^* - s_i\ell_i^*$

by definition, so that $\frac{1}{n}\sum_i m_i^* = -R_0$ holds.

Social welfare can now be decomposed as follows:

$$\begin{aligned}
W(x;\mathscr{S}) &= \phi^{-1}\{\frac{1}{n}\sum_{i\in I}\phi(m_i - m_i^*)\}, \\
&= \frac{1}{n}\sum_{i\in I} m_i - \underbrace{\frac{1}{n}\sum_{i\in I} m_i^*}_{-R_0} + \phi^{-1}\{\frac{1}{n}\sum_{i\in I}\phi(m_i - m_i^* - \frac{1}{n}\sum_{i\in I}(m_i - m_i^*))\}, \\
&= [R_0 - R(x;\mathscr{S})] - [-\frac{1}{n}\sum_{i\in I} m_i - R(x;\mathscr{S})] - \\
&\qquad [-\phi^{-1}\{\frac{1}{n}\sum_{i\in I}\phi(m_i - m_i^* - \frac{1}{n}\sum_{i\in I}(m_i - m_i^*))\}], \\
&= RS(x;\mathscr{S}) - EB(x;\mathscr{S}) - UI(x;\mathscr{S}),
\end{aligned}$$

as required.

C Data selection

We describe the data selection for both the hourly gross wage estimation, the preference estimation, and the simulations. For the preference estimation and the simulations, we work with our basic sample of singles without children, between 18 and 65 years old, not living with their parents, who are either unemployed (job-seeking) or employed (but not self-employed) with a wage higher than the 2016 minimum wage (9.11 euro), and who do not receive any disability allowance. For the hourly gross wage estimation, we extended the sample to individuals with a partner (and possibly children) to get more precise estimates.[33]

D Hourly gross wages

For individuals who worked in 2016, we observe their yearly gross earnings and the numbers of months worked part-time and full-time. We compute their hourly gross wage rate as

$$\text{hourly gross wage} = \frac{\text{yearly gross earnings}}{24 \times \frac{52}{12} \times \#\text{part-time months} + 40 \times \frac{52}{12} \times \#\text{full-time months}},$$

where the numbers 24 and 40 correspond with the median number of hours worked per week (as reported by part-time and full-time working individuals in 2017). We exclude individuals with a

[33] The sample extension also allowed us to estimate a Heckman selection model, but in the end we did not use it as the results were very similar to the simpler OLS model.

gross wage rate below the minimum wage (9.11 euro).

We estimate a standard OLS regression model to predict the hourly gross wage rate of individuals who did not work in 2016. The estimation results are presented in Table 8.

Table 8: OLS estimates of the log of hourly gross wages

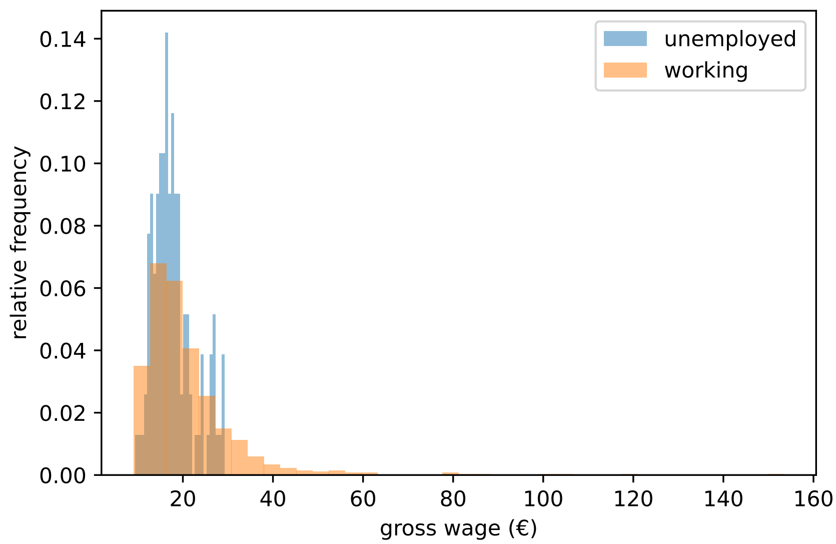
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.3312	0.0249	93.47	0.0000
gender-male	0.1002	0.0187	5.35	0.0000
educ-secondary	0.1399	0.0171	8.16	0.0000
educ-tertiary	0.4688	0.0166	28.28	0.0000
area-middle-density	-0.0159	0.0120	-1.32	0.1855
area-rural	-0.0382	0.0161	-2.38	0.0176
has-partner	-0.0140	0.0158	-0.89	0.3736
gender-male × has-partner	0.0797	0.0225	3.55	0.0004
experience	0.0259	0.0017	14.81	0.0000
experience^2	-0.0004	0.0000	-8.50	0.0000
nat-EU	0.0541	0.0192	2.82	0.0048
nat-not-EU	-0.1401	0.0311	-4.50	0.0000
RMSE	0.3239			
adjusted R^2	0.3422			

Notes: because of the extended sample (see Appendix B), we included a dummy 'has-partner' as a covariate.

Ceteris paribus, males have higher wages (+10%), higher education levels lead to higher wages (+14% for a secondary degree and +47% for a tertiary degree as highest degree relative to lower than secondary degrees), wages in rural areas are lower than in urban areas (–4%), having a partner leads to higher wages for males (+7%), but not for females, work experience leads to higher wages at a decreasing rate, and (non-Belgian) EU-citizens earn higher wages (+5%), while non-EU citizens earn lower wages (–14%) compared to Belgians.

The distribution of hourly gross wages—computed for the working and predicted for the unemployed—is displayed in Figure 8 for the (limited) sample of singles without children that we use in the simulations. Predictions below the minimum wage lead to exclusion, so both distributions are truncated to the left at the minimum wage (9.11 euro). The mean hourly gross wage rate is 21.16 euro for the working and 17.70 euro for the unemployed. Moreover, the wage dispersion is larger among the working.

Figure 8: Distribution of hourly gross wages for working and unemployed singles without children



E Estimation of preferences

Utility is specified as

$$u(c_i(\ell), \ell; z_i) + \epsilon_i(\ell), \tag{15}$$

with $u(c_i(\ell), \ell; z_i)$ the deterministic utility part and $\epsilon_i(\ell)$ a random utility term (independent and identically distributed over individuals and choices according to an extreme value type I distribution).

The deterministic utility term is specified as the sum of (i) the log of (augmented) net income, (ii) labour dummies, and (iii) a taste-for-work shifter, i.e.,

$$u(c_i(\ell), \ell; z_i) = \alpha \log(c_i(\ell) + \kappa) + 1[\ell = 24]\beta + 1[\ell = 38]\gamma + 1[\ell = 51]\delta + 1[\ell \neq 0]t(z_i), \tag{16}$$

with κ a prespecified positive constant (introduced to avoid negative numbers for the logarithmic function),[34] $1[\cdot]$ a dummy variable that is equal to one if the expression between brackets is true and zero otherwise, α, β, γ, and δ preference parameters (to be estimated), and t a taste-for-work function of the covariates (see below). As in van Soest (1995), we use a simulated maximum likelihood procedure to account for the uncertainty in the prediction of the wage rates of the unemployed.[35]

[34] We also experimented with polynomial utility-of-consumption functions, but these were outperformed in terms of fit (using Akaike's information criterion) by the simple log-specification with κ set to 5000 (the number that is also used for the estimations and simulations).

[35] For the unemployed, we use the OLS regression estimates of Table 8 and add 50 i.i.d. errors terms from a

E.1 No rationing

In case there is no rationing, each individual chooses labour hours from the discrete choice set $L = \{0, 24, 38, 51\}$ to maximize utility. The taste-for-working function is specified as

$$\begin{aligned} t(z_i) \; = \; & \kappa_{10} 1[\text{gender} = \text{male}] + \kappa_{11} 1[\text{educ} = \text{secondary}] + \kappa_{12} 1[\text{educ} = \text{tertiary}] + \\ & \kappa_{13} 1[\text{gender} = \text{male}] \times 1[\text{educ} = \text{secondary}] + \kappa_{14} 1[\text{gender} = \text{male}] \times 1[\text{educ} = \text{tertiary}] + \\ & \kappa_2 \frac{\text{age}}{100} + \kappa_3 \left(\frac{\text{age}}{100}\right)^2 + \\ & \kappa_{40} 1[\text{area} = \text{middle}] + \kappa_{41} 1[\text{area} = \text{rural}] + \\ & \kappa_{50} 1[\text{nationality} = \text{EU}] + \kappa_{51} 1[\text{nationality} = \text{not EU}]. \end{aligned}$$

Table 9 presents the estimation results.

Table 9: Estimation results for preferences (no rationing)

Parameter	Estimate	Std. Error	Pr(>|z|)
α	13.6492	3.1620	0.0000
β	-3.0194	0.3215	0.0000
γ	-2.6509	0.1827	0.0000
δ	-5.7496	0.6862	0.0000
κ_{10}	-0.9917	0.4716	0.0355
κ_{11}	0.1847	0.5829	0.7513
κ_{12}	0.8075	1.1006	0.4632
κ_{13}	0.2643	0.7702	0.7315
κ_{14}	0.4001	1.4910	0.7884
κ_2	23.9396	0.6811	0.0000
κ_3	-36.0606	1.1982	0.0000
κ_{40}	0.7697	0.6227	0.2164
κ_{41}	0.5692	0.9502	0.5492
κ_{50}	-0.1663	1.0318	0.8719
κ_{51}	0.2612	1.6762	0.8762

The coefficients α, β, γ, and δ are significant and have the expected sign. In particular, the negative coefficients for β, γ, and δ indicate a disutility for working (for the reference type). Among the other covariates only gender (κ_{10}) and age (κ_2,κ_3) are significant. Males (κ_{10}) have a lower taste for working. Being older (κ_2,κ_3) initially increases the taste for working (up to the age of 33 years), but decreases it afterwards.

normal distribution $N(0, \sigma_\epsilon^2)$ to simulate their log-likelihood contribution. The variance σ_ϵ^2 is approximated by the variance of the error terms of the OLS regression.

E.2 Rationing

Let $O \subseteq L$ denote an opportunity set and let $\mathscr{O}$ collect all subsets of L that contain the unemployment alternative 0.[36] The probability that an individual chooses an amount of labour $\ell \in L$ is now defined as

$$P(\ell|z_i) = \sum_{O \in \mathscr{O} | \ell \in O} P(O|z_i) \times \frac{\exp(u(c_i(\ell), \ell; z_i))}{\sum_{\ell' \in O} \exp(u(c_i(\ell'), \ell'; z_i))}, \tag{17}$$

where the first factor to the right-hand side of equation (17) is the probability that individual i faces opportunity set O and the second factor is the probability that individual i chooses ℓ from this opportunity set. We further impose that the probability that an opportunity ℓ is available in someone's opportunity set, is independent of other opportunities.[37] The probability that an individual faces opportunity set O is therefore defined as

$$P(O|z_i) = \Pi_{\ell \in O} p(\ell|z_i) \Pi_{\ell \notin O} (1 - p(\ell|z_i)),$$

with $p(\ell|z_i)$ the probability that ℓ is in the opportunity set of individual i (where $p(0|z_i) = 1$ for all individuals).

To estimate the probabilities, we use a logit specification of the form:

$$p(\ell|z_i) = \frac{\exp(f(\ell|z_i))}{1 + \exp(f(\ell|z_i))},$$

with

$$\begin{aligned} f(\ell|z_i) =& \rho_0^\ell + \rho_1^\ell 1[educ = \text{secondary}] + \rho_2^\ell 1[educ = \text{tertiary}] + \rho_3^\ell 1[area = \text{middle-density}] + \\ & \rho_4^\ell 1[area = \text{rural}] + \rho_5^\ell 1[region = \text{Flanders}] + \rho_6^\ell 1[region = \text{Wallonia}] + \\ & \rho_7^\ell 1[nationality = \text{EU}] + \rho_8^\ell 1[nationality = \text{Non-EU}] + \rho_9^\ell 1[gender = \text{male}]. \end{aligned}$$

For the preference specification, we now assume that the tastes-for-working function is only a function of age and gender[38], i.e.,

$$t(z_i) = \kappa_{10} 1[gender = \text{male}] + \kappa_2 \frac{age}{100} + \kappa_3 (\frac{age}{100})^2.$$

The results of the estimation are displayed in Table 10. The coefficients α, β, γ, and δ are again significant, but now the coefficients β, γ, and δ are positive, indicating that the reference

[36] Given four elements in $L = \{0, 24, 38, 51\}$ and given the assumption that not working is available in each opportunity set, we are left with eight possible subsets in $\mathscr{O}$.

[37] We also estimated a version where these probabilities were not independent, but information criteria such as the AIC criterion favor the current specification.

[38] These were the only significant variables in the specification without rationing, see Table 9.

type prefers working to being unemployed, ceteris paribus. Gender (κ_{13}) and age (κ_2,κ_3) are again significant and their sign is the same as in the case without rationing (Table 9). For the estimation of opportunities, better education (ρ_1, ρ_2) improves work opportunities (for working full-time or more), living in an urban area (the reference group for ρ_3, ρ_4) worsens work opportunities, living in Flanders (ρ_5) improves work opportunities (except for working more than full-time) relative to living in Brussels, living in Wallonia (ρ_6) worsens work opportunities (except for working half-time) relative to living in Brussels, nationality (ρ_7, ρ_8) does not seem to play a significant role, and being male (ρ_9) worsens opportunities for working half-time, but improves opportunities for working full-time and more.

Table 10: Estimation results for preferences and opportunities

Parameter	Estimate	Std. Error	Pr(>|z|)
α	1.4894	0.1286	0.0000
β	2.0619	0.1054	0.0000
γ	0.4682	0.1787	0.0088
δ	6.8668	0.9995	0.0000
κ_{10}	-1.3599	0.2251	0.0000
κ_2	23.4902	0.3165	0.0000
κ_3	-42.3195	0.5441	0.0000
ρ_0^{24}	0.8625	0.1764	0.0000
ρ_1^{24}	-0.5701	0.2714	0.0356
ρ_2^{24}	0.5554	0.3411	0.1034
ρ_3^{24}	0.0117	0.3237	0.9711
ρ_4^{24}	0.8407	0.5336	0.1151
ρ_5^{24}	0.7297	0.3556	0.0402
ρ_6^{24}	-0.0213	0.2959	0.9426
ρ_7^{24}	-0.8578	0.5473	0.1170
ρ_8^{24}	0.5671	0.6988	0.4171
ρ_9^{24}	-1.7220	0.2245	0.0000
ρ_0^{38}	0.0563	0.1508	0.7091
ρ_1^{38}	0.4703	0.2196	0.0322
ρ_2^{38}	2.1472	0.4492	0.0000
ρ_3^{38}	0.7135	0.2990	0.0170
ρ_4^{38}	0.6452	0.4264	0.1302
ρ_5^{38}	1.3774	0.5135	0.0073
ρ_6^{38}	-0.6296	0.2108	0.0028
ρ_7^{38}	-0.2480	0.4001	0.5354
ρ_8^{38}	0.1679	0.6092	0.7828
ρ_9^{38}	0.6895	0.2132	0.0012

Parameter	Estimate	Std. Error	Pr(>|z|)
ρ_0^{51}	-3.4837	0.2303	0.0000
ρ_1^{51}	0.7679	0.3724	0.0392
ρ_2^{51}	2.7129	0.3413	0.0000
ρ_3^{51}	0.7506	0.3657	0.0401
ρ_4^{51}	2.5794	0.7876	0.0011
ρ_5^{51}	-0.1865	0.3420	0.5855
ρ_6^{51}	-2.8667	0.5696	0.0000
ρ_7^{51}	-0.1934	0.6968	0.7814
ρ_8^{51}	0.1355	1.0673	0.8989
ρ_9^{51}	1.4420	0.2969	0.0000