PREFERENCE RESPONSIBILITY VERSUS POVERTY REDUCTION IN THE TAXATION OF LABOR INCOMES

Lancelot Henry de Frahan, François Maniquet

REPRINT | 3199

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email:lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/core-reprints.html

Journal of Public Economics 197 (2021) 104386

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Preference responsibility versus poverty reduction in the taxation of labor incomes ☆

Lancelot Henry de Frahan [a,*], François Maniquet [b]

[a] Becker Friedman Institute at the University of Chicago, Chicago 60637, USA
[b] CORE/LIDAM, UCLouvain, Louvain-La-Neuve 1348, Belgium, and LISER, Esch-sur-Alzette 4366, Luxembourg

ARTICLE INFO

Article history:
Received 26 April 2020
Revised 8 February 2021
Accepted 10 February 2021
Available online 10 April 2021

JEL classification:
H21
D63

Keywords:
optimal taxation
responsibility
poverty reduction

ABSTRACT

We study the tax schemes that maximize social welfare functions built on axioms of responsibility for one's preferences (the requirement that the social welfare function should treat identically agents with the same wage, independently of their preferences) and poverty reduction. We find zero and negative marginal tax rates on low incomes at the optimum and bunching at the income level of the most hard-working minimum wage households. When preferences are iso-elastic, we derive the optimal tax formula, which we calibrate to the US economy. Our formula approximates the shape of the current US tax function for households with at least one child. This result suggests that a fairness-based approach, and these axioms in particular, can help close the gap between the recommendations of optimal tax theory and actual policies.

[…] a pivotal part of this economic plan is increasing the earned-income tax credit which, more than anything else we could do, will reward work and family and responsibility […] this will be the first time in the history of our country when we will be able to say that if you work 40 hours a week and you have children in your home, you will be lifted out of poverty."

President Bill Clinton, July 29, 1993 (shortly before expanding the Earned Income Tax Credit)

1. Introduction

Income taxation in general, and labor income taxation in particular, is the ultimate policy instrument to reduce income inequality. Since Mirrlees (1971)'s seminal contribution, most of the literature has embraced the view that full income equality is not a legitimate objective. The literature on optimal taxation is mainly welfarist: the utilitarian objective, especially when individual utility functions are concave, has long been seen appropriate to reduce inequality without eliminating it. Several authors, however, have recently expressed dissatisfaction with the utilitarian objective, in particular when preference heterogeneity is taken into account (see, for instance, Boadway, 2012, or Piketty and Saez, 2013). In addition, as reflected in the above quotation of President Bill Clinton, public tax policy discussions rarely revolve around utilitarian principles and often invoke some notion of "fairness" or other non-welfarist considerations. Finally, Weinzierl (2014) also finds survey-based evidence that most Americans reject important implications of utilitarianism.

An alternative approach to utilitarianism has been recently proposed, after the introduction of the ethics of responsibility into normative debates (see detailed accounts in Roemer, 1998, and Fleurbaey, 2008). This ethics recently emerged in political philosophy and yielded theories of equality of resources and equality of opportunities, to list just two main examples (see Roemer, 1998, and Fleurbaey, 2008 for discussions and references). In a nutshell, the ethics of responsibility is grounded on the assumption that not all inequalities are unjust. As a consequence, the social objective should be to identify and eliminate the unjust inequalities and to remain neutral towards the other inequalities. Several authors have studied the consequences on income taxation of considering that income inequalities due to differences in labor time are not

☆ François Maniquet's work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ ERC n° 269831. This research advanced while Maniquet was visiting the department of economics of University College London. Generous support from the Leverhulme Foundation is gratefully acknowledged. Declarations of interest: none.
* Coreesponding author.
E-mail addresses: lancelot@uchicago.edu (L. Henry de Frahan), francois.maniquet@uclouvain.be (F. Maniquet).

https://doi.org/10.1016/j.jpubeco.2021.104386
0047-2727/

unjust (see Fleurbaey and Maniquet, 2006, 2007 and Lockwood and Weinzierl, 2015). To put it differently, under the latter view, agents with the same wage rate should be free to choose their labor time and redistributing incomes among them is not legitimate. In addition to existing references to this view by politicians (see, for instance, the quotation of President Bill Clinton above), it seems to reflect views of citizens. Using questionnaire-based experiments, Konow (2001) presents hypothetical redistribution situations to American subjects. Redistributing production among equally skilled individuals having exerted different levels of efforts is largely viewed as unfair, whereas redistributing from high-skill to low-skill individuals is viewed as fair.[1]

One may argue, however, that freeness to choose may lead to freeness to lose, if there is not enough redistribution to lift agents out of poverty. Also using questionnaire experiments, Weinzierl (2014) investigates the popular support to redistributing towards the poor. Subjects have to express their opinion over actual and counterfactual tax-transfer schemes in the US. Weinzierl observes that a majority of subjects agree to subsidize the poor but less than what utilitarian or maximin welfarist preferences would recommend. That raises the question of the compatibility between the fairness principles of responsibility and poverty reduction.

In this paper, we derive the formula of the optimal labor income tax scheme consistent with the fairness principles of responsibility and poverty reduction and we calibrate the formula to the US economy. More specifically, we first build a family of social welfare functions that combine neutrality towards inequalities caused by different labor time choices with the goal of reducing poverty. Second, we derive an optimal tax formula from the maximization of these unconventional social welfare functions under incentive and budget constraints.[2] Third, using Current Population Survey (CPS) data, in combination with an estimate of the labor supply elasticity from Chetty (2012), we calibrate the formula to the US economy. Finally, we compare the resulting calibrations with the current system.

There are three main lessons to draw from our paper. First, it is possible to define a social objective combining the three goals of poverty alleviation, responsibility and efficiency and to derive the resulting optimal tax. Surprisingly enough, the most basic tension among these three goals is the conflict between poverty reduction and Pareto efficiency. If being poor means consuming a bundle of goods below some poverty line, the main issue comes from the fact that agents may prefer bundles below the poverty line to bundles above it if the labor time associated to the latter is too large. This issue echoes analyses of optimal labor income tax that reduces poverty when poverty is merely defined in terms of income (see, for instance, Kanbur et al., 1994a,b; Wane, 2001), and the resulting optimal tax schemes are Pareto inefficient. We propose a definition of "being poor" that is consistent with Pareto efficiency. Somewhat surprisingly, it can easily be further adjusted to also be compatible with the requirement of responsibility (that is the requirement that the social welfare function should be such that, absent any informational constraint, no redistribution take place among agents with the same wage). Being poor, under this definition, means consuming a bundle on an indifference curve that lies everywhere below the poverty line.

The second main lesson of our paper is that it is possible to fully characterize the optimal income tax derived from an unconventional social objective that embodies notions of fairness. Indeed, on the one hand, the literature on optimal taxation has successfully obtained and calibrated expressions detailing the optimal tax as a function of behavioral responses, distribution of income (or types) and a set of weights capturing the normative preferences of society. It has provided very little guidance, however, on how to choose those weights.[3] When the objective is utilitarian, for instance, weights depend on an arbitrary cardinalization of the utility function used to describe behavioral responses.

On the other hand, a long tradition in the social choice literature has highlighted how different fairness principles can be called for to characterize social welfare functions (see, among many others, the surveys of Roemer, 1998 or Fleurbaey and Maniquet, 2011). Though some criteria for the taxation of labor income have been previously derived from these social welfare functions, this literature has not derived the precise formula of the tax function, leading to somewhat of a disconnect with the optimal tax literature (in addition to not allowing for precise calibrations). As a result, we view this study as an example of how to bridge the divide between the (essentially axiomatic) fairness approach to optimal tax and a classical social welfare function based approach.

The third main lesson of our paper is that the optimal tax function that is consistent with our fairness principles shares some salient features of the current income tax schedule in the US. As can be seen on Fig. 1, the optimal tax corresponding to the combination of Pareto efficiency, responsibility and poverty reduction exhibits zero and negative marginal tax rates on low-incomes as well as a discrete jump to positive marginal tax rates around $16,000. These features are somewhat unconventional in the optimal taxation literature and yet approximate some important characteristics of the current US income tax. The jump mimics the phaseout of the Earned Income Tax Credit (EITC) and of various welfare programs. The existence of non-positive marginal tax rates on low incomes and the steep jump towards positive rates is a very robust result under our normative objective and does not depend on the specific parameterizations of the poverty line, the agents' preferences or the distribution of types. Finally, depending on the specific parametrization of the poverty line, optimal marginal tax rates on high-incomes may be lower than would be implied by setting the marginal welfare weights to zero in the upper tail (a common normative assumption in the optimal tax literature under utilitarianism).[4] We discuss more precisely the similarities and differences between our calibrations and the current US tax schedule in Section 5.

To be precise, we don't claim that the current US tax schedule is the outcome of maximizing a normative objective similar to ours. Of course a tax schedule is the outcome of a series of reforms inspired by different normative and electoral objectives. What we do claim is that the tax schedule that emerges as optimal in our study *is* reasonable, given its similarity with an existing one. In addition, we believe that our analysis shows that switching from the traditional utilitarian normative assumption to a fairness-based approach can help close the gap between the recommendations of optimal tax models and the reality of policy.

The rest of the paper is organized as follows. In Section 2, we define the model and the basic properties. In Section 3, we introduce a definition of poverty reduction that is compatible, as a nor-

[1] Schokkaert and Devooght (2003) confirm this finding for different countries. See Gaertner and Schokkaert (2012), chapter 4, for a survey. Saez and Stantcheva (2016, Online Appendix C) also report survey evidence according to which citizens are more enclined to be generous towards those who work hard at a low wage than towards those who work less at a higher wage, even if they earn the same income.

[2] To be clear, we consider the optimal tax function under the assumption that the planner only observe agents' incomes. Because of the latter informational constraint, and despite the normative goal of responsibility, agents with the same wage rate but different incomes may still end up being taxed differently at the solution. This would not be the case in a first-best context.

[3] An exception is given by the so-called Rawlsian (or maximin) objective, in which all weights are set to zero except those of the lowest-skill agents. This Rawlsian objective has no straightforward generalization when preferences differ.

[4] Lockwood and Weinzierl (2016) have previously noted that existing tax policies are consistent with less redistributive preferences that is traditionally assumed in the literature.

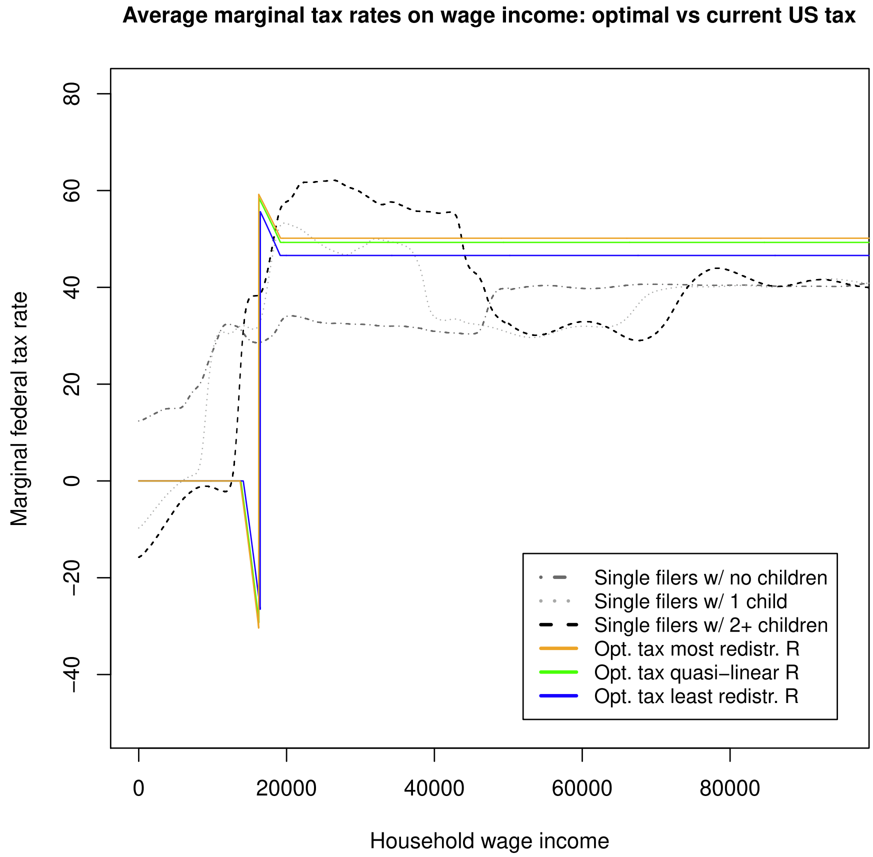

Fig. 1. The solid lines depict the optimal labor income tax according to three versions of our social objective when the poverty line is set at $18,850 with an average slope of $6/h. The dashed and dotted lines depict the actual marginal tax rates in the US for different household structures. The current average marginal tax rate includes both the actual federal rate and the implicit tax rate from the phaseout of various welfare programs. Details on the calibration and estimation of the current tax system using CPS data can be found in Online Appendix F.

mative goal, with Pareto efficiency and responsibility for agents' preferences and we define a prominent family of social welfare functions combining the three goals we are interested in as well as a fourth, auxiliary, property. In Section 4, assuming that preferences are quasi-linear and iso-elastic (customary assumptions in the literature), we derive the formula of the second-best tax schemes that maximize our social objective. In Section 5, we estimate the distribution of types using Current Population Survey (CPS) microdata and calibrate the tax formula to the U.S. economy. We also use the CPS data and NBER's TAXSIM to obtain a description of the current tax system and put our results into perspective.[5] Finally, in Section 6, we compare the ability of our results to mimic the U.S. system with that of other optimal tax results in the literature. In Section 7, we give some concluding comments.

2. The model and the basic axioms

There are two goods, labor time, denoted ℓ, and consumption, denoted c. A bundle (of goods) is a pair $z = (\ell, c) \in X = \mathbb{R}^2_+$. There is a finite set N of agents. Each agent $i \in N$ is characterized by their wage, $w_i \in [\underline{w}, \infty)$, and their preferences $R_i \in \mathscr{R}$ over bundles. Wages are assumed to be bounded from below. Preferences are assumed to be continuous, increasing in consumption, decreasing in labor time, convex, and such that a best bundle exists in each budget.[6] An economy is a list of characteristics $E = (w_N, R_N) \in \mathscr{E} = \left([\underline{w}, \infty)_+ \times \mathscr{R}\right)^N$, one pair of characteristics for each agent. An allocation is a list of bundles $z_N = (z_i)_{i \in N}$, one bundle for each agent.

We are interested in ranking allocations as a function of the parameters of the economy. Formally, a social welfare function, in short a SWF, is a function $\mathbf{R}$, such that for each economy $E \in \mathscr{E}, \mathbf{R}(E)$ is an ordering on the set of allocations X^N. For two allocations $z_N, z'_N \in X^N$, we write $z_N \mathbf{R}(E) z'_N$, resp. $z_N \mathbf{P}(E) z'_N, z_N \mathbf{I}(E) z'_N$, to denote that z_N is socially as good as z'_N, resp. strictly better, indifferent.

The following notation will prove useful. For any preferences $R \in \mathscr{R}$ and set $A \subset X$, we write $m(R, A)$ to denote the set of best bundles in A according to R:

$$m(R, A) = \{z \in A | \forall z' \in A : zRz'\}.$$

In case $m(R, A)$ contains more than one bundle, all bundles of $m(R, A)$ are indifferent for R. For any $z = (\ell, c) \in X$ and $w \in [\underline{w}, \infty)$, we write $B(z, w)$ to denote the budget of slope w going through z:

$$B(z, w) = \{z' = (\ell', c') \in X | c' - w\ell' \leqslant c - w\ell\}.$$

We write $B(z, w) + \Delta$ to denote the budget set obtained by translating $B(z, w)$ by an amount Δ, that is $\{z' = (\ell', c') \in X | c' - w\ell' \leqslant c - w\ell + \Delta\}$.

Finally, for any $z \in X, R \in \mathscr{R}$ and $w \in [\underline{w}, \infty)$, we write $IB(z, R, w)$ to denote the implicit budget of an agent (R, w) at z, that is the budget of slope w that leaves this agent indifferent between z and maximizing over that budget:

[5] Regarding TAXSIM, see Daniel and Coutts (1993).

[6] Formally, the latter condition requires that for all $w_i \in [\underline{w}, \infty), R_i \in \mathscr{R}$ and $b \in \mathbb{R}$, there exists $z^* = (\ell^*, c^*) \in X$ such that $c^* \leqslant b + w_i\ell^*$ and $z^* R_i z$ for all $z = (\ell, c)$ such that $c \leqslant b + w_i\ell$.

$IB(z,R,w) = B(z',w)$ such that $z' I z$ and $z' \in m(R, IB(z,R,w))$.

We complete this section with the definition of two well-known axioms that we impose on our SWFs. Strong Pareto is the requirement that if everyone considers his bundle in z_N at least as good as in z'_N, then z_N must be socially as good as z'_N. If, in addition, at least one agent strictly prefers his bundle in z_N, z_N must be deemed a strict social improvement.

Axiom 1. *strong Pareto*

For all $E = (w_N, R_N) \in \mathscr{E}$, all$z_N, z'_N \in X^N$,

$$[z_i R_i z'_i, \forall i \in N] \Rightarrow [z_N \mathbf{R}(E) z'_N]$$

and $\left[z_i R_i z'_i, \forall i \in N, \text{ and } \exists j \in N : z_j P_j z'_j\right] \Rightarrow [z_N \mathbf{P}(E) z'_N]$

We now define the axiom capturing the fairness principle of responsibility. It captures the idea that income differences merely due to labor time differences should not be deemed unjust. Another way of stating this principle is by reference to full income (that is, income computed when leisure is valued at wage rate): two agents with the same wage rates should obtain the same full income. In terms of tax and transfer, it means that these two agents should ideally receive the same transfer or pay the same tax. Fleurbaey and Maniquet (2005, 2006) have proposed to formalize this principle as follows: take two agents with the same wage rate. Ideally, they should have the same full income, that is the bundles they consume should be their preferred bundle in the same budget. Assume, to the contrary, that the bundles they consume do not arise from the same budget. The following axiom says that such a *budget* inequality is unjust. Consequently, a budget inequality reducing transfer (as opposed to a bundle inequality reducing transfer) among them is a social improvement.

Axiom 2. *equal-wage transfer*

For all $E = (w_N, R_N) \in \mathscr{E}, z_N, z'_N \in X^N$, if there exist $j,k \in N$ and $\Delta \in \mathbb{R}_+$ such that $w_j = w_k$,

$$z'_j \in m\left(R_j, B\left(z'_j, w_j\right)\right), z_j \in m(R_j, B(z_j, w_j)),$$

$$z'_k \in m(R_k, B(z'_k, w_k)), z_k \in m(R_k, B(z_k, w_k)),$$

and

$$B\left(z'_j, w_j\right) = B(z_j, w_j) - \Delta \supseteq B(z'_k, w_k) = B(z_k, w_k) + \Delta,$$

and for all $i \neq j,k : z_i = z'_i$, then $z'_N \mathbf{P}(E) z_N$.

In the complete absence of taxation, that is at a laisser-faire allocation, all agents are treated identically by the tax system, as they do not receive nor pay anything. Moreover, the resulting allocation is Pareto efficient. This means that there exist SWFs, which consider laisser-faire allocations as social optima, satisfying both *strong Pareto* and *equal-wage transfer*, implying that these two axioms are compatible with each other. The drawback of laisser-faire, however, is that some agents may end up in material deprivation, and the more so if they have small wages. The main goal of this paper is to study the consequences of combining *strong Pareto* and *equal-wage transfer* with the idea that society should not let agents end up at too low of a material standard. We introduce such an axiom of poverty reduction in the next section.

3. Poverty reduction

Studying poverty reduction first requires introducing a poverty line in the model. We indeed assume that there is a thin, connected and strictly increasing set of bundles in the consumption set that captures the desire by society to let all agents live a materially decent life. Let $PL \subset X$ satisfy the following properties:

- for all $z = (\ell, c), z' = (\ell', c') \in PL$, either $z = z'$ or $\ell < \ell' \iff c < c'$,
- for all $\ell \in \mathbb{R}$, there exists $c \in \mathbb{R}$ such that $(\ell, c) \in PL$,
- the set of bundles above PL is convex.

One comment is in order. Our poverty line is strictly increasing, that is, the minimal consumption level that makes a life materially decent is assumed to increase with the amount of labor performed by the agent. This can be interpreted in two ways. First, as labor time rises the agent's basic needs increase: more food is required by the organism to afford those efforts, clothes wear out faster, etc. Second, leisure can be considered as a basic need so that less leisure needs to be compensated by more income (we may think for instance of the cost parents incur for hiring someone to take care of their children when they work full time). All results derived in the paper hold for poverty lines arbitrarily close to an horizontal line so that our conclusions are not a byproduct of the assumption that PL is increasing in labor time. To further illustrate this point, we include the limiting case of a flat poverty line in the calibrations of Online Appendix E. Finally, note that, besides its shape, how high PL lies in the consumption set of the agents also determines the generosity of society towards the poor.

We write $z = (\ell, c) > PL$, for any $z \in X$, to describe the situation in which there is $(\ell, c') \in PL$ such that $c > c'$, and $A > PL$, for any $A \subset X$, to describe the situation in which $z > PL$ for all $z \in A$. We refer to indifference curves as to $I(z,R)$, that is, for any $z \in X, R \in \mathscr{R}$,

$$I(z,R) = \{z' \in X | z' I z\}.$$

We now define an axiom reflecting the objective of poverty reduction. Basically, we state that a lump-sum transfer of consumption from a non-poor to a poor agent must be a social improvement. This raises the question of who is (non-) poor. As we look for an axiom of poverty reduction that is compatible with *strong Pareto* and *equal-wage transfer*, we face constraints on the way we can define poverty. The main difficulty is this one: if poverty is defined as consuming a bundle below the poverty line, then this contradicts *strong Pareto* because an agent can be indifferent between a bundle below (in which case she is poor) and a bundle above (in which case she is non-poor) the poverty line.

The difficulty to define poverty reduction in a way that is consistent with *strong Pareto* also echoes the literature initiated by Fleurbaey and Trannoy (2003). This literature has explored the conditions under which a bundle transfer axiom is compatible with strong Pareto. Applying the results from this literature (see, for instance, the survey in Fleurbaey and Maniquet, 2011, chapter 2), we assume that an agent is poor (resp. non-poor) if her indifference curve lies below (resp. above) the poverty line, and we require that a bundle inequality reducing transfer of consumption from a non-poor to a poor agents that have the same labor time is a social improvement.

Axiom 3. *poverty-reduction transfer*

For all $E = (w_N, R_N) \in \mathscr{E}$, all $\Delta \in \mathbb{R}_+$, all $j,k \in N$, all $z_N = (l_N, c_N), z'_N = (l'_N, c'_N) \in X^N$ such that $l_j = l'_j = l_k = l'_k, c'_j = c_j - \Delta, c'_k = c_k + \Delta$ and for all $i \neq j,k : z_i = z'_i$,

$$\left[I\left(z'_j, R_j\right) > PL > I(z'_k, R_k)\right] \Rightarrow [z'_N \mathbf{R}(E) z_N].$$

This axiom requires a number of comments. First, the definition of being poor conveyed by this axiom is much weaker than a definition in terms of consuming a bundle below the poverty line. As a result, an agent is neither poor nor non-poor if her indifference curve crosses the poverty line. Depending on the preferences of the agent, it may be the case that she is neither poor nor non-poor in a very large part of her indifference map. This does not imply, however, that only a tiny fraction of the agents will end up being poor. Our definition, indeed, can still be made as demanding as one wishes, by setting the poverty line higher in the consumption set. In the calibrations below, we make both the slope and the intercept of the poverty line vary.

Second, this definition can be given a welfarist flavor. Given that poverty is not defined based on the bundle one consumes but based on her indifference curve, we can think of being poor as having a utility level below some threshold. In such a welfarist frame, one should bear in mind that the utility threshold below which one qualifies as poor and the utility threshold above which one qualifies as non-poor typically do not coincide. As a consequence, this definition does not equip us with much comparability among agents. The welfare of one agent can be declared larger than that of another agent only if the former agent's utility level is above her individual non-poverty threshold whereas the latter agent's utility level is below her individual poverty threshold.

Third, the tension between the objectives of responsibility and poverty reduction, as declared in the title of the paper, is illustrated by the weakness of the notion of poverty embedded in this axiom. For the sake of completeness, we discuss in Appendix A stronger and maybe more natural axioms of poverty reduction that fail to be compatible with *strong Pareto* and *equal-wage transfer*.

Finally, on a more technical note, this allows us to give an additional justification for our somehow unconventional assumption of a poverty line that is strictly increasing in labor time. As it is customary in the optimal taxation literature, we assume that there is no bound on labor time. Finiteness of individually optimal labor times is then deduced from the assumption of an increasing disutility of labor. With a flat poverty line, therefore, we would have faced the difficulty that nobody ever has her indifference curve entirely below the poverty line (except for agents with flat indifference curves), so that no agent would have been declared poor according to our definition. Having a strictly increasing poverty line is, therefore, a natural assumption to study our weak axiom of poverty reduction in a model without a natural bound on the labor supply.

The combination of *strong Pareto*, *equal-wage transfer*, *poverty reduction transfer* and an auxiliary axiom of *separability*, which we formally define in Appendix B, leads us to consider the family of SWFs which we call $\mathbf{R}^{\tilde{R}\,lex}$. A $\mathbf{R}^{\tilde{R}\,lex}$ SWF first computes a well-being level for each agent at their assigned bundle, and then it applies the leximin criterion $\geqslant_{lex}$ to vectors of well-being levels.[7] The well-being measure associated to a $\mathbf{R}^{\tilde{R}\,lex}$ SWF is parameterized by some individual reference preferences $\tilde{R} \in \mathscr{R}$. These reference preferences need to be chosen among the preferences that contain the poverty line as an indifference curve: for all $z = (\ell, c), z' = (\ell', c') \in PL$,

$$z \, \tilde{I} \, z'.$$

Fig. 2 illustrates how well-being levels are computed. The well-being level of an agent (w_i, R_i) consuming bundle z_i is computed by looking at the bundle that $\tilde{R}$ prefers in the implicit budget of (w_i, R_i) at z_i. Formally, Let $\tilde{u} : X \to \mathbb{R}$ be a utility representation of $\tilde{R}$. Then, for $z_N, z'_N \in X^N$,

$$z_N \, \mathbf{R}^{\tilde{R}\,lex} z'_N \iff$$

$$(b(z_i, w_i, R_i))_{i \in N} \geqslant_{lex} \left(b(z'_i, w_i, R_i)\right)_{i \in N},$$

where for all $z \in X, w \in [\underline{w}, \infty)$, and $R \in \mathscr{R}$,

$$b(z, w, R) = \tilde{u}\left(m\left(\tilde{R}, IB(z, w, R)\right)\right).$$

The fact that $\mathbf{R}^{\tilde{R}\,lex}$ satisfies *strong Pareto* follows from $b(\cdot, w, R)$ being a utility representation of R and the leximin criterion being strictly increasing in all its arguments. The fact that $\mathbf{R}^{\tilde{R}\,lex}$ satisfies *equal-wage transfer* comes from the fact that the well-being index $b(\cdot, w, R)$ is computed using implicit budgets. As a result, if $w_j = w_k$, as soon as $z_j \in m(R_j, B(z_j, w_j))$ and $z_k \in m(R_k, B(z_k, w_k))$, we have

$$b(z_j, w_j, R_j) \geqslant b(z_k, w_k, R_k) \iff B(z_j, w_j) \supset B(z_k, w_k),$$

and the leximin criterion will give priority to the agent with the smallest budget.

The fact that $\mathbf{R}^{\tilde{R}\,lex}$ satisfies *poverty-reduction transfer* is related to the specific property satisfied by $\tilde{R}$. Indeed, let us assume that $I(z_j, R_j) > PL > I(z_k, R_k)$. It is sufficient to prove that $b(z_j, w_j, R_j) > b(z_k, w_k, R_k)$ because the leximin criterion gives priority to the lowest well-being level. First, $I(z_j, R_j) > PL$ implies that $IB(z_j, w_j, R_j)$ contains a part of PL. As a result, maximizing $\tilde{R}$ over $IB(z_j, w_j, R_j)$ must yield a utility level larger than at PL (remember PL is one of the indifference curves of $\tilde{R}$). On the contrary, $PL > I(z_k, R_k)$ implies that $IB(z_k, w_k, R_k)$ is everywhere below PL, so that maximizing $\tilde{R}$ over $IB(z_k, w_k, R_k)$ must yield a utility level lower than at PL, the desired result.

This proves that the $\mathbf{R}^{\tilde{R}\,lex}$ SWFs satisfy the three axioms we have defined. We can also prove that all SWFs satisfying these three axioms and *separability* rank allocations around the poverty line in the same way as $\mathbf{R}^{\tilde{R}\,lex}$ SWFs, that is they all rank allocations in the same way as $R^{\tilde{R}\,lex}$ when an increase in well-being of an agent whose implicit budget lies below the poverty line is accompanied by a decrease in that of an agent above the poverty line. In other words, the combination of the aforementioned axioms *locally* characterize social preferences around the poverty line. Given that this proof is similar to existing proofs in the literature, we send it to Section B in the appendix. This result, though, calls for several remarks.

First, the *separability* axiom that we use for this result formalizes the idea that indifferent agents should not influence the social ranking. This is a classical axiom in social choice theory and it is now well established that combining *strong Pareto*, *separability* and some inequality reduction axiom similar to *poverty reduction transfer* leads to SWFs of the leximin type, that is, the combination of these axioms strengthens a requirement of inequality aversion (like our axiom of *poverty reduction transfer*) into a requirement of infinite aversion to this inequality (see, for instance, the survey in Fleurbaey and Maniquet, 2011, chapter 3).

Second, per this local characterization, the definition of poor and non-poor agents has changed, to encompass many more situ-

[7] The leximin criterion consists in applying the maximin criterion lexicographically, that is a vector weakly dominates another vector according to the leximin if the minimal component of the latter vector is larger than the minimal component of the former vector, or they are equal and the second minimal component of the latter vector is larger than the second minimal component of the former vector, etc., or they are equal.

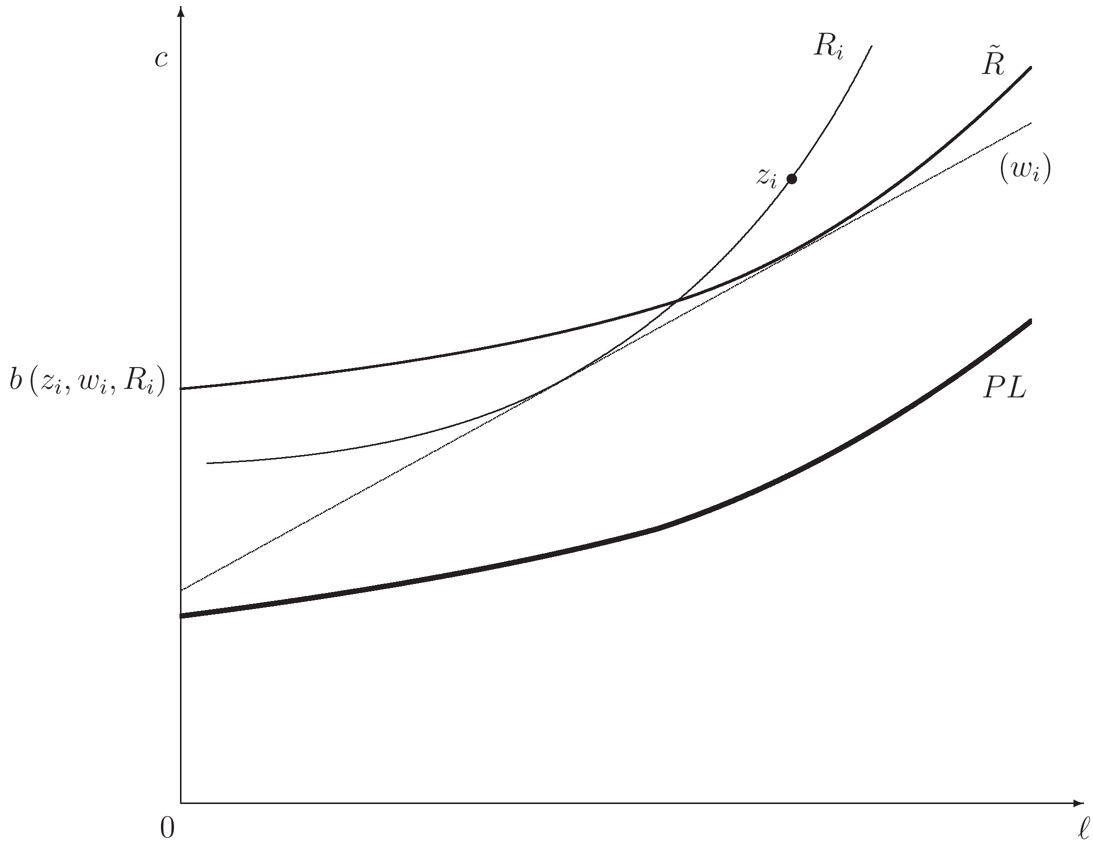

Fig. 2. Illustration of the SWF $\mathbf{R}^{\tilde{R}\,lex}$. Agent i consumes bundle z_i. Her implicit budget is the budget of slope w_i that is tangent to her indifference curve going through z_i. The reference preferences $\tilde{R}$ contain PL as an indifference curve. The utility associated with the best bundle for $\tilde{R}$ in i's implicit budget gives the well-being index for $i : b(z_i, w_i, R_i)$. Finally, $\mathbf{R}^{\tilde{R}\,lex}$ applies the leximin to vectors of $b(.)$.

ations. It is now sufficient to look at the implicit budget of the agents. As soon as the frontier of the implicit budget of an agent is everywhere below the poverty line, this agent qualifies as poor, whether or not this agent's entire indifference curve lies below it. As soon as the frontier of the implicit budget of an agent is somewhere above the poverty line, this agent qualifies as non-poor, whether or not this agent's entire indifference curve lies above the poverty line. Observe that with this new definition, agents are either poor or non-poor at an allocation; that is there is no subset of consumption bundles at which an agent is neither non-poor nor poor.

Third, all $\mathbf{R}^{\tilde{R}\,lex}$ SWFs identify in exactly the same way poor and non-poor agents because this only depends on whether their implicit budget lies below the poverty line or not. Consequently, we can deduce from this result that any allocation in which poverty is eliminated is socially preferable to any allocation in which at least one agent's implicit budget is still below the poverty line. This teaches us what the program of the planner has to be: designing a tax scheme that eliminates poverty in the sense that the implicit budget of no agent lies below the poverty line. This is formally stated and proven in the appendix. Of course, poverty being eliminated in this sense does not guarantee that all agents' *bundles* are above the poverty line.

This conclusion is illustrated in Fig. 3. Agent j consumes z_j, which lies above the poverty line, and her indifference curve through z_j crosses the poverty line. If the wage of j is low, for instance if $w_j = w$, then her implicit budget, drawn in the figure, is everywhere below the poverty line (slopes of implicit budget lines are mentioned within parentheses below the lines). Agent j qualifies as poor, and a transfer from a non-poor agent to agent j is a social improvement. If, on the other hand, her wage is high, for instance if $w_j = w'$, then her implicit budget crosses the poverty line and she qualifies as non-poor. That illustrates the fact that well-being indices depend on wage. In the case of agent j, at the same bundle, her well-being is an increasing function of wage.

The contrary happens for agent k, when she consumes z_k, below the poverty line. We have $b(z_k, w, R_k) > b(z_k, w', R_k)$ because her implicit budget lies below the poverty line when $w_k = w'$ but not when $w_k = w$. The difference between agents j and k comes from the fact that agent j has preferences that are more work averse than the reference preferences (illustrated by the fact that her indifference curve crosses the poverty line from below) and the contrary is true for agent k. That implies that high wage agents may have low well-being levels, if their preferences are sufficiently less directed towards leisure than the reference preferences. This will have potentially drastic consequences on the shape of the optimal tax on high incomes. In our calibrations, for poverty lines that are steep enough, marginal tax rates in the right-tail are lower than the rate that maximizes revenue raised from high incomes.

4. Optimal tax functions

In the remaining of the paper, we study the properties of the tax schemes that maximize the SWFs we have justified in the previous sections. We adopt the classical Mirrlees setting. The planner does not observe the characteristics of the agents. She only observes their income. She has to design a redistributive tax scheme $\tau : \mathbb{R}_+ \rightarrow \mathbb{R}$. As a result, an agent with wage $w \in [\underline{w}, \infty)$ and a labor time of ℓ earns income $y = w\ell$, pays income tax $\tau(y)$ and consumes $c = y - \tau(y)$.

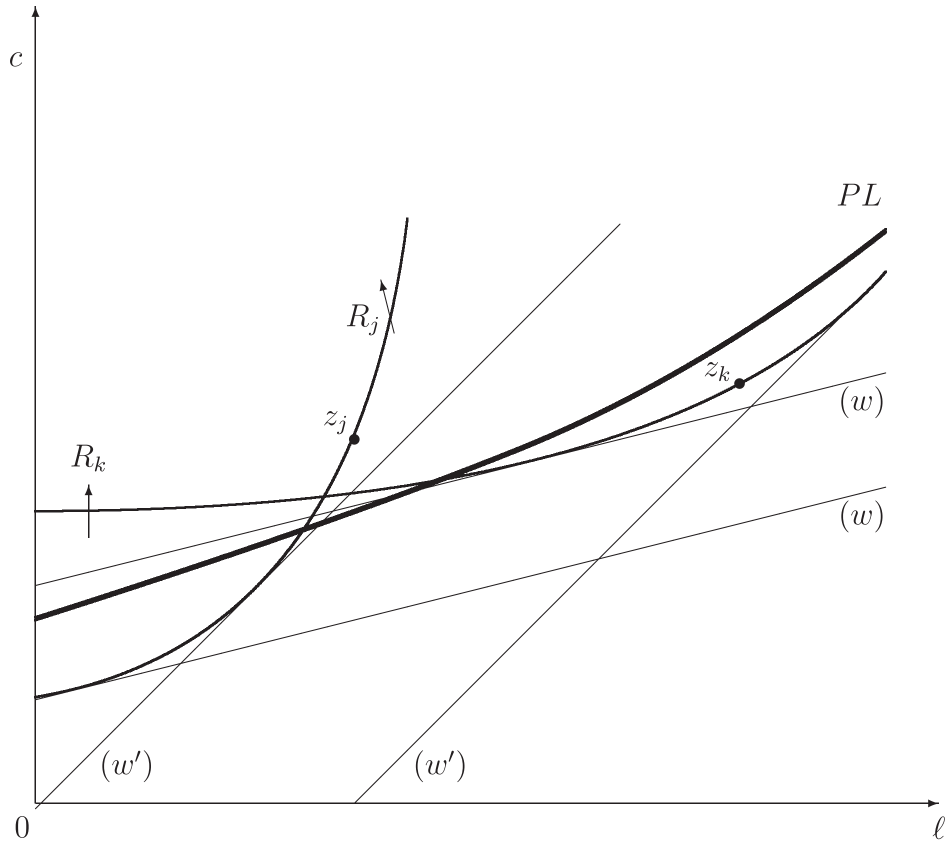

Fig. 3. Who is (non-) poor? Agent j consumes bundle z_j above the poverty line. She is poor when her wage is $w_j = w$ but non-poor when $w_j = w'$. Agent k consumes bundle z_k below the poverty line. She is poor when her wage $w_k = w'$ but non-poor when $w_k = w$. This figure illustrates two important points. First, agents may consume a bundle under the poverty line, yet be non-poor according to our well-being index. Second, well-being indices depend on wages.

Given the complexity of the problem, we assume that agents' preferences are quasi-linear and iso-elastic: they are represented by

$$u_i(c, \ell) = c - \frac{\epsilon}{1+\epsilon}\left(\frac{\ell}{\theta_i}\right)^{\frac{1+\epsilon}{\epsilon}}. \quad (1)$$

Agents differ in their taste for leisure only through parameter $\theta_i \in [\underline{\theta}, \overline{\theta}]$. We define $n_1 = \underline{w}\overline{\theta}$, the type of the agent with the lowest skill and the lowest taste for leisure, that is, n_1 is the parameter of all agents who are behaviorally identical to the hardworking poor agent. Their earning level is denoted by y_1. We further assume that the poverty line itself, PL, has the shape of an iso-elastic indifference curve.

$$PL = \left\{ z = (c, \ell) \in X = \mathbb{R}^2_+ | \bar{P}_0 = c - \frac{\epsilon}{1+\epsilon}\left(\ell/\tilde{\theta}\right)^{\frac{1+\epsilon}{\epsilon}} \right\}.$$

Studying tax schemes forces us to study the income/consumption space instead of the labor time/consumption space. That requires some rewriting. Abusing notation, however, we keep u_i to denote the utility function of agent i with wage w_i in the income/consumption space, and we write

$$u_i(c, y) = c - \frac{\epsilon}{1+\epsilon}\left(\frac{y}{w_i\theta_i}\right)^{\frac{1+\epsilon}{\epsilon}}. \quad (2)$$

This assumption, also used by Lockwood and Weinzierl (2015), reduces the two dimensions of heterogeneity to a single index in the income consumption space: $n_i = w_i\theta_i$.[8] All agents with the same parameter n_i will be behaviorally equivalent (that is they all have the same preferences in the income/consumption space) even if they differ in their wage.

The economies we are interested in are economies with a finite but large number of agents. As is customary in the optimal tax literature, we capture the assumption of a large economy by assuming that there is a continuum of types. We then assume that $n_i \sim F(n_i)$ in the population and the associated density is denoted by $f(n_i)$. Notice that $F(n_i)$ is a marginal distribution generated by the joint distribution of θ_i and w_i. We assume, in addition, that the joint probability density of (w_i, θ_i) is strictly positive everywhere on $[\underline{w}, \infty) \times [\underline{\theta}, \overline{\theta}]$. An important implication is that $\forall n \in [\underline{n}, n_1]$, there exists a non-zero density of agents earning the minimum wage $\underline{w}$.[9]

The constraints of the SWF maximization are well-known and we review them quickly. We begin with the incentive compatibility constraints. The planner chooses income-consumption bundles $(y(n), c(n))$ for each "behavioral" type n. We write $u(n)$ the utility of agents of type n at bundles $(y(n), c(n))$. Incentive compatibility constraints require:

$$u(n) = \max_{n' \in [\underline{n}, \infty)} c(n') - \frac{\epsilon}{1+\epsilon}\left(\frac{y(n')}{n}\right)^{\frac{1+\epsilon}{\epsilon}}$$
$$= c(n) - \frac{\epsilon}{1+\epsilon}\left(\frac{y(n)}{n}\right)^{\frac{1+\epsilon}{\epsilon}}. \quad (3)$$

[8] More generally, collapsing multiple dimensions of heterogeneity to one parameter in the context of optimal taxation has been suggested by Mirrlees (1976) and used by Brett and Weymark (2003) and Choné and Laroque (2010).

[9] Similarly, $\forall n \in [n_1, \infty)$, there exist some agents with $\theta_i = \overline{\theta}$ and $w_i = n_i/\overline{\theta}$.

In addition to incentive compatibility, the planner maximizes the SWF under the constraint that at least some revenue S be raised to finance public goods, which, using Eq. (2), can be written

$$\int_{\underline{n}}^{\infty}\left(y(n)-\left(u(n)+\frac{\epsilon}{1+\epsilon}\left(\frac{y(n)}{n}\right)^{\frac{1+\epsilon}{\epsilon}}\right)\right)dF(n)\geqslant S. \quad (4)$$

We now turn to the definition of the social objective. Agents with the same n may not all have the same well-being index $b(z_i, w_i, R_i)$, because it depends on their underlying wage w_i and preference parameter θ_i. For any $\mathbf{R}^{\tilde{R}\,lex}$, however, we can restrict our attention to the minimum well-being among agents of type n at utility level $u(n)$, which we denote $\underline{b}(u(n); n)$.

4.1. Quasi-linear $\tilde{R}$

We know from Section 3 that $\mathbf{R}^{\tilde{R}\,lex}$ is such that the reference preferences $\tilde{R}$ contain the poverty line as an indifference curve. While $\tilde{R}$ are thus characterized at the poverty line, the remaining indifference curves need to be drawn. For now, we consider the case where $\tilde{R}$ are quasi-linear. We will see in Section 4.2 how the solution to this case can be amended so as to provide a solution to two other polar choices of $\mathbf{R}^{\tilde{R}\,lex}$ as well. All three polar choices are considered in the calibrations of Section 5.

Under quasi-linear reference preferences, $\tilde{R}$ can be represented by the following utility function:

$$\tilde{u}(c,\ell)=c-\frac{\epsilon}{1+\epsilon}\left(\frac{l}{\bar{\theta}}\right)^{\frac{1+\epsilon}{\epsilon}}. \quad (5)$$

The problem of the planner consists in maximizing $\mathbf{R}^{\tilde{R}\,lex}$, in which $\tilde{R}$ is represented by Eq. (5), under incentive constraints (3) and resource constraint (4). The following proposition, illustrated in Fig. 4, completely characterizes the optimal tax scheme under the assumption that the distribution of types has an upper tail Pareto index α.

Proposition 1. *If $F(n)$ has a Pareto index α in the upper tail and if the lower-bound on utilities binds on one interval above y_1^*, then there exist $n_u \leqslant n_{\underline{b}} \leqslant n_{\bar{b}} \leqslant n_l \leqslant n' \in [\underline{n}, \infty)$, such that $\tau'(y(n))$ satisfies*

1. $\tau'(y(n)) = 0 \, \forall n \in [\underline{n}, n_u]$
2. $\frac{\tau'(y(n))}{1-\tau'(y(n))} = \left(\frac{1+\epsilon}{\epsilon}\right)\frac{F(n_u)-F(n)}{nf(n)} \forall n \in \left[n_u, n_{\underline{b}}\right]$
3. $y(n) = y_1^* \forall n \in \left[n_{\underline{b}}, n_{\bar{b}}\right]$
4. $\frac{\tau'(y(n))}{1-\tau'(y(n))} = \left(\frac{1+\epsilon}{\epsilon}\right)\dfrac{[F(n_l)-F(n)]+\dfrac{1-\left[1-\left(\frac{\underline{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}}{\left(\frac{1+\epsilon}{\epsilon}\right)\left[1-\left(\frac{\underline{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}}n_l f(n_l)}{nf(n)} \forall n \in \left[n_{\bar{b}}, n_l\right]$
5. If $\frac{1+\epsilon}{1+\epsilon+\alpha\epsilon} \geqslant 1-\left[1-\left(\frac{\underline{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$: then
$$\tau'(y(n)) = 1-\left[1-\left(\frac{\underline{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}} \forall n \in [n_l, \infty]$$
6. *Otherwise,*
$$\tau'(y(n)) = 1-\left[1-\left(\frac{\underline{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}} \forall n \in [n_l, n'] \text{ and}$$
$$\frac{\tau'(y(n))}{1-\tau'(y(n))}=\frac{1-F(n)}{nf(n)}\left(\frac{1+\epsilon}{\epsilon}\right)\forall n\in[n',\infty)$$
and the marginal tax rate converges to $\tau'(y(n)) \rightarrow \frac{1+\epsilon}{1+\epsilon+\alpha\epsilon}$ as $n \rightarrow \infty$.

This proposition, whose proof is relegated to Appendix D, gives the formula that will be used in the next section for the calibrations. Per the leximin nature of $\mathbf{R}^{\tilde{R}\,lex}$, the formula features intervals of n, $[\underline{n}, n_u]$ and $[n_l, n']$, over which the well-being index $\underline{b}(u(n); n)$ is equalized. On the remaining intervals, efficiency considerations govern the optimal tax rate. We give a more detailed explanation of why these different intervals appear at the solution and how the values of their endpoints are determined in Online Appendix Section I.2. Several comments are in order to interpret the result.

1. Very low incomes, that is incomes in the range $[y(\underline{n}), y(n_u)]$, are taxed at a zero marginal rate. First, it is important to note that this result does not depend on the assumption that preferences are quasi-linear and iso-elastic. As it is transparent from the proof, this result is fully general, and holds even if both the extensive and the intensive reaction to taxation are taken into account.[10] Second, if it turns out that $y(\underline{n}) = 0$ (which is the relevant case), there may be some bunching at this earning level.
2. Incomes in the range $\left[y(n_u), y(n_{\underline{b}})\right]$ face a negative income tax rate. The result that marginal tax rates should be zero or negative below $y^*(n_1)$ is more related to the fact that the social welfare function satisfies *equal-wage transfer* than *poverty reduction transfer*. Indeed, it is the combination of *strong Pareto* and *equal-wage transfer* that forces us to measure well-being by using implicit budgets, with the result that well-being is equalized among minimal wage agents if they all maximize their utility over the same real (first-best) budget, as it is the case with a zero marginal tax rate. Surprisingly, the fact that the social welfare function also satisfies *poverty-reduction transfer* is not seen in the shape of the marginal tax rates on low incomes.[11]
3. The next feature of the optimal tax is that there is bunching at the earning level of agent 1. Indeed, all types of agents in the range $\left[n_{\underline{b}}, n_{\bar{b}}\right]$ earn the same income. The solution shares some similarities with the "selfishly optimal nonlinear income tax schedules" studied by Brett and Weymark (2017). In both cases, the tax schedule is one that maximizes redistribution towards agents with some skill (n_1 under our notation) in the middle of the distribution (under the constraint that no other agent end up with a level of well-being below that of n_1). Ignoring the constraint that $y(n)$ be increasing in n_1, this leads to negative marginal tax rates to the left and positive marginal tax rates to the right of n_1. Bunching occurs at n_1 to maintain the monotonicity of $y(n)$. On the other hand, the tax rates we obtain in the $[\underline{n}, n_u]$ and $\left[n_{\bar{b}}, n_l\right]$ intervals differ from what Brett and Weymark (2017) obtain because they directly follow from our constraint that the well-being of no agent should be lower than that of agent 1.
4. Because the SWF is a leximin, the optimal tax rate above $y_1^* = y(n_1)$ is the one that equalizes the well-being index of agents with $\theta_i = \bar{\theta}$ (they are the worst-off within each n) unless there exists a Pareto improvement. The possibility of a Pareto improvement is the reason why the revenue-maximizing tax rate plays a key role in the proposition. The revenue-maximizing tax rate is described by $\frac{\tau'(y(n))}{1-\tau'(y(n))} = \frac{1-F(n)}{nf(n)}\left(\frac{1+\epsilon}{\epsilon}\right)$ which implies $\tau'(n) \rightarrow \frac{1+\epsilon}{1+\epsilon+\alpha\epsilon}$ in the upper-tail when $F(n)$ has Pareto index α. Let us first consider large incomes: we compare the

[10] In order to underscore this point, we formally derive this generalization in Online Appendix J.

[11] This also explains the similarity with the optimal tax rates on low incomes derived in Fleurbaey and Maniquet (2007), which studies a social objective combining *equal-wage transfer* with an axiom of equality in utility among agents with the same preferences. Under the assumption that labor time is bounded above, they obtain a zero marginal tax rate below $y^*(n_1)$. Proposition 1 generalizes their result, making intervals $\left[y(n_u), y\left(n_{\underline{b}}\right)\right]$ and $\left[n_{\bar{b}}, n_l\right]$ appear.

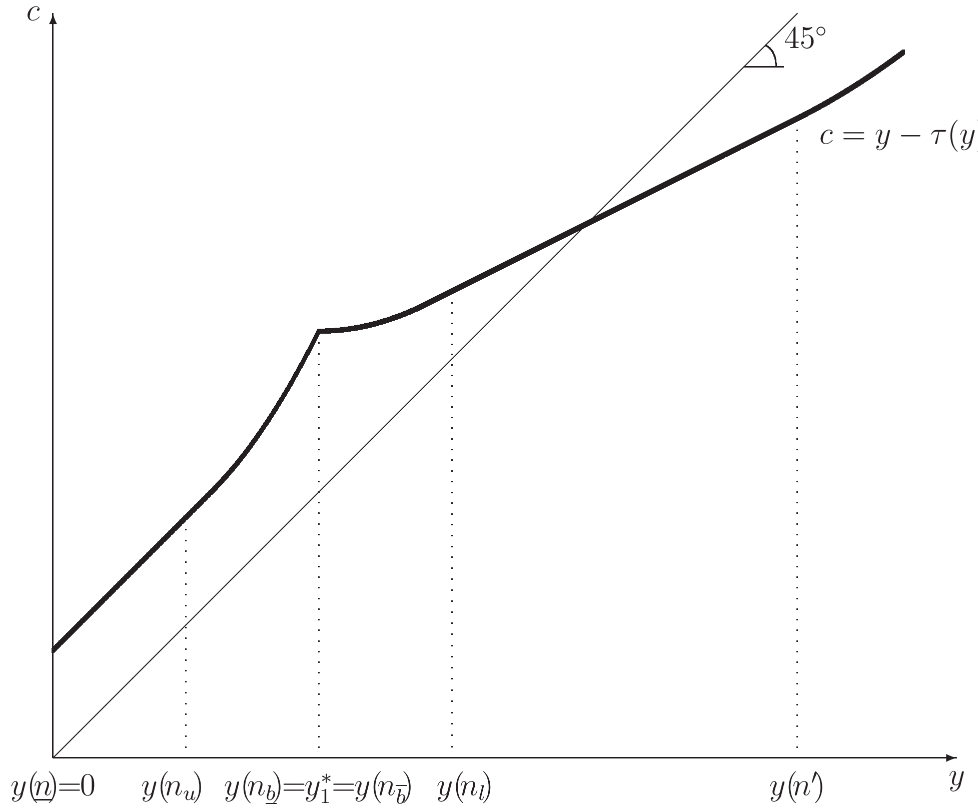

Fig. 4. Illustration of Proposition 1. The income range is divided in 5 intervals. Marginal tax rates are zero in the first interval $[y(\underline{n}), y(n_u))$, negative and decreasing in the second one $[y(n_u), y(n_{\underline{b}})]$, positive and decreasing in the third one $[y(n_{\bar{b}}), y(n_l))$, constant in the fourth one $[y(n_l), y(n'))$, and decreasing in the last one $[y(n'), \infty)$.

revenue-maximizing tax rate in the tail to the well-being equalizing tax rate, $1 - \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$. If the former is lower than the latter, taxing revenue in the upper-tail at $\frac{1+\epsilon}{1+\epsilon+\alpha\epsilon}$ is a Pareto improvement over equalizing well-being since it raises more revenue while agents in that part of the distribution also face a lower tax rate. A similar reasoning applies to incomes right above $y(n_1)$. Recall that agents with $n \in (n_1, n_{\bar{b}}]$ all choose income $y(n_1)$ and have well-being index strictly larger than 1's. As a result, the lower-bound on utilities derived from the leximin nature of the social objective is not binding at $n_{\bar{b}}$.

5. We can intuitively sum up the influence of each axiom embedded in the social objective on the optimal tax as follows. First, *equal-wage transfer* implies that the marginal rate is zero on very low incomes and negative on the following interval. Second, *poverty reduction transfer* is key in 1) determining the amount of transfer to the lowest income earners, and a higher poverty line implies a larger transfer, and 2) determining the marginal tax rates on incomes above $y(n_{\bar{b}})$, where a steeper poverty line yields a lower tax rate. Finally, *strong Pareto* is responsible for the shape of the tax function around $y(n_1)$ and on very large incomes.

4.2. Three polar choices for $\tilde{R}$

Proposition 3 in the appendix characterizes social preferences locally around the poverty line. The fact that the characterization is only local is reflected in that the reference preferences $\tilde{R}$ must have an indifference curve that coincides with the poverty line but the rest of the indifference map is not characterized. A corollary of Proposition 3 is that, under *strong Pareto*, *equal-wage transfer*, *poverty reduction transfer* and *separability*, the tax schedule must prioritize lifting all implicit budgets above the poverty line (see Corollary 1 in appendix Section C).

Assume that we have identified the allocation that just achieves this. It is typically the case that some additional surplus can still be produced. The question becomes how do we allocate this additional surplus? Maximizing a SWF from the $\mathbf{R}^{\tilde{R}\,lex}$ family ensures that such potential surplus is redistributed in a way that is consistent with the aforementioned four axioms. On the other hand, the $\mathbf{R}^{\tilde{R}\,lex}$ family encompasses many choices of reference preferences above the poverty line. The shape of these indifference curves govern how the surplus beyond poverty eradication is to be redistributed across the distribution of n. We consider three polar cases.

The most redistributive $\mathbf{R}^{\tilde{R}\,lex}$ consists in maximizing the surplus that is redistributed to agents with low n while the least redistributive $\mathbf{R}^{\tilde{R}\,lex}$ minimizes redistribution above what is necessary to lift all implicit budgets above the poverty line. The third, intermediary, way of allocating the surplus consists in distributing it equally among agents. When preferences are quasi-linear, equal distribution means that we equalize the utility gain above the allocation at which poverty is eliminated. This is accomplished by choosing the $\tilde{R}$ that is itself quasi-linear. Proposition 1 was derived under the latter assumption. In appendix H, we provide a detailed derivation of the optimal tax function in the other two cases. We briefly describe the corresponding solutions in the next two paragraphs. In addition, Fig. 5 presents the shape of the indifference curves above the poverty line for the least- and most redistributive $\tilde{R}$. Finally, and most importantly, note that in the calibrations, all three polar cases end up corresponding to surprisingly similar optimal tax schemes.

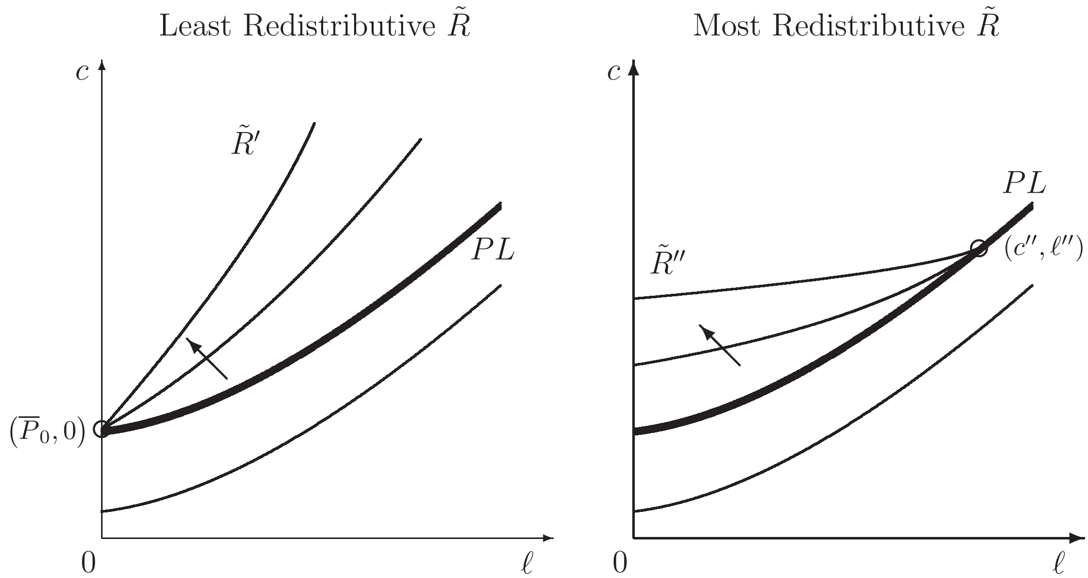

Fig. 5. Illustration of the least- and most redistributive reference preferences. The left panel shows some indifference curves of the least redistributive $\tilde{R}$. Above the poverty line, the curves rotate around $(\overline{P}_0, 0)$. The right panel shows some indifference curves of the most redistributive $\tilde{R}$. Above the poverty line, the portion of each curve over $[0, \ell'')$ rotate around (c'', ℓ''). There is a kink at (c'', ℓ'') and the remaining portion (over $[\ell'', \infty)$) follows PL. In both cases, $\tilde{R}$ are quasi-linear under PL.

From Proposition 1, we know that steeper reference preferences (i.e. a steeper poverty line) lead to less redistribution. As a result, the least redistributive reference preferences consistent with our axioms are such that the indifference curves become labor averse as quickly as possible above the poverty line. In order to obtain such preferences, we rotate the relevant indifference curves around $(c, \ell) = (\overline{P}_0, 0)$ by letting $\tilde{\theta}$ become smaller. Under these reference preferences, the formula in Proposition 1 describes the optimal tax by replacing $\tilde{\theta}$ with $\tilde{\theta}'$. The latter is the lowest $\tilde{\theta}$ so that the constraint that everybody (and in particular agents with $n = n_1$) be lifted out of poverty is satisfied.

The opposite reasoning applies to the most redistributive $\mathbf{R}^{\tilde{R}lex}$: we let the reference preferences become as hardworking as possible above the poverty line. Given that there is no upper bound on labor time, there is no natural bundle around which to rotate the reference indifference curves, unlike in the least redistributive case. Therefore, we fix a sufficiently large ℓ'' and we define the reference preferences in such a way that indifference curves rotate around bundle (c'', ℓ'') on the poverty line. As a result, there is a kink in the preferences at ℓ'' that can be seen on Fig. 5. The resulting optimal tax function resembles the one described in Proposition 1 in that there are intervals over which the minimal well-being is equalized across n and intervals over which the lower-bound is not binding. The lower-bound changes in the following ways. First, it remains the same as the allocation that just eliminates poverty for agents with $n \geqslant n''$ where $\frac{n''}{\overline{\theta}}$ is the wage level at which preferences characterized by $\tilde{\theta}$ would choose a labor time of ℓ''. In other words, ℓ'' and n'' are defined so that, along the lower-bound, the utility of agents with $n \geqslant n''$ is at the poverty line and they do not reap any share of the surplus above the allocation that eradicates poverty. Second, there is a portion $[n_0, n'']$ along which the marginal tax rate is decreasing. This corresponds to the wage levels at which the reference preferences are tangent to the budget set at ℓ''. Third, let $\tilde{\theta}''$ be the value associated to the highest reference indifference curve for which the resource constraint is binding, for all agents with $n \in [n_1, n_0]$, the marginal tax rate along the lower-bound is $1 - \left[1 - \left(\frac{\tilde{\theta}''}{\overline{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$.[12] As far as the optimal allocation is concerned, Proposition 1 applies, $\forall n \in [\underline{n}, n_0)$, with $\tilde{\theta} = \tilde{\theta}''$. For $[n_0, \infty)$, the solution is the well-being equalizing allocation previously described. In some cases, and similar to Proposition 1, there exists some $n' \in [n_0, \infty)$ after which applying the revenue maximizing tax rate generates a Pareto improvement.

5. Calibration to the US economy

The optimal tax formula that we obtained in the previous section leaves us with a number of questions that cannot be answered theoretically and that require some calibrations. In general, we are interested in the magnitudes implied by our optimal formula as well as the differences between the three polar choices of $\tilde{R}$. More specifically, we focus on four questions - introduced below - that calibrations help us answer. Before proceeding to the analysis, let us briefly describe the data sources and methods used to calibrate.

The distribution of types n in the economy is estimated via Maximum Likelihood using CPS microdata (Flood et al., 2018). We estimate a four parameter double Pareto Lognormal distribution (DPLn). For each household in the data, we simulate the marginal federal tax rate they face using NBER's TAXSIM and estimate the implicit marginal tax rate from the phaseout of government assistance. Conditional on a calibrated labor supply elasticity parameter, we can recover agents' types n_i by inverting the first-order conditions of the agents' problem. The parameters of the DPLn distribution are then estimated by maximizing the log-likelihood of this sample of n_i. We pick a labor supply elasticity $\epsilon = 0.33$ which is the preferred estimate in Chetty (2012). We use information about hours worked and hourly wage rates in our sample of CPS data to discipline $\overline{\theta}$. Finally, we need to calibrate the level and slope of the poverty line. These are normative choices. For this reason, we present the results for many different choices of these parameters in Online Appendix E. In the main text,

[12] More precisely, as $\tilde{\theta}''$ increases, so does the marginal tax rate along the lower-bound over $[n_1, n'')$. Holding the utility of agents with $n = n''$ at the poverty line, the utility of all agents with $n < n''$ is increased as $\tilde{\theta}''$ is raised. A larger $\tilde{\theta}''$ therefore requires larger aggregate resources in the economy. At the solution, $\tilde{\theta}''$ is such that the resource constraint is just binding.

we start by assuming that $\overline{P}_0 = \$18,850$. According to the 2015 CPS (March Supplement), the average household size in the US is 2.54. The weighted average of the 2014 poverty lines for households of size 3 is $18,850 (DeNavas-Walt and Proctor, 2015). It is slightly more challenging to calibrate the average slope of the poverty line but the federal minimum wage ($\$7.25/h$) may be a useful benchmark. We start the calibrations with a value of $\$6/h$. More details about the calibrations are provided in Online Appendix F.

Fig. 6 and the middle panel of Table 1 show the calibrated optimal tax rates and consumption schedules according to all three choices of $\widetilde{R}$. We focus on four characteristics of the calibrated optimal allocations. First, the optimal consumption of agents with zero earnings range from $\$11,009 - \$12,025$. As previously noted, poverty can be eradicated in the sense that all implicit budgets are at or above the poverty line, yet some agents consume a bundle that is located below the line. In these calibrations, this is the case for all agents earning no income which illustrates the trade-off between poverty and responsibility. Second, the formula recommends a sharp increase in the marginal taxation rates around the income earned by those who work full time at the minimum wage (income y_1^* in Proposition 1). The second question we would like to answer is: how big should this increase be? It is very similar across the three polar choices of $\widetilde{R}$. The tax rates jump from around -0.29 to around 0.58. Third, the formula also requires marginal tax rates to be decreasing in intervals just before (between the income levels earned by n_u and $n_{\underline{b}}$ in the proposition) and just after (between the income levels earned by $n_{\underline{b}}$ and n_l) this sharp increase. How large should these intervals be and how decreasing should the rates be over this interval is our third question. Across all three $\widetilde{R}$, these two intervals are located around $\$13,900 - \$16,300$ and $\$16,300 - \$19,160$.

Consistently with what we said when deriving this formula, we do not think, however, that how marginal rates evolve in these two intervals is essential because it crucially depends on the specific shape of the utility function that we assumed. More important is the fourth question. The formula recommends that, quickly after the decrease in marginal tax rates that follows the sharp increase around the minimum wage, rates remain stable on a rather large interval (interval $[n_l, n']$ in the proposition). Around which level should rates stabilize and how large should this interval be? The answer is remarkably consistent across the three polar cases: $0.47 - 0.50$. Notice that these rates apply from $y(n_l)$ to all larger incomes. In these calibrations, tax rates on large incomes are determined by equalizing the well-being $\underline{b}(u(n), n)$ across agents with $n \in [n_l, \infty)$. This is mainly due to the average slope of the poverty line being set at $\$6/h$. For flatter poverty lines, tax rates on the largest incomes are sometimes determined so as to maximize revenue collected to be re-distributed.

An important lesson from these calibrations is that, while in theory there may be large differences between the three polar cases of $\widetilde{R}$, empirically, at these values of $\overline{P}_0$ and $\tilde{\theta}$, optimal tax rates and consumption levels are not very sensitive to the choice of $\widetilde{R}$. This conclusion is supported for instance by the small differences in $c(0)$ or in the tax rate on large incomes. Intuitively, if the poverty line is sufficiently high relative to aggregate resources in the economy, the focus on poverty reduction and responsibility almost completely pins down the optimal allocation even though social preferences are only fully characterized in the neighborhood of the poverty line. As we can see in the top panel of Table 1, choices of lower poverty lines leave more room for an additional normative choice to be made about how to redistribute the surplus above the allocation that just eliminates poverty. On the contrary, if the poverty line is so high that not everyone is lifted out of poverty at the optimum, the three cases collapse to a unique solution. This is the case in the bottom panel of Table 1.

Fig. 1 plots the calibrated optimal tax rates alongside the current marginal tax rates for single filers with 0, 1 and 2 or more children respectively. We show the marginal tax rates for joint filers in Online Appendix F. In the US, households with at least one child face negative marginal tax rates on incomes below $6,000 to $13,000 depending on the number of children and filing status. This is driven by the Earned Income Tax Credit - though some of the work incentives from the EITC are undone by the implicit positive marginal tax rate generated by the phaseout of welfare programs such as SSI, TANF and SNAP. Marginal rates tend to increase steeply roughly around $10,000 and reach up to 60% for households with more than 2 children.[13] Marginal tax rates then tend to decline (after a plateau) and stabilize around 40% for high incomes. Households with no children do not face negative marginal tax rates, nor do they face as high positive marginal tax rates around $10,000 to $45,000 because both the EITC and the welfare system are less generous towards families without children.

Interestingly, our calibrated optimal tax rates resemble the tax schedule faced by households with at least one child. Our formula recommends a sharp increase in marginal rates around the income earned by those working full time at the minimal wage and this is what we observe for households with children.[14] Our formula also recommends that marginal rates stay around zero before this sharp increase and stay more or less constant afterwards. This is again consistent with the way households with children are taxed in the US. Besides these similarities, there are four major differences. First, we cannot rationalize negative marginal rates in the neighborhood of zero earning. The rates should be zero according to our formula. This is not a huge difference, but the current scheme is better explained in this interval by arguments in terms of extensive margin effects (people deciding to move from unemployment to employment), like in Saez (2002), or in terms of incentives to go beyond overestimated costs of finding jobs, such as in Lockwood Benjamin, 2020.

The second difference between our optimal tax scheme and the current one in the US is that it is hard to rationalize a brutal decrease in marginal tax rates after the phasing out of the EITC, followed by a plateau. Our results push towards a smoothing of this part of the tax system, where the constraint that the well-being of agents earning incomes in this bracket should not be lower than the well-being of the hardworking poor is likely to be binding.

Third, in the calibrations, tax rates on larger incomes ($0.47 - 0.49$) are about 10 percent higher than the current US federal schedule implying more redistribution towards everybody else. On the other hand, they remain lower than those that would maximize revenue raised at the top (0.54).[15] It is worth noting that, unlike the other differences and similarities mentioned here, these values are somewhat sensitive to the particular choice of poverty line and reference preferences chosen for the calibration.

Finally, the fourth difference is that the US tax system for childless households does not look at all like the one our axioms

[13] In Online Appendix F, we show, for each household type, the break down of the marginal tax rate between the federal income tax and the phaseout of welfare payments.

[14] The exact income, y_1^*, at which the jump happens is determined endogeneously at the solution and depends on the calibrated value of $n_1 = \underline{w}\overline{\theta}$. We pick $\underline{w} = \$7.25$ (the federal minimum wage). We calibrate $\overline{\theta}$ using the 95th percentile of the empirical distribution of θ_i in a sample of hourly workers in the CPS. More details are provided in Online Appendix F.

[15] The revenue maximizing rate is likely an underestimate. This is because our estimate, in CPS data, of the upper-tail Pareto parameter implies a thinner tail of the income distribution than studies using administrative data. We discuss this point more in depth in appendix G and, importantly, we show that the results of our main calibrations are not sensitive to assuming a thicker tail.

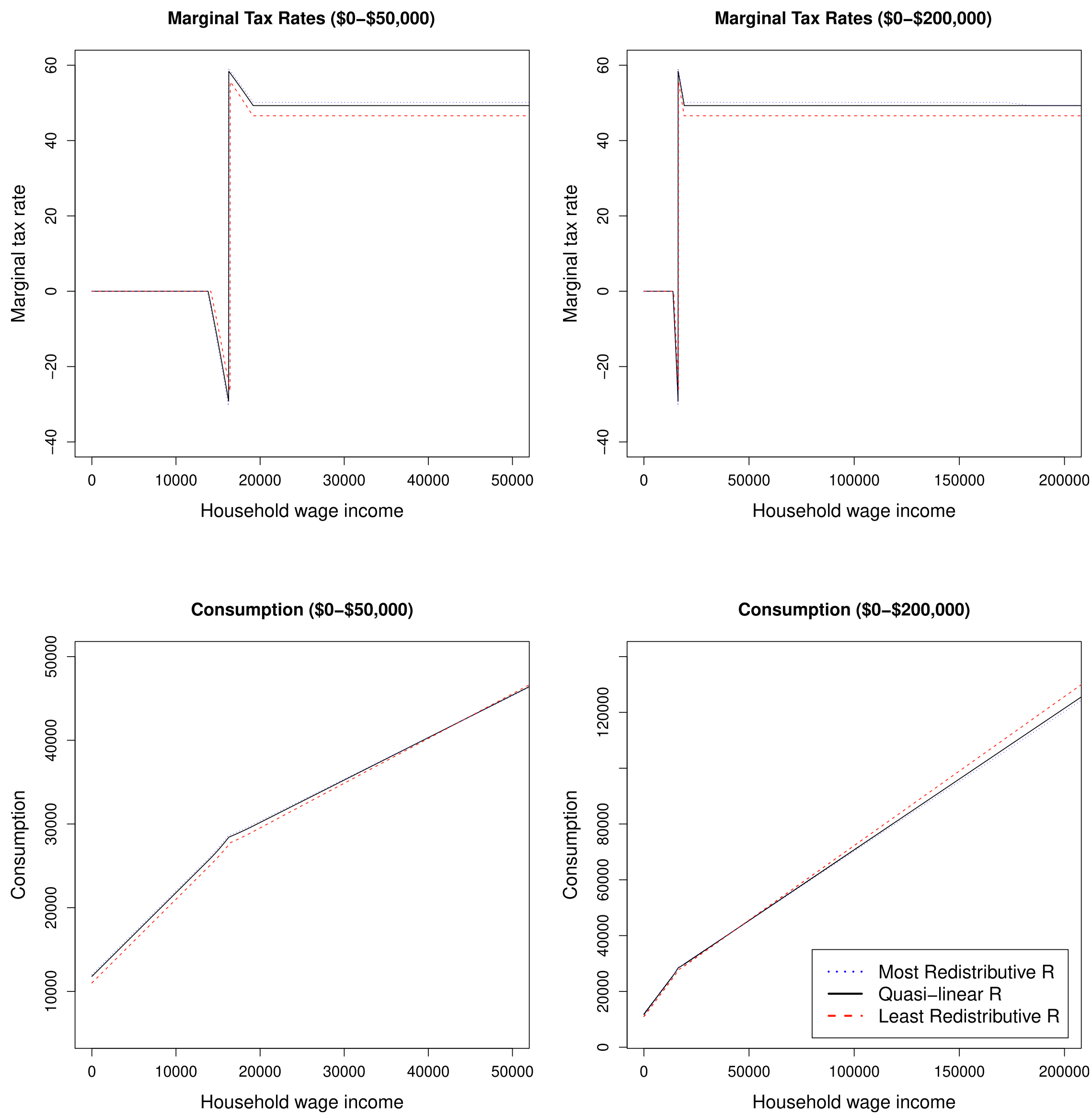

Fig. 6. Comparing optimal tax rates and consumption under the 3 polar cases ($PL = \$18,850$; Slope $= \$6/h$). The overall amount of redistribution is similar across the 3 versions of the SWF. Labor supply elasticity is $\epsilon = 0.33$. The preference parameter of the most hardworking agents ($\overline{\theta} = 222.28$) is calibrated using hours worked in a year by hourly workers in the CPS. The distribution of types n is estimated (using CPS data) via Maximum likelihood and assuming a Double Pareto Lognormal distribution (DPLn).

rationalize. It is therefore clear that responsibility and poverty reduction may be seen as related to the tax system only for what concerns households with children. Furthermore, this observation is consistent with Bill Clinton's quote at the beginning of this article which only mentions families with children.

6. Relationship to the literature

In this section, we compare the shape of our optimal tax scheme with optimal tax schemes derived from other SWFs.

First, when agents are assumed to have identical preferences, in which case the SWF is often written as utilitarian, marginal tax rates are typically large on low incomes and decreasing after, contrary to what we obtain (see Diamond, 1998, for a survey of the consequences of the utilitarian objective). This is the case under the maximin objective in incomes as well. This maximin objective also leads to different conclusions than ours regarding the taxation of large incomes, when our lower bound on utilities is binding. At the risk of underlining something obvious, let us recall that a utilitarian optimal tax depends 1) on the cardinalization of the utility functions representing preferences, whereas our result does not, and 2) on the shape of the density function of types, whereas our result does not, except on intervals in which the optimal tax is determined by efficiency only (intervals $[n_u, n_{\underline{b}}]$, $[n_{\overline{b}}, n_l]$ and $([n', \infty))$.

Many authors, however, have reached similar conclusions as ours about the optimality of non-positive tax rates on low earning levels from different viewpoints. First, negative marginal tax rates are optimal when labor responses on the extensive margin are important either because of a fixed cost of labor participation (see Saez, 2002; Diamond, 1980 or Blundell and Shephard, 2012), or a present bias (see Lockwood, Forthcoming).

Second, negative marginal tax rates can be justified under the maximization of a utilitarian objective when preferences differ and social weights are a function of these preferences (see Boadway et al., 2002).

Third, many studies have already identified a connection between poverty reduction and negative marginal tax rates on

Table 1
Calibrated optimal tax schedule.

$\overline{P}_0 = \$15,379$	$c(0)$ (\$)	$y(n_u)$ (\$)	y_1^* (\$)	$y(n_l)$ (\$)	$T'(y_1^* - \varepsilon)$	$T'(y_1^* + \varepsilon)$	$T'(y(n_l))$	$T'(y \to \infty)$
Most Red.	12,778	13,326	16,015	19,228	−0.34	0.63	0.53	0.49
Quasi-linear	11,794	13,819	16,280	19,175	−0.29	0.58	0.49	0.49
Least Red.	8,632	14,951	16,859	19,022	−0.21	0.48	0.39	0.39
$\overline{P}_0 = \$18,850$								
Most Red.	12,025	13,711	16,223	19,187	−0.30	0.59	0.50	0.49
Quasi-linear	11,794	13,819	16,280	19,175	−0.29	0.58	0.49	0.49
Least Red.	11,009	14,144	16,450	19,135	−0.27	0.56	0.47	0.47
$\overline{P}_0 \geqslant \$20,036$								
All	11,794	13,819	16,280	19,175	−0.29	0.58	0.49	0.49

Note: Each panel presents the calibrated optimal tax according to each choice of $\widetilde{R}$ (most redistributive, quasi-linear and least redistributive) for poverty lines at, respectively, \$15,379, \$18,850, and above \$20,036. The first two values correspond to weighted averages of the 2015 official US poverty lines for households of size 2, and 3 . The average slope of the poverty line is calibrated at \$6/h. The second column, $c(0)$, is the level of consumption for agents with zero income. The fourth column, y_1^*, is the income level at which there is a discrete jump in marginal tax rates. The third, fourth, and fifth columns indicate the endpoints of the intervals over which rates are decreasing. The sixth and seventh column report the marginal tax rates just before and right after y_1^*. The tax rate on incomes larger than $y(n_l)$ is reported in column eight. The last column indicates the marginal tax rate in the tail.

low incomes. Saez and Stantcheva (2016) reach that conclusion from a SWF that minimizes the poverty gap when it is defined in a way that is compatible with Pareto efficiency. Maniquet and Neumann (forthcoming) find negative marginal rates on earnings that lead to an after-tax income below the poverty line (but they do not derive the formula of the optimal tax), under the requirement that redistributing from rich to poor is desirable only if they have the same labor time. In both cases, negative marginal rates are immediately related to the objective of poverty reduction, whereas it is more related to the objective of responsibility in our case.

Two rather different SWFs, capturing a goal of poverty reduction that conflicts with Pareto efficiency, lead to negative tax rates on low incomes. Wane (2001) models aggregate poverty as an externality that individuals have a willingness to pay to reduce. As a result, society's concern for poverty reduction can be addressed through pigouvian earning subsidies to the poor. At the optimal tax, the poorest are worse-off than if the negative externality caused by poverty is ignored in the SWF. In Kanbur et al. (1994a), the social planner minimizes an Atkinson index of poverty that only depends on incomes. As a result, individuals at the bottom of the income distribution are incentivized to work longer than required by efficiency so as to increase their income and decrease poverty.

Related to our approach, Fleurbaey and Maniquet (2006, 2007) obtain non-positive marginal tax rates from social objectives that are also justified by fairness principles. The SWF used in Fleurbaey and Maniquet (2006) does not satisfy our responsibility axiom but a weaker axiom based on the idea that no redistribution should take place in economies in which all agents would have the same wage rate. As a result, the objective of the planner should be to maximize the average income subsidy rate over low incomes and to maximize the tax return on all incomes above the one of the hardworking poor agent. The SWFs used in Fleurbaey and Maniquet (2007) are closer to ours, as they satisfy the same responsibility axiom as ours (our axiom of *equal-wage transfer*). The axiom with which it is combined is logically unrelated to our axiom of poverty reduction, but the resulting SWFs are similar to ours. Their result that marginal tax rates should be zero over an interval of low incomes is similar to ours, as it also derived from their SWFs satisfying responsibility (see Remark 2 after Proposition 1 above). Importantly, Fleurbaey and Maniquet (2006, 2007) do not derive the precise formula of the optimal tax function, but we believe the method we use to solve our problem could be used to provide the formula of the optimal tax for their SWF too. On a more technical note, the results they obtained are grounded on the simplifying assumption that there is an upper bound on the labor supply, a (natural) assumption that is not typical in the optimal tax literature and that we don't impose in this paper.

Some authors have also studied SWFs based on responsibility. Saez and Stantcheva (2016) present the results of their income weights approach when the objective consists of the equality of opportunity objective of Roemer et al. (2003). By assumptions, incomes are determined by agents' socio-economic background and by their efforts. The objective is then to equalize as much as possible the incomes of agents exerting the same level of effort. As a consequence, income weights are proportional to the share of the population coming from low background and earning that income level. As this share decreases with income but does not converge to zero, the optimal tax is U-shaped, tax rates are lower than what the classical utilitarianism would yield and tax rates on very high incomes are lower than under the income maximin objective. The latter feature of their result is the only one similar to ours.

Lockwood and Weinzierl (2015) study a similar objective of responsibility as ours. They do not axiomatize their SWF, but they show that if a fixed heterogeneity of types is more and more due to differences in preferences, which do not call for redistribution, then the optimal tax is less progressive, with lower but yet typically positive tax rates on low incomes.

To conclude this section, we can add that none of the optimal formulas derived in this literature would reproduce the US labor income tax of households with children as closely as our formula does. The closest to the current shape of the US tax is given by the formula of Saez and Stantcheva (2016), according to which tax rates should be negative until the income of those agents whose consumption is equal to the poverty line, after which taxe rates jump to a very high level. However, the earning threshold at which the jump takes place is lower than under the current US tax, and tax rates beyond this threshold are those that maximize tax revenues, much larger than the current ones.

7. Conclusion

This paper starts an analysis of optimal taxation at the level of axioms that the social welfare function should satisfy and it completes the analysis at the level of the full characterization of the optimal tax formula, a formula that is then calibrated to the US economy.

The axioms that we started with embed fairness principles. For this reason, this paper follows a current trend of the literature in which the emphasis shifts from the classical utilitarian objective, in which the social optimum is defined in terms of allocation of subjective utility, to the fairness of the resulting allocation of resources, in this case labor time and consumption.

The two main fairness norms that we have studied are responsibility for one's preferences and poverty reduction. The former is related to the requirement that, absent any information constraint, no redistribution should take place among agents having the same ability to earn income, a requirement that seems to be central to the many approaches to fairness in labor income taxation (see Fleurbaey and Maniquet, 2018, for a survey). The latter, poverty reduction, has already received attention in the literature, but in a way that is incompatible with Pareto efficiency, with the consequence that, contrary to what we obtain here, an increase in the social objective can be concomitant with a decrease in the utility of the poor. The axiom of poverty reduction that we have studied here avoids this paradox, but at the price of weakening the definition of being poor. Rather than defining poverty as consuming a bundle below the poverty line, we defined it as consuming a bundle on an indifference curve entirely below the poverty line.

As a result, a tax scheme may be optimal in our sense and, yet, leave some agents consume bundles below the poverty line. What the optimal tax scheme accomplishes, though, is to avoid redistribution among the low-wage agents. As a result, the marginal tax rates are zero or even negative up to the earning level of the low-wage agent choosing the longest labor time. Consequently, the budget of these agents crosses the poverty line from below, so that agents end up being lifted out of poverty, that is consuming bundles above the poverty line, provided their labor time is sufficient.

Another consequence of our axioms is that, in spite of our social preference taking the shape of an extremely egalitarian one, the optimal tax rates on high incomes may differ from the ones that would maximize the tax revenue (the so-called Rawlsian tax scheme). This is even more true as we choose a steeper or lower poverty line. This has to do with the way we define the well-being index which the social preference tries to maximin. Indeed, it is typically not true that an agent with a larger wage ends up with a larger well-being level (whereas in the classical framework an agent with a larger wage always ends up with a higher utility). This comes from the responsibility axiom, which forces us to look at implicit budgets rather than consumption bundles.

Finally, we compare our calibrated tax rates to the current US tax system. Despite a few differences, it is surprisingly close to the one that we rationalize with our axioms, at least as far as households with one or more children are concerned. This shows that switching from the traditional utilitarian normative assumption to a fairness-based approach can help close the gap between the recommendations of optimal tax models and the reality of policy.

A recent development - and attractive feature - of the modern optimal tax literature has been to derive formulae that are somewhat robust to model specification and are expressed in terms of sufficient statistics. Though our results do not share this feature, we view it as an unavoidable consequence of the normative stance taken in this paper: we assume that society aims at correcting for certain sources of inequalities, the ones generating poverty, but remain neutral towards others, the ones emerging from different labor time choices. This requires a *model* of what the sources of inequality are and it is therefore not surprising that the corresponding tax formula is *model dependent*. This relates to a long tradition in ethics positing that normative judgments must be context dependent.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank two anonymous referees, the editor, Marianne Bertrand, Stéphane Bonhomme, Richard Blundell, Nathaniel Hendren, Hilary Hoynes, Guy Laroque, Casey Mulligan, Doron Ravid, Raaj Sah, Johannes Spinnewyn, Matthew Weinzierl, Glen Weyl and John Weymark as well as seminar participants at the LSE, IFS, LISER, the Universities of Chicago, Cologne, Manchester, Oslo, the Delhi School of Economics and conference attendees at the 2016 and 2018 Social Choice and Welfare meetings at Lund and Seoul, 2017 RIDGE conference in public finance, Montevideo, for their comments.

Appendix A. Axioms of poverty reduction

Axiom 4. *poverty-reduction transfer 1*

For all $E=(w_N,R_N)\in\mathscr{E}$, all $\Delta\in\mathbb{R}_+$, all $j,k\in N$, all $z_N=(l_N,c_N), z'_N=(l'_N,c'_N)\in X^N$ such that $l_j=l'_j=l_k=l'_k, c'_j=c_j-\Delta, c'_k=c_k+\Delta$ and for all $i\neq j,k: z_i=z'_i$,

$$\left[z'_j>PL>z'_k\right]\Rightarrow\left[z'_N\mathbf{R}(E)z_N\right].$$

Axiom 5. *poverty-reduction transfer 2*

For all $E=(w_N,R_N)\in\mathscr{E}$, all $\Delta\in\mathbb{R}_+$, all $j,k\in N$, all $z_N=(l_N,c_N), z'_N=(l'_N,c'_N)\in X^N$ such that $l_j=l'_j=l_k=l'_k, c'_j=c_j-\Delta, c'_k=c_k+\Delta$ and for all $i\neq j,k: z_i=z'_i$,

$$\left[I\left(z'_j,R_j\right)>PL>z'_k\right]\Rightarrow\left[z'_N\mathbf{R}(E)z_N\right].$$

These three axioms are illustrated in Fig. 7. It should be transparent that the first version of the axiom is logically stronger than the second one, which is logically stronger than Axiom 3 defined in Section 3 (henceforth referred to as *poverty-reduction transfer 3*). Moreover, we have the following (in)compatibilities with the other axioms. To state these (in)compatibilities, we recall a well-known implication of strong Pareto, Pareto indifference. It is the requisite that if all agents are indifferent between two allocations, then social preferences must also be indifferent.

Axiom 6. *Pareto indifference*

For all $E=(w_N,R_N)\in\mathscr{E}$, and $z_N=(l_N,c_N), z'_N=(l'_N,c'_N)\in X^N$, if $z_iI_iz'_i$ for all $i\in N$, then $z_N\mathbf{I}(E)z'_N$.

Proposition 2.

1. *No SWF satisfies Pareto indifference and poverty-reduction transfer 1.*
2. There are SWFs satisfying *Pareto indifference* and *poverty-reduction transfer 2*.
3. *No SWF satisfies Pareto indifference, equal-wage transfer and poverty-reduction transfer 2.*

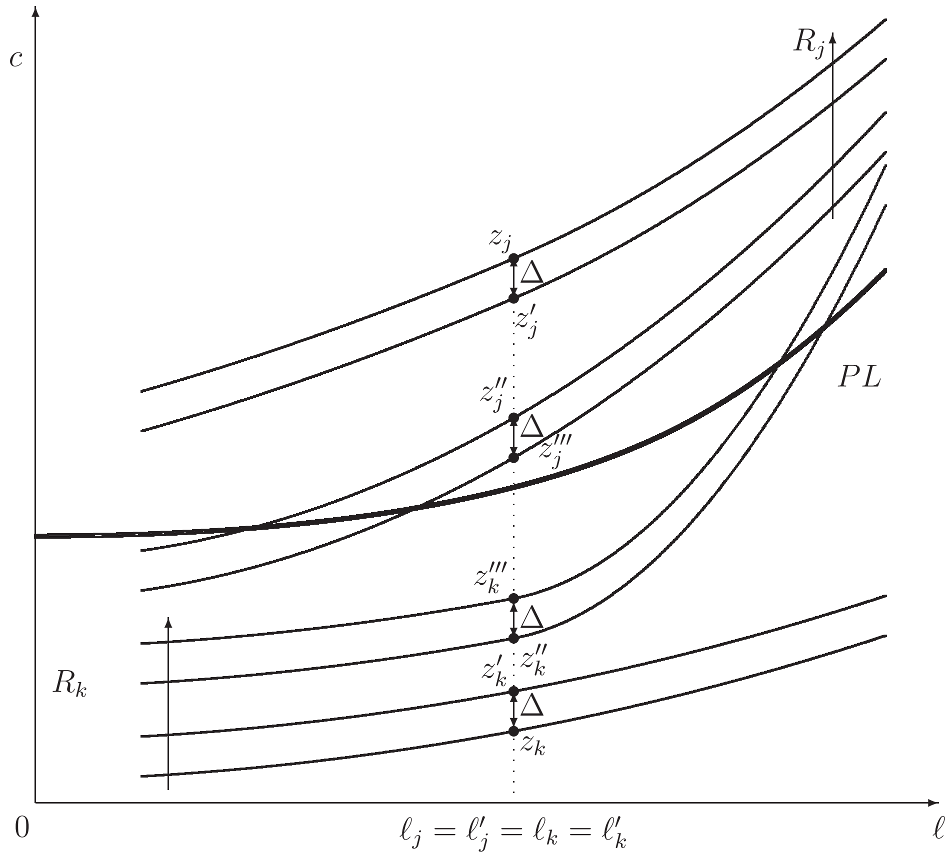

Fig. 7. Illustration of the axioms of *poverty-reduction transfer* 1, 2 and 3: according to *poverty-reduction transfer* 1, any transfer of Δ from either z_j or z''_j to either z_k or z''_k is a strict social improvement; according to *poverty-reduction transfer* 2, any transfer of Δ from z_j to either z_k or z''_k is a strict social improvement; according to *poverty-reduction transfer* 3, only a transfer of Δ from z_j to z_k is a strict social improvement.

4. *There are SWFs satisfying strong Pareto, equal-wage transfer and poverty-reduction transfer 3.*

The proof of the four statements of Proposition 2 is rather intuitive. We present it in the text. Statement 1 is reminiscent of the proof of the impossibility of a Paretian egalitarian in Fleurbaey and Trannoy (2003). It is illustrated in Fig. 8. The key feature of the figure is that both agents j and k have preferences that cross the poverty line, but in opposite direction. As a result, when they both choose a low labor time, say $\ell_j = \ell_k$, then agent j has a consumption level above the poverty line whereas k's consumption is below the poverty line. As a result, $\left(z'_j, z'_k\right)$ is socially preferred to (z_j, z_k). When they both choose a large labor time, say $\ell''_j = \ell''_k$, at the same satisfaction levels, k's consumption level is above the poverty line contrary to j's consumption level. As a result, $\left(z'''_j, z'''_k\right)$ is socially preferred to $\left(z''_j, z''_k\right)$. By *Pareto indifference*, which is implied by *strong Pareto*, (z_j, z_k) is socially indifferent to $\left(z'''_j, z'''_k\right)$ and $\left(z'_j, z'_k\right)$ is socially indifferent to $\left(z''_j, z''_k\right)$, creating an intransitivity.

The proof of statement 3 is illustrated in Fig. 9. *Poverty-reduction transfer 2* requires that $\left(z'_j, z'_k\right)$ be socially preferred to (z_j, z_k). *Equal-wage transfer* requires that $\left(z'''_j, z_k\right)$ be socially preferred to $\left(z''_j, z'_k\right)$. Finally, *Pareto indifference* requires that (z_j, z_k) be socially indifferent to $\left(z'''_j, z_k\right)$ and $\left(z'_j, z'_k\right)$ to $\left(z''_j, z'_k\right)$, creating an intransitivity.

Statement 3 is at the heart of the ethical dilemma that this paper addresses. Agents j and k have the same wage, represented in the figure by the slope of their budgets. The responsibility goal pushes towards no redistribution among them two. However, agent k enjoys a bundle below the poverty line, whereas agent j's indifference curve is everywhere above the poverty line. This is why even the weak *poverty-reduction transfer 2* requires to prefer a redistribution between the two agents when they don't work hard, that is, when their labor times are $\ell_j = \ell_k$. The two transfers go in opposite directions, whereas agent j is indifferent between the two involved bundles, hence the incompatibility.[16]

Statements 2 and 4 are proven by way of examples. Statement 4 is proven in the main text in Section 3 ($\mathbf{R}^{\tilde{R}\,lex}$ satisfies *strong Pareto*, *equal-wage transfer* and *poverty-reduction transfer 3*). We now prove statement 2. We first need to introduce some terminology. Let $\mathcal{U}$ denote the set of utility functions representing preferences in $\mathcal{R}$. This example works by choosing a utility representation for each individual preference relation and apply the leximin criterion to the utility vector corresponding to the evaluated allocations. Let $U : \mathcal{R} \to \mathcal{U}$ be a representation function satisfying the property that for each $R \in \mathcal{R}, U(R)$ is a utility function representing R and if we write $u = U(R)$ then for all $z \in X$

$$u(z) = 1 \Longleftrightarrow z\,I\,m(R, PL), \tag{6}$$

that is, all agents have the same utility level at their preferred bundle on the poverty line (and this utility level is equal to 1). Social

[16] Observe that this incompatibility does not depend on our assumption that the poverty line is strictly increasing. Even a flat poverty line would display the same basic problem.

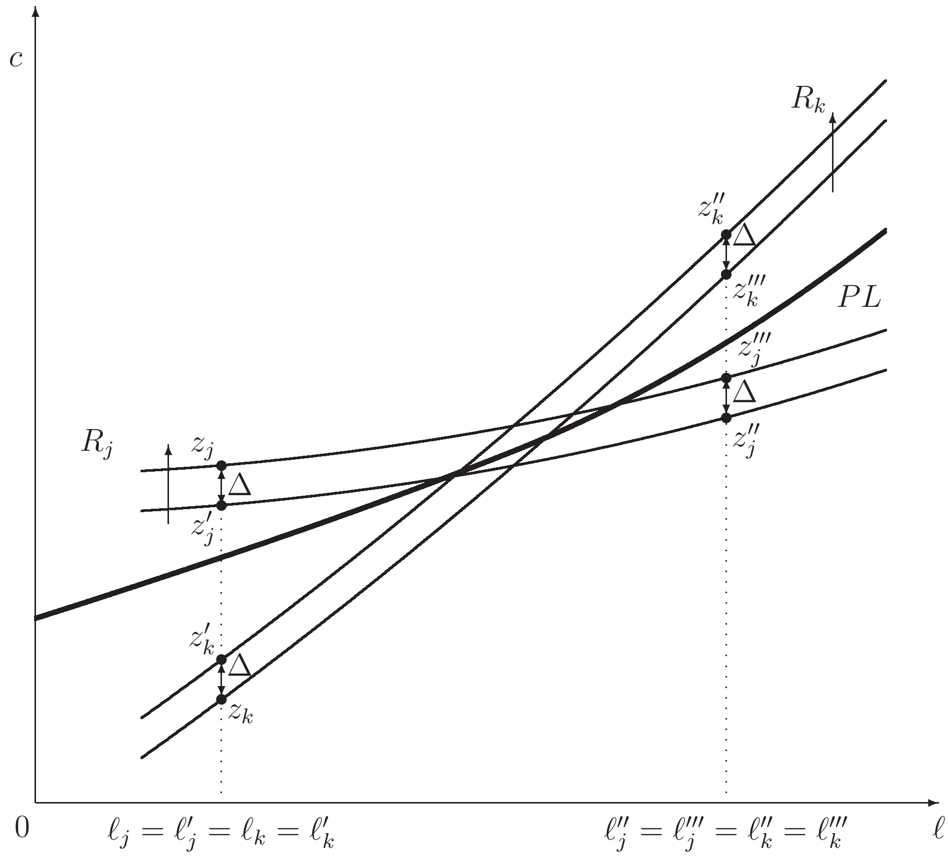

Fig. 8. Illustration of Proposition 2, statement 1. According to *poverty-reduction transfer 1*, $\left(z'_j, z'_k\right)$ is socially preferred to (z_j, z_k) and $\left(z'''_j, z'''_k\right)$ is preferred to $\left(z''_j, z''_k\right)$. By *Pareto indifference*, (z_j, z_k) is socially indifferent to $\left(z'''_j, z'''_k\right)$ and $\left(z'_j, z'_k\right)$ is indifferent to $\left(z''_j, z''_k\right)$. This leads to an intransitivity showing that no SWF satisfies *Pareto indifference* and *poverty-reduction transfer 1*.

preferences R^U are defined as follows: for all $E = (w_N, R_N) \in \mathscr{E}$, all $z_N, z'_N \in X^N$,

$$z_N R^U z'_N \iff (u_i(z_i))_{i\in N} \geqslant_{lex} \left(u_i(z'_i)\right)_{i\in N}$$

where for all $i \in N, u_i = U(R_i)$. The fact that R^U satisfies *strong Pareto* follows from u_i being a utility representation of R_i and the leximin criterion being strictly increasing in all its arguments. The fact that R^U satisfies *poverty-reduction transfer 2* follows Eq. (6) satisfied by the utility representation. Indeed, as long as $I\left(z'_j, R_j\right) > PL > z'_k$, we have $u_j\left(z'_j\right) > 1 > u_k(z'_k)$, so that the leximin criterion gives priority to agent k, with the consequence that z'_N is strictly preferred to z_N.

Appendix B. Proof of Proposition 3 on SWF's

We begin by defining *separability* formally. Assume the social preference relation is set between two allocations, z_N and z'_N, whereas each agent in a subset M of the population is allocated exactly the same bundle in both allocations. The requirement is that agents in M should not matter in the social welfare function, which is captured by the requirement that the social preference relation should remain unaffected if the preferences of agents in M change and if their bundles change in such a way that they are still exactly the same in the two resulting allocations.

Axiom 7. *separability*

For all $E = (w_N, R_N), E' = (w'_N, R'_N) \in \mathscr{E}$, all z_N, z'_N, z''_N and $z'''_N \in X^N$, if $w_i = w'_i$ for all $i \in N$ and there exists $M \subset N$ such that

$$\forall i \in M : z_i = z'_i \text{ and } \quad z''_i = z'''_i,$$
$$\forall i \in N \setminus M : R_i = R'_i,$$
$$\forall i \in N \setminus M : z_i = z''_i \text{ and } \quad z'_i = z'''_i,$$

then

$$z_N \mathbf{R}(E) z'_N \iff z''_N \mathbf{R}(E') z'''_N.$$

We need to introduce the following terminology, which corresponds to the implicit budget terminology of Samuelson (1974). The equivalent income is the intercept of an implicit budget. Formally, for (z, w, R), we let $t(z, w, R)$ denote the intercept of $IB(z, w, R)$, that is,

$$t(z, w, R_i) = \min\{t \in \mathbb{R} | \exists (\ell, c) \in X,\ c = t + w\ell,\ (\ell, c) R z\}.$$

With the same abuse of notation as above, we define

$$t(w, PL) = \min\{t \in \mathbb{R} | \exists (\ell, c) \in PL,\ c = t + w\ell\}.$$

Slightly abusing on notation, we write $IB(w, PL)$ to denote the implicit budget of slope w that is tangent from below to the poverty line.

Proposition 3. *If a SWF satisfies strong Pareto, equal-wage transfer, poverty-reduction transfer and separability, then for all $E = (w_N, R_N) \in \mathscr{E}$ such that $|N| \geqslant 4$, all $z_N, z'_N \in X^N$, if there exist $j, k \in N$ such that*

- $IB(z_j, w_j, R_j) \supset IB\left(z'_j, w_j, R_j\right) \supset IB(w_j, PL)$,
- $IB(w_k, PL) \supset IB(z'_k, w_k, R_k) \supset IB(z_k, w_k, R_k)$,

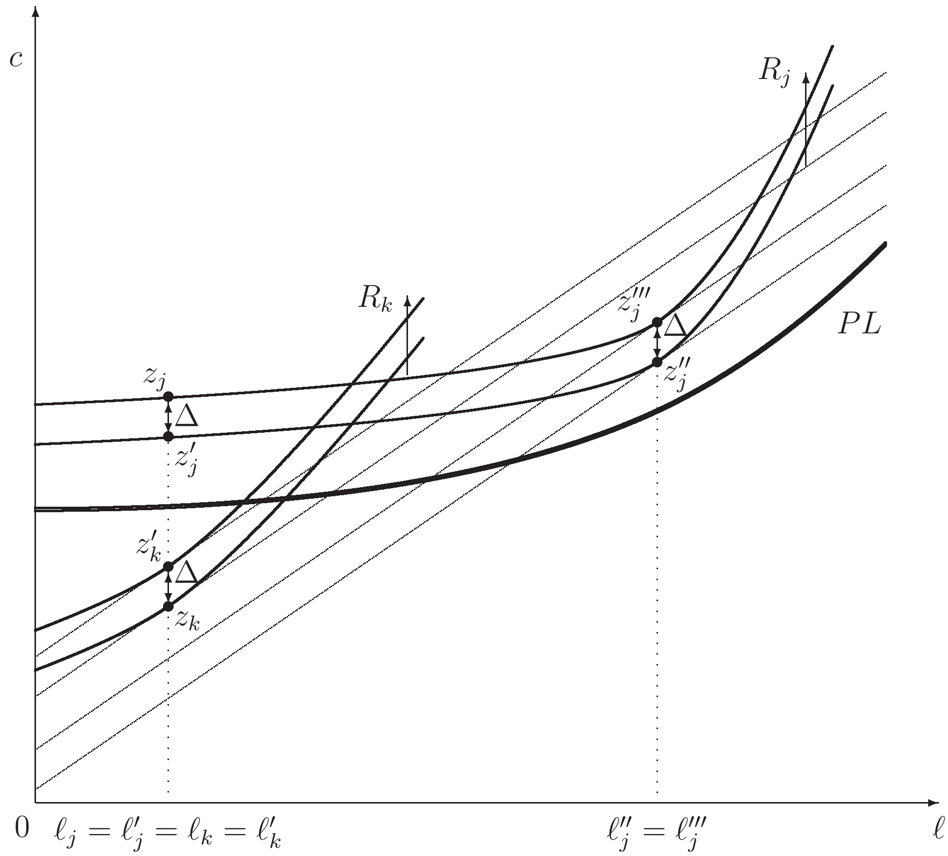

Fig. 9. Illustration of Proposition 2, statement 3. By *poverty-reduction transfer 2*, $\left(z_j', z_k'\right)$ is socially preferred to (z_j, z_k). By *Pareto indifference*, (z_j, z_k) is socially indifferent to $\left(z_j''', z_k\right)$ and $\left(z_j', z_k'\right)$ is indifferent to $\left(z_j'', z_k'\right)$. Finally, by *equal-wage transfer*, $\left(z_j''', z_k\right)$ is socially preferred to $\left(z_j'', z_k'\right)$. This leads to an intransitivity showing that no SWF satisfies *Pareto indifference*, *equal-wage transfer* and *poverty-reduction transfer 2*.

and for all $i \neq j, k : z_i = z_i'$*, then* $z_N' \mathbf{P}(E) z_N$.

Proof: In the proof, we make use of *Pareto indifference*, previously defined as Axiom 6. It is a well-known implication of strong Pareto. We prove the theorem in three steps. The first step is a preliminary result about the existence of specific preference relations which are illustrated in Fig. 10. Take any two $w, w' \in [\underline{w}, \infty)$, and any $t, t', t'', t''' \in \mathbb{R}$ such that.

- $t < t' < t(w, PL)$,
- $t(w', PL) < t'' < t'''$.

There necessarily exists $R^\circ \in \mathscr{R}$ and $z = (\ell, c), z' = (\ell', c'), z'' = (\ell'', c''), z''' = (\ell''', c''') \in X$ such that

- $\ell = \ell' = \ell'' = \ell'''$,
- $c + c''' = c' + c''$,
- $t(z, w, R^\circ) = t$,
- $t(z', w', R^\circ) = t'$,
- $t(z'', w', R^\circ) = t''$,
- $t(z''', w', R^\circ) = t'''$.

Instead of proving the claim algebraically, we illustrate its simple intuition in Fig. 10. The distance between the two budgets of slope $\tilde{w}'$ is much larger than that between the budgets of slope $\tilde{w}$. In spite of that, it is possible to construct indifference curves that are tangent to the former budgets and yet become arbitrarily close to each other.

We now turn to the second step. We prove the theorem in a specific case. Let $E = (w_N, R_N) \in \mathscr{E}, z_N, z_N' \in X^N$, and $j, k \in N$ satisfy the conditions of the statement. Assume, moreover, that there exist $t_j > t(w_j, PL)$ and $t_k < t(w_k, PL)$ such that

$$\begin{aligned} t\left(z_j', w_j, R_j\right) - t_j &= t(z_j, w_j, R_j) - t\left(z_j', w_j, R_j\right), \\ t_k - t(z_k', w_k, R_k) &= t(z_k', w_k, R_k) - t(z_k, w_k, R_k). \end{aligned}$$

As $|N| \geqslant 4$, we assume $|N| = 4$ to save on notation. Let $N = \{j, k, a, b\}$. We need to prove that $\left(z_j', z_k', z_a, z_b\right) \mathbf{P}(E) (z_j, z_k, z_a, z_b)$. By *Pareto indifference*, we can do as if z_j and z_j' (resp. z_k and z_k') were the best bundles of j (resp. k) in the corresponding implicit budget. If it is not the case, indeed, then we can identify those best bundles and replace z_j, z_j', z_k and z_k' by those bundles. Let $R^\circ \in \mathscr{R}$ and $z = (\ell, c), z' = (\ell', c'), z'' = (\ell'', c'')$, $z''' = (\ell''', c''') \in X$ be such that

- $\ell = \ell' = \ell'' = \ell'''$,
- $c + c''' = c' + c''$,
- $t(z, w_k, R^\circ) = t(z_k', w_k, R_k)$,
- $t(z', w_k, R^\circ) = t_k$,
- $t(z'', w_j, R^\circ) = t_j$,
- $t(z''', w_j, R^\circ) = t\left(z_j', w_j, R_j\right)$.

Let $E' \in \mathscr{E}$ be defined by

$$E' = \left((w_j, w_k, w_j, w_k), (R_j, R_k, R^\circ, R^\circ)\right).$$

By *equal-wage transfer* applied twice, first between agents j and a and then between agents k and b,

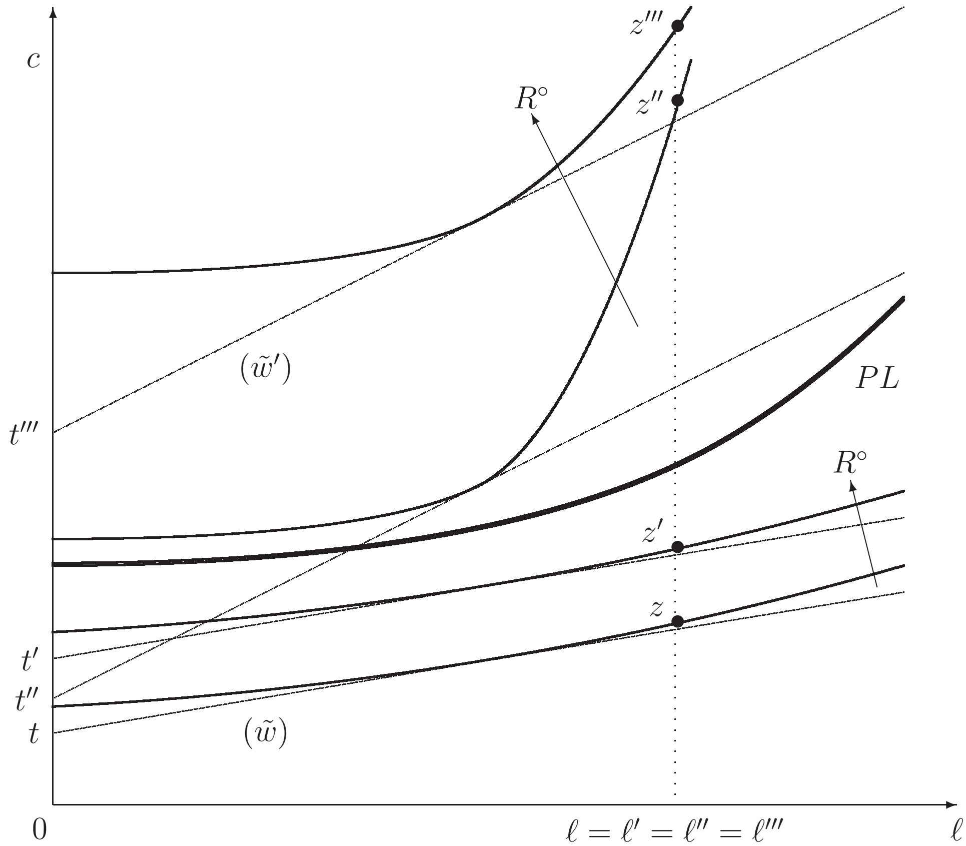

Fig. 10. Illustration of the existence of R°. Four indifference curves of R° are represented. The two indifference curves above the poverty line are such that despite the distance between implicit budgets at wage $\tilde{w}'$ being large, they are very close to each other at $\ell = \ell' = \ell'' = \ell'''$. In fact, $c''' - c'' = c' - c$. For any given distance between budgets of the same slope, there always exists such pair of indifference curves that are tangent to both budgets, yet become arbitrarily close to each other. The existence of such preferences is key in proving Proposition 3.

$$\left(z_j', z_k', z''', z\right) \mathbf{P}(E') \left(z_j, z_k, z'', z'\right).$$

By *poverty-reduction* applied between agents a and b,

$$\left(z_j', z_k', z'', z'\right) \mathbf{P}(E') \left(z_j', z_k', z''', z\right).$$

By transitivity,

$$\left(z_j', z_k', z'', z'\right) \mathbf{P}(E') \left(z_j, z_k, z'', z'\right).$$

By *separability*,

$$\left(z_j', z_k', z_a, z_b\right) \mathbf{P}(E) \left(z_j, z_k, z_a, z_b\right),$$

the desired outcome. Of course, it is unlikely that such t_j and t_k exist, as $t\left(z_j', w_j, R_j\right)$ can be arbitrarily close to $t(w_j, PL)$ and $t(z_k', w_k, R_k)$ arbitrarily close to $t(w_k, PL)$.

As a third and final step, we deal with the general case. We can find $t_j > t(w_j, PL)$ and $t_k < t(w_k, PL)$ and an integer n such that

$$\begin{aligned} t\left(z_j', w_j, R_j\right) - t_j &= \tfrac{1}{n}\left(t(z_j, w_j, R_j) - t\left(z_j', w_j, R_j\right)\right), \\ t_k - t(z_k', w_k, R_k) &= \tfrac{1}{n}\left(t(z_k', w_k, R_k) - t(z_k, w_k, R_k)\right). \end{aligned}$$

Then, we can define two series of $n-1$ bundles $z_j^1, \ldots, z_j^{n-1}$ and $z_k^1, \ldots, z_k^{n-1}$ such that for all $p \in \{1, \ldots, n-1\}$,

$$t\left(z_j^p, w_j, R_j\right) = t(z_j, w_j, R_j) - p\left(t\left(z_j', w_j, R_j\right) - t_j\right), \quad (7)$$

$$t(z_k^p, w_k, R_k) = t(z_k, w_k, R_k) + p\left(t_k - t(z_k', w_k, R_k)\right). \quad (8)$$

For all $p \in \{1, \ldots, n\}$, we can apply step 2 and prove that

$$\left(z_j^p, z_k^p, z_a, z_b\right) \mathbf{P}(E) \left(z_j^{p-1}, z_k^{p-1}, z_a, z_b\right),$$

with the obvious convention that $z_j^0 = z_j, z_j^n = z_j'$, and $z_k^0 = z_k, z_k^n = z_k'$. The proof then goes on by transitivity applied to those n steps, to reach the final outcome that

$$\left(z_j', z_k', z_a, z_b\right) \mathbf{P}(E) \left(z_j, z_k, z_a, z_b\right). \qquad \square$$

Appendix C. Proof of Corollary 1 on the planner's objective

The following corollary proves that the objective of any planner interested in maximizing $\mathscr{R}^{\tilde{R}\,lex}$ should be to lift all agents out of poverty in the sense that the implicit budget of no agent should lie below the poverty line.

Corollary 1. *If a SWF satisfies strong Pareto, equal-wage transfer, poverty-reduction transfer and separability, then for all* $E = (w_N, R_N) \in \mathscr{E}$ *such that* $|N| \geqslant 4$, *all* $z_N, z_N' \in X^N$, *if there exists* $j \in N$ *such that*

$$IB(w_j, PL) \supset IB(z_j, w_j, R_j)$$

whereas for all $i \in N$

$$IB(z_i', w_i, R_i) \supset IB(w_i, PL),$$

then $z_N' \mathbf{P}(E) z_N$.

Proof. Let $E = (w_N, R_N) \in \mathscr{E}, z_N, z'_N \in X^N$, and $j \in N$ satisfy the conditions of the corollary. Let us assume, contrary to the claim, that $z_N \mathbf{R}(E) z'_N$. Let $M \subset N$ be defined by

$$M = \{i \in N | i \neq j, z_i P_i z'_i\}.$$

If $M = \emptyset$, then, by *strong Pareto*, we have reached a contradiction. Let us then assume that $M \neq \emptyset$. Let $z''_N \in X^N$ be such that

$$\forall i \in M, IB(w_i, z_i, R_i), IB(w_i, z'_i, R_i) \supset IB(w_i, z''_i, R_i) \supset IB(w_i, PL),$$
$$\forall i \notin M, i \neq j,\ z''_i = z_i,$$
$$IB(w_j, PL) \supset IB\left(w_j, z''_j, R_j\right) \supset IB(w_j, z_j, R_j).$$

Existence of such a z''_N is guaranteed by the conditions imposed on z_N and z'_N in the statement of the corollary.

Let $r > 0$ and $z^0_j, z^1_j, \ldots, z^m_j \in X, m = |M|$ be such that

$$\begin{aligned} IB\left(w_j, z^0_j, R_j\right) &= IB(w_j, z_j, R_j), \\ IB\left(w_j, z^k_j, R_j\right) &= IB\left(w_j, z^{k-1}_j, R_j\right) + (0, r), \forall k \in \{1, \ldots, m\}, \\ IB\left(w_j, z^m_j, R_j\right) &= IB\left(w_j, z''_j, R_j\right). \end{aligned}$$

By applying Proposition 3 m times, we can prove that $z''_N \mathbf{P}(E) z_N$. Indeed, at each step, z^{k-1}_j is replaced by z^k_j, which means that an agent whose implicit budget is strictly below the poverty line is made better-off, and z_i is replaced with z''_i, which means that an agent strictly above the poverty line is made worse-off, which implies a strict social preference. By transitivity, $z''_N \mathbf{P}(E) z'_N$, which contradicts *strong Pareto*, because $z'_i R_i z''_i$ for all $i \in N$ and the preference is strict for some agents, including j. □

Appendix D. Proof of Proposition 1

D.1. Overview

We prove Proposition 1 in four steps described below. We first provide a sketch of the proof describing each step and how they relate to each other. A formal statement and derivation of each step follows.

In the first step, we quickly prove a useful property: for each n, those with the lowest wage are the worst-off among all agents with $n_i = n$. Step 1 is useful because, by the leximin property of $\mathscr{R}^{\tilde{R}\,lex}$, we can focus on agents with $w_i = \underline{w}\ \forall n \in [\underline{n}, \underline{w}\overline{\theta}]$ and with $w_i = n_i/\overline{\theta}\ \forall n_i \in [\underline{w}\overline{\theta}, \infty)$. We define the well-being index of the worst-off agents at n_i as $\underline{b}(u(n_i); n_i)$.

In the second step, we prove that agents with $\theta_i = \overline{\theta}$ and $w_i = \underline{w}$ (the most-hardworking minimum wage agents) are among the worst-off at the second-best optimum. We define $n_1 = \underline{w}\overline{\theta}$. To be clear, step 1 is a comparison within n while step 2 compares across n. It is also important to note that step 2 proves that agents with $\theta_i = \overline{\theta}$ and $w_i = \underline{w}$ are among the worst-off at the optimum but other types $n \neq n_1$ may very well have just as low a well-being index at the solution. The result derived in step 2 is important because it allows to re-write the maximization problem in a more tractable way. The original maximization problem consists in maximizing the minimum well-being across n_i under incentive constraints (3) and the resource constraint (4). Assume that the maximized objective takes value b^*. From step 2, we know that $b^* = \underline{b}(u^*(n_1); n_1)$ where $u^*(n)$ denotes the utility for type n at the solution. We can consider, instead of the original maximization problem, its dual which consists in maximizing a budget surplus under the constraint that $\underline{b}(u(n); n) \geqslant b^*$ for all n (in addition to incentive compatibility constraints). As $\underline{b}(u(n); n)$ is a utility representation for lowest wage agents within type n, it is strictly increasing in the first argument. As a result, the constraint can be rewritten directly in terms of a lower-bound on utilities:

$$u(n_1) = \underline{u}_1 \quad (9)$$
$$u(n) \geqslant \underline{u}(n) \quad \forall n \in [\underline{n}, \infty) \quad (10)$$

where $\underline{u}_1 = u^*(n_1)$ and $\underline{u}(n)$ is the utility assignment as a function of n such that $\underline{b}(u(n); n) = b^*$ for all n. Notice that (9) is a consequence of the result proven in step 2. Indeed, from the point of view of the dual problem, we established in that step that the lower-bound on utilities is always binding for agents with $n = n_1$.

In the third step, we set up an Hamiltonian corresponding to the (dual) problem described above and derive the first-order conditions. Constraint (9) provides the initial condition: the state variable $u(n)$ is anchored at n_1. Since typically $\underline{n} < n_1 < \infty$, we divide the problem in two subproblems: the lower sub-problem which consists in solving the optimal trajectory, $\{y^*(n), u^*(n)\}$, on $[\underline{n}, n_1]$ (with initial condition (9) at the upper-end) and the upper sub-problem which consists in solving the optimal trajectory on $[n_1, \infty)$ (with initial condition (9) at the lower-end). The second-order conditions of the incentive compatibility constraints, which are equivalent to $y(n)$ being monotonic everywhere, are binding at n_1 leading to some bunching.[17] The solution within each sub-problem is characterized by intervals over which the lower-bound on utilities (10) is binding and intervals over which it is not. The former intervals are delimited by threshold values of $n : \underline{n}, n_u, n_l$ and n'. Consequently, these threshold values correspond to the values of n at which the lower-bound on utilities becomes binding. By definition of the lower-bound, over these intervals, the marginal tax rate is such that the well-being index of the worst-off agent at each n is equalized across n. In the lower-subproblem, that is $\forall n \in [\underline{n}, n_1]$, this consists in equalizing the well-being of agents earning the minimum wage $\underline{w}$ but having different preferences θ_i. For $\mathscr{R}^{\tilde{R}\,lex}$, the incentive compatible allocation that equalizes the well-being of agents with the same wage corresponds to a zero marginal tax rate. As a result, the optimal tax on low-incomes under $\mathscr{R}^{\tilde{R}\,lex}$ is always such that there is, for some $n_u \in [\underline{n}, n_1]$, an interval $[\underline{n}, n_u]$ with zero marginal tax rates followed by an interval $\left[n_u, n_{\underline{b}}\right)$ over which marginal tax rates are negative.[18] In the upper sub-problem, that is $\forall n \in [n_1, \infty)$, characterizing the exact shape of the incentive compatible allocation that equalizes well-being across the worst-off within each n (agents with the same preference parameter $\overline{\theta}$ but different wages) requires parameterizing agents' (and reference) preferences.

This brings us to step 4. The assumption of iso-elastic and quasi-linear preferences really only is needed in this final step. In step 4, we first derive Lemma 2, an algebraic expression for the well-being index of the worst-off within each $n : \underline{b}(u(n); n)$. Using this algebraic expression, we are able to calculate the optimal marginal tax rate on intervals for which the lower-bound (10) is binding. Under our assumptions on preferences, the marginal tax rate is constant on these intervals. We also derive an expression for the solution on intervals where the lower-bound is not binding.

After introducing some notation, we formally prove steps 1–4. A few additional details and more technical derivations are left for Online Appendix I. In Online Appendix J, we also present some generalizations of steps 1–3 and the consequent robustness of certain key features of the optimal tax schedule to allowing for more general preferences.

[17] Bunching at $y^*(n_1)$ happens on interval $\left[n_{\underline{b}}, n_{\overline{b}}\right]$. It occurs for reasons similar to Brett and Weymark (2017). The optimal tax problem described in this section shares some similarities with the maximization problem they consider with the important difference that we face an additional constraint: the lower-bound on utilities (10).

[18] The remaining types $\left[n_{\underline{b}}, n_1\right]$ are bunched with n_1 at income y^*_1.

D.2. Introducing some notation

Studying tax schemes forces us to study the income/consumption space instead of the labor time/consumption space. As a general rule, we underline variables when they are converted from the former to the latter space. Let $E = (w_N, R_N)$ be an economy. For each $i \in N$, we redefine preferences R_i as $\underline{R}_i$ as follows:

$$(y,c)\,\underline{R}_i\,(y',c') \iff \left(\frac{y}{w_i},c\right) R_i \left(\frac{y'}{w_i},c\right).$$

We now refer to (second-best) budgets as $\underline{B}(\tau)$, defined as

$$\underline{B}(\tau) = \{\underline{z} = (y,c) \in \underline{X} | c \leqslant y - \tau(y)\}.$$

All implicit budgets in the (y,c) space have slopes equal to 1. We now refer to them as follows:

$$\underline{IB}(\underline{z}_i, \underline{R}_i) = \{\underline{z} = (y,c) \in \underline{X} | \exists t \in \mathbb{R} : c - y \leqslant t \text{ and } \underline{z}_i\,\underline{I}_i\,\underline{m}(\underline{R}_i, \underline{IB}(\underline{z}_i, \underline{R}_i))\}.$$

Let τ be a tax scheme. Let $\underline{z}_N$ be the allocation generated by the tax scheme, that is for all $i \in N$,

$$\underline{z}_i \in \underline{m}(\underline{R}_i, \underline{B}(\tau)).$$

We say that τ is feasible if it generates $\underline{z}_N = (y_i, c_i)_{i \in N}$ such that

$$\sum_{i \in N} c_i \leqslant \sum_{i \in N} y_i.$$

We say that a feasible tax scheme τ is the optimal tax scheme in economy $E = (w_N, R_N)$ if τ generates allocation $\underline{z}_N$ corresponding to allocation z_N in the labor time/consumption space, and for all other feasible tax scheme τ', generating allocation $\underline{z}'_N$ corresponding to allocation z'_N in the labor time/consumption space we have

$$z_N\,\mathbf{R}^{R\sim lex}\,z'_N.$$

D.3. Step 1: The worst-off within each n

Lemma 1. *Among agents of type n, agents with the lowest wage are always the worst-off according to $b(z_i, w_i, \theta_i)$. That is, $\forall j,k$ with $n_j = n_k, \underline{z}_j = \underline{z}_k, w_j < w_k$,*

$$b(z_j, w_j, \theta_j) \leqslant b(z_k, w_k, \theta_k)$$

Proof. Indeed, let us consider two agents, j and k with $n_j = n_k$, such that $\underline{R}_j = \underline{R}_k$, but $w_j < w_k$. Because they have the same preferences in the income-consumption space, they choose the same $\underline{z}_j = (y_j, c_j) = \underline{z}_k = (y_k, c_k)$, with the resulting property that $\underline{IB}(\underline{z}_j, \underline{R}_j) = \underline{IB}(\underline{z}_k, \underline{R}_k)$. Therefore, shifting our attention from the (y,c)-space to the (ℓ, c)-space, we note that $w_j < w_k$ implies $IB(z_j, w_j, R_j) \subset IB(z_k, w_k, R_k)$, where $z_j = (y_j/\theta_j, c_j)$ and $z_k = (y_k/\theta_k, c_k)$. Consequently, $b(z_j, w_j, R_j) \leqslant b(z_k, w_k, R_k)$. The argument is illustrated in Fig. 11. □

D.4. Step 2: Who is the worst-off across n?

Proposition 4. *If tax scheme τ is optimal in economy $E = (w_N, R_N)$ for a social welfare function $\mathbf{R}^{\tilde{R}\,lex}$, then allocation $\underline{z}_N$ generated by τ has the feature that the corresponding allocation z_N in the labor time/consumption space is such that agent 1 has the lowest well-being level:*

$$b(z_1, w_1, R_1) \leqslant b(z_i, w_i, R_i), \forall i \in N.$$

Proof. Fig. 12 illustrates the reasoning. Let $E = (w_N, R_N)$. Let τ be optimal in E for $\mathbf{R}^{\tilde{R}\,lex}$. Let $\underline{z}_N$ be generated by τ. Let z_N corresponds to $\underline{z}_N$ in the (ℓ, c) space. Without loss of generality, we restrict our attention to functions $y - \tau(y)$ that follow the lower envelope of the union of the upper contours sets at all bundles $\underline{z}_i$. Formally, we assume that $\forall y \in \mathscr{R}_+$, there exists $i \in N$ such that $(y, y - \tau(y))\underline{I}_i\,\underline{z}_i$. If it were not the case, then at least some interval of incomes would not be chosen by any agents. Modifying the tax scheme, on that latter interval, so that it follows the lower envelope of the agents' upper contours sets does not modify any agent's behavior nor well-being so that the change is irrelevant.

Let us assume, contrary to the claim, that

$$b(z_1, w_1, R_1) > b(z_j, w_j, R_j) = b$$

for some $j \in N$. Let $\hat{\underline{z}}_1 = (\hat{y}_1, \hat{c}_1) \in \underline{X}$ be defined by:

$$b\left(\left(\frac{\hat{y}_1}{w_1}, \hat{c}_1\right), w_1, R_1\right) = b$$

and

$$\hat{\underline{z}}_1 \in \underline{IB}(\hat{\underline{z}}_1, \underline{R}_1).$$

Note that by assumption, $\underline{z}_1\,\underline{P}_1\,\hat{\underline{z}}_1$. Later on in the proof, we will use the following fact:

$$\forall y \leqslant \hat{y}_1, \tau(y) \leqslant \hat{y}_1 - \hat{c}_1. \tag{11}$$

Let us prove this fact. Assume it is not true. Then there exists $y^* < \hat{y}_1$, such that $\tau(y^*) = \max_{y < \hat{y}_1} \tau(y)$ and $\tau(y^*) > \tau(\hat{y}_1)$. By the assumption made in the beginning of the proof, there exists $i \in N$ such that $\underline{z}_i\,\underline{I}_i\,(y^*, y^* - \tau(y^*))$. Note that among all the feasible bundles that leave i indifferent to $(y^*, y^* - \tau(y^*))$, this bundle itself maximizes the paid tax (or minimizes the received transfer), so that, if τ is supposed to be optimal, it is i's bundle: $\underline{z}_i = (y^*, y^* - \tau(y^*))$. Indeed, if it were not the case, by replacing $\underline{z}_i$ with $(y^*, y^* - \tau(y^*))$, the planner would not change anything in the well-being of the agents, all the incentive constraints would remain satisfied and there would be a budget surplus, which could be redistributed. As a consequence, and because we have assumed that $\forall n \in [\underline{n}, n_1]$ there exist agents with the minimum wage, there exists $j \in N$ such that $w_j = w_1$ and $\underline{z}_j = (y^*, y^* - \tau(y^*))$. As a result,

$$\underline{IB}(\underline{z}_j, \underline{R}_j) = \{\underline{z} = (y,c) \in \underline{X} | c = y - \tau(y^*)\}.$$

That implies

$$\underline{IB}(\underline{z}_j, \underline{R}_j) \subset \underline{IB}(\hat{\underline{z}}_1, \underline{R}_1),$$

with the consequence that

$$b\left(\left(\frac{y^*}{w_j}, y^* - \tau(y^*)\right), w_j, R_j\right) < b,$$

the desired contradiction, proving the fact stated above.

We now define an income threshold $\hat{y}$ by distinguishing two cases.

1. $\hat{c}_1 > \hat{y}_1 - \tau(\hat{y}_1)$: $\hat{y}$ is defined by
$$(\hat{y}, \hat{y} - \tau(\hat{y}))\,\underline{I}_1\,\hat{\underline{z}}_1,$$
and in case several such $\hat{y}$ exist, we take the largest.
2. $\hat{c}_1 > \hat{y}_1 - \tau(\hat{y}_1)$: $\hat{y}$ is defined by
$$\tau(\hat{y}) = \hat{y}_1 - \hat{c}_1,$$
and in case several such $\hat{y}$ exist, we take the largest.

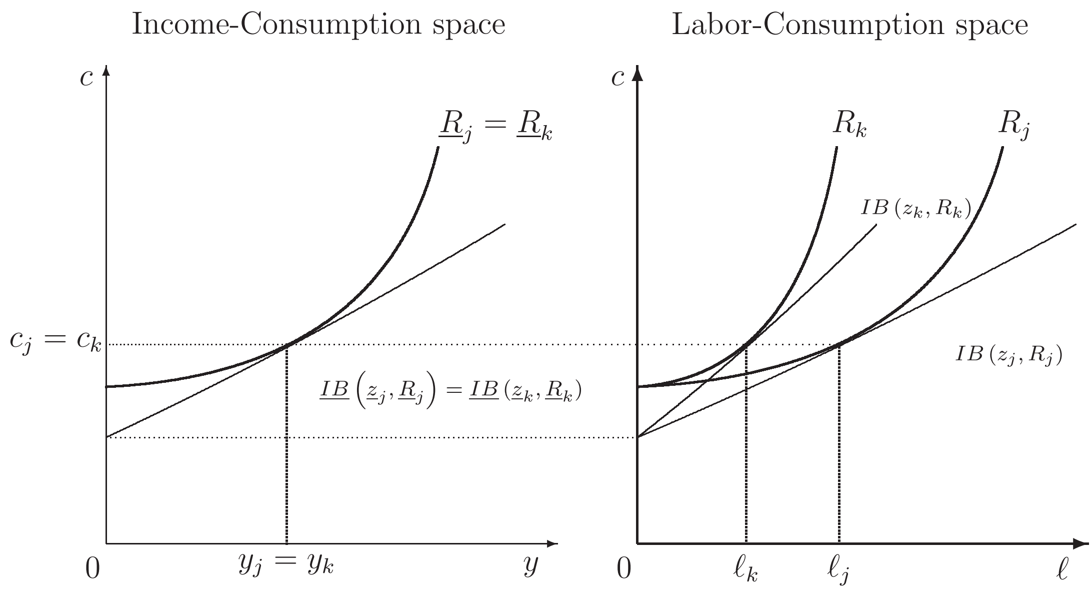

Fig. 11. Illustration of Lemma 1. In the (y, c) space, agents j and k have the same preferences $\underline{R}$, choose the same bundle $\underline{z}$ and have the same implicit budget $\underline{IB}$. They have different wages $w_j < w_k$. To be represented in the (ℓ, c) space, each agent's indifference curve R, bundle z and implicit budget IB needs to be re-scaled by their wage. Indeed, $\ell_k = y_k / w_k < \ell_j = y_j / w_j$. Clearly, $IB(z_j, w_j, R_j) \subset IB(z_k, w_k, R_k)$. Therefore, $b(z_j, w_j, R_j) \leqslant b(z_k, w_k, R_k)$.

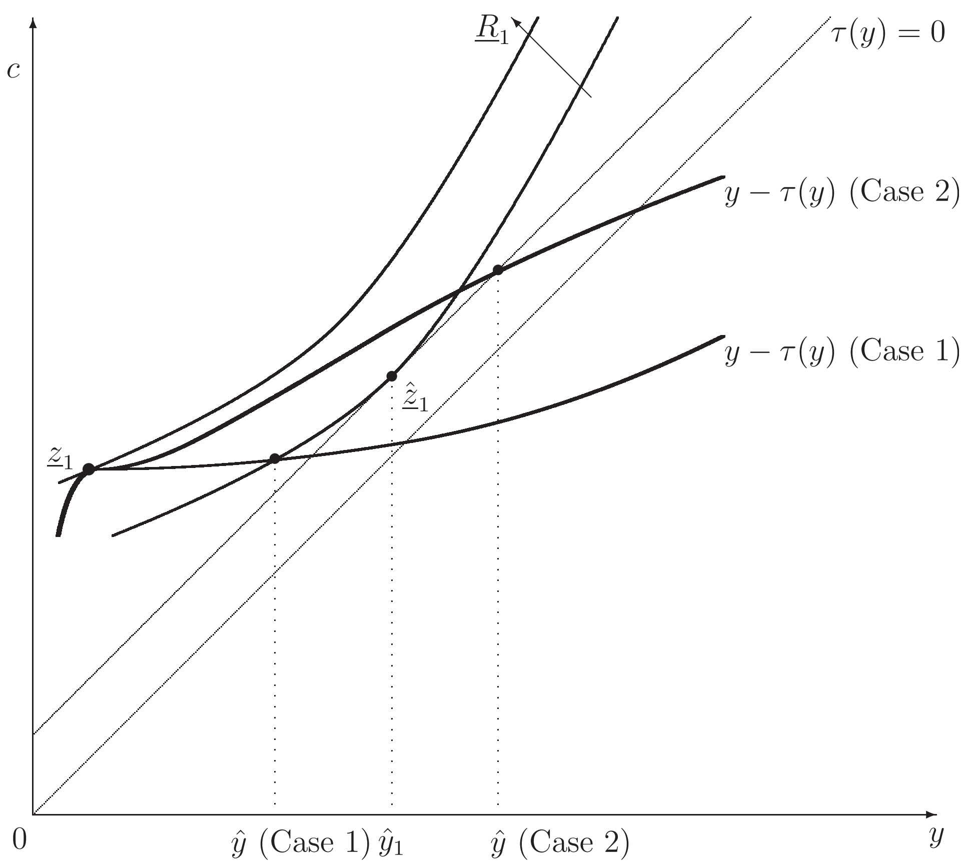

Fig. 12. Illustration of the proof of Proposition 4. The result is proven by contradiction. Contrary to the claim, assume that τ is optimal for $\mathbf{R}^{\tilde{R}lex}$ but some agent $j \neq 1$ has a well-being index $b(z_j, w_j, R_j) = b < b(z_1, w_1, R_1)$. The proof involves bundle $\hat{\underline{z}}_1$ at which 1's well-being index is b and $\hat{\underline{z}}_1$ is on the frontier of 1's implicit budget at that level of well-being. There are two possible cases: either τ crosses 1's indifference curve (at well-being b) below $\hat{\underline{z}}_1$ (case 1) or it crosses above it (case 2). In each case we can define some $\hat{y}$ such that alternative tax scheme $\tau'(y) = \max\{\tau(y), \tau(\hat{y})\}$ generates a budget surplus and no agents has well-being below b. This contradicts the optimality of τ.

Let τ' be defined by: for all $y \geqslant 0$

$$\tau'(y) = \max\{\tau(y), \tau(\hat{y})\}. \qquad (12)$$

and let $\underline{z}'_N$ be the allocation generated by τ'.

Claim 1: τ' collects a budget surplus: $\sum_{i \in N} \tau'(y'_i) > 0$. By construction, $\underline{B}(\tau') \subset \underline{B}(\tau)$. Therefore, if $\tau'(y_i) = \tau(y_i)$, that is, τ' and τ coincide at y_i, then $\underline{z}'_i = \underline{z}_i$, that is agent i does not change her choice of bundle. If $\tau'(y_i) \neq \tau(y_i)$, because $\tau(y_i) < \tau(\hat{y})$, then $\underline{z}_i' \neq \underline{z}_i$ but by construction of $\tau', \tau'(y_i') \geqslant \tau(y_i)$. In any case, $\tau'(y_1') > \tau(y_1)$, which proves the claim.

Claim 2: $\min_{i\in N} b(z_i', w_i, R_i) = b$. Let $k \in N$ be such that

$$b(z_k', w_k, R_k) = \min_{i\in N} b(z_i', w_i, R_i).$$

We consider two cases in turn. Case 1: $w_k = w_1$. Because $R_k \leqslant R_1$ since $(\theta_k \leqslant \overline{\theta} = \theta_1), y_k' \leqslant y_1'$. Gathering Eq. (12), and the fact that $\tau(\hat{y}) \leqslant \hat{y}_1 - \hat{c}_1$, we obtain

$$\forall y \leqslant \hat{y}_1, \tau'(y) \leqslant \hat{y}_1 - \hat{c}_1.$$

This implies

$$\underline{IB}(\underline{z}_k, \underline{R}_k) \subseteq \underline{IB}(\hat{\underline{z}}_1, \underline{R}_1),$$

with the immediate consequence that $b(z_k', w_k, R_k) \geqslant b$, the desired outcome. Case 2: $w_k > w_1$. If $y_k' \geqslant \hat{y}$, then $y_k' = y_k, \underline{z}_k' = \underline{z}_k$, nothing changes in the well-being levels and the claim is proven. If $y_k' < \hat{y}$, then

$$\underline{IB}(\underline{z}_k, \underline{R}_k) \supseteq \underline{IB}(\hat{\underline{z}}_1, \underline{R}_1),$$

so that, because $w_k > w_1$,

$$IB(z_k', R_k, w_k) \supset IB(z_1', R_1, w_1),$$

so that

$$b(z_k', R_k, w_k) \geqslant b(z_1', R_1, w_1),$$

the desired outcome that completes the proof of Claim 2.

Gathering both claims, τ' does not decrease the minimal well-being level whereas it collects a budget surplus. Redistributing that surplus among all agents would strictly increase the minimal well-being level, contradicting the assumption that τ is optimal. □

D.5. Step 3: Setting-up the maximization problem and deriving the first-order conditions

We consider the "dual" of the original problem:

$$V \equiv \max_{\{u(n),y(n)\}} \int_{\underline{n}}^{\infty} \left(y(n) - \left(u(n) + \frac{\epsilon}{1+\epsilon} \left(\frac{y(n)}{n} \right)^{\frac{1+\epsilon}{\epsilon}} \right) \right) dF(n) \tag{13}$$

subject to (3), (9) and (10)

Constraints (9) and (10) are the lower-bound on utilities at the solution presented in the overview of the proof and constraint (3) is the incentive compatibility constraint. The first-order conditions of (3) being satisfied at $n' = n$ implies:

$$u'(n) = \left(\frac{y(n)}{n} \right)^{\frac{1+\epsilon}{\epsilon}} \left(\frac{1}{n} \right) \tag{14}$$

The second-order conditions of (3) are equivalent to (see Mirrlees, 1976):

$$y'(n) \geqslant 0 \forall n \in [\underline{n}, \infty) \tag{15}$$

We divide (13) into two independent subproblems:

$$V^L(y_1) \equiv \max_{\{u(n),y(n)\}} \int_{\underline{n}}^{n_1} \left(y(n) - \left(u(n) + \frac{\epsilon}{1+\epsilon} \left(\frac{y(n)}{n} \right)^{\frac{1+\epsilon}{\epsilon}} \right) \right) dF(n)$$

subject to (9), (10), (14), (15) and

$$y(n) \leqslant y_1 \forall n$$

and

$$V^U(y_1) \equiv \max_{\{u(n),y(n)\}} \int_{n_1}^{\infty} \left(y(n) - \left(u(n) + \frac{\epsilon}{1+\epsilon} \left(\frac{y(n)}{n} \right)^{\frac{1+\epsilon}{\epsilon}} \right) \right) dF(n)$$

subject to (9), (10), (14), (15) and

$$y(n) \geqslant y_1 \forall n$$

We denote by $y^*(n_1)$ the income of agent 1 at the solution. The requirement that $y(n) \leqslant y^*(n_1) \forall n \in [\underline{n}, n_1]$ in the lower subproblem and $y(n) \geqslant y^*(n_1) \forall n \in [n_1, \infty)$ in the upper subproblem follows from the constraint that $y(n)$ be everywhere increasing in n (including at n_1). This last constraint is likely to be binding at n_1 (see Online Appendix I.1 for a detailed explanation).

In practice, both $V^L(y)$ and $V^U(y)$ can be solved independently for a given y_1 using Hamiltonians. By definition of $y^*(n_1)$ being optimal, it solves:

$$\frac{\partial V^L}{\partial y_1}(y^*(n_1)) + \frac{\partial V^U}{\partial y_1}(y^*(n_1)) = 0 \tag{16}$$

where the derivatives of $V^L(y)$ and $V^U(y)$ can be found using the envelope theorem.

In Online Appendix I.3, we define the Hamiltonians corresponding to each subproblem and derive the first-order conditions which can be re-arranged as:

$$\frac{\tau'(y(n))}{1-\tau'(y(n))} = A(n)D(n)C(n) \quad \forall n \in [\underline{n}, n_{\underline{b}}] \tag{17}$$

$$\frac{\tau'(y(n))}{1-\tau'(y(n))} = A(n)B(n)C(n) \quad \forall n \in [n_{\overline{b}}, \infty) \tag{18}$$

with

$$A(n) = \left[\frac{1+\epsilon}{\epsilon} \right] \tag{19}$$

$$B(n) = 1 - F(n) - \int_n^{\infty} \mu(t)dt \tag{20}$$

$$C(n) = \frac{1}{nf(n)} \tag{21}$$

$$D(n) = -F(n) + \int_{\underline{n}}^{n} \mu(t)dt \tag{22}$$

where $\mu(n)$ is the lagrange multiplier associated with the lower-bound on utilities (constraint (10)) at n. In addition, $n_{\underline{b}} \in [\underline{n}, n_1]$ and $n_{\overline{b}} \in [n_1, \infty)$ are the end-points of the interval of types over which agents are bunched and earn $y^*(n_1)$. For any n at which the constraint $u(n) \geqslant \underline{u}(n)$ is binding, $\mu(n) > 0$. Otherwise, $\mu(n) = 0$.

D.6. Step 4: Optimal tax formula as a function of economic and normative primitives

Lemma 2. *Under the assumption that reference preferences can be represented by the utility function in (5), the worst-off agent at type n_i has well-being index*

$$\underline{b}(u(n_i); n_i) = u(n_i) + \frac{1}{1+\epsilon} \left[\left(n_i \frac{\tilde{\theta}}{\overline{\theta}} \right)^{1+\epsilon} - (n_i)^{1+\epsilon} \right] \text{if } n_i \geqslant n_1 \tag{23}$$

$$= u(n_i) + \tfrac{1}{1+\epsilon} \left[\left(\underline{w}\tilde{\theta} \right)^{1+\epsilon} - (n_i)^{1+\epsilon} \right] \text{if } n_i \leqslant n_1.$$

Proof. The intercept of the implicit budget of agent i is the smallest T_i such that

$$T_i + \max_{\ell_i} \left\{ w_i \ell_i - \frac{\epsilon}{1+\epsilon} \left(\frac{\ell_i}{\theta_i} \right)^{\frac{1+\epsilon}{\epsilon}} \right\} \geqslant u_i$$

Therefore:

$$T_i = u_i - \frac{1}{1+\epsilon}(w_i\theta_i)^{1+\epsilon}$$

Maximizing utility function (5) (the reference preferences) over i's implicit budget, we obtain the following indirect utility:

$$b(u_i, w_i, \theta_i) = u_i + \frac{1}{1+\epsilon}\left[\left(w_i\tilde{\theta}\right)^{1+\epsilon} - (w_i\theta_i)^{1+\epsilon}\right]$$

Finally, per Lemma 1, the minimum well-being within each n_i is:

$$\begin{aligned}\underline{b}(u(n_i); n_i) = &\quad u(n_i) + \tfrac{1}{1+\epsilon}\left[\left(n_i\tfrac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon} - (n_i)^{1+\epsilon}\right]\text{if } n_i \geqslant n_1\\ = &\quad u(n_i) + \tfrac{1}{1+\epsilon}\left[\left(\underline{w}\tilde{\theta}\right)^{1+\epsilon} - (n_i)^{1+\epsilon}\right]\text{if } n_i \leqslant n_1\end{aligned}$$

Using the expression for the well-being index in Lemma 2, we first derive the marginal tax rate on intervals where the lower-bound on utilities (10) is binding. Equalizing the well-being of the worst-off for each n implies $\frac{d}{dn}\underline{b}(u(n); n) = 0 \Rightarrow$:

$$u'(n) = n^{\epsilon} \quad \forall n \in [\underline{n}, n_1) \tag{25}$$

$$u'(n) = \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right] n^{\epsilon} \quad \forall n \in [n_1, \infty) \tag{26}$$

Combining Eq. (25) and (26) with the first-order conditions of the incentive compatibility constraint, we find:

$$y(n) = n^{1+\epsilon} \quad \forall n \in [\underline{n}, n_1) \tag{27}$$

$$y(n) = \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{\epsilon}{1+\epsilon}} n^{1+\epsilon} \quad \forall n \in [n_1, \infty) \tag{28}$$

Finally, using the fact that $1 - \tau'(y(n)) = \left(\frac{y(n)}{n}\right)^{\frac{1}{\epsilon}}\left(\frac{1}{n}\right)$:

$$\tau'(y(n)) = 0 \quad \forall n \in [\underline{n}, n_1) \tag{29}$$

$$\tau'(y(n)) = 1 - \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}} \quad \forall n \in [n_1, \infty) \tag{30}$$

Let us first focus on $[\underline{n}, n_1)$ and fully characterize the solution on that interval. From (15), $y(n)$ is increasing. Therefore, $y^*(n) \leqslant y_1 = y^*(n_1)$ can only bind (if at all) on $[n_{\underline{b}}, n_1]$ for some $n_{\underline{b}} \leqslant n_1$. If such an interval exists, then $y^*(n) = y^*(n_1) \forall n \in [n_{\underline{b}}, n_1]$.

Re-arranging the first-order conditions (17), we have

$$\int_{\underline{n}}^{n} \mu(t)dt = \frac{1+\epsilon}{\epsilon}\frac{\tau'(y(n))}{1-\tau'(y(n))} nf(n) + \int_{\underline{n}}^{n} f(t)dt$$

As a result, we can write

$$\mu(n) = \frac{d}{dn}\left[\frac{1+\epsilon}{\epsilon}\frac{\tau'(y(n))}{1-\tau'(y(n))} nf(n)\right] + f(n)$$

Whenever $\mu(n) > 0$, we have $\tau'(y(n)) = 0$ and $\tau''(y(n)) = 0$ because of (29). As a result, either $\mu(n) = 0$ or $\mu(n) = f(n)$. This implies that $\int_{\underline{n}}^{n} f(t)dt \geqslant \int_{\underline{n}}^{n} \mu(t)dt$ which, combined with (17), means that tax rates are (weakly) negative on $[\underline{n}, n_1]$. Notice that if marginal tax rates are everywhere zero or negative, the lower-bound (along which the marginal tax rate is zero) can be binding on at most one interval from $\underline{n}$ to some n_u (this proves Claim 1 in Proposition 1). Finally, because the lower-bound is only binding on interval $[\underline{n}, n_u]$, we have that $\int_{\underline{n}}^{n}\mu(t)dt = \int_{\underline{n}}^{n_u}\mu(t)dt = \int_{\underline{n}}^{n_u} f(t)dt \; \forall n \in [n_u, n_1]$. Plugging back into (17) proves Claim 2 of Proposition 1.

We now turn our attention to $[n_1, \infty)$. First, because of (15), the constraint that $y(n) \geqslant y_1 = y^*(n_1)$ can only bind (if at all) on interval $[n_1, n_{\bar{b}})$ for some $n_{\bar{b}} \in [n_1, \infty)$. This directly proves Claim 3 of Proposition 1. Second, by assumption there is at most one interval above n_1 where (10) binds. We call that interval $[n_l, n']$ with $n_l \leqslant n' \in (n_{\bar{b}}, \infty)$. The marginal tax rate on that interval is given by (30).

Regarding the optimal tax rate in the upper tail, notice that when $\int_{n}^{\infty}\mu(t)dt = 0$, the optimal tax rate is the revenue-maximizing one defined by:

$$\frac{\tau'(y(n))}{1-\tau'(y(n))} = \frac{1-F(n)}{nf(n)}\left(\frac{1+\epsilon}{\epsilon}\right) \tag{31}$$

and converges to $\tau'(y(n)) \to \frac{1+\epsilon}{1+\epsilon+\alpha\epsilon}$ as $n \to \infty$.[19] Denote by $\tilde{\tau}'(y(n))$ the marginal tax rate defined by (31). Two scenarii are possible. If $\frac{1+\epsilon}{1+\epsilon+\alpha\epsilon} \leqslant 1 - \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$, there exist some n' such that $\tilde{\tau}'(n') = 1 - \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$ and $\tilde{\tau}'(n) \leqslant 1 - \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}} \forall n \in [n', \infty)$. In that case, applying $\tau(y(n)) = \tilde{\tau}(y(n)) \; \forall n \in [n', \infty)$ is a Pareto improvement compared to applying (30). This proves Claim 6 in Proposition 1. In that case, we also have:

$$\int_{n}^{\infty}\mu(t)dt = \int_{n_l}^{n'}\mu(t)dt = 1 - F(n_l) - \frac{1 - \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}}{\left(\frac{1+\epsilon}{\epsilon}\right)\left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}} n_l f(n_l)$$

$\forall n \in [n_1, n_l]$.[20] Combining the latter with Eq. (18) yields an expression for marginal tax rates over $[n_{\bar{b}}, n_l]$ (Claim 4 in Proposition 1).

If $\frac{1+\epsilon}{1+\epsilon+\alpha\epsilon} \geqslant 1 - \left[1 - \left(\frac{\tilde{\theta}}{\bar{\theta}}\right)^{1+\epsilon}\right]^{\frac{1}{1+\epsilon}}$, the lower-bound on utilities remains binding in the tail. In that case, the marginal tax rate over $[n_{\bar{b}}, n_l]$ is the same as in the previous case but the tax rate over interval $[n_l, \infty)$ is also given by Eq. (30).[21] This demonstrates Claim 5 in Proposition 1 and concludes the proof. □

Appendix E. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jpubeco.2021.104386.

References

Blundell, Richard, Shephard, Andrew, 2012. Employment, hours of work and the optimal taxation of low-income families. Rev. Econ. Stud. 79 (2), 481–510.

Boadway, R., 2012. From Optimal Tax Theory to Tax Policy. MIT Press, Cambridge and New York.

Boadway, R., Jacquet, L., 2008. Optimal marginal and average income taxation under maximin. J. Econ. Theory 143, 425–441.

Boadway, Robin, Marchand, Maurice, Pestieau, Pierre, Racionero, Maria del Mar, 2002. Optimal redistribution with heterogenous preferences for leisure. J. Public Econ. Theory 4(4), 475–498.

[19] See Boadway and Jacquet (2008) for a discussion of the Rawlsian marginal tax rates under quasi-linear preferences and various distributions of wages.

[20] For more details, see paragraph "**Case 2**" in Online Appendix I.3.4.

[21] For more details, see paragraph "**Case 1**" in Online Appendix I.3.4.

Brett, Craig, Weymark, John A., 2003. Financing education using optimal redistributive taxation. J. Public Econ. 87, 2549–2569.
Brett, Craig, Weymark, John A., 2017. Voting over selfishly optimal nonlinear income tax schedules. Games Econ. Behav. 101, 172–188.
Chetty, Raj, 2012. Bounds on elasticities with optimization frictions: a synthesis of micro and macro evidence on labor supply. Econometrica, Econometric Soc. 80 (3), 969–1018.
Choné, Philippe, Laroque, Guy, 2010. Negative marginal tax rates and heterogeneity. Am. Econ. Rev. 100 (5), 2532–2547.
DeNavas-Walt, Carmen, Proctor, Bernadette D., 2014. U.S. Census Bureau, Current Population Reports, P60–252, Income and Poverty in the United States: 2014, U. S. Government Printing Office, Washington, DC.
Diamond, Peter, 1980. Income taxation with fixed hours of work. J. Public Econ. 13 (1), 101–110.
Diamond, A., 1998. Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. Am. Econ. Rev. 88 (1), 83–95.
Daniel, Feenberg, Coutts, Elisabeth, 1993. An introduction to the TAXSIM model. J. Policy Anal. Manage. 12 (1), 189–194.
Fleurbaey, M., 2008. Fairness, Responsibility, and Welfare. Oxford University Press, Oxford.
Fleurbaey, M., Maniquet, F., 2005. Fair social orderings with unequal production skills. Soc. Choice Welfare 24, 1–35.
Fleurbaey, Marc, Maniquet, François, 2006. Fair income tax. Rev. Econ. Stud. 73, 55–83.
Fleurbaey, Marc, Maniquet, François, 2007. Help the low-skilled, of let the hardworking thrive? A study of fairness in optimal income taxation. J. Public Econ. Theory 9, 467–500.
Fleurbaey, M., Maniquet, F., 2011. A Theory of Fairness and Social Welfare. Cambridge University Press, Cambridge.
Fleurbaey, M., Maniquet, F., 2018. Optimal income taxation theory and principles of fairness. J. Econ. Lit. forthcoming.
Fleurbaey, M., Trannoy, A., 2003. The impossibility of a Paretian Egalitarian. Soc. Choice Welfare 21, 243–464.
Flood, S., King, M., Rodgers, R., Ruggles, S., Warren, J.R., 2018. Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset]. IPUMS, Minneapolis, MN.
Gaertner, W., Schokkaert, E., 2012. Empirical Social Choice. Cambridge University Press, Cambridge.
Kanbur, R.M., Keen, M., Tuomala, M., 1994a. Optimal non-linear income taxation for the alleviation of income-poverty. Eur. Econ. Rev. 38, 1613–1632.
Kanbur, R., Keen, M., Tuomala, M., 1994b. Labor supply and targeting in poverty alleviation programs. World Bank Econ. Rev. 8, 191–211.
Konow, J., 2001. Fair and square: the four sides of distributive justice. J. Econ. Behav. Org. 46, 137–164.
Lockwood Benjamin, B., 2020. Optimal income taxation with present bias. Am. Econ. J.: Econ. Policy 12 (4), 298–327.
Lockwood, Benjamin B., Weinzierl, Matthew, 2015. De Gustibus Non Est Taxandum: heterogeneity in preferences and optimal redistribution. J. Public Econ. 124, 74–80.
Lockwood, Benjamin B., Weinzierl, Matthew, 2016. Positive and normative judgments implicit in U.S. tax policy, and the costs of unequal growth and recessions. J. Monetary Econ. 77 (2016), 30–47.
Maniquet, Francois, Neumann, Dirk. forthcoming. Well-being, poverty and labor income taxation: theory and application to Europe and the U.S. Am. Econ. J.: Microeconomics.
Mirrlees, James A., 1971. An exploration in the theory of optimum income taxation. Rev. Econ. Stud. 38 (2), 175–208.
Mirrlees, James A., 1976. Optimal tax theory: a synthesis. J. Public Econ. 6, 327–358.
Piketty, T., Saez, E., 2013. Optimal labor income taxation. In: Auerbach, Alan J., Chetty, Raj, Feldstein, Martin, Saez, Emmanuel (eds.), Handbook of Public Economics, Vol. 5, pp. 391–474, Amsterdam: Elsevier, North Holland.
Roemer, John E., 1998. Equality of Opportunity. Harvard University Press, Cambridge.
Roemer, John E., Aaberge, Rolf, Colombino, Ugo, Fritzell, Johan, Jenkins, Stephen P., Lefranc, Arnaud, Marx, Ive, Page, Marianne, Pommer, Evert, Ruiz-Castillo, Javier, Segundo, Maria Jesus San, Tranaes, Torben, Trannoy, Alain, Wagner, Gert G., Zubiri, Ignacio, 2003. To what extent do fiscal regimes equalize opportunities for income acquisition among citizens?. J. Public Econ. 87 (3–4), 539–565.
Saez, Emmanuel, 2002. Optimal income transfer programs: intensive versus extensive labor supply responses?. Q. J. Econ. 117 (3), 1039–1073.
Saez, E., Stantcheva, S., 2016. Generalized social marginal welfare weights for optimal tax theory. Am. Econ. Rev. 106, 24–45.
Samuelson, Paul A., 1974. Complementarity – An Essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. J. Econ. Lit. 12, 1255–1289.
Schokkaeert, E., Devooght, K., 2003. Responsibility-sensitive fair compensation in different cultures. Soc. Choice Welfare 21, 207–242.
Wane, W., 2001. The optimal income tax when poverty is a public bad. J. Public Econ. 82, 271–299.
Weinzierl, Matthew, 2014. The promise of positive optimal taxation: normative diversity and a role for equal sacrifice. J. Public Econ. 118, 128–142.